Online Learning and Decision-Making under Generalized Linear Model with High-Dimensional Data

Xue Wang* Mike Mingcheng Wei* Tao Yao†

*Penn State University, Industrial and Manufacturing Engineering, xzw118@psu.edu
*University at Buffalo, School of Management, mcwei@buffalo.edu
†Penn State University, Industrial and Manufacturing Engineering, tyy1@engr.psu.edu

We propose a minimax concave penalized multi-armed bandit algorithm under the generalized linear model (G-MCP-Bandit) for a decision-maker facing high-dimensional data in an online learning and decision-making process. We demonstrate that the G-MCP-Bandit algorithm asymptotically achieves the optimal cumulative regret in the sample size dimension T, $O(\log T)$, and further attains a tight bound in the covariate dimension d, $O(\log d)$. In addition, we develop a linear approximation method, the 2-step weighted Lasso procedure, to identify the MCP estimator for the G-MCP-Bandit algorithm under non-iid samples. Under this procedure, the MCP estimator matches the oracle estimator with high probability and converge to the true parameters at the optimal convergence rate. Finally, through experiments based on synthetic data and two real datasets (warfarin dosing dataset and Tencent search advertising dataset), we show that the G-MCP-bandit algorithm outperforms other benchmark algorithms in terms of cumulative regret and that the benefits of the G-MCP-Bandit algorithm seem to increase with the data's sparsity level and the size of the decision set.

Key words: Multi-armed bandit, minimax concave penalty, high-dimensional data, online learning and decision-making, generalized linear model.

1. Introduction

Individual-level data have become increasingly accessible in the Internet era, and decision-makers have accelerated data accumulation with extraordinary speed in a variety of industries, including health care, retail, advertising, etc. The growing availability of user-specific data, such as demographics, geographics, medical records, and searching/browsing history, provides decision-makers with unprecedented opportunities to tailor decisions to individual users. For example, doctors can personalize treatments for patients based on their medical history, clinical tests, and biomarkers; search engines can offer personalized advertisements for users based on their queries, demographics, and geographics. These user-specific data are often collected sequentially over time, during which decision-makers adaptively learn to predict the expected rewards based on users' responses to each

available decision as a function of the user-specific data (i.e., the user's covariates) and optimally adjust decisions to maximize their rewards – an *online* learning and decision-making process.

This online learning and decision-making process requires a thoughtful balance between exploration and exploitation. Consider a decision-maker who selects decisions for incoming users and obtains rewards based on users' responses to these decisions. To maximize his expected rewards, the decision-maker first needs an accurate predictive model for users' responses, which is typically uncertain at the beginning but can be partially learned through collecting samples of users' responses. On the one hand, the decision-maker could select a decision that yields the "highest", based on his best knowledge so far, expected reward (i.e., exploitation). Yet, this decision can be suboptimal, as the selection is based on the rough prediction of users' responses due to limited samples. Even worse, the decision-maker could incorrectly estimate the expected reward of the true optimal decision to be low and never have a chance to correct such a mistake (as the decision-maker will not select the true optimal decision due to the current low reward prediction, he will not generate additional samples to be able to learn and correct his incorrect estimation). On the other hand, the decision-maker can improve his predictive ability and learn users' responses by collecting more response samples, which often are obtained through random clinical trials and/or user experiments and are typically costly (i.e., exploration). The exploration and exploitation dilemma has been extensively studied in the multi-armed bandit model (Robbins 1952), but the growing dimensionality and availability of data have added another layer of complexity to the bandit model.

In practice, individual-level data are typically presented in a high-dimensional fashion, which poses significant computational and statistical challenges in the online learning and decision-making process. Traditional statistical methods, such as Ordinary Least Squares (OLS), require a large number of samples (e.g., the sample size must be larger than the covariate dimension) to be deemed computationally feasible. Under high-dimensional settings, learning the accurate predictive models requires a substantial amount of samples, which are obtained, if possible, through costly trials or experiments. Take the search advertising industry for example. Search advertising occurs when an Internet user searches certain keyword(s) (i.e., a query) in an online search engine and then the search engine displays both search results, in response to the user's query, and some sponsored ads, in response to the query and user-specific information. In order to select the ad that maximizes its revenue, the search engine must have accurate estimations on users' clicking probabilities in response to the displayed ads – Click-Through Rate (CTR).

However, the search engine's ability to accurately predict CTR is often crippled by the high-dimensional search advertising data coupled with limited samples. Counting more than three quarters of a million distinct words and their combinations (OxfordDictionaries 2018), there are nearly infinite possible queries the user can submit to the search engine. For example, from 2003 to 2012,

Google answered 450 billion unique queries, and it has estimated that 16% to 20% of queries submitted every day have never been used before (Mitchell 2012). Hence, to accurately estimate a single ad's CTR to these queries, the search engine requires billions, if not trillions, of samples. The craving for samples will be further intensified if the search engine practices personalized advertising by taking users' individual information (such as demographics and geographics) into consideration. However, the available samples for the search engine to learn and predict CTR are greatly limited. Consider a 45 days new marketing campaign promoting a sales event or merchandise, during which time an average ad is expected to reach approximately one third of a million users (WordStream 2017, Shewan 2017). Among these users, a very small portion can be selected to perform costly experiments to learn CTR, and that number is much smaller comparing to the size of queries and individual data.

In this paper, we propose a new algorithm, the G-MCP-Bandit algorithm, for online learning and decision-making processes in high-dimensional settings. Our algorithm follows the ideas of the bandit model and develops a ϵ-decay random sampling method to balance the exploration-and-exploitation trade-off. We allow the decision-maker's reward function to follow the generalized linear model (McCullagh and Nelder 1989), which is a large class of models including the linear model, the logistic model, the Poisson regression model, etc., and we adopt the Minimax Concave Penalized (MCP) method (Zhang et al. 2010) to improve the parameter estimations and predict the expected rewards in high-dimensional settings.

In the high-dimensional statistics literature, MCP is developed to explore and recover the latent sparse data structure for high-dimensional data. Compared to traditional statistical methods (e.g., OLS), MCP uses significantly fewer data samples and delivers better performance in high-dimensional settings (Zhang et al. 2010). Although it is statistically favorable to adopt MCP, solving the MCP estimator (an NP-complete problem) could be computationally challenging. We propose a linear approximation method, the 2-step weighted Lasso procedure (2sWL), under the bandit setting as an efficient approach to tackle this challenge. We show that the MCP estimator solved by the 2sWL procedure matches the oracle estimator with high probability and converges to the true parameter with the optimal convergence rate. Since the bandit model mixes the exploitation and exploration phases, samples generated under the exploitation phase may be non-iid. Therefore, we adopt a matrix perturbation technique to derive new oracle inequalities for the MCP estimator under non-iid samples. To the best of our knowledge, this work is the first one that applies MCP to handle non-iid samples.

We theoretically demonstrate that the G-MCP-Bandit algorithm can significantly improve the cumulative regret bound in high-dimensional settings comparing to existing bandit algorithms. In particular, we benchmark the G-MCP-Bandit algorithm to an oracle policy, in which all parameter

vectors are common knowledge, and adopt the expected cumulative regret (i.e., the difference in rewards achieved by the oracle policy and the G-MCP-Bandit algorithm) as the performance measure. We show that the cumulative regret of the G-MCP-Bandit algorithm over T users (i.e., a sample size of T) is at most $O(\log T)$, which is the optimal/lowest theoretical bound for all possible algorithms (Goldenshluger and Zeevi 2013). Further, we show that the G-MCP-Bandit algorithm also attains a tight bound in the covariate dimension d, $O(\log d)$. We believe that our work is the first one in high-dimensional settings that attains the logarithmic dependence on both the sample size dimension and the covariate dimension, which are of particular importance in high-dimensional data with limited samples and suggest that the G-MCP-Bandit algorithm can bring substantial regret reduction comparing to existing bandit algorithms.

Through two synthetic-data-based experiments, we benchmark the G-MCP-Bandit algorithm's performance to other state-of-the-art bandit algorithms designed both in low-dimensional settings, OLS-Bandit by Goldenshluger and Zeevi 2013 and OFUL by Abbasi-Yadkori et al. 2011, and in high-dimensional settings, Lasso-Bandit by Bastani and Bayati 2015. We find that the G-MCP-Bandit algorithm performs favorably in both experiments. In particular, when the sample size is not extremely small[1], the G-MCP-Bandit algorithm appears to be able to accurately learn the parameter estimations with limited samples and therefore have the lowest cumulative regret. Furthermore, we observe that the benefits of the G-MCP-Bandit algorithm over other benchmark algorithms seems to increase with the data's sparsity level and the size of the decision set.

Finally, we evaluate the G-MCP-Bandit algorithm's performance through two real-data-based experiments, warfarin dosing data and Tencent search advertising data, where the technical assumptions specified for the theoretical analysis of the G-MCP-Bandit algorithm's expected cumulative regret may not hold. We observe that the G-MCP-Bandit algorithm continues to perform favorably in both experiments. In particular, in the warfarin dosing experiment (formulated as a 3-armed bandit problem with 93 covariates), the G-MCP-Bandit algorithm needs the fewest patient samples (i.e., merely 50 patients) to provide better dosing decisions than actual physicians. Similarly, in the Tencent search advertising experiment (formulated as a 3-armed bandit problem with hundreds of thousands of covariates), the G-MCP-Bandit algorithm, after observing 140 users, can consistently generate better average revenue than other benchmark algorithms under the linear model. Further, we observe that the choice of the underlying reward model can significantly influence the G-MCP-Bandit algorithm's performance. In particular, under the logistic model, which is a special case of the generalized linear model, the G-MCP-Bandit algorithm merely needs 20 users to outperform other benchmark algorithms. This observation suggests that understanding the context of

[1] When the sample size is extremely small, the decision-maker has little information to learn. Therefore, all algorithms perform equally poorly.

the underlying managerial problem and identifying the appropriate model for the G-MCP-Bandit algorithm can be critical and bring the decision-maker substantial revenue improvement.

2. Literature Review

This research is closely related to the exploration-exploitation trade off in the multi-armed bandit literature. Rigollet and Zeevi (2010), Slivkins (2014) follow the non-parametric approach and consider that the arm reward can be any smooth non-parametric function. Under this approach, the expected cumulative regret has an exponential dependence on the covariate dimension d, which is undesirable under high-dimensional settings where d can be extremely large. Such exponential dependence can be improved by following the parametric approach. Auer (2002) proposes the UCB algorithm for a linear bandit model, where the arm reward can be approximated by a linear combinations of covariates. Since Auer (2002), other UCB-type algorithms(e.g., Dani et al. 2008, Rusmevichientong and Tsitsiklis 2010, Abbasi-Yadkori and Szepesvari 2012, Deshpande and Montanari 2012) and Bayesian-type algorithms (e.g., Agrawal and Goyal 2013, Russo and Van Roy 2014) have been proposed and shown to improve on the expected cumulative regret. Yet, allowing the adversary and without regulating the sample generating process, the statistical performance of the parameter vector estimation in the learning process may suffer. As a result, the expected cumulative regret bound typically has a sublinear dependence on the sample size dimension T (e.g., $O(\sqrt{T})$) and a polynomial dependence on the covariate dimension d. However, in high-dimensional settings, where the covariate dimension and the sample size dimension can be exceedingly large, these algorithms can perform poorly.

By introducing a forced sampling approach to the linear bandit model, Goldenshluger and Zeevi (2013) ensure that enough samples generated in their algorithm possess desired iid property and show that their proposed OLS-Bandit algorithm can achieve $O(\log T)$ dependence on the sample size dimension T in low-dimensional settings. Following a similar approach, Bastani and Bayati (2015) propose the Lasso-Bandit algorithm, which attains a poly-logarithmic dependence on the sample size dimension $O(\log^2 T)$ and the covariate dimension $O(\log^2 d)$ in high-dimensional settings. In this paper, we allow the reward function to follow the generalized linear model, which contains a wide family of models that includes the linear bandit model. We propose a ϵ-decay random sampling method and show that our proposed G-MCP-Bandit algorithm continues to achieve the optimal cumulative regret bound on the sample size dimension $O(\log T)$ and attain a tight bound in the covariate dimension $O(\log d)$ in high-dimensional settings. We believe that our work is the first one that attains the logarithmic dependence on both the sample size dimension and the covariate dimension in high-dimensional settings.

Our research is also connected to the statistical learning literature. In high-dimensional statistics, Lasso type methods (Tibshirani 1996) have become the golden standard for high-dimensional

learning (Meinshausen et al. 2006, 2009, Zhang et al. 2008, Van de Geer et al. 2008). Yet, Lasso-type regularizations may lead to estimation bias, and strong conditions are needed for analyzing its theoretical performance guarantee (Fan et al. 2014a). Recently, Zhang et al. (2010) proposes MCP, a non-convex penalty method, which entails better statistical properties, such as the unbiasedness and a strong oracle property for high-dimensional sparse estimation, and requires weaker conditions than Lasso (Zou 2006, Fan et al. 2014b, Meinshausen et al. 2006). Although it is statistically favorable to adopt MCP, solving the MCP estimator (an NP-complete problem) could be computationally challenging (Liu et al. 2017, 2016). Various approximation methods have been developed in the literature. For example, Fan and Li (2001) use the local quadratic approximation, Fan et al. (2014b, 2018), Zou (2006), Zhao et al. (2014) adopt the local linear approximation, Zhang et al. (2010) choose the path following algorithm, and Liu et al. (2017) propose the second-order approximation. Our proposed solution procedure (the 2sWL procedure) is analogous to the local linear approximation and guarantees that the solution has desirable statistical properties for theoretical analysis and can be efficiently solved. In the literature, the theoretical analysis of MCP's statistical properties relies on the assumption that all samples are iid, which is hardly the case under bandit models. This paper also contribute to the statistical learning literature by deriving new oracle inequalities for MCP under non-iid samples.

3. Model Settings

Consider a sequential arrival process $t \in \{1, 2, ..., T\}$. At each time step t, a single user (e.g., consumer or patient), described by a high-dimensional feature covariate vector $\boldsymbol{x}_t \in \mathbb{R}^{1\times d}$, arrives. The covariate vector combines all available (but not necessarily valuable for the decision-maker to base his decision on) user-specific data, such as demographics, geographics, browsing/shopping history, and medical records. Upon arrival, users' covariate vectors $\{\boldsymbol{x}_t\}_{t\geq 0}$ become observable to the decision-maker and are iid distributed according to an unknown distribution $\mathcal{P}_x$.

Based on the user's covariate vector $\boldsymbol{x}$, the decision-maker will select a decision from a decision set $\mathcal{K} = \{1, 2, ..., K\}$ to maximize his expected reward. The user will respond to the chosen decision $k \in \mathcal{K}$, and such response will generate a reward for the decision-maker. Take the search advertising for example. The search engine can recommend one of K different ads to the user; the user can respond to the recommended ad by clicking, which generates revenue for the search engine. We denote this reward under the chosen decision k as R_k, which follows a distribution $\mathbb{P}(R_k|\boldsymbol{x}^T\boldsymbol{\beta}_k^{true})$, where $\boldsymbol{x}$ is the user's covariate vector and $\boldsymbol{\beta}_k^{true}$ is the unknown parameter vector corresponding to decision k.

We present the reward function in terms of the generalized linear model (McCullagh and Nelder 1989), which is a large class of models including the linear model, the logistic model, the Poisson

regression model, etc. For example, if we assume that R_k is a σ-gaussian random variable with mean $\boldsymbol{x}^T\boldsymbol{\beta}_k^{true}$, then we can define the density function of the distribution $\mathbb{P}(R_k|\boldsymbol{x}^T\boldsymbol{\beta}_k^{true})$ as $g(R_k = r|\boldsymbol{x}^T\boldsymbol{\beta}_k^{true}) = (1/\sqrt{2\pi\sigma^2})\exp(-\frac{(r-\boldsymbol{x}^T\boldsymbol{\beta}_k^{true})^2}{2\sigma^2})$, which is the standard setting for the classic linear multi-armed bandit model where the reward takes a linear form: $R_k(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}_k^{true} + \epsilon$ (Auer 2002, Agrawal and Goyal 2013). The cumulative regret performance of the linear bandit algorithms has been extensively studied by Dani et al. (2008) and Goldenshluger and Zeevi (2013), among others, under low-dimensional settings and by Bastani and Bayati (2015) under high-dimensional settings. The generalized linear model adopted in this paper facilitates us to go beyond the classic linear bandit model, as the reward may take a nonlinear form in practice. For instance, the search engine collects revenue only when a user has clicked the recommended ad; otherwise, the search engine earns nothing – a logistic model by nature. By specifying R_k as a binary random variable (e.g., $R_k \in \{0, 1\}$), we can define the mass function of the distribution $\mathbb{P}(R_k|\boldsymbol{x}^T\boldsymbol{\beta}_k^{true})$ as $g(R_k = 1|\boldsymbol{x}^T\boldsymbol{\beta}_k^{true}) = 1/(1+\exp(-\boldsymbol{x}^T\boldsymbol{\beta}_k^{true}))$ and $g(R_k = 0|\boldsymbol{x}^T\boldsymbol{\beta}_k^{true}) = \exp(-\boldsymbol{x}^T\boldsymbol{\beta}_k^{true})/(1+\exp(-\boldsymbol{x}^T\boldsymbol{\beta}_k^{true}))$, which is a logistic bandit model with the binary reward (Elmachtoub et al. 2017, Scott 2015, 2010).

The parameter vector $\boldsymbol{\beta}_k^{true}$ is high-dimensional with latent sparse structure, and we denote $\mathcal{S}_k = \{j : \beta_{k,j}^{true} \neq 0\}$ as the index set for significant covariates, which have non-zero coefficient parameters and therefore are important for the decision-maker to predict the user's response. This index set is also unknown to the decision-maker. We define the number of significant covariates as $|\mathcal{S}_k|$, which is typically much smaller than the dimension of the covariate vector.

The decision-maker's objective is to maximize his expected cumulative reward. Denote the decision-maker's current policy as $\pi = \{\pi_t\}_{t\geq 0}$, where $\pi_t \in \mathcal{K}$ is the decision prescribed by policy π at time t. To benchmark the performance of policy π, we first introduce an *oracle policy* $\pi^* = \{\pi_t^*\}_{t\geq 0}$ under which the decision-maker knows the true parameter vector values $\boldsymbol{\beta}_k^{true}$ for all $k \in \mathcal{K}$ and chooses the best decision to maximize his expected reward:

$$\pi_t^* = \arg\max_{k\in\mathcal{K}} \left\{\mathbb{E}[R_k|\boldsymbol{x}_t, \boldsymbol{\beta}_k^{true}]\right\} = \arg\max_{k\in\mathcal{K}} \left\{\int_{-\infty}^{+\infty} r_k dG(r_k|\boldsymbol{x}_t^T\boldsymbol{\beta}_k^{true})\right\},$$

where $G(r_k|\boldsymbol{x}_t^T\boldsymbol{\beta}_k^{true})$ is the cumulative distribution function for R_k. Note that in practice, the parameter vector $\boldsymbol{\beta}_k^{true}$ is unknown to the decision-maker, and therefore the construction and definition of the oracle policy directly imply that the decision-maker's reward under policy π is upper-bounded by that of the oracle policy. We therefore define the decision-maker's expected cumulative regret up to time T under the policy π as follows:

$$R^C(T) = \sum_{t=1}^{T} \mathbb{E}[R_t^{\pi_t^*} - R_t^{\pi_t}],$$

which is the expected reward difference between the optimal policy π^* and the decision-maker's alternative policy π. To maximize his expected cumulative reward, the decision-maker is equivalent to explore for the policy π that minimizes the cumulative regret up to time T.

Before presenting the proposed G-MCP-Bandit algorithm, we will first state five technical assumptions necessary for the theoretical analysis of the decision-maker's expected cumulative regret. The first three assumptions are adopted directly from the multi-armed bandit literature, and the last two assumptions from the high-dimensional statistics literature.

A. 1 (Parameter set) There exist positive constants $x_{\max}$, s, $R_{\max}$, $\beta_{\min}$ and b such that for any t and $k \in \mathcal{K}$, we have $\|\boldsymbol{x}_t\|_\infty \leq x_{\max}$, $|\mathcal{S}_k| \leq s$, $|R_k| \leq R_{\max}$, $\beta_{\min} \leq \min_{j \in \mathcal{S}_k, k \in \mathcal{K}} |\beta_{k,j}^{true}|$, $\|\boldsymbol{\beta}_k^{true}\|_1 \leq b$ and all feasible β satisfies $\|\boldsymbol{\beta}\|_1 \leq b$.

The first assumption is a standard assumption in the bandit literature (Rusmevichientong and Tsitsiklis 2010) and ensures that both the covariate vector $\boldsymbol{x}$ and the coefficient vector $\boldsymbol{\beta}_k$ are upper bounded so that the maximum regret at every time step will also be upper bounded to avoid trivial decisions. Most real world applications, including two real data experiments in §6.2 and §6.3, satisfy this assumption.

A. 2 (Margin condition) There exists a $C > 0$ such that $\mathbb{P}(0 < |\mathbb{E}[R_i|\boldsymbol{x}, \boldsymbol{\beta}_i^{true}] - \mathbb{E}[R_j|\boldsymbol{x}, \boldsymbol{\beta}_j^{true}]| \leq \gamma) \leq C R_{\max} \gamma$ for $i \neq j$ and $i, j \in \mathcal{K}$.

The second assumption is first introduced in the classification literature by Tsybakov et al. (2004). Goldenshluger and Zeevi (2013) and Bastani and Bayati (2015) adopt this assumption to the linear bandit model, under which the Margin Condition ensures only a fraction of covariates can be drawn near the boundary hyperplane $\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true} - \boldsymbol{\beta}_j^{true}) = 0$ in which rewards for both arms are nearly equal. Clearly, if a large proportion of covariates are drawn from the vicinity of the boundary hyperplane, then for any bandit algorithm, a small estimation error in the decision parameter vectors may lead the decision-maker to choose the suboptimal decision and perform poorly (Bastani and Bayati 2015). Therefore, this margin condition ensures that given a user's covariate vector, decisions can be properly separated from each other and ordered based on their rewards.

A. 3 (Arm optimality) There exists a partition $\mathcal{K}_o$ and $\mathcal{K}_s$ for $\mathcal{K}$. For $k_1 \in \mathcal{K}_s$, we will have $\mathbb{E}[R_{k_1}|\boldsymbol{x}, \boldsymbol{\beta}_{k_1}^{true}] + h < \max_{k \neq k_1} \mathbb{E}[R_k|\boldsymbol{x}, \boldsymbol{\beta}_k^{true}]$ for a positive constant h for every $\boldsymbol{x}$. For $k_2 \in \mathcal{K}_o$, these exists another positive constant p^* such that $\min \mathbb{P}(\boldsymbol{x} \in U_{k_2}) \geq p^*$, where $U_{k_2} \doteq \{\boldsymbol{x} | \mathbb{E}[R_{k_2}|\boldsymbol{x}, \boldsymbol{\beta}_{k_2}^{true}] > \max_{k \neq k_2} \mathbb{E}[R_k|\boldsymbol{x}, \boldsymbol{\beta}_k^{true}] + h, k \in \mathcal{K}\}$.

The arm optimality condition (Goldenshluger and Zeevi 2013, Bastani and Bayati 2015) ensures that as the sample size increases, the parameter vectors for optimal decisions can eventually be learned. In particular, this condition separates decisions to an optimal decision subset $\mathcal{K}_o$ and a suboptimal decision subset $\mathcal{K}_s$. Decision i in $\mathcal{K}_o$ is strictly optimal for some users' covariate vectors

(denoted by set U_i); otherwise, decision j in $\mathcal{K}_s$ must be strictly suboptimal for all users' covariate vectors. Therefore, even if there is a small estimation error for decision i in $\mathcal{K}_o$, the decision-maker will be more likely to choose decision i for a user with a covariate vector draw from the set U_i. Accordingly, as sample size T increases, decision-makers can improve their estimations for optimal arms' parameter vectors.

These first three assumptions are directly adopted from the multi-armed bandit literature and have been shown to be satisfied for all discrete distributions with finite support and a very large class of continuous distributions (see Bastani and Bayati 2015 for detailed examples and discussions).

A. 4 (Restricted eigenvalue condition) There exists $\kappa > 0$ such that for all feasible $\boldsymbol{\xi}$ satisfying $\|\xi\|_1 \leq b$ and $\boldsymbol{u}$ such that $\|\boldsymbol{u}^c_{\mathcal{S}_k}\|_1 \leq 3\|\boldsymbol{u}_{\mathcal{S}_k}\|_1$, we have $\frac{\kappa}{s}\|\boldsymbol{u}_{\mathcal{S}_k}\|_1^2 \leq \boldsymbol{u}^T\mathbb{E}[\nabla^2\mathcal{L}(\boldsymbol{\xi})]\boldsymbol{u}$, where $\mathcal{L}$ is the log likelihood function, $\mathcal{L}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{j=1}^{n} -\log g(r_j|\boldsymbol{x}_j^T\boldsymbol{\beta})$, and $\{\boldsymbol{x}_j, j = 1, 2, ..., n\}$ are iid random samples with $\boldsymbol{x}_j \in U_k, k \in \mathcal{K}$.

The restricted eigenvalue condition assumption is a standard assumption in high-dimensional statistics and is necessary for the identifiability and consistency of high-dimensional estimators (Fan et al. 2018, 2014b). This assumption considers the local geometry of the log likelihood function $\mathcal{L}$ with iid samples in U_k. To intuit, note that under low-dimensional settings, the literature (Montgomery et al. 2012) requires that $\mathcal{L}$ is strongly convex around the true parameter vector $\boldsymbol{\beta}^{true}$ (e.g., the Hessian matrix in OLS estimator is positive-definite and invertible) in order to achieve identifiability of the parameter vector. However, the strong convexity assumption is typically violated in high-dimensional settings, as the sample size can be much smaller than the covariate dimension. Therefore, a weaker condition is adopted: The $\mathcal{L}$ exhibits local strongly convex behavior only in some restricted subspace of u. In high-dimensional linear models, the restricted eigenvalue condition assumption is analogous to the compatibility condition (Bastani and Bayati 2015, Bühlmann and Van De Geer 2011), restrict strongly convexity condition (Negahban et al. 2009, Loh and Wainwright 2013), and sparse eigenvalue condition (Zhang et al. 2012, Fan et al. 2018).

A. 5 (Density function) The negative logarithm of the reward density function $f(r|y) \doteq -\log g(r|y)$ is (i) convex with smooth gradient and hessian in y, and (ii) there exists positive constants σ, σ_2 and σ_3 such that $|f^{'}(r|y)| \leq \sigma$, $f^{''}(r|y) < \sigma_2$ and $|f^{'''}(r|y)| \leq \sigma_3$.

The density function assumption enables us to use the estimated expected reward to statistically infer the true expected reward. Specifically, under this assumption, when the parameter estimator $\boldsymbol{\beta}$ is close enough to the underlying true parameter vector $\boldsymbol{\beta}^{true}$, the negative logarithm of the reward density function under the estimator $\boldsymbol{\beta}$, $g(\boldsymbol{x}^T\boldsymbol{\beta})$, will converge to that under the true parameter vector $\boldsymbol{\beta}^{true}$, $g(\boldsymbol{x}^T\boldsymbol{\beta}^{true})$. The density function assumption is a fairly weak technical assumption. Many common distributions, such as sub-Gaussian distribution and Bernoulli distribution, satisfy this density function assumption.

4. G-MCP-Bandit Algorithm

One of the major challenges for online learning and decision-making problems is discovering the underlying sparse data structure and estimating the parameter vector for high-dimensional data with limited samples. Lasso (Tibshirani 1996) has been proposed as an efficient statistical learning method and adopted in the multi-armed bandit literature (Bastani and Bayati 2015) to hurdle this challenge. However, the Lasso estimator can be biased and performs inadequately, especially when the magnitude of true parameters is not too small (Fan and Li 2001). One way to address this performance issue is to construct new penalty functions that could render unbiased estimators and improve the sparse structure discovery under high-dimensional data with limited samples. In this research, we will adopt the novel MCP method.

4.1. Parameter Vector Estimation

For notation convenience, we will omit parameters' subscripts corresponding to the choice of arms, as long as doing so will not cause any misinterpretation. Consider an oracle estimator for an arbitrary arm, $\boldsymbol{\beta}^{oracle}$, which is the parameter estimator when the decision-maker has perfect knowledge of the index set for significant covariates $\mathcal{S}$. In other words, the oracle estimator can be determined by setting $\beta_j = 0$ for $j \in \mathcal{S}^c$ and solving

$$\boldsymbol{\beta}^{oracle}(\boldsymbol{X}, \boldsymbol{r}) \doteq \arg \min_{\substack{\boldsymbol{\beta}_{\mathcal{S}^c}=0 \\ \boldsymbol{\beta}_{\mathcal{S}}}} \left\{ \frac{1}{|\mathcal{A}|} \sum_{j \in \mathcal{A}} f(r_j | \boldsymbol{x}_j^T \boldsymbol{\beta}) \right\}, \tag{1}$$

where $\mathcal{A}$ is the available historical data samples and $f(\cdot|\cdot)$ is the negative logarithm of the reward density function defined early. When solving for the oracle estimator, the decision-maker can directly ignore insignificant covariates by forcing their corresponding coefficients to be zero and essentially reduce the high-dimensional problem to a low-dimensional counterpart. The statistical performance of the oracle estimator is provided in the following lemma.

LEMMA 1. *Let n be the sample size. Under assumption A.1, A.4, and A.5, the following inequality for the oracle estimator holds*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{\frac{8s^2\sigma^2 x_{\max}^2}{\mu_0^2 n}} \right) \geq 1 - \delta_1(n), \tag{2}$$

where $\delta_1(n) \doteq 2\exp(-\frac{C_h n \mu_0}{2s x_{\max}^2}) + s\exp(-\frac{\mu_0 n}{8s\sigma_2 x_{\max}^2})$, and C_h and μ_0 are positive constants.

Since there are only $|\mathcal{S}|$ significant covariates, which is upper-bounded by s, are free to change in Equation (1), the optimal statistical performance of the likelihood estimation is commonly recognized as $O(\sqrt{s/n})$ in the literature (Fan et al. 2018, Zhao et al. 2018), which doesn't include the dependence of the largest eigenvalue in the objective function's Hessian matrix. In Equation (2), we explicitly include its influence and can directly verify that the largest eigenvalue in the objective

function's Hessian matrix is universally upper bounded by $\sigma_2 s x_{\max}^2$ and therefore Equation (2) reduces to $O(\sqrt{s/n})$ dependence. In other words, the oracle estimator attains the optimal statistical performance.

However, the significant covariates index set $\mathcal{S}$ is typically unknown to the decision-maker in practice, and we will rely on the MCP method to recover this latent sparse structure. To better understand the rationale behind the MCP method, we start with the following weighted Lasso estimator:

$$\boldsymbol{\beta}^W(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{w}) \doteq \arg\min_{\beta} \left\{ \frac{1}{|\mathcal{A}|} \sum_{j \in \mathcal{A}} f(r_j | \boldsymbol{x}_j^T \boldsymbol{\beta}) + \sum_{i=1}^{d} w_i |\beta_i| \right\}, \tag{3}$$

where $\boldsymbol{w} = (w_1, w_2, ..., w_d)$ is a positive weights vector chosen by the decision-maker. Note that when we set $w_i = \lambda$ for all i, $\boldsymbol{\beta}^W(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{w})$ reduces to the standard Lasso estimator, which can be biased when the magnitude of true parameters is not too small. To recover the sparse structure and provide an unbiased parameter estimator, an ideal way to select $\{w_i\}$ is to set $w_i = \lambda > 0$ for all $i \in \mathcal{S}^c$ and $w_j = 0$ for all $j \in \mathcal{S}$. By doing so, when the weight λ is large enough, the weighted Lasso estimator converges to the oracle estimator $\boldsymbol{\beta}^{oracle}(\boldsymbol{X}, \boldsymbol{r})$. The benefits of the weighted Lasso method have attracted considerable attention recently, and various mechanisms have been proposed in the literature aiming to improve the weight selection process (Zou 2006, Huang et al. 2008, Candes et al. 2008). The MCP method, adopted in our paper, reflect such a process.

In particular, we define the following MCP penalty function:

$$P_{\lambda,a}(x) \doteq \int_0^{|x|} \max\left(0, \lambda - \frac{1}{a}|t|\right) dt,$$

where a and λ are positive parameters selected by the decision-maker, and the MCP estimator can be presented as follows:

$$\boldsymbol{\beta}^{MCP}(\boldsymbol{X}, \boldsymbol{r}, \lambda) \doteq \arg\min_{\beta} \mathcal{L}_{\mathcal{A}_k}(\boldsymbol{\beta}) = \arg\min_{\beta} \left\{ \frac{1}{|\mathcal{A}|} \sum_{j \in \mathcal{A}} f(r_j | \boldsymbol{x}_j^T \boldsymbol{\beta}) + \sum_{i=1}^{d} P_{\lambda,a}(\beta_i) \right\}. \tag{4}$$

Denote the index set for non-zero coefficients solutions in Equation (4) as $\mathcal{J} \doteq \{j : \hat{\beta}_j \neq 0\}$. If the absolute value of the MCP estimator in $\mathcal{J}$ is greater than $a\lambda$, then $P_{\lambda,a}(\beta_j)$ become constant parameters for all $j \in \mathcal{J}$. Therefore, we will have $P_{\lambda,a}(\beta_j) = \frac{1}{2}a\lambda^2$ for $j \in \mathcal{J}$ and $P_{\lambda,a}(\beta_j) = 0$ otherwise. In other words, the statistical performance of solving the MCP estimator is equivalent to solving the following problem: $\arg\min_{\boldsymbol{\beta}_{\mathcal{J}^c}=0, \boldsymbol{\beta}_{\mathcal{J}}} \{\frac{1}{|\mathcal{A}|} \sum_{j \in \mathcal{A}} f(r_j | \boldsymbol{x}_j^T \boldsymbol{\beta})^2\}$. Hence, if $\mathcal{J} = \mathcal{S}$, then the MCP estimator converges to the oracle estimator.

Solving the MCP estimator can be challenging. Liu et al. (2017) have shown that it is an NP-complete problem to find the MCP estimator by globally solving Equation (4). In the next subsection, we propose a local linear approximation method, the 2-step Weighted Lasso (2sWL) procedure, to tackle this challenge, and we demonstrate that the estimator solved by the 2sWL procedure will match the oracle estimator $\boldsymbol{\beta}^{oracle}$ with high probability.

4.2. 2-Step Weighted Lasso Procedure

The 2sWL procedure consists of two steps. We first solve a standard Lasso problem by setting all positive weights in Equation (3) to a given parameter λ_0. Then, we use the Lasso estimator obtained in the first step to update the weights vector $\boldsymbol{w}$ by taking the first-order derivatives of the MCP penalty function, and applying this updated weight vector, we re-solve the weighted Lasso problem in Equation (3) to obtain the MCP estimator. The procedures of 2sWL at time t can be described as follows:

2-Step Weighted Lasso (2sWL) Procedure:

Require: input parameters a and λ

Step 1: solve a standard Lasso problem

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}^W(\boldsymbol{X}, \boldsymbol{y}, \lambda);$$

Step 2: update $w_j = \begin{cases} P'_{a,\lambda}(|\beta_{1,j}|) & , \text{for } \beta_{1,j} \neq 0 \\ \lambda & , \text{for } \beta_{1,j} = 0 \end{cases}$

and solve a weighted Lasso Problem

$$\hat{\boldsymbol{\beta}}_{2sWL} = \boldsymbol{\beta}^W(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{w}).$$

As the 2sWL procedure is equivalent to solving the Lasso problem twice, the worst-case computation complexity for 2sWL is on same order as for the standard Lasso problem. In practice, we can initialize the second step procedure with a warm start from the first step of the Lasso solution, which further reduces the computation time.

The following proposition shows that the MCP estimator identified by the 2sWL procedure can recover the oracle estimator with high probability.

PROPOSITION 1. *Under assumptions A.1, A.4, and A.5, if* $\min\{|\beta_j^{true}|, \beta^{true} \neq 0, j = 1, 2, ..., d\} \geq \left(\frac{96s}{\kappa} + a\right)\lambda$, $a > \frac{96s}{\kappa}$, *the MCP estimator solved under the 2sWL procedure,* $\boldsymbol{\beta}^{MCP}$ *satisfies the following inequality*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{\frac{8s^2\sigma^2 x_{\max}^2}{\mu_0^2 n}}\right) \geq 1 - \delta_1(n) - \delta_2(n, n, \lambda) - \delta_3(n), \tag{5}$$

where $\delta_2(n, n_1, \lambda) \doteq d \exp\left(-\frac{n\lambda^2}{2x_{\max}^2}\left(\left(\frac{1}{4} - \frac{24ns}{n_1 \kappa a}\right) \min\left\{1, \frac{n\mu_0}{8n_1 s x_{\max}^2}\right\}\right)^2\right)$, $\delta_3(n) \doteq \exp(-C_1 n)$, μ_0 *and* C_1 *are positive constants.*

Comparing to the oracle estimator β^{oracle} in Lemma 1, the probability bound on the MCP estimator under the 2sWL procedure has two extra terms $\delta_2(n, n, \lambda)$ and $\delta_3(n)$, which depend on the covariate dimension d and the sample size n. Note that as the sample size increases, these two extra terms decrease to 0 at exponential rates. In other words, as the sample size increases, $\boldsymbol{\beta}^{MCP}$ matches the oracle parameters with high probability and converges to the true parameters at the optimal convergence rate.

4.3. ϵ-decay Random Sampling Method

As bandit models involve exploitation and exploration, samples generated under exploitation typically are not iid. These non-iid samples pose challenges to the existing MCP literature, which relies on the assumption that samples are iid in establishing the convergence rate and regret bounds (see the proof of Proposition 1 in §4.2).

In this research, to ensure that there are some iid samples generated in the online learning and decision-making process, we propose a ϵ-decay random sampling method, in which the decision-maker draws random samples, with decreasing probability, by randomly selecting decisions from the decision set with equal probability. In particular, the ϵ-decay random sampling method can be described as follows:

ϵ-decay Random Sampling Method: At time t, the decision-maker will draw a random sample, with probability $\min\{1, t_0/t\}$, where t_0 is a pre-determined positive constant. If the seller has decided to draw a random sample at time t, then the decision-maker will randomly select a decision from his decision set with equal probability. Otherwise, the decision-maker will follow a bi-level decision structure, which will be specified later, to determine the optimal decision to maximize his expected reward.

The ϵ-decay random sampling method can balance the exploitation and exploration trade-off by ensuring that the decision-maker does not explore too much to significantly sacrifice his revenue performance (as the number of random samples decays in time) but has sufficient random samples to guarantee the quality of the parameter vector estimation. In particular, we can bound the random sample size in the following proposition.

PROPOSITION 2. *Let $C_0 \geq 10$, $T > \frac{(t_0+1)^2}{e^2}$, and $t_0 = 2C_0|\mathcal{K}|$. Under the ϵ-decay random sampling method, the random sample size n_k for arm $k \in \mathcal{K}$ up to time T is bounded by*

$$C_0(1+\log(T+1)-\log(t_0+1)) \leq n_k \leq 3C_0(1+\log(T)-\log(t_0))$$

with probability at least $1-2/(T+1)$.

4.4. G-MCP-Bandit Algorithm

After establishing the MCP estimator's statistical property and the ϵ-decay random sampling method, we are ready to present the proposed G-MCP-Bandit algorithm. The execution of the G-MCP-Bandit algorithm can be summarized as follows:

G-MCP-Bandit Algorithm

Require: Input parameters $t_0, h, \lambda_{1,0}, \lambda_{2,0}, a$.
Initialize $\boldsymbol{\beta}_i^{random}(0) = \boldsymbol{\beta}_i^{whole}(0) = \mathbf{0}$, and $\mathcal{R}_{\pi_0} = \mathcal{W}_{\pi_0} = \phi$ for all $i \in \mathcal{K}$.
For $t = 1, 2, \ldots.$ **do**
Observe $\boldsymbol{x}_t$.
Draw a binary random variable $\mathcal{D}_t$, where $\mathcal{D}_t = 1$ with probability $\min\{1, t_0/t\}$.
If $\mathcal{D}_t = 1$
Assign π_t to a random decision $k \in \mathcal{K}$ with probability $\mathbb{P}(\pi_t = k) = 1/|\mathcal{K}|$.
Play decision π_t, observe r_t, and update $\mathcal{R}_{\pi_t} = \mathcal{R}_{\pi_{t-1}} \cup \{\boldsymbol{x}_t, r_t\}$ and $\mathcal{W}_{\pi_t} = \mathcal{W}_{\pi_{t-1}} \cup \{\boldsymbol{x}_t, r_t\}$.
Else
Construct the optimal decision set:
$\Pi_t = \left\{ i : \mathbb{E}[R_i|\boldsymbol{x}_t, \boldsymbol{\beta}_i^{random}(t-1)] \geq \max_{j \in \mathcal{K}} \mathbb{E}[R_j|\boldsymbol{x}_t, \boldsymbol{\beta}_j^{random}(t-1)] - \frac{1}{2}h,\ i \in \mathcal{K} \right\}$.
If Π_t is a singleton
Set $\pi_t = \Pi_t$.
Else
Set $\pi_t = \arg\max_{k \in \Pi_t} \mathbb{E}[R_k|\boldsymbol{x}_t, \boldsymbol{\beta}_k^{whole}(t-1)]$.
End If
Play decision π_t, observe r_t, and update $\mathcal{W}_{\pi_t} = \mathcal{W}_{\pi_{t-1}} \cup \{\boldsymbol{x}_t, r_t\}$.
End If
For all $k \in \mathcal{K}$, set $\lambda_1(t) = \lambda_{1,0}\sqrt{1 + \frac{\log d}{\log(t+1)}}$ and $\lambda_2(t) = \lambda_{2,0}\sqrt{\frac{\log(t+1)+\log d}{t+1}}$.
Update parameters $\boldsymbol{\beta}_k^{random}(t)$ via the 2sWL procedure with $(\mathcal{R}_{\pi_t}, \lambda_1(t))$.
Update parameters $\boldsymbol{\beta}_k^{whole}(t)$ via the 2sWL procedure with $(\mathcal{W}_{\pi_t}, \lambda_2(t))$.
End for

Specifically, the decision-maker will start by assigning values for system parameters (t_0, $\mathcal{K}$, $s_{\max}$, and h), which can be optimized through tuning, and initialing two parameter vector estimators ($\boldsymbol{\beta}^{random}$ and $\boldsymbol{\beta}^{whole}$) and two sample datasets ($\mathcal{R}_{\pi_0}$ and $\mathcal{W}_{\pi_0}$, which represent the random sample set and the whole sample set, respectively). Then, for an incoming user at time t, the decision-maker will draw a random sample with probability $\min\{1, t_0/t\}$. There are two possibilities:

- If the decision-maker decides to draw a random sample, then he will randomly choose a decision k from his decision set $\mathcal{K}$ with equal probability of $1/|\mathcal{K}|$; then, he will implement the chosen decision (i.e., $\pi_t = k$), observe the user's response, and claim the corresponding reward; finally, the decision-maker will include the user's covariate vector and the corresponding reward $\{\boldsymbol{x}_t, r_t\}$ in both sample datasets, $\mathcal{R}_{\pi_t}$ and $\mathcal{W}_{\pi_t}$.

- If the decision-maker decides not to draw a random sample on this incoming user, then he will use the bi-level decision structure to determine his decision. In the upper-level decision-making process, the decision-maker will first construct an optimal decision set Π_t. Specifically, all decisions in the optimal decision set Π_t are estimated, based on the random sample MCP estimator $\boldsymbol{\beta}^{random}$, to yield expected rewards within $h/2$ of the maximum possible reward. If there is only one decision in the optimal decision set Π_t, then the decision-maker will implement this decision as the optimal decision; otherwise, the decision-maker will perform the lower-level decision-making process, in which the decision-maker will estimate, by using the whole sample MCP estimator $\boldsymbol{\beta}^{whole}$, the

rewards for all decisions in the optimal decision set Π_t and select the decision that generates the highest expected reward. Then, observing the user's response to the optimal decision and collecting the corresponding reward, the decision-maker will only update the whole sample dataset $\mathcal{W}_{\pi_t}$ by appending the user's covariate vector and the corresponding reward $\{\boldsymbol{x}_t, r_t\}$.

Finally, the decision-maker will reset two parameters, λ_1 and λ_2, and use the 2sWL procedure to update the random sample parameter vector estimator $\boldsymbol{\beta}^{random}$ and the whole sample parameter vector estimator $\boldsymbol{\beta}^{whole}$, based on sample data sets $\mathcal{R}_{\pi_t}$ and $\mathcal{W}_{\pi_t}$, respectively.

The expected cumulative regret upper bound for the G-MCP-Bandit algorithm can be established in the following theorem.

THEOREM 1. *Under assumptions A.1-A.5, let* $t_0 = 2C_0|\mathcal{K}|$, $T \geq T_0$, $\lambda_{1,0} = \frac{\beta_{\min} p^* \kappa}{(2304s + ap^*\kappa)\sqrt{1+\log d}}$, $\lambda_{2,0} = \frac{\sqrt{2}x_{\max}^2}{\frac{1}{4} - \frac{192}{p^*\kappa a}\min\{1, \frac{\mu_0}{p^* s x_{\max}^2}\}}$, *and* $a \geq \frac{2304s}{\kappa p^*}$. *The cumulative regret of the G-MCP-Bandit algorithm up to time* T *is upper bounded:*

$$\begin{aligned} R^C(T) &\leq R_{\max}(T_0 + |\mathcal{K}|) + (6R_{\max}|\mathcal{K}|C_0 + 31R_{\max}|\mathcal{K}| + 2e^{4\sigma x_{\max} b} C R_{\max}^3 |\mathcal{K}| x_{\max}^2 C_{\boldsymbol{\beta}} s^3) \log(T+1) \\ &= O(|\mathcal{K}| s^2 (s + \log d) \log T), \end{aligned}$$

where T_0, C_0, C_h, μ_0 *and* C_β *are constants independent of* T.

Theorem 1 shows that the expected cumulative regret of the G-MCP-Bandit algorithm over T users is upper-bounded by $O(\log T)$. Goldenshluger and Zeevi (2013) have shown that under low-dimensional settings, the expected cumulative regret for a linear bandit model is lower-bounded by $O(\log T)$, which is directly applicable to the high-dimensional settings. Further, note that the linear model is a special case of the generalized linear model. Therefore, the expected cumulative regret of the G-MCP-Bandit algorithm is also lower-bounded by $O(\log T)$. In other words, the G-MCP-Bandit algorithm achieves the optimal expected cumulative regret in the sample size dimension. This result comes from the facts that we can ensure $O(\log T)$ random samples at time T via the ϵ-decay random sampling method (Proposition 2) and that the MCP estimator is able to match the oracle estimator with high probability (Proposition 1). Further, when compared to the Lasso-Bandit algorithm proposed by Bastani and Bayati (2015) for the linear model under high-dimensional settings, the G-MCP-Bandit algorithm reduces the dependence of the expected cumulative regret on the sample size dimension from $O(\log^2 T)$ to $O(\log T)$. As the G-MCP-Bandit algorithm achieves the optimal expected cumulative regret and improves on the cumulative regret performance from existing high-dimensional bandit algorithms in the sample size dimension, we expect that the G-MCP-Bandit algorithm will be able to improve the learning process of the parameter vector estimation with limited samples and perform favorably in the cumulative regret performance even in sample-poor regions.

Theorem 1 also demonstrates that the cumulative regret of the G-MCP-Bandit algorithm in the high-dimensional covariate vector d is upper-bounded by $O(\log d)$. This bound presents a significant improvement over other classic bandit algorithms (Goldenshluger and Zeevi 2013, Abbasi-Yadkori and Szepesvari 2012, Dani et al. 2008), which yield polynomial dependence on d, and is also a tighter bound than the Lasso-type algorithm (i.e., $O(\log^2 d)$ in Bastani and Bayati 2015). This improvement is of particular importance in high-dimensional settings, in which the covariate dimension can be extremely large, and it suggests that the G-MCP-Bandit algorithm can bring substantial regret reduction comparing to existing bandit algorithms, which we will illustrate through experiments in §6.

5. Key Steps of Regret Analysis for the G-MCP-Bandit Algorithm

In this section, we provide the abridged technical proofs for Theorem 1 – the main theorem in this paper. Specifically, we briefly lay out four key steps in establishing the expected cumulative regret upper bound for the G-MCP-Bandit algorithm. In the first step, we highlight the influence of non-iid data, inherited from the multi-armed bandit model, and provide the statistical convergence property for the MCP estimator under partially iid samples. Applying these results to the G-MCP-Bandit algorithm, in the second and third steps, we establish the convergence properties for both the random sample estimator, which is based on samples generated only through the ϵ-decay random sampling method, and the whole sample estimator, which uses all available samples. Finally, in the last step, we establish the total expected cumulative regret by separating the regret up to time T into three segments and providing a bound for each segment. The main structure and sequence of our proving steps described above are first introduced by Bastani and Bayati (2015), which presents their expected regret analysis for a linear bandit model (i.e., LASSO-Bandit algorithm) in a similar sequence. We will largely follow their presentation structure, but with different steps, proving techniques, and convergence properties, to illustrate the key steps in analyzing the G-MCP-Bandit algorithm.

5.1. General Non-iid Sample Estimator

Note that the restricted eigenvalue condition (A.4 in §3) for high-dimensional statistics is typically established for iid samples in the literature. Yet, in this research, we consider the G-MCP-Bandit algorithm, under which only part of the samples are iid, so we first show that the restricted eigenvalue condition continues to hold for partially iid samples (Lemma EC.6 in E-Companion). Then, we can establish some general results for the MCP estimator under non-iid data.

We denote $\mathcal{W}$ as the whole sample set that contains all users' covariate vectors $\boldsymbol{X}$ and the corresponding rewards $\boldsymbol{r}$ for an arbitrary decision $k \in \mathcal{K}$ up to time T, and $\boldsymbol{\beta}^{MCP}$ as the MCP estimator for the parameter vector corresponding to decision k. Note that as samples in $\mathcal{W}$ are not

iid, standard MCP convergence results (Fan et al. 2014b, 2018) cannot be directly applied. Recall that we proposed the ϵ-decay random sampling method and that samples generated under this method are iid. Therefore, there exists a subset $\mathcal{A} \subseteq \mathcal{W}$ such that all samples in this subset are iid from the distribution $\mathcal{P}_{\boldsymbol{X}}$. The next step is to show that when the cardinality of $\mathcal{A}$ (i.e., $|\mathcal{A}|$) is large enough, $\boldsymbol{\beta}^{MCP}$ will converge to the true parameters $\boldsymbol{\beta}^{true}$.

PROPOSITION 3. *Denote the whole sample size as n and the sub-sample set, containing only iid random samples, as $\mathcal{A}$. Under assumptions A.1, A.4, and A.5, if $\beta_{\min} \geq (\frac{96ns}{\kappa|\mathcal{A}|} + a)\lambda$ and $a > \frac{96ns}{\kappa|\mathcal{A}|}$, then for $\zeta \leq \frac{\mu_0|\mathcal{A}|\sqrt{C_2\lambda}}{2n}$, the following inequality hold for the MCP estimator under the 2sWL procedure $\boldsymbol{\beta}^{MCP}$*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \leq \frac{2n\zeta}{|\mathcal{A}|\mu_0}\right) \geq 1 - \delta_2(n, |\mathcal{A}|, \lambda) - \delta_3(|\mathcal{A}|) - \delta_4(n, |\mathcal{A}|, \zeta). \tag{6}$$

Moreover, if $|\mathcal{A}| \geq \frac{2s^2x_{\max}^2}{\mu_0}$, then we have the following result

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{\frac{8s^2\sigma^2x_{\max}^2 n}{\mu_0^2|\mathcal{A}|^2}}\right) \geq 1 - \delta_1(|\mathcal{A}|) - \delta_2(n, |\mathcal{A}|, \lambda) - \delta_3(|\mathcal{A}|), \tag{7}$$

where C_2 and μ_0 are positive constants and $\delta_4(n, |\mathcal{A}|, \zeta) \doteq s\exp\left(-\frac{|\mathcal{A}|\mu_0}{8\sigma_2 s x_{\max}^2}\right) + s\exp\left(-\frac{n\zeta^2}{2\sigma^2x_{\max}^2}\right)$.

Proposition 3 describes the statistical properties of the non-iid MCP estimators under the 2sWL procedure. First, if we don't require the iid sample size $|\mathcal{A}|$ to be sufficiently large, then the MCP estimator's statistical performance is given by Equation (6). If we set ζ to be on the order of $O(s/\sqrt{n})$, then $\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|$ is on the order of $O(\sqrt{s^2n/|\mathcal{A}|^2})$, which matches the result of Equation (7). Meanwhile, however, $\delta_4(n, |\mathcal{A}|, \zeta)$ in Equation (6) becomes a positive constant asymptotically, which implies that when $|\mathcal{A}|$ is not large enough, the MCP estimator may not warrant good statistical performance. Yet, when we have sufficient iid samples (i.e., $|\mathcal{A}| \geq \frac{2s^2x_{\max}^2}{\mu_0}$), Equation (7) suggests that the MCP estimator not only guarantees a better statistical convergence ($O(\sqrt{s^2n/|\mathcal{A}|^2})$) but also attains probability 1 when the whole sample size n and the iid sample size $|\mathcal{A}|$ go to infinity.

Moreover, Proposition 3 shows the necessity of generating iid random samples in high-dimension bandit settings. Non-iid samples are inevitable in online learning and decision-making process, so ensuring desired asymptotical performance of the parameter vector estimation in high-dimensional settings can only be achieved through generating sufficient number of iid samples, as shown in Proposition 3. We will show in next two subsections that the size of iid samples generated under the ϵ-decay random sampling method is on the order of $O(\log T)$ and that the size can be further improved to the order of $O(T)$ under the bi-level decision structure in the G-MCP-Bandit algorithm.

5.2. Estimator from Random Samples up to Time T

In Proposition 3, we show that the MCP estimator will converge to the oracle parameter as long as the sample set contains a sufficient number of iid samples. Recall that in our proposed G-MCP-Bandit algorithm, samples generated by the ϵ-decay random sampling method are iid, and the size of these iid samples is on the order of $O(\log(T))$; see Proposition 2. Combining these observations, we can establish the statistical performance of the MCP estimator under the G-MCP-Bandit algorithm in the following proposition.

PROPOSITION 4. *Let* $t_0 = 2C_0|\mathcal{K}|$, $T \geq \max\{(t_0+1)^2/e^2 - 1, e\}$, $\lambda = C_5\sqrt{1+\log d/\log(T+1)}$ *and* $a > 2304s/p^*\kappa$. *If assumptions A.1, A.3, A.4, and A.5 hold, then the MCP estimator under the G-MCP-Bandit algorithm* $\boldsymbol{\beta}^{MCP}$ *will satisfy the following inequality*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_1 \leq \min\left\{\frac{1}{\sigma x_{\max}}, \frac{h}{4e\sigma R_{\max} x_{\max}}\right\}\right) \geq 1 - \frac{7}{T+1},$$

where C_0 *and* C_5 *are positive constants.*

5.3. Estimator from Whole Samples up to Time T

In addition to the iid samples generated by the ϵ-decay random sampling method, other samples can also be iid and used to improve the statistical performance of the MCP estimator. To intuit, recall that in the G-MCP-Bandit algorithm, when the user is not selected to perform a random sampling, the decision-maker will use the bi-level structure to determine the optimal decision to maximize his expected reward. In the upper-level decision-making process, only iid samples will be used (as $\boldsymbol{\beta}^{random}$ is the MCP estimator based on samples generated only by the ϵ-decay random sampling method) to determine the candidate(s) for the optimal decision set. From Proposition 4, we know that this random sample MCP estimator will not be far away from its true parameter values. In other words, if we define the event that the random sample MCP estimator at time t is within a given distance from its true parameter as event $\mathcal{E}_6$:

$$\mathcal{E}_6 \doteq \left\{\|\boldsymbol{\beta}_k^{random}(t) - \boldsymbol{\beta}_k^{true}\|_1 \leq \min\left\{\frac{1}{\sigma x_{\max}}, \frac{h}{4e\sigma R_{\max} x_{\max}}\right\},\ k \in \mathcal{K}\right\}, \tag{8}$$

then event $\mathcal{E}_6$ will happen with high probability. Further, conditioning on event $\mathcal{E}_6$, we can directly verify that for any $\boldsymbol{x} \in U_k$, $k \in \mathcal{K}$, the following inequality holds:

$$\mathbb{E}(R_k|\boldsymbol{x}, \boldsymbol{\beta}_k^{random}(t)) \geq \max_{j \neq k} \mathbb{E}(R_j|\boldsymbol{x}, \boldsymbol{\beta}_j^{random}(t)) + \frac{h}{2}. \tag{9}$$

Therefore, if using Equation (9) as the selecting criterion, the decision-maker will be able to choose the optimal decision k for any $x \in U_k, k \in \mathcal{K}$ with high probability. Formally, we can bound the total number of times under which event $x \in U_k$ and event $\mathcal{E}_6$ happen simultaneously. In particular,

we define $M(i) \doteq \mathbb{E}\left[\sum_{j=1}^{T+1} \mathbb{1}(\boldsymbol{x}_j \in U_k, \mathcal{E}_6, x_j \notin \mathcal{R}_k)|\mathcal{F}_i\right]$ for $i \in \{0,1,2,..,T+1\}$, where $\mathcal{F}_i = \{(\boldsymbol{x}_j, r_j)$ for $j \leq i\}$ and $\mathcal{R}_k$ is the set containing iid samples generated through the ϵ-decay random sampling method for arm k. Then, $\{M(i)\}$ is a martingale with bounded difference $|M(i) - M(i+1)| \leq 1$ for $i = 0,1,2,...,T$, and we can bound the value of $M(T+1)$ in the following proposition:

PROPOSITION 5. *If* $T \geq \max\{14, 4C_0|\mathcal{K}|\}$, *then* $\mathbb{P}\left(M(T+1) \leq \frac{p^*(T+1)}{8}\right) \leq \exp\left(-\frac{(p^*)^2 T}{128}\right)$.

Intuitively, Proposition 5 suggests that with high probability, the actual iid sample size in U_k for decision k will be on the order of $O(T)$ instead of $O(\log T)$. This improvement is the reason why the whole sample MCP estimator $\boldsymbol{\beta}^{whole}$ used in the lower-level decision-making process has a better statistical performance, compared to the random sample MCP estimator $\boldsymbol{\beta}^{random}$ used in the upper-level decision-making process. Specifically, we can establish the convergence property for the whole sample MCP estimator in the following proposition.

PROPOSITION 6. *Let* $t_0 = 2C_0|\mathcal{K}|$, $T > T_0$, $\lambda = C_4\sqrt{\frac{\log(T+1)+1+\log d}{T+1}}$, *and* $a > \frac{2304s}{p^*\kappa}$. *If assumptions A.1, A.3, A.4, and A.5 hold, then at time T the whole sample MCP estimator under the G-MCP-Bandit algorithm* $\boldsymbol{\beta}^{whole}$ *will satisfy the following inequality:*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{whole}(T) - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{C_{\boldsymbol{\beta}}\frac{s^2}{T+1}}\right) \geq 1 - \frac{12}{T+1},$$

where C_0, T_0, C_4, *and* C_β *are positive constants.*

5.4. Cumulated Regret Up To Time T

Finally, to bound the cumulative regret for the G-MCP-Bandit algorithm, we need to divide the time, up to time T, into three groups and provide a upper bound for each group.

The first group contains all samples before time T_0 and all random samples up to time T. Note that before time T_0 (the explicit expression for T_0 is given in the proof of Theorem 1 in E-Companion), the decision-maker does not have sufficient samples to accurately estimate covariate parameter vectors. Hence, the reward under the G-MCP-Bandit algorithm will suffer and be sub-optimal compared to that of the oracle case. We can bound the cumulative regret by the worst case performance: $R_{\max}T_0 + R_{\max}|\mathcal{K}|(2 + 6C_0 \log T)$, where the first part of this cumulative regret is for all samples before time T_0 and the second part is for all random samples up to time T.

Next, we will segment the $t > T_0$ case into two groups, depending on whether we can accurately estimate covariate parameter vectors by using only random samples. In particular, the second group includes cases where $t > T_0$ and the random-sample-based estimators are not accurate (i.e., event $\mathcal{E}_6$ doesn't hold). Under those scenarios, inevitably, the decision-maker's decisions will be suboptimal with high probability. However, note that as the size of iid samples increases in t, the

probability of event $\mathcal{E}_6$ not occurring decreases. We can bound the cumulative regret for the second group by $7R_{\max}|\mathcal{K}|\log(T+1)$.

The last group includes scenarios where $t > T_0$ and the random sample estimators are accurate enough. Benefiting from the improved estimation accuracy (Proposition 6), we can bound the cumulative regret for the last group by $(24R_{\max}|\mathcal{K}| + 4e^{4\sigma^2 x_{\max} b} C R_{\max}^3 |\mathcal{K}| x_{\max}^2 C_\beta s^3)\log(T)$. Combining the cumulative regret for all three groups, Theorem 1 directly follows.

6. Empirical Experiments

In this section, we will benchmark the G-MCP-Bandit algorithm to OFUL (Abbasi-Yadkori et al. 2011), OLS-Bandit (Goldenshluger and Zeevi 2013), and Lasso-Bandit (Bastani and Bayati 2015). In particular, we seek answers to the following two questions: How does the performance of the G-MCP-Bandit algorithm compare to other bandit algorithms? And how is the performance of the G-MCP-Bandit algorithm influenced by the data availability (T), the data dimensions (s and d), and the size of the decision set (K)?

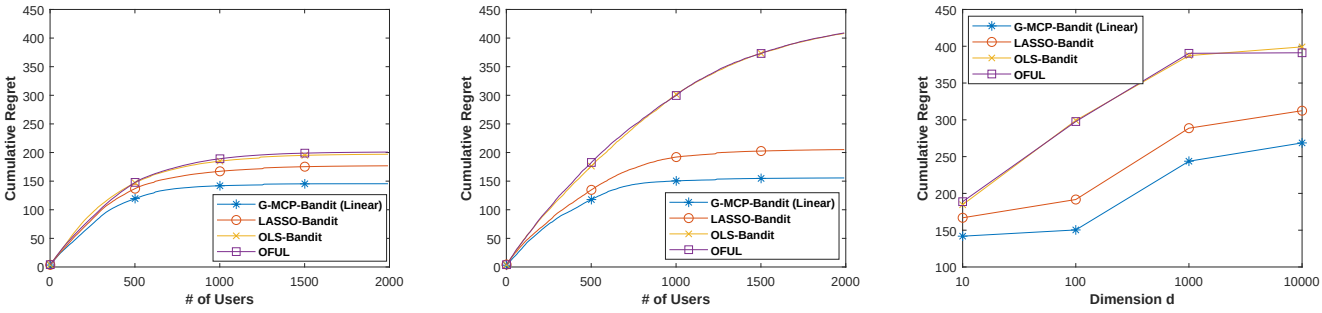
To this end, we start with two synthetic-data-based experiments in §6.1 and conduct two additional experiments based on real datasets, the warfarin dosing patient data in §6.2 and the Tencent search advertising data in §6.3, respectively. Note that the algorithms and theoretical bounds of OFUL, OLS-Bandit, and Lasso-Bandit are developed under the assumption that the reward function follows the linear model, which is a special case in the G-MCP-Bandit algorithm. Therefore, for fair comparison, we specify the underlying reward function for the G-MCP-Bandit algorithm to follow the same linear model (i.e., the reward under decision k for a user with covariate vector $\boldsymbol{x}$ takes the form of $R_k(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\beta}_k^{true} + \epsilon$, where ϵ is a σ-gaussian random variable) in all experiments, except the Tencent search advertising data experiment, in which we explore the performance of the G-MCP-Bandit model under both the linear model and the logistic model.

6.1. Synthetic Data (Linear Model)

In the first synthetic data experiment, we fix the size of the decision set K and focus on the impacts of the data dimensions, s and d, and the data availability, T, on learning algorithms' cumulative regret performance. In particular, we consider a two-arm bandit setting (i.e., $K = 2$). To simulate different sparsity levels, we vary the covariate dimension $d = \{10, 10^2, 10^3, 10^4\}$ and keep the dimension for significant covariates unchanged at $s = 5$. Therefore, as the covariate dimension d increases, the data become sparser. The underlying true parameter vectors for covariates are arbitrarily set to be $\boldsymbol{\beta}_1 = (1, 2, 3, 4, 5, 0, 0, ...)$ for the first arm and $\boldsymbol{\beta}_2 = 1.1 \cdot \boldsymbol{\beta}_1$ for the second arm. For each incoming user, we randomly draw her covariate vector from $N(0, I_{d\times d})$ and the error term in the linear model ϵ from $N(0, 1)$. Finally, we use the same parameter λ value in both the Lasso-Bandit algorithm and the G-MCP-Bandit algorithm and select the unique parameter for

the G-MCP-Bandit algorithm a at 2. For each algorithm, we perform 100 trials and report the average cumulative regret for OFUL, OLS-Bandit, Lasso-Bandit, and G-MCP-Bandit (under the linear model) in Figure 1.

Figure 1 Synthetic study 1: The impact of T and d on the cumulative regret, where $K = 2$ and $s = 5$.



(a) d=10 (b) d=100 (c) T=1000

Overall, we observe that the G-MCP-Bandit algorithm significantly outperforms OFUL, OLS-Bandit, and Lasso-Bandit and achieves the lowest cumulative regret. Facing only two decisions/arms, the decision-maker can easily identify the optimal arm, and therefore OFUL and OLS-Bandit, both of which are not specifically designed for high-dimensional settings, perform nearly identically. Lasso-Bandit and G-MCP-Bandit could benefit from their abilities to recover the sparse structure and identify the significant covariates. Therefore, compared to OFUL and OLS-Bandit, Lasso-Bandit and G-MCP-Bandit can improve their parameters estimations, especially under high-dimensional settings, and perform substantially better. Further, the improvement of the cumulative regret performance of G-MCP-Bandit over Lasso-Bandit follows from the facts that the MCP estimator is unbiased and could improve the sparse structure discovery. Next, we will discuss the influence of sample size T and the covariate dimension d on these algorithms' cumulative regret performance.

Figure 1(a) and 1(b) illustrate the influence of the sample size T on the cumulative regret for the cases where $d = 10$ and $d = 100$ (other cases exhibit a similar pattern and are therefore omitted)[2]. As we have proven that G-MCP-Bandit provides the optimal time dependence under both low-dimensional and high-dimensional settings (Theorem 1), G-MCP-bandit strictly improves on the cumulative regret performance from Lasso-Bandit, especially when T is not too small. Note that facing insufficient samples, all algorithms fail to accurately learn parameter vectors and

[2] In all four experiments where $d \in \{10, 10^2, 10^3, 10^4\}$, we simulated the sample size up to $10,000$ and observe that the G-MCP-Bandit algorithm's cumulative regret seems to be stabilized before $T = 2000$. Therefore, we only plot for the first 2000 samples to avoid duplications.

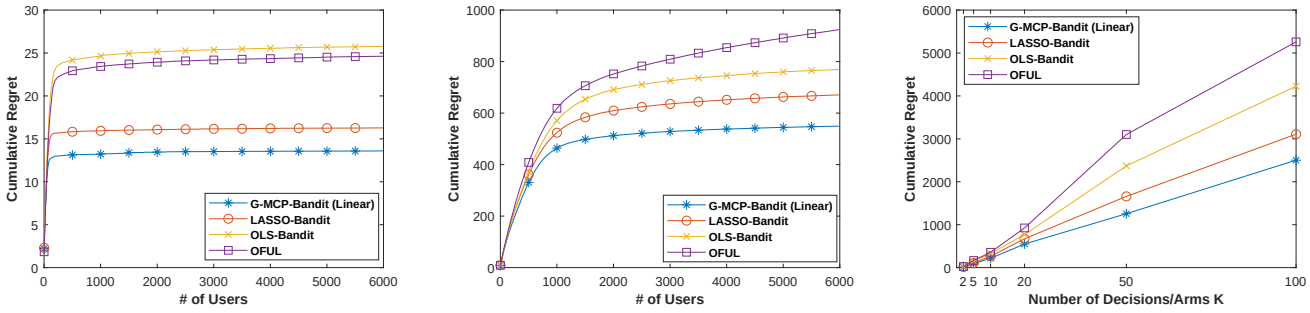
therefore perform poorly. As the sample size increases, the G-MCP-bandit algorithm is able to, in an expeditious fashion, unveil the underlying sparse data structure, accurately estimate parameter vectors, and outperform all other benchmarks. For example, in Figure 1(b), we observe that the regret reduction of G-MCP-Bandit over all other algorithms is larger than 10% when the sample size T is larger than 350. This observation echoes our theoretical findings that the G-MCP-Bandit algorithm attains the optimal regret bound in sample size dimension $O(\log T)$.

We also observe that the benefits of G-MCP-Bandit over other three algorithms appear to increase in the data sparsity level. Figure 1(c) presents the influence of the covariate dimension d on the cumulative regret for the case where $T = 1000$. Recall that we fixed the dimension for significant covariates $s = 5$. Therefore, as the covariate dimension d increases, the data become sparser (i.e., d/s increases). As expected, the cumulative regret for all four algorithms increases in the covariate dimension d, but at different rates. On the one hand, both OLS-Bandit and OFUL lack the ability to recover the sparse data structure and are ill suited for high-dimensional problems. On the other hand, Lasso-Bandit and G-MCP-Bandit, which adopt different statistical learning methods for the sparse structure discovery and are designed for high-dimensional problems, have lower cumulative regret that increases in d at a slower rate. Further, we notice that the G-MCP-Bandit algorithm has the least increase in cumulative regret among all four algorithms, which confirms our theoretical finding in Theorem 1: The G-MCP-Bandit algorithm has a better dependence on the covariate dimension $O(\log d)$ than Lasso-Bandit $O(\log^2 d)$, OFUL, and OLS-Bandit (the last two algorithms have polynomial bounds in d).

In the second synthetic data experiment, we study the influence of the size of decision set by varying $K = \{2, 5, 10, 20, 50, 100\}$ and keeping the data dimensions unchanged ($s = 5$ and $d = 100$). For each decision, we randomly draw the parameter vector for the significant covariates from a uniform distribution, $U(0, 1)$. Finally, we keep other parameters the same as in the first synthetic data experiment. Figure 2 plots the average cumulative regret for OFUL, OLS-Bandit, Lasso-Bandit, and G-MCP-Bandit (under the linear model).

We observe that the benefits of adopting G-MCP-Bandit over the other three algorithms increases with the size of the decision set. In particular, as K increases, the cumulative regret gap between G-MCP-Bandit and any other algorithm grows; see Figure 2(c). This observation is as expected. To intuit, note that as we add more possible decisions into the decision set, the complexity and difficulty for the decision-maker to select the optimal decision grow for two main reasons. First, the decision-maker will need more samples to identify the significant covariates and estimate the parameter vectors. Second, as the number of decisions increases, the process of comparing the expected rewards among all decisions and selecting the optimal decision becomes more vulnerable to estimation errors. Therefore, we should expect that as the number of arms increases, the amount

Figure 2 **Synthetic study 2: The impact of T and K on the cumulative regret, where $d = 100$ and $s = 5$.**



(a) K=2 (b) K=20 (c) T=6000

of samples required for these algorithms to accurately learn the parameter vectors and select the optimal decision will increase as well.

Figure 2(a) and Figure 2(b) plot the cumulative regret for the case of two arms and twenty arms, respectively. Clearly, the decision-maker needs far more samples before his cumulative regret can be stabilized in the case of twenty arms than in the case of two arms. Therefore, the cumulative regret performance under all algorithms suffers from the increasing size of the decision set. As discussed earlier, the G-MCP-Bandit algorithm attains the optimal bound in the sample size dimension and is able to learn the sparse data structure and provide accurate unbiased estimators for parameter vectors. Hence, we observe that the benefits of adopting the G-MCP-Bandit algorithm over other algorithms are amplified as the number of arms increases, as illustrated in Figure 2(c).

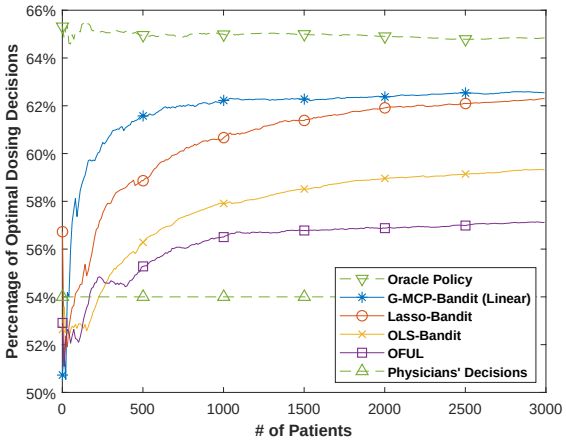
6.2. Warfarin Dosing Patient Data (Linear Model)

In the first real-data-based experiment, we considers a health care problem in which physicians determine the optimal personalized warfarin dosage for incoming patients (Consortium et al. 2009). Using the same dataset, Bastani and Bayati (2015) demonstrate that the Lasso-Bandit algorithm outperforms other existing bandit algorithms, including OFUL-LS (Abbasi-Yadkori et al. 2011), OFUL-EG (Abbasi-Yadkori and Szepesvari 2012), and OLS-Bandit (Goldenshluger and Zeevi 2013). The warfarin dosing patient data contains detailed covariates (the size of covariates used in our experiment is 93) for 5,700 patients, including demographic, diagnosis, and genetic information that can be used to predict the optimal warfarin dosage.

We apply the G-MCP-Bandit algorithm to the warfarin dosing patient dataset to evaluate its performance in practical decision-making contexts where the technical assumptions specified early in §3 may not hold. Following Bastani and Bayati (2015), we formulate this problem as a 3-armed bandit with covariates under the linear model.

Figure 3 compares the average fraction of optimal/correct dosing decisions under G-MCP-Bandit (under the linear model) to those under OFUL, OLS-Bandit, Lasso-Bandit, actual physicians'

Figure 3 **Warfarin dosing experiment: The percentage of optimal warfarin dosing decisions.**



decisions, and the oracle policy. We observe that as long as the sample size is not too small (e.g., the number of patients exceeds 40), the G-MCP-Bandit algorithm will outperform physicians' decisions, OLS-Bandit, Lasso-Bandit, and OFUL. However, when there are very limited samples (< 40 patients), the physicians' static decisions (i.e., always recommend medium dose) perform the best, with a stable optimal percentage of 54%. This is because that without sufficient samples, all learning algorithms are unable to accurately learn the parameter vectors for patients' covariates, and consequently they behave suboptimally.

As the sample size increases, all learning algorithms are able to update their estimation of parameter vectors and eventually outperform the physicians' static decisions. Among all learning algorithms, the G-MCP-Bandit algorithm requires the fewest samples (i.e., $T > 40$ for G-MCP-Bandit, $T > 90$ for Lasso-Bandit, $T > 180$ for OFUL, $T > 220$ for OLS-Bandit) to provide better dosing decisions than physicians.

6.3. Tencent Search Advertising Data (Linear & Logistic Models)

In the last experiment, we scale up the dataset's dimensionality to consider a search advertising problem at Tencent. The Tencent search advertising dataset is collected by Tencent's proprietary search engine, soso.com, and it documents the interaction sessions between users and the search engine (Tencent 2012). In the dataset, each session contains a user's demographic information (age and gender), the query issued by the user (combinations of keywords), ads information (title, URL address, and advertiser ID), the user's response (click or not), etc. This dataset is high-dimensional with sparse data structure and contains millions of observations and covariates. To put the size of the dataset into perspective, it contains $149,639,105$ session entries, more than half a million ads, more than one million unique keywords, and more than 26 million unique queries.

For illustration purposes, we focus on a three-ad experiment[3] (with ad IDs 21162526, 3065545, and 3827183). Each of these three ads has an average CTR higher than 2% and more than $100,000$ session entries, which provide reasonably accurate estimation for parameter vectors (see next paragraph for more discussions). In total, there are $849,338$ session entries with $169,744$ unique queries and 8 covariates for users' demographic information. As the search engine receives payment from advertisers only when the user has clicked the sponsored ad, we arbitrarily assume that advertisers will award the search engine \$1, \$5, and \$10 for each clicked ad, respectively.

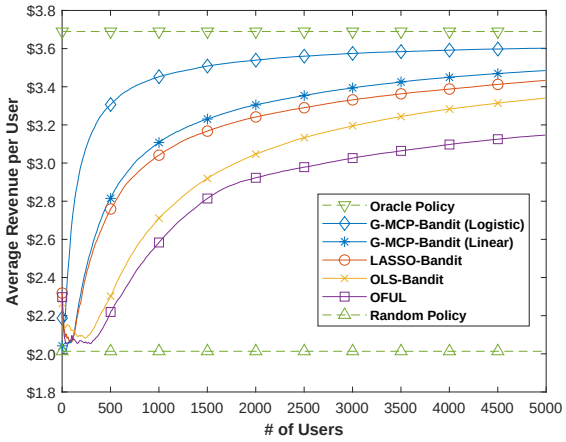
Figure 4 plots the the average revenue performance under OFUL, OLS-Bandit, Lasso-Bandit, a random policy, the oracle policy, and G-MCP-Bandit (under both linear and logistic models). It is worth noting that the "true" oracle policy is impossible to implement, as the true parameter vectors are unknown, or at least have considerable variance even when all session entries in the dataset are used for estimation. Therefore, the oracle policy in the experiment represents the scenario when the search engine has access to all data to estimate these parameter vectors and make ad selection decisions. In addition, we introduce the random policy as another benchmark to simulate the scenario in which the search engine will randomly recommend an ad with equal probability to an incoming user. Finally, note that the CTR prediction is binary in nature (i.e., click or not). We therefore include the G-MCP-Bandit algorithm under the logistic model and compare it to the G-MCP-Bandit algorithm under the linear model to study the influence of the underlying model choice. In the experiment, we simulate incoming users by permuting their covariate vectors randomly. For each algorithm, we perform 100 trials and report the average revenue with 5000 users, which seems to be sufficient for the G-MCP-Bandit algorithm to converge.

We can show that all learning algorithms generate higher average revenue than the random policy for any number of users and that the G-MCP-Bandit algorithm outperforms other algorithms under most scenarios. Specifically, when comparing all algorithms under the same linear model, we observe that the G-MCP-Bandit algorithm (under the linear model) has better average revenue performance than OFUL, OLS-Bandit, and Lasso-Bandit as soon as there are more than 140 users. This observation is consistent with that in warfarin dosing experiment in §6.2 and suggests that compared to other benchmark algorithms, the G-MCP-Bandit algorithm can improve the parameter vector estimation under high-dimensional data with limited samples and achieve better revenue performance.

Further, we find the choice of underlying models can significantly influence the G-MCP-Bandit algorithm's average revenue performance. Note that the advertisers award the search engine only

[3] We have extended the experiment to include more ads, but we find that doing so will not qualitatively change our observations and insights but considerably increases the computation time. Therefore, we decide to focus on this three-ad experiment in the paper.

Figure 4 **Tencent search advertising experiment: The average revenue under different algorithms.**



when users have clicked the recommended ads. Therefore, the search engine's reward function is binary in nature. When comparing the G-MCP-Bandit algorithm under the logistic model to that under the linear model, both of which are special cases of the G-MCP-Bandit algorithm, we observe that the former always dominates the latter for any number of users. In addition, the G-MCP-Bandit algorithm under the logistic model merely needs 20 users to outperform the other three algorithms. This observation suggests that understanding the underlying managerial problem and identifying the appropriate model for the G-MCP-Bandit algorithm can be critical and bring substantial revenue improvement for the decision-maker.

7. Conclusion

In this research, we develop the G-MCP-Bandit algorithm for online learning and decision-making processes in high-dimensional settings under limited samples. We adopt the matrix perturbation technique to derive new oracle inequality for the MCP estimator under non-iid samples and further propose a linear approximation method, the 2sWL procedure, to overcome the computational and statistical challenges associated with solving the MCP estimator (an NP-complete problem) under the bandit setting. We demonstrate that the MCP estimator solved by the 2sWL procedure matches the oracle estimator with high probability and converges to the true parameters with the optimal convergence rate. Further, we show that the cumulative regret of the G-MCP-Bandit algorithm over the sample size T is bounded by $O(\log T)$, which is the lowest theoretical bound for all possible algorithms under both low-dimensional and high-dimensional settings. In the covariate dimension d, the cumulative regret of the G-MCP-Bandit algorithm is bounded by $O(\log d)$, which is also a tighter bound than existing bandit algorithms. Finally, we illustrate that compared to other benchmark algorithms, the G-MCP-Bandit algorithm performs favorably in synthetic-data-based and real-data-based experiments.

Implementing the G-MCP-Bandit algorithm under high-dimensional data with a large decision set in an online setting can be challenging in practice, and addressing these challenges can extend this research to several directions. One of the major challenges is the computation time, especially when the covariate dimension and the decision set are extremely large. In particular, during a collaboration with a leading online marketplace, we adopted the G-MCP-Bandit algorithm, aiming to improve its product recommendation system. Using its datasets (with 5 million covariates and 30 million products), we showed that the G-MCP-Bandit algorithm improved the prediction of the conversion rate by 15% and the expected revenue by 5% on average, but a single server could take hours to execute the algorithm. We can implement the G-MCP-Bandit algorithm in a hybrid online-offline setting, where we recommend products by following the bi-level decision structure for every user but update the parameter vector estimation $\boldsymbol{\beta}^{random}$ and $\boldsymbol{\beta}^{whole}$ in batches every a couple of hours. Yet, in order to implement the G-MCP-Bandit algorithm in *online* settings, where we also update the parameter vector estimation for every incoming user, parallel computation techniques must be developed to tremendously reduce the computation time. Other challenges for the G-MCP-Bandit algorithm are how to simultaneously recommend multiple products and how to dynamically update the recommendation if the user did not click the recommended products but kept refreshing the recommendation page. Tackling these challenges requires an integration of the assortment optimization and Bayesian learning into the G-MCP-Bandit algorithm.

References

Abbasi-Yadkori Y, Pál D, Szepesvári C (2011) Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 2312–2320.

Abbasi-Yadkori Y, Szepesvari C (2012) *Online learning for linearly parametrized control problems* (University of Alberta).

Agrawal S, Goyal N (2013) Thompson sampling for contextual bandits with linear payoffs. *International Conference on Machine Learning*, 127–135.

Auer P (2002) Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research* 3(Nov):397–422.

Bastani H, Bayati M (2015) Online decision-making with high-dimensional covariates .

Bühlmann P, Van De Geer S (2011) *Statistics for high-dimensional data: methods, theory and applications* (Springer Science & Business Media).

Candes EJ, Wakin MB, Boyd SP (2008) Enhancing sparsity by reweighted 1 minimization. *Journal of Fourier analysis and applications* 14(5-6):877–905.

Consortium IWP, et al. (2009) Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009(360):753–764.

Dani V, Hayes TP, Kakade SM (2008) Stochastic linear optimization under bandit feedback .

Deshpande Y, Montanari A (2012) Linear bandits in high dimension and recommendation systems. *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 1750–1754 (IEEE).

Elmachtoub AN, McNellis R, Oh S, Petrik M (2017) A practical method for solving contextual bandit problems using decision trees. *Proceedings of the Thirty-third Conference on Uncertainty in Artificial Intelligence (UAI)* (AUAI Press).

Fan J, Han F, Liu H (2014a) Challenges of big data analysis. *National science review* 1(2):293–314.

Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96(456):1348–1360.

Fan J, Liu H, Sun Q, Zhang T (2018) I-lamm for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Annals of statistics* 46(2):814.

Fan J, Xue L, Zou H (2014b) Strong oracle optimality of folded concave penalized estimation. *Annals of statistics* 42(3):819.

Goldenshluger A, Zeevi A (2013) A linear response bandit problem. *Stochastic Systems* 3(1):230–261.

Huang J, Ma S, Zhang CH (2008) Adaptive lasso for sparse high-dimensional regression models. *Statistica Sinica* 1603–1618.

Liu H, Yao T, Li R, Ye Y (2017) Folded concave penalized sparse linear regression: Sparsity, statistical performance, and algorithmic theory for local solutions. *Mathematical Programming* 134, URL `http://dx.doi.org/10.1007/s10107-017-1114-y`.

Liu H, Yao T, Li R, et al. (2016) Global solutions to folded concave penalized nonconvex learning. *The Annals of Statistics* 44(2):629–659.

Loh PL, Wainwright MJ (2013) Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Advances in Neural Information Processing Systems*, 476–484.

McCullagh P, Nelder J (1989) *Generalized linear models* (Chapman and Hall/CRC).

Meinshausen N, Bühlmann P, et al. (2006) High-dimensional graphs and variable selection with the lasso. *The annals of statistics* 34(3):1436–1462.

Meinshausen N, Yu B, et al. (2009) Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics* 37(1):246–270.

Mitchell J (2012) How google search really works. `https://readwrite.com/2012/02/29/interview_changing_engines_mid-flight_qa_with_goog/#awesm=~oiNkM4tAX3xhbP`, accessed: Oct 22nd, 2018.

Montgomery DC, Peck EA, Vining GG (2012) *Introduction to linear regression analysis*, volume 821 (John Wiley & Sons).

Negahban S, Yu B, Wainwright MJ, Ravikumar PK (2009) A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Advances in Neural Information Processing Systems*, 1348–1356.

OxfordDictionaries (2018) How many words are there in the english language? `https://en.oxforddictionaries.com/explore/how-many-words-are-there-in-the-english-language/`, accessed: Oct 22nd, 2018.

Rigollet P, Zeevi A (2010) Nonparametric bandits with covariates. *arXiv preprint arXiv:1003.1630* .

Robbins H (1952) Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society* 58(5):527–535.

Rudelson M, Vershynin R, et al. (2013) Hanson-wright inequality and sub-gaussian concentration. *Electron. Commun. Probab* 18(82):1–9.

Rusmevichientong P, Tsitsiklis JN (2010) Linearly parameterized bandits. *Mathematics of Operations Research* 35(2):395–411.

Russo D, Van Roy B (2014) Learning to optimize via posterior sampling. *Mathematics of Operations Research* 39(4):1221–1243.

Scott SL (2010) A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry* 26(6):639–658.

Scott SL (2015) Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry* 31(1):37–45.

Shewan D (2017) The comprehensive guide to online advertising costs. `https://www.wordstream.com/blog/ws/2017/07/05/online-advertising-costs`, accessed: Oct 22nd, 2018.

Slivkins A (2014) Contextual bandits with similarity information. *The Journal of Machine Learning Research* 15(1):2533–2568.

Tencent (2012) Predict the click-through rate of ads given the query and user information. `https://www.kaggle.com/c/kddcup2012-track2`, accessed: Oct 22nd, 2018.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

Tropp JA, et al. (2015) An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* 8(1-2):1–230.

Tsybakov AB, et al. (2004) Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics* 32(1):135–166.

Van de Geer SA, et al. (2008) High-dimensional generalized linear models and the lasso. *The Annals of Statistics* 36(2):614–645.

WordStream (2017) Average ctr (click-through rate): Learn how your ctr compares. `https://www.wordstream.com/average-ctr`, accessed: Oct 22nd, 2018.

Zhang CH, Huang J, et al. (2008) The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics* 36(4):1567–1594.

Zhang CH, Zhang T, et al. (2012) A general theory of concave regularization for high-dimensional sparse estimation problems. *Statistical Science* 27(4):576–593.

Zhang CH, et al. (2010) Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics* 38(2):894–942.

Zhao T, Liu H, Zhang T (2014) Pathwise coordinate optimization for sparse learning: Algorithm and theory. *arXiv preprint arXiv:1412.7477* .

Zhao T, Liu H, Zhang T, et al. (2018) Pathwise coordinate optimization for sparse learning: Algorithm and theory. *The Annals of Statistics* 46(1):180–218.

Zou H (2006) The adaptive lasso and its oracle properties. *Journal of the American statistical association* 101(476):1418–1429.

Electronic Companion to "Online Learning and Decision-Making under Generalized Linear Model with High-Dimensional Data"

To simplify the notation in the E-companion, we denote $\nabla_{\mathcal{A}} F(\boldsymbol{x})$ as the vector with $(\nabla_{\mathcal{A}} F(\boldsymbol{x}))_i = (\nabla F(\boldsymbol{x}))_i,\ i \in \mathcal{A}$, where $(\cdot)_i$ is the i-th element in the vector. Similarly we denote $\nabla^2_{\mathcal{A},\mathcal{B}} F(\boldsymbol{x})$ as the matrix with $(\nabla^2_{\mathcal{A},\mathcal{B}} F(\boldsymbol{x}))_{ij} = (\nabla^2 F(\boldsymbol{x}))_{ij},\ i \in \mathcal{A}, j \in \mathcal{B}$, where $(\cdot)_{ij}$ is the element in i-th column and j-th row. We denote $\lambda_{\min}(\boldsymbol{X})/\lambda_{\max}(\boldsymbol{X})$ as the smallest/largest eigenvalue of matrix $\boldsymbol{X}$.

Proof of Lemma 1 Lemma 1 directly follows Lemma EC.2 in "Appendix: Supplemental Lemmas and Proofs" at the end of this Electronic Companion by setting $|\mathcal{A}| = n$.

Proof of Proposition 1 Proposition 1 follows Proposition 3 by setting $|\mathcal{A}| = n$.

Proof of Proposition 2 Under the ϵ-decay random sampling method, the probability of randomly drawing arm k at time t is $\min\{1, t_0/t\}/|\mathcal{K}|$, where $|\mathcal{K}|$ is the number of arms. Hence, at time T, the expected total number of times at which arm k were randomly drawn is

$$\mathbb{E}[n_k] = \frac{1}{|\mathcal{K}|} \sum_{t=1}^{T} \min\left\{1, \frac{t_0}{t}\right\}.$$

When $T > t_0$,

$$\mathbb{E}[n_k] = \frac{1}{|\mathcal{K}|}\left(t_0 + \sum_{t=t_0+1}^{T} \frac{t_0}{t}\right) = \frac{t_0}{|\mathcal{K}|}\left(1 + \sum_{t=t_0+1}^{T} \frac{1}{t}\right). \tag{EC.1}$$

Since the function $f(t) = 1/t$ is decreasing in t , it can be bounded as follows.

$$\int_t^{t+1} \frac{1}{t} dt < \frac{1}{t} < \int_{t-1}^{t} \frac{1}{t} dt,\ t \geq 2.$$

As $t_0 \geq 1$, for any t from $t_0 + 1$ to T, we have

$$\log(T+1) - \log(t_0+1) < \sum_{t=t_0+1}^{T} \frac{1}{t} < \log(T) - \log(t_0). \tag{EC.2}$$

Combining (EC.1) and (EC.2), we can bound $\mathbb{E}[n_k]$ as follows.

$$\frac{1}{|\mathcal{K}|} t_0 (1 + \log(T+1) - \log(t_0+1)) < \mathbb{E}[n_k] < \frac{1}{|\mathcal{K}|} t_0 (1 + \log(T) - \log(t_0)). \tag{EC.3}$$

Since $n_k = \sum_{t=1}^{T} \mathbb{1}\{\text{random sampling for arm } k \text{ at } t\}$, we can view n_k as the summarization of bounded iid random variables. Via Chernoff bound, we can build the connect between n_k and $\mathbb{E}[n_k]$.

$$\mathbb{P}\left(\frac{1}{2}\mathbb{E}[n_k] \leq n_k \leq \frac{3}{2}\mathbb{E}[n_k]\right) > 1 - 2\exp\left(-\frac{1}{10}\mathbb{E}[n_k]\right). \tag{EC.4}$$

We then relax the $\mathbb{E}[n_k]$ in (EC.4) with the upper and lower bounds provided in (EC.3) and the following result is attained.

$$\mathbb{P}\left(\frac{t_0(1+\log(T+1)-\log(t_0+1))}{2|\mathcal{K}|} \leq n_k \leq \frac{3t_0(1+\log(T)-\log(t_0))}{2|\mathcal{K}|}\right) \geq 1 - 2\left(\frac{t_0+1}{e(T+1)}\right)^{\frac{t_0}{10|\mathcal{K}|}}. \tag{EC.5}$$

When $t_0 = 2C_0|\mathcal{K}|,\ C_0 \geq 10$, and $T > \frac{(t_0+1)^2}{e^2}$, we can simplify the right-hand size of (EC.5).

$$1 - 2\left(\frac{t_0+1}{e(T+1)}\right)^{\frac{t_0}{10|\mathcal{K}|}} \geq 1 - 2\left(\frac{e\sqrt{T+1}}{e(T+1)}\right)^{C_0/5} \geq 1 - \frac{2}{T+1}. \tag{EC.6}$$

Proof of Proposition 3 In the first step of 2sWL procedure, we are essentially solving the Lasso problem. From Lemma EC.7, we have $\|\boldsymbol{\beta}^{lasso} - \boldsymbol{\beta}^{true}\|_1 \leq \frac{96ns\lambda}{|\mathcal{A}|\kappa}$ which high probability. As we assume $\beta_{\min} \geq \left(\frac{96ns}{|\mathcal{A}|\kappa} + a\right)\lambda$ and $\|\boldsymbol{\beta}^{lasso} - \boldsymbol{\beta}^{true}\|_\infty \leq \|\boldsymbol{\beta}^{lasso} - \boldsymbol{\beta}^{true}\|_1$ we have the follow statements hold.

$$|\beta_i^{lasso}| \geq a\lambda,\ i \in \mathcal{S} \text{ and } |\beta_i^{lasso}| \leq \frac{96ns\lambda}{|\mathcal{A}|\kappa},\ i \in \mathcal{S}^c, \tag{EC.7}$$

where we ignore the subscript in $\mathcal{S}_k$ to simplify the notation. Combining (EC.7) and $P_\lambda'(|x|) = \max\{0, \lambda - |x|/a\}$, we have the following two results.

$$P_\lambda'(|\beta_i^{lasso}|) = 0 \quad i \in \mathcal{S}, \tag{EC.8}$$

$$P_\lambda'(|\beta_i^{lasso}|) \geq P_\lambda'\left(\frac{96ns\lambda}{|\mathcal{A}|\kappa}\right) = \left(\lambda - \frac{96ns\lambda}{|\mathcal{A}|\kappa a}\right) \quad i \in \mathcal{S}^c. \tag{EC.9}$$

Define the event $\mathcal{E}_2$ as follows

$$\mathcal{E}_2 = \left\{\|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{oracle})\|_\infty < \lambda - \frac{96ns\lambda}{|\mathcal{A}|\kappa a}\right\}. \tag{EC.10}$$

From the convexity of $\mathcal{L}(\boldsymbol{\beta})$, we can build a lower bound on the optimal objective function value in the second step of 2sWL.

$$\mathcal{L}(\boldsymbol{\beta}^*) + \sum_j P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^*| \geq \mathcal{L}(\boldsymbol{\beta}^{oracle}) + \nabla\mathcal{L}(\boldsymbol{\beta}^{oracle})^T(\boldsymbol{\beta}^* - \boldsymbol{\beta}^{true}) + \sum_j P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^*|, \tag{EC.11}$$

where $\boldsymbol{\beta}^*$ is the optimal solution of the second step of the 2sWL procedures. From the definition of oracle solution, we have

$$\boldsymbol{\beta}^{oracle} = \arg\min_{\boldsymbol{\beta}_{\mathcal{S}^c}=0} \mathcal{L}(\boldsymbol{\beta}) \Rightarrow 1)\ \nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{oracle}) = 0 \text{ and } 2)\ \boldsymbol{\beta}_{\mathcal{S}^c} = 0. \tag{EC.12}$$

Combining (EC.8), (EC.9), (EC.11), and (EC.12), we have

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}^*) + \sum_{j\in\mathcal{S}^c} P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^*| &\geq \mathcal{L}(\boldsymbol{\beta}^{oracle}) + \nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{oracle})^T(\boldsymbol{\beta}^*_{\mathcal{S}^c} - \boldsymbol{\beta}^{oracle}_{\mathcal{S}^c}) + \sum_{j\in\mathcal{S}^c} P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^*| \\
&= \mathcal{L}(\boldsymbol{\beta}^{oracle}) + \sum_{j\in\mathcal{S}^c}\left(\nabla_j\mathcal{L}(\boldsymbol{\beta}^{oracle})(\beta_j^* - 0) + P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^*|\right) \\
&= \mathcal{L}(\boldsymbol{\beta}^{oracle}) + \sum_{j\in\mathcal{S}^c} P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^{oracle}| \\
&\quad + \sum_{j\in\mathcal{S}^c}\left(\nabla_j\mathcal{L}(\boldsymbol{\beta}^{oracle})\text{sign}(\beta_j^*) + P_\lambda'(|\beta_j^{lasso}|)\right)|\beta_j^*|.
\end{aligned} \tag{EC.13}$$

Using $\mathcal{E}_2$ defined in (EC.10), (EC.13) can be simplified as follows.

$$\mathcal{L}(\boldsymbol{\beta}^*) + \sum_{j\in\mathcal{S}^c} P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^*| \geq \mathcal{L}(\boldsymbol{\beta}^{oracle}) + \sum_{j\in\mathcal{S}^c} P_\lambda'(|\beta_j^{lasso}|)\cdot|\beta_j^{oracle}| + c_0\sum_{j\in\mathcal{S}^c}|\beta_j^*|, \tag{EC.14}$$

where c_0 is a positive constant. Since $\boldsymbol{\beta}^*$ is the optimal solution of the second step in 2sWL, per (EC.14) we must have $\beta_j^* = 0$ for all $j \in \mathcal{S}^c$. Together with the uniqueness of the solution of (1), $\boldsymbol{\beta}^{oracle}$ is also the

unique optimal solution to the second step in 2sWL, i.e, $\boldsymbol{\beta}^{MCP} = \boldsymbol{\beta}^{oracle}$. Therefore once event $\mathcal{E}_2$ happens, with high probability $\boldsymbol{\beta}^{MCP}$ becomes the oracle solution, which enjoy the optimal statistical performance. We then need to consider the chance that $\mathcal{E}_2$ happens and the result is summarized in Lemma EC.11. Per Lemma EC.11, the following $\mathcal{E}_3, \mathcal{E}_4$ and $\mathcal{E}_5$ implies $\mathcal{E}_2$.

$$\begin{aligned}
\mathcal{E}_3 &= \left\{ \|\nabla_{\mathcal{S}^c} \mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty \leq \left(1 - \frac{96ns}{|\mathcal{A}|\kappa a}\right) \frac{\lambda}{4} \right\}, \\
\mathcal{E}_4 &= \left\{ \|\nabla_{\mathcal{S}} \mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty \leq \left(1 - \frac{96ns}{|\mathcal{A}|\kappa a}\right) \frac{\mu_0 |\mathcal{A}| \lambda}{8sn x_{\max}^2} \right\}, \\
\mathcal{E}_5 &= \left\{ \|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{C_2 \lambda} \right\},
\end{aligned}$$

where C_2 is a positive constant. Now, we can bound the probability of events $\mathcal{E}_3$, $\mathcal{E}_4$, and $\mathcal{E}_5$ happen simultaneously. From Assumption **A.5** and Hoeffding bound we have the following inequality for $t_1 > 0$

$$\mathbb{P}\left(\|\nabla_{\mathcal{S}} \mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty \geq t_1\right) = \mathbb{P}\left(\frac{1}{n} \sum_{j=1}^{n} x_{j\mathcal{S}}^T f'(r_j | \boldsymbol{x}_{j,\mathcal{S}}^T \boldsymbol{\beta}^{true})\|_\infty \geq t_1 \right) \leq s \exp\left(-\frac{n t_1^2}{2\sigma^2 x_{\max}^2} \right). \tag{EC.15}$$

Similarly for $t_2 > 0$, we have the following result.

$$\mathbb{P}\left(\|\nabla_{\mathcal{S}^c} \mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty \geq t_2\right) \leq (d-s) \exp\left(-\frac{n t_2^2}{2\sigma^2 x_{\max}^2} \right). \tag{EC.16}$$

By setting $t_1 = t_2 = (\frac{1}{4} - \frac{24ns}{|\mathcal{A}|\kappa a}) \min\left\{1, \frac{\mu_0 |\mathcal{A}|}{8sn x_{\max}^2}\right\} \lambda$, we have

$$\mathbb{P}\left((\mathcal{E}_4')^c \cup (\mathcal{E}_5')^c\right) \leq d \exp\left(-\frac{n\lambda^2 \left((\frac{1}{4} - \frac{24ns}{|\mathcal{A}|\kappa a}) \min\left\{1, \frac{\mu_0 |\mathcal{A}|}{8sn x_{\max}^2}\right\} \right)^2}{2x_{\max}^2} \right). \tag{EC.17}$$

We can further bound event $\mathcal{E}_5$ via Lemma EC.2. We can have the following result by setting t in Lemma EC.2 satisfying $t \leq \frac{\mu_0 |\mathcal{A}| \sqrt{C_2 \lambda}}{2n}$.

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{C_2 \lambda} \right) \geq 1 - s \exp\left(-\frac{\mu_0 |\mathcal{A}|}{8s\sigma_2 x_{\max}^2} \right) - s \exp\left(-\frac{nt^2}{2s\sigma^2 x_{\max}^2} \right). \tag{EC.18}$$

Moreover, from (EC.55) in Lemma EC.2, the following result hold for $|\mathcal{A}| \geq \frac{2s^2 x_{\max}^2}{\mu_0}$:

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{\frac{8s^2\sigma^2 x_{\max}^2 n}{\mu_0^2 |\mathcal{A}|^2}} \right) \geq 1 - s \exp\left(-\frac{\mu_0 |\mathcal{A}|}{8s\sigma_2 x_{\max}^2} \right) - 2 \exp\left(-\frac{C_h |\mathcal{A}| \mu_0}{2s x_{\max}^2} \right). \tag{EC.19}$$

Combining Lemma EC.7, (EC.17) and (EC.18) , we have the following inequality for $\zeta \leq \frac{\mu_0 |\mathcal{A}| \sqrt{C_2 \lambda}}{2n}$.

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \leq \frac{2n\zeta}{|\mathcal{A}|\mu_0} \right) \geq 1 - \delta_2(n, |\mathcal{A}|, \lambda) - \delta_3(|\mathcal{A}|) - \delta_4(n, |\mathcal{A}|, \zeta). \tag{EC.20}$$

Similarly, by $|\mathcal{A}| \geq \frac{2s^2 x_{\max}^2}{\mu_0}$, the following result comes directly from Lemma EC.7, (EC.17) and (EC.19).

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \leq \sqrt{\frac{8s^2\sigma^2 x_{\max}^2 n}{\mu_0^2 |\mathcal{A}|^2}} \right) \geq 1 - \delta_1(|\mathcal{A}|) - \delta_2(n, |\mathcal{A}|, \lambda) - \delta_3(|\mathcal{A}|). \tag{EC.21}$$

Proof of Proposition 4 Directly from Lemma EC.9.

Proof of Proposition 5 Since $\{M(i)\}$ is a martingale with bounded difference 1, we can use $M(0)$ to bound the value of $M(T+1)$ with Azuma's inequality as follow:

$$\begin{aligned}\mathbb{P}\left(|M(T+1)-M(0)|\geq\frac{1}{2}M(0)\right)&\leq\exp\left(\frac{-M(0)^2/4}{2(T+2)}\right)\\ \Rightarrow\mathbb{P}\left(M(T+1)\leq\frac{1}{2}M(0)\right)&\leq\exp\left(\frac{-M(0)^2/4}{2(T+2)}\right).\end{aligned}$$

The term $M(0)$ can be expressed as follows

$$\begin{aligned}M(0)&=\mathbb{E}\left[\sum_{i=1}^{T+1}\mathbb{1}(\boldsymbol{x}_i\in U_k,\mathcal{E}_6,\boldsymbol{x}\notin\mathcal{R}_k))\right]\\&=\sum_{i=1}^{T+1}\mathbb{P}(\boldsymbol{x}_i\in U_k,\mathcal{E}_6,x\notin\mathcal{R}_k).\end{aligned}\tag{EC.22}$$

As $\{\boldsymbol{x}\in U_k\}$ is independent of $\{\mathcal{E}_6,\boldsymbol{x}\notin\mathcal{R}_k\}$ and $\{\boldsymbol{x}\notin\mathcal{R}_k\}$ is independent on $\{\mathcal{E}_6\}$, (EC.22) implies the following inequality

$$\begin{aligned}M(0)&=\sum_{i=1}^{T+1}\mathbb{P}(\boldsymbol{x}_i\in U_k)\mathbb{P}(\mathcal{E}_6)\mathbb{P}(\boldsymbol{x}\notin\mathcal{R}_k)\\&\geq\sum_{i=1}^{T+1}p^*(1-\frac{7}{T+1})(1-\frac{2C_0|\mathcal{K}|}{T+1}),\end{aligned}\tag{EC.23}$$

where (EC.23) uses assumption **A.3**, Proposition 4 and Proposition 2.

When $T\geq\max\{14,4C_0|\mathcal{K}|\}$, we have

$$\frac{7}{T+1}\leq\frac{1}{2}\tag{EC.24}$$

$$\frac{2C_0|\mathcal{K}|}{T+1}\leq\frac{1}{2},\tag{EC.25}$$

which implies that

$$M(0)\geq\sum_{i=1}^{T+1}\frac{p^*}{4}=\frac{p^*(T+1)}{4}.\tag{EC.26}$$

Therefore, the following inequalities hold

$$\begin{aligned}&\mathbb{P}\left(M(T+1)\leq\frac{p^*(T+1)}{8}\right)\leq\mathbb{P}\left(M(T+1)\leq\frac{1}{2}M(0)\right)\leq\exp\left(\frac{-(p^*)^2(T+1)^2/64}{2(T+2)}\right)\\ \Rightarrow&\mathbb{P}\left(M(T+1)\leq\frac{p^*(T+1)}{8}\right)\leq\exp\left(-\frac{(p^*)^2((T+2)^2+1-2(T+2))}{128(T+2)}\right)\\ \Rightarrow&\mathbb{P}\left(M(T+1)\leq\frac{p^*(T+1)}{8}\right)\leq\exp\left(-\frac{(p^*)^2T}{128}-\frac{p^*}{128(T+2)}\right)\\ \Rightarrow&\mathbb{P}\left(M(T+1)\leq\frac{p^*(T+1)}{8}\right)\leq\exp\left(-\frac{(p^*)^2T}{128}\right)\end{aligned}\tag{EC.27}$$

Proof of Proposition 6 According to Lemma EC.10, when event $\mathcal{E}_6$ defined by (8) happens, the following inequality must hold for any $\boldsymbol{x}\in U_k$,

$$\mathbb{E}(R_k|\boldsymbol{x},\boldsymbol{\beta}_k^{random}(t))\geq\max_{j\neq k}\mathbb{E}(R_j|\boldsymbol{x},\boldsymbol{\beta}_j^{random}(t))+\frac{h}{2}.$$

Therefore, the lower-level decision-making process of the algorithm, in which the decision-maker will successfully select arm i for x by using the random sample estimator, will maintain the iid property of $\boldsymbol{x}$ since it can be viewed as rejection sampling. From Proposition 5, we have

$$\mathbb{P}\left(M(T+1) \leq \frac{p^*(T+1)}{8}\right) \leq \exp\left(-\frac{(p^*)^2 T}{128}\right). \tag{EC.28}$$

Since $M(T+1) = \mathbb{E}\left[\sum_{j=1}^{T+1} \mathbb{1}(\boldsymbol{x}_j \in U_k, \mathcal{E}_6, \boldsymbol{x}_j \notin \mathcal{R}_k) | \mathcal{F}_{T+1}\right] = \sum_{j=1}^{T+1} \mathbb{1}(\boldsymbol{x}_j \in U_k, \mathcal{E}_6, \boldsymbol{x}_j \notin \mathcal{R}_k)$, the amount of iid samples among the whole sample for arm k up to time $T+1$ will be lower bounded by $M(T+1)$. Denote $\mathcal{A}$ and n as the set of iid samples belonging to U_K in the whole sample set and size of the whole sample respectively. The follow inequality holds.

$$\mathbb{P}\left(|\mathcal{A}| \geq \frac{p^*(T+1)}{8}\right) \geq 1 - \exp\left(-\frac{(p^*)^2 T}{128}\right), \ n \leq T+1. \tag{EC.29}$$

Consider $|\mathcal{A}| \geq \frac{p^*(T+1)}{8}$, $n \leq (T+1)$, $\lambda = C_4\sqrt{\frac{\log(T+1)+\log d}{T+1}}$, and $T \geq T_0$, where $C_4 = \frac{\sqrt{2}x_{\max}}{(\frac{1}{4} - \frac{192s}{p^*\kappa a})\min\{1, \frac{\mu_0}{p^* s x_{\max}^2}\}}$ and $T_0 = \max\left\{14, 4C_0|\mathcal{K}|, \frac{128}{(p^*)^2}, \frac{64}{C_1 p^*)^2}, \frac{256 s^2 x_{\max}^4}{(C_h p^*)^2}, \frac{64 s^2 \sigma^2 x_{\max}^4 (1+\log s)^2}{(\mu_0 p^*)^2}\right\}$, the following results can be obtained:

$$|\mathcal{A}| \geq \frac{2s^2 x_{\max}^2}{\mu_0}, \ a > \frac{96ns}{\kappa|\mathcal{A}|} \text{ and } \beta_{\min} \geq (\frac{96ns}{\kappa|\mathcal{A}|} + a)\lambda.$$

We then have the following result via Proposition 3.

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\| \geq \sqrt{\frac{512 s^3 \sigma^2 x_{\max}^2}{\mu_0^2 (p^*)^2 (T+1)}}\right) \leq \delta_1\left(\frac{p^*(T+1)}{8}\right) + \delta_2\left(T+1, \frac{p^*(T+1)}{8}, \lambda\right) + \delta_3\left(\frac{p^*(T+1)}{8}\right) \tag{EC.30}$$

Combining $T > T_0$, $\lambda = C_4\sqrt{\frac{\log(T+1)+\log d}{T+1}}$ and the fact $T+1 \geq \sqrt{T+1}\log(T+1)$ for $T > 0$, we have

$$\delta_1\left(\frac{p^*(T+1)}{8}\right) + \delta_2\left(T+1, \frac{p^*(T+1)}{8}, \lambda\right) + \delta_3\left(\frac{p^*(T+1)}{8}\right) \leq \frac{4}{T+1} \tag{EC.31}$$

$$\mathbb{P}\left(|\mathcal{A}| \leq \frac{p^*(T+1)}{8}\right) \leq \frac{1}{T+1}. \tag{EC.32}$$

Set $C_\beta = \frac{512\sigma^2 x_{\max}}{\mu_0^2 (p^*)^2}$, and Proposition 6 directly follows by combining (EC.31), (EC.30), (EC.32) and $\mathbb{P}(\mathcal{E}_6^c) \leq \frac{7}{T+1}$ from Lemma EC.10.

Proof of Theorem 1 We divide the time, up to time T, into three groups and derive the cumulative regret bound for each group separately. Consider the following three groups:

1. $x_i \in \mathcal{R}_k, k \in \mathcal{K}$ and $T \leq T_0$.
2. $x_i \notin \mathcal{R}_k, k \in \mathcal{K}$, $T > T_0$ and $\mathcal{E}_6$ doesn't hold,
3. $x_i \notin \mathcal{R}_k, k \in \mathcal{K}$ $T > T_0$ and $\mathcal{E}_6$ holds.

Before going to the detail proof, we first state the choice of T_0 and C_0 such that the requirements of Proposition 4-6 are satisfied.

$$T_0 = \max\left\{\frac{(t_0+1)^2}{e^2} - 1, 14, 4C_0|\mathcal{K}|, \frac{128}{(p^*)^2}, \frac{64}{C_1 p^*)^2}, \frac{256 s^2 x_{\max}^4}{(C_h p^*)^2}, \frac{64 s^2 \sigma^2 x_{\max}^4 (1+\log s)^2}{(\mu_0 p^*)^2}\right\}$$

$$C_0 = \max\left\{10, \frac{16}{p^*}, \frac{4}{p^* C_1}, \frac{4x_{\max}^2}{C_5^2}\left((\frac{1}{4} - \frac{576s}{p^*\kappa a})\min\left\{1, \frac{\mu_0 p^*}{192 s x_{\max}^2}\right\}\right)^{-2}, \frac{32\sigma_2 s x_{\max}^2 (1+\log s)}{p^* \mu_0}, \frac{4\sigma^2 x_{\max}^2 (1+\log s)}{t^2}\right\},$$

where $t \le \min\left\{\frac{\mu_0 p^*\sqrt{\tilde{C}_2\lambda}}{48}, \frac{p^*\mu_0}{48\sigma\sqrt{s}x_{\max}}, \frac{hp^*\mu_0}{192e\sigma\sqrt{s}R_{\max}x_{\max}}\right\}$, $C_1 = \min\left\{1, \kappa^2/\left(192s\sigma_2 x_{\max}^2(3+2\sqrt{\sigma_2}x_{\max})\right)^2\right\}$, $\tilde{C}_2 = \frac{\mu_0 p^*}{2\sigma_3 s x_{\max}^3(\mu_0 p^*+48sx_{\max}^2)}$ and $C_5 = \frac{\beta_{\min}p^*\kappa}{(2304s+ap^*\kappa)\sqrt{1+\log d}}$.

Regret in part 1: Denote the regret for the first part as $R_1(T)$.

$$R_1(T) \le R_{\max}\left(\sum_{i=T_0}^{T}\mathbb{1}(x_i \in \mathcal{R}_k, k \in \mathcal{K}) + T_0\right) \le R_{\max}\left(\sum_{k\in\mathcal{K}} n_k + T_0\right). \tag{EC.33}$$

From Proposition 2, we know that

$$\mathbb{P}\left(n_k \le \frac{3t_0(1+\log(T)-\log(t_0))}{2|\mathcal{K}|}\right) \ge 1 - \frac{2}{T+1}. \tag{EC.34}$$

If we require $t_0 = 2C_0|\mathcal{K}|$, $C_0 \ge 10$, and $T \ge \max\{(t_0+1)^2/e^2 - 1, 14\}$, then the above equation can be simplified to

$$\mathbb{P}\left(n_k \le 6C_0\log T\right) \ge 1 - \frac{2}{T+1} \Rightarrow \mathbb{P}\left(n_k > 6C_0\log T\right) \le \frac{2}{T+1} \tag{EC.35}$$

which implies

$$\mathbb{P}\left(\sum_{k\in\mathcal{K}} n_k > 6C_0|\mathcal{K}|\log T\right) \le \mathbb{P}\left(\cup_{k\in\mathcal{K}}(n_k > 6C_0\log T)\right) \le \sum_{k\in\mathcal{K}}\mathbb{P}\left(n_k > 6C_0\log T\right) \le \frac{2|\mathcal{K}|}{T+1}, \tag{EC.36}$$

and

$$\begin{aligned}
R_1(T) \le R_{\max}\left(\sum_{k\in\mathcal{K}} n_k + T_0\right) &= R_{\max}\left(\sum_{k\in\mathcal{K}} n_k | \sum_{k\in\mathcal{K}} n_k > 6C_0|\mathcal{K}|\log T\right)\mathbb{P}\left(\sum_{k\in\mathcal{K}} n_k > 6C_0|\mathcal{K}|\log T\right)\\
&\quad + R_{\max}\left(\sum_{k\in\mathcal{K}} n_k | \sum_{k\in\mathcal{K}} n_k \le 6C_0|\mathcal{K}|\log T\right)\mathbb{P}\left(\sum_{k\in\mathcal{K}} n_k \le 6C_0|\mathcal{K}|\log T\right)\\
&\quad + R_{\max}T_0\\
&\le R_{\max}T\frac{2|\mathcal{K}|}{T+1} + R_{\max}6C_0|\mathcal{K}|\log T\left(1 - \frac{2|\mathcal{K}|}{T+1}\right) + R_{\max}T_0\\
&\le 2R_{\max}|\mathcal{K}| + 6R_{\max}C_0|\mathcal{K}|\log T + R_{\max}T_0\\
&\le R_{\max}|\mathcal{K}|(2 + 6C_0\log T) + R_{\max}T_0.
\end{aligned} \tag{EC.37}$$

Regret in part 2: Denote the regret for the second part as $R_2(T)$.From Lemma EC.9, we know that

$$\begin{aligned}
&\mathbb{P}\left(\|\boldsymbol{\beta}^{random}(t) - \boldsymbol{\beta}^{true}\|_1 \le \min\left\{\frac{1}{\sigma x_{\max}}, \frac{h}{4e\sigma R_{\max}x_{\max}}\right\}\right) \ge 1 - \frac{7}{T+1},\ k \in \mathcal{K}\\
\Rightarrow &\mathbb{P}(\mathcal{E}_6(T)) \ge 1 - \frac{7|\mathcal{K}|}{T+1}.
\end{aligned} \tag{EC.38}$$

Therefore, $R_2(T)$ can be bounded as follows

$$\begin{aligned}
R_2(T) &\le \mathbb{E}[\sum_{i=1}^{T}\mathbb{1}(\mathcal{E}_6(i)^c)R_{\max}]\\
&= \sum_{i=1}^{T}\mathbb{E}[\mathbb{1}(\mathcal{E}_6(i)^c)]R_{\max}\\
&= \sum_{i=1}^{T}\mathbb{P}(\mathcal{E}_6(i)^c)R_{\max}\\
&\le 7R_{\max}|\mathcal{K}|\log(T+1).
\end{aligned} \tag{EC.39}$$

Regret in part 3: Denote the regret for the third part as $R_3(T)$. Without loss of generality, we assume that arm i is true optimal arm at time t. Then, the regret at time t can be bounded as follows

$$\begin{aligned} r_t &= \mathbb{E}\left(\mathbb{1}\left(j=\arg\max_{k\in\mathcal{K}}\mathbb{E}[R_k|\boldsymbol{x}_t,\boldsymbol{\beta}_k^{whole}(t)]\right)\left(\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]-\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{true}]\right)\right) \\ &\le \mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right)\left(\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]-\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{true}]\right)\right). \end{aligned} \tag{EC.40}$$

Denote $\mathcal{E}(t,\delta)_{8,k}=\{\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]>\mathbb{E}[R_k|\boldsymbol{x}_t,\boldsymbol{\beta}_k^{true}]+\delta\},\ k\neq i,k\in\mathcal{K}$. Then we have the following bound.

$$\begin{aligned} & r_t \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left\{\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right\}\cap\mathcal{E}(t,\delta)_{8,j}\right)\left(\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]-\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{true}]\right)\right) \\ +&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left\{\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right\}\cap\mathcal{E}(t,\delta)_{8,j}^c\right)\left(\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]-\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{true}]\right)\right) \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left\{\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right\}\cap\mathcal{E}(t,\delta)_{8,j}\right)(2R_{\max})\right) \qquad \text{(EC.41)} \\ +&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left\{\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right\}\cap\mathcal{E}(t,\delta)_{8,j}^c\right)(\delta)\right). \qquad \text{(EC.42)} \end{aligned}$$

The term in (EC.42) can be bounded as follows

$$\begin{aligned} &\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left\{\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right\}\cap\mathcal{E}(t,\delta)_{8,j}^c\right)(\delta)\right) \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\mathcal{E}(t,\delta)_{8,j}^c\right)(\delta)\right) \\ =&\sum_{j\neq i}\mathbb{P}\left(\mathcal{E}(t,\delta)_{8,j}^c\right)\delta \\ =&(|\mathcal{K}|-1)CR_{\max}\delta^2\le CR_{\max}|\mathcal{K}|\delta^2, \end{aligned} \tag{EC.43}$$

where the last inequality comes from assumption **A.2**. Now we consider the term in (EC.41), which can be bounded as follows

$$\begin{aligned} &\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left\{\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]\right\}\cap\mathcal{E}(t,\delta)_{8,j}\right)(2R_{\max})\right) \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]-\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{true}]>\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]-\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]+\delta\right)(2R_{\max})\right) \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\left|\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{whole}(t)]-\mathbb{E}[R_j|\boldsymbol{x}_t,\boldsymbol{\beta}_j^{true}]\right|>-\left|\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{whole}(t)]-\mathbb{E}[R_i|\boldsymbol{x}_t,\boldsymbol{\beta}_i^{true}]\right|+\delta\right)(2R_{\max})\right) \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(R_{\max}\sigma e^{2\sigma x_{\max}b}x_{\max}\|\boldsymbol{\beta}_k^{true}-\boldsymbol{\beta}_k^{whole}(t)\|_1>-R_{\max}\sigma e^{2\sigma x_{\max}b}x_{\max}\|\boldsymbol{\beta}_i^{true}-\boldsymbol{\beta}_i^{whole}(t)\|_1+\delta\right)(2R_{\max})\right) \\ \le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\|\boldsymbol{\beta}_k^{true}-\boldsymbol{\beta}_k^{whole}(t)\|_1+\|\boldsymbol{\beta}_i^{true}-\boldsymbol{\beta}_i^{whole}(t)\|_1\ge\frac{\delta}{R_{\max}\sigma e^{2\sigma x_{\max}b}x_{\max}}\right)(2R_{\max})\right), \end{aligned} \tag{EC.44}$$

where the second last inequality comes from the first part of the Lemma EC.10 and $\|\boldsymbol{\beta}\|_1 \le b$ in assumption **A.1**. From Proposition 6, we have the following inequality.

$$\mathbb{P}\left(\|\boldsymbol{\beta}_k^{whole}(t)-\boldsymbol{\beta}_k^{true}\|_2 \ge \sqrt{C_{\boldsymbol{\beta}}\frac{s^2}{T}}\right) \le \frac{12}{T+1}. \tag{EC.45}$$

As $\|\boldsymbol{\beta}_k^{whole}(t)-\boldsymbol{\beta}_k^{true}\|_2 \ge \frac{1}{\sqrt{s}}\|\boldsymbol{\beta}_k^{whole}(t)-\boldsymbol{\beta}_k^{true}\|_1$, (EC.45) implies

$$\mathbb{P}\left(\|\boldsymbol{\beta}_k^{whole}(t)-\boldsymbol{\beta}_k^{true}\|_1 \ge \sqrt{C_{\boldsymbol{\beta}}\frac{s^3}{T+1}}\right) \le \frac{12}{T+1}. \tag{EC.46}$$

Denote event $\mathcal{E}_9$ as follows

$$\mathcal{E}_9 = \{\|\boldsymbol{\beta}_k^{whole}(t)-\boldsymbol{\beta}_k^{true}\|_1 \ge \frac{\delta}{2R_{\max}\sigma e^{2\sigma x_{\max} b}x_{\max}},\ k\in\mathcal{K}\}. \tag{EC.47}$$

Combining (EC.44) and (EC.46), we have:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\|\boldsymbol{\beta}_j^{true}-\boldsymbol{\beta}_j^{whole}(t)\|_1\|\boldsymbol{\beta}_i^{true}-\boldsymbol{\beta}_i^{whole}(t)\|_1 \ge \frac{\delta}{R_{\max}\sigma e^{2\sigma x_{\max} b}x_{\max}}\right)(2R_{\max})\right)\\
=&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\|\boldsymbol{\beta}_j^{true}-\boldsymbol{\beta}_j^{whole}(t)\|_1+\|\boldsymbol{\beta}_i^{true}-\boldsymbol{\beta}_i^{whole}(t)\|_1 \ge \frac{\delta}{R_{\max}\sigma e^{2\sigma x_{\max} b}x_{\max}}\middle|\mathcal{E}_9\right)\mathbb{1}(\mathcal{E}_9)(2R_{\max})\right)\\
+&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\|\boldsymbol{\beta}_j^{true}-\boldsymbol{\beta}_j^{whole}(t)\|_1+\|\boldsymbol{\beta}_i^{true}-\boldsymbol{\beta}_i^{whole}(t)\|_1 \ge \frac{\delta}{R_{\max}\sigma e^{2\sigma x_{\max} b}x_{\max}}\middle|\mathcal{E}_9^c\right)\mathbb{1}(\mathcal{E}_9^c)(2R_{\max})\right)\\
\le&\mathbb{E}\left(\sum_{j\neq i}\mathbb{1}\left(\frac{1}{2}\delta+\frac{1}{2}\delta\ge\delta\middle|\mathcal{E}_9\right)\mathbb{1}\left(\mathcal{E}_9\right)\left(2R_{\max}\right)\right)+0\\
=&\mathbb{E}\left(\mathbb{1}\left(\mathcal{E}_9(t)\right)\left(2R_{\max}\right)\right) \le 2R_{\max}\mathbb{P}(\mathcal{E}_9).
\end{aligned} \tag{EC.48}$$

Furthermore, by setting $\delta = 2R_{\max}\sigma e^{2\sigma x_{\max} b}x_{\max}\sqrt{C_{\boldsymbol{\beta}}\frac{s^3}{T+1}}$, we have the following result:

$$r_t \le 2R_{\max}\mathbb{P}(\mathcal{E}_9)+CR_{\max}|\mathcal{K}|\delta^2 \le \frac{24R_{\max}|\mathcal{K}|}{T+1}+CR_{\max}|\mathcal{K}|\frac{4R_{\max}^2\sigma^2 e^{4\sigma x_{\max} b}x_{\max}^2 C_{\boldsymbol{\beta}}s^3}{T+1} = \frac{C_{R_3}}{T+1} \tag{EC.49}$$

where $C_{R_3} = 24R_{\max}|\mathcal{K}|+4e^{4\sigma_2 x_{\max} b}CR_{\max}^3|\mathcal{K}|x_{\max}^2 C_{\boldsymbol{\beta}}s^3$. Hence, the third part of the regret can be bounded as follows:

$$R_3(T) = \sum_{i=1,i\in\mathcal{R}(T)}^{T} r_t \le \sum_{i=1}^{T}\frac{C_{R_3}}{T} \le \int_1^T \frac{C_{R_3}}{t}dt \le C_{R_3}\log(T) \tag{EC.50}$$

Finally, the total regret bound can be obtained by combining the bounds for these three parts:

$$\begin{aligned}
R_1(T)+R_2(T)+R_3(T) &\le R_{\max}[|\mathcal{K}|(2+6C_0\log T)+T_0]+7R_{\max}|\mathcal{K}|\log(T+1)+C_{R_3}\log(T)\\
&\le R_{\max}(T_0+|\mathcal{K}|)+(6R_{\max}|\mathcal{K}|C_0+31R_{\max}|\mathcal{K}|+4\sigma^2 e^{4\sigma_2 x_{\max} b}CR_{\max}^3|\mathcal{K}|x_{\max}^2 C_{\boldsymbol{\beta}}s^3)\log(T+1)\\
&= O(|\mathcal{K}|s^2(s+\log d)\log T).
\end{aligned}$$

Appendix: Supplemental Lemmas and Proofs

LEMMA EC.1. *Let $\mathcal{A}$ be the set of iid samples. Under assumption A.1 and A.5, there exists a constant $\mu_0 > 0$ such that for all feasible $\boldsymbol{\xi}$ defined in assumption A.4 we have*

$$\mathbb{P}\left(\lambda_{\min}(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\xi})) \geq \frac{|\mathcal{A}|}{2n}\mu_0\right) \leq 1 - s\exp\left(-\frac{|\mathcal{A}|\mu_0}{8s\sigma_2 x^2_{\max}}\right). \tag{EC.51}$$

Proof of Lemma EC.1 Note that $f(\cdot|\cdot)$ is convex and has smooth gradient. We denote $\boldsymbol{z}'_j = \boldsymbol{x}_{j,\mathcal{S}}\sqrt{f''(r_j|\boldsymbol{x}^T_{j,\mathcal{S}}\boldsymbol{\xi}_{\mathcal{S}})}$. Combine with $f(\cdot|\cdot) = -\log g(\cdot|\cdot)$ and we have

$$\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,\mathcal{S}}x^T_{i,\mathcal{S}}f''(r_i|\boldsymbol{x}^T_{i,\mathcal{S}}\boldsymbol{\xi}_{\mathcal{S}}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{z}'_i(\boldsymbol{z}'_i)^T \succeq \lambda_{\min}\left(\frac{1}{n}\sum_{j\in\mathcal{A}}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T\right) I.$$

Then, we bound $\lambda_{\min}\left(\frac{1}{n}\sum_{j\in\mathcal{A}^c}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T\right)$ via Theorem 5.1.1 in Tropp et al. (2015) with $\epsilon = 1/2$:

$$\mathbb{P}\left(\lambda_{\min}(\frac{1}{n}\sum_{j\in\mathcal{A}}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \frac{1}{2}\lambda_{\min}(\mathbb{E}[\frac{1}{n}\sum_{j\in\mathcal{A}}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T])\right) \leq s\left(\frac{\exp(-1/2)}{\sqrt{1/2}}\right)^{\lambda_{\min}(\mathbb{E}[\frac{1}{n}\sum_{j\in\mathcal{A}}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T])/(s\sigma_2 x^2_{\max}/n)} \tag{EC.52}$$

$$\Rightarrow \mathbb{P}\left(\lambda_{\min}(\frac{1}{n}\sum_{j\in\mathcal{A}}^n\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \frac{|\mathcal{A}|}{2n}\lambda_{\min}(\mathbb{E}[\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T])\right) \leq s\exp\left(-\frac{\log(e/2)n\lambda_{\min}(\frac{|\mathcal{A}|}{n}\mathbb{E}[\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T])}{2s\sigma_2 x^2_{\max}}\right), \tag{EC.53}$$

where (EC.52) uses $0 \leq \lambda_{\min}(\frac{1}{n}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \lambda_{\max}(\frac{1}{n}\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \frac{s}{n}(z'_{\max})^2 = \frac{s}{n}\sigma_2 x^2_{\max}$ and the last inequality comes from the assumption **A.1**. As we only consider the significant dimensions, under assumption **A.4**, we can verify that there exists a $\mu_0 > 0$ such that $\mathbb{E}[\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T] = \mathbb{E}[\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}_{\mathcal{A}}(\boldsymbol{\xi})] \succeq \mu_0 I$. Then, we have $\mathbb{1}\left(\lambda_{\min}(\frac{1}{n}\sum_{j\in\mathcal{A}}^n\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \frac{|\mathcal{A}|}{2n}\lambda_{\min}(\mathbb{E}[\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T])\right) \geq \mathbb{1}\left(\lambda_{\min}(\frac{1}{n}\sum_{j\in\mathcal{A}}^n\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \frac{|\mathcal{A}|}{2n}\mu_0\right)$. Thus (EC.53) implies

$$\mathbb{P}\left(\lambda_{\min}(\frac{1}{n}\sum_{j\in\mathcal{A}}^n\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T) \leq \frac{|\mathcal{A}|}{2n}\mu_0\right) \leq s\exp\left(-\frac{\log(e/2)|\mathcal{A}|\lambda_{\min}(\mathbb{E}[\boldsymbol{z}'_j(\boldsymbol{z}'_j)^T])}{2s\sigma_2 x^2_{\max}}\right).$$

Combining with the fact $\log(e/2)/2 \geq 1/8$, Lemma EC.1 follows immediately.

LEMMA EC.2. *Let the whole sample size be n and iid random sample set be $\mathcal{A}$. If assumptions* **A.1**,**A.4** *and* **A.5** *hold, there exist $\mu_0 > 0$ such that for $t > 0$ we have*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\| \geq \frac{2nt}{|\mathcal{A}|\mu_0}\right) \leq s\exp\left(-\frac{|\mathcal{A}|\mu_0}{8s\sigma_2 x^2_{\max}}\right) + s\exp\left(-\frac{nt^2}{2s\sigma^2 x^2_{\max}}\right). \tag{EC.54}$$

Furthermore, if $|\mathcal{A}| \geq \frac{2s^2x^2_{\max}}{\mu_0}$ we have

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \geq \sqrt{\frac{8s^2\sigma^2x^2_{\max}n}{\mu_0^2|\mathcal{A}|^2}}\right) \leq s\exp\left(-\frac{\mu_0|\mathcal{A}|}{8s\sigma_2 x^2_{\max}}\right) + 2\exp\left(-\frac{C_h|\mathcal{A}|\mu_0}{2s\sigma_2 x^2_{\max}}\right), \tag{EC.55}$$

where C_h is a positive constant.

Proof of Lemma EC.2 From the definition of oracle solution, we know

$$\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{oracle}) = 0. \tag{EC.56}$$

Expanding (EC.56) at $\boldsymbol{\beta}^{true}$ we will have the following result for some $\boldsymbol{\xi} \in \{\tau\boldsymbol{\beta}^{oracle} + (1-\tau)\boldsymbol{\beta}^{true}, \tau \in [0,1]\}$.

$$\begin{aligned}
\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}) + \nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\xi)(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}) &= 0 \\
\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\xi)(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}) &= -\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}) \\
(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})^T\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\xi)(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}) &= -(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})^T\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}) \\
\lambda_{\min}(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\xi))\|(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\|_2^2 &\le \|(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\|_2\|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_2 \\
\lambda_{\min}(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\xi))\|(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\|_2 &\le \|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_2. \qquad \text{(EC.57)}
\end{aligned}$$

The $\lambda_{\min}(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\xi))$ term on the left hand side of (EC.57) can be lower bounded away 0 via Lemma EC.1 with high probability. Thus we only need to construct the upper bound for right-hand side of (EC.57) that can be expanded as follows

$$\|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_2 = \left\|\frac{1}{n}\sum_{j=1}^{n} x_{j\mathcal{S}}^T f^{'}(r_j|\boldsymbol{x}_{j,\mathcal{S}}^T\boldsymbol{\beta}^{true})\right\|_2. \qquad \text{(EC.58)}$$

Under assumption **A.5**, we have $|f^{'}(r_j|\boldsymbol{x}_{j,\mathcal{S}}^T\boldsymbol{\beta}^{true})| \le \sigma$. Combining with $\mathbb{E}[f^{'}(r_j|\boldsymbol{x}_{j,\mathcal{S}}^T\boldsymbol{\beta}^{true})] = 0$, we can verify that $f^{'}(r_j|\boldsymbol{x}_{j,\mathcal{S}}^T\boldsymbol{\beta}^{true})$ is a σ-subgaussian random variable. From Hoeffding inequality, there exists a $t > 0$ such that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} x_{ji}^T f^{'}(r_j|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})\right| \ge t\right) \le \exp\left(-\frac{nt^2}{2\sigma^2 x_{\max}^2}\right) \quad \forall i \in \mathcal{S}. \qquad \text{(EC.59)}$$

Hence, we have

$$\begin{aligned}
\mathbb{P}\left(\|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_2 \ge t\right) = \mathbb{P}\left(\left\|\frac{1}{n}\sum_{j=1}^{n} x_{ji}^T f^{'}(r_j|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})\right\|_2 \ge t\right) &\le \mathbb{P}\left(\sqrt{|\mathcal{S}|}\left\|\frac{1}{n}\sum_{j=1}^{n} x_{ji}^T f^{'}(r_j|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})\right\|_\infty \ge t\right) \\
&\le s\exp\left(-\frac{nt^2}{2s\sigma^2 x_{\max}^2}\right), \qquad \text{(EC.60)}
\end{aligned}$$

where the inequality in (EC.60) follows from $|\mathcal{S}| \le s$. Combining (EC.60), (EC.57) and Lemma EC.1, the statement in (EC.54) follows.

Now, the first half of Lemma EC.2 has been proven, and we switch to the second half. Denote $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, ..., \epsilon_n]$ where $\epsilon_j = f^{'}(r_j|\boldsymbol{x}_{j,\mathcal{S}}^T\boldsymbol{\beta}^{true}),\ j = 1, 2, .., n$. Then $\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{\boldsymbol{true}})$ can be rewritten as $\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{\boldsymbol{true}}) = \frac{1}{n}\boldsymbol{X}_{\mathcal{S}}\boldsymbol{\epsilon}$ with $\boldsymbol{X}_{\mathcal{S}} = [\boldsymbol{x}_{1,\mathcal{S}}, ..., \boldsymbol{x}_{n,\mathcal{S}}]$. Using the Hanson-Wright inequality (Theorem 1.1 in Rudelson et al. 2013), we have

$$\begin{aligned}
&\mathbb{P}\{|\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}]| > \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T X_{\mathcal{S}})\boldsymbol{\epsilon}]\} \\
\le& 2\exp\left(-C_h\min\left\{\frac{\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}]}{\sigma^2\|\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}}\|_2}, \frac{(\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}])^2}{\sigma^4\|\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}}\|_F^2}\right\}\right) \\
\le& 2\exp\left(-C_h\min\left\{\frac{\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}\frac{\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}]}{\sigma^2}, \frac{\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})^2}{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})^2}\frac{\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}]^2}{s\sigma^4}\right\}\right) \\
\le& 2\exp\left(-C_h\min\left\{n\frac{\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}, \frac{n^2}{s}\frac{\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})^2}{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})^2}\right\}\right) \\
\le& 2\exp\left(-n\frac{C_h\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}\right), \qquad \text{(EC.61)}
\end{aligned}$$

where C_h is a positive constant and $\mathbb{E}_{\boldsymbol{\epsilon}}$ denote the expectation with respect to $\boldsymbol{\epsilon}$. The last inequality, (EC.61), holds when $n \geq s\frac{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}{\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}$. Define the event $\mathcal{E}_1$ as follows

$$\mathcal{E}_1 = \left\{ |\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon} - \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}]| \leq \mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}] \right\}. \tag{EC.62}$$

Under event $\mathcal{E}_1$, we have

$$\|\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}\boldsymbol{\epsilon}\|_2 \leq \sqrt{\frac{1}{n}\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}} \leq \sqrt{\frac{2}{n}\mathbb{E}_{\boldsymbol{\epsilon}}[\boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}]}. \tag{EC.63}$$

Let $\boldsymbol{P}_j = \boldsymbol{X}_{\mathcal{S}}^T(\boldsymbol{X}_{\mathcal{S}}\boldsymbol{X}_{\mathcal{S}}^T)^{-1}\boldsymbol{X}_{\mathcal{S}}$. We have $(\boldsymbol{P}_j\boldsymbol{\epsilon})^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})(\boldsymbol{P}_j\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\boldsymbol{\epsilon}$, and (EC.63) implies the following result.

$$\|\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}\boldsymbol{\epsilon}\|_2 \leq \sqrt{\frac{2}{n}\mathbb{E}_{\boldsymbol{\epsilon}}[(\boldsymbol{P}_j\boldsymbol{\epsilon})^T(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})(\boldsymbol{P}_j\boldsymbol{\epsilon})]} \leq \sqrt{\frac{2}{n}\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\mathbb{E}_{\boldsymbol{\epsilon}}[\|\boldsymbol{P}_j\boldsymbol{\epsilon}\|_2^2]} \leq \sqrt{\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})\frac{2s\sigma^2}{n}}, \tag{EC.64}$$

where the last inequality comes the facts that $\mathbb{E}_{\boldsymbol{\epsilon}}[\|\boldsymbol{P}_j\boldsymbol{\epsilon}\|_2^2] = s\sigma^2$ in which $\boldsymbol{P}_j$ can be viewed as a projection matrix from n dimension to s dimension and ϵ_j is a σ-subguassian random variable. Therefore, from $\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{\boldsymbol{true}}) = \frac{1}{n}\boldsymbol{X}_{\mathcal{S}}\boldsymbol{\epsilon}$, $\lambda_{\max}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}}) \leq sx_{\max}^2$ and (EC.61)-(EC.64), we have the following inequalities.

$$\mathbb{P}\left(\|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_2 \leq \sqrt{\frac{2s^2\sigma^2x_{\max}^2}{n}} \right) \geq 1 - 2\exp\left(-n\frac{C_h\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})}{sx_{\max}^2} \right). \tag{EC.65}$$

Since $\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}} = \frac{1}{n}\sum_{j=1}^n \boldsymbol{x}_{j,\mathcal{S}}\boldsymbol{x}_{j,\mathcal{S}}^T \succeq \frac{1}{n}\sum_{j=1}^n \boldsymbol{x}_{j,\mathcal{S}}\boldsymbol{x}_{j,\mathcal{S}}^T\frac{f''(r_j|\boldsymbol{x}_{j,\mathcal{S}}^T\boldsymbol{\xi})}{\sigma_2} = \frac{1}{\sigma_2}\nabla_{\mathcal{S},\mathcal{S}}^2\mathcal{L}(\boldsymbol{\xi})$. We then may apply the Lemma EC.1 to further lower bound $\lambda_{\min}(\frac{1}{n}\boldsymbol{X}_{\mathcal{S}}^T\boldsymbol{X}_{\mathcal{S}})$ by $\frac{|\mathcal{A}|\mu_0}{2n\sigma_2}$ for some $\mu_0 > 0$ with high probability and then (EC.55) follows.

LEMMA EC.3. *If there exists K and σ_0 such that $K^2\left(E[\exp(\boldsymbol{z}_{t,i}^2/K^2) - 1]\right) \leq \sigma_0^2$, then the following probability bound will hold for all $t > 0$:*

$$P\left\{ \|\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T - E[\boldsymbol{z}_j\boldsymbol{z}_j^T]\|_\infty \geq 2K^2t + 2K\sigma_0\sqrt{2t} + 2K\sigma_0\lambda\left(\frac{K}{\sigma_0}, n, \binom{d}{2}\right) \right\} \leq \exp\left(-nt\right) \tag{EC.66}$$

where $\lambda\left(\frac{K}{\sigma_0}, n, \binom{d}{2}\right) = \sqrt{\frac{2\log(d(d-1))}{n}} + \frac{K\log(d(d-1))}{n}$.

Proof of EC.3 From the exercise 14.3 in Bühlmann and Van De Geer (2011).

LEMMA EC.4. *If there exist κ_0, $\mathcal{S}$, and $\boldsymbol{z}_j$, $j = 1, 2, .., n$ such that $\|u_S\|_1^2 \leq \frac{|\mathcal{S}|}{\kappa_0}u^T\mathbb{E}[\boldsymbol{z}_j\boldsymbol{z}_j^T]u$ holds for all $u \in \mathcal{U} \doteq \{u : \|u_{S^c}\|_1 \leq 3\|u_S\|\}$ and $\left\|\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T - \mathbb{E}[\boldsymbol{z}_j\boldsymbol{z}_j^T]\right\| \leq \frac{\kappa}{32|S|}$, then for all $u \in \mathcal{U}$, the follow inequality holds:*

$$\|u_S\|_1^2 \leq \frac{|S|}{\kappa_0/2}u^T\left[\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T\right]u \tag{EC.67}$$

Proof of EC.4 From Corollary 6.8 in Bühlmann and Van De Geer (2011).

LEMMA EC.5. *Let $\boldsymbol{x}_j$, $j=1,2,...,n$, be random iid samples. Under assumptions* **A.4** *and* **A.5**, *the follow inequality holds for all $\boldsymbol{u}$ such that $\|\boldsymbol{u}_{\mathcal{S}^c}\|_1 \le 3\|\boldsymbol{u}_{\mathcal{S}}\|_1$:*

$$\mathbb{P}\left(\frac{\kappa}{2s}\|\boldsymbol{u}_{\mathcal{S}}\|_1^2 \le u^T\nabla^2\mathcal{L}(\boldsymbol{\beta})\boldsymbol{u}\right) \ge 1-\exp(-C_1 n), \tag{EC.68}$$

where $C_1 = \min\left\{1, \kappa^2/\left(192s\sigma_2 x_{\max}^2(3+2\sqrt{\sigma_2}x_{\max})\right)^2\right\}$.

Proof of EC.5 From the definition of $\mathcal{L}(\boldsymbol{\beta})$, we have $\nabla^2\mathcal{L}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{j=1}^n \boldsymbol{x}_j\boldsymbol{x}_j^T f''(r_j, \boldsymbol{x}_j^T\boldsymbol{\xi})$. Under assumption **A. 5**, we know that f is convex with smooth gradient. We may denote $\boldsymbol{z}_j = \boldsymbol{x}_j\sqrt{f''(r_j,\boldsymbol{x}_j^T\boldsymbol{\xi})}$ and then get $\nabla^2\mathcal{L}(\boldsymbol{\xi}) = \frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T$. Furthermore, under assumption **A.1** and **A.5**, we have $|f''(r_j|\boldsymbol{x}_j^T\boldsymbol{\xi})| \le \sigma_2$ and $\|\boldsymbol{x}\|_\infty \le x_{\max}$, which implies that $\boldsymbol{z}_j$ is element-wise bounded by $z_{\max} = \|\boldsymbol{z}_j\|_\infty = \left\|\boldsymbol{x}_j\sqrt{f''(r_j|\boldsymbol{x}_j^T\boldsymbol{\xi})}\right\|_\infty \le \sqrt{\sigma_2}x_{\max}$. Since $\boldsymbol{z}_j$ is bounded, it will satisfy the definition of the subguassian random variable. We can use the Lemma EC.3 as a bridge to connect the sample matrix $\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T$ to its population counterpart $\mathbb{E}[\boldsymbol{z}_j\boldsymbol{z}_j^T]$. Let $K = z_{\max}$ and $\sigma_0 = \sqrt{2}z_{\max}$ and we will have $K^2\left(E[\exp(\boldsymbol{z}_{t,i}^2/K^2)-1]\right) \le z_{\max}^2(e-1) \le \sigma_0^2$ for all $t \ge 0$ and $i = 1,2,...,d$. Therefore, under Lemma EC.3, for $t>0$, we have

$$P\left\{\left\|\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T - E[\boldsymbol{z}_j\boldsymbol{z}_j^T]\right\|_\infty \ge 2z_{\max}^2 t + 4z_{\max}^2\sqrt{t} + \sqrt{8}z_{\max}^2\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right)\right\} \le \exp(-nt), \tag{EC.69}$$

where $\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right) = \sqrt{\frac{2\log(d(d-1))}{n}} + \frac{z_{\max}\log(d(d-1))}{n}$. (EC.69) indicates that when the sample size is large enough, $\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T$ will not be far away from $\mathbb{E}[\boldsymbol{z}_j\boldsymbol{z}_j^T]$ element-wise with high probability.

Now we only need to show that if $\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T$ is close enough to $\mathbb{E}[\boldsymbol{z}_j\boldsymbol{z}_j^T]$, $\nabla^2\mathcal{L}$ satisfies (EC.68). To this end, we need Lemma EC.4. We set $n \ge \log d/C_1$ and $t = C_1$ in (EC.69). Then the following inequalities hold.

$$2z_{\max}^2 t + 4z_{\max}^2\sqrt{t} \le 2z_{\max}^2\sqrt{C_1} + 4z_{\max}^2\sqrt{C_1} = 6z_{\max}^2\sqrt{C_1} \tag{EC.70}$$

$$\sqrt{8}z_{\max}^2\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right) \le \sqrt{8}z_{\max}^2\left(\sqrt{\frac{2\log(d^2)}{n}} + \frac{z_{\max}\log(d^2)}{n}\right) \le 8\sqrt{2}z_{\max}^2(1+z_{\max})\sqrt{C_1}, \tag{EC.71}$$

where (EC.70) and (EC.71) use $\log d/n \le C_1 \le 1$. Combining (EC.70) and (EC.71), we have

$$\begin{aligned} 2z_{\max}^2 t + 4z_{\max}^2\sqrt{t} + \sqrt{8}z_{\max}^2\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right) &\le 2z_{\max}^2\left(3+4\sqrt{2}(1+z_{\max})\right)\sqrt{C_1} \\ &\le 6z_{\max}^2\left(3+2z_{\max}\right)\sqrt{C_1} \le \frac{\kappa}{32s}, \end{aligned} \tag{EC.72}$$

where (EC.72) uses $\sqrt{2} \le \frac{3}{2}$ and $C_1 \le \kappa^2/\left(192s\sigma_2 x_{\max}^2\left(3+2\sqrt{\sigma_2}x_{\max}\right)\right)^2 \le \kappa^2/\left(192sz_{\max}^2\left(3+2z_{\max}\right)\right)^2$. Then, (EC.69) can satisfy the following inequality.

$$\mathbb{P}\left\{\left\|\frac{1}{n}\sum_{j=1}^n \boldsymbol{z}_j\boldsymbol{z}_j^T - E[\boldsymbol{z}_j\boldsymbol{z}_j^T]\right\|_\infty \le \frac{\kappa}{32s}\right\} \ge 1-\exp(-C_1 n). \tag{EC.73}$$

The statement of Lemma EC.5 follows by combining (EC.73) with Lemma EC.4.

LEMMA EC.6. *Let $\mathcal{A}_k^{iid}$ be the index set such that for all $i \in \mathcal{A}_k^{iid}$, $\boldsymbol{x}_i$ are random iid samples. If for all $\boldsymbol{u}$ such that $\|\boldsymbol{u}_{S^c}\|_1 \leq 3\|\boldsymbol{u}_S\|_1$, we have $\frac{\kappa}{2s}\|\boldsymbol{u}_S\|_1^2 \leq \boldsymbol{u}^T\nabla^2\mathcal{L}_{\mathcal{A}_k^{iid}}(\boldsymbol{\xi})\boldsymbol{u}$, then the follow inequality holds:*

$$\frac{|\mathcal{A}^{iid}|\kappa}{2ns}\|\boldsymbol{u}_S\|_1^2 \leq \boldsymbol{u}^T\nabla^2\mathcal{L}(\boldsymbol{\xi})\boldsymbol{u}, \tag{EC.74}$$

where $\mathcal{L}_{\mathcal{A}}(\boldsymbol{\beta})$ denotes the likelihood function with samples only in $\mathcal{A}^{iid}$.

proof of EC.6 We can rewrite $\nabla\mathcal{L}(\boldsymbol{\xi})$ with $\boldsymbol{z}_j = \boldsymbol{x}_j\sqrt{f''(r_j|\boldsymbol{x}_j^T\boldsymbol{\xi})}$ as follow.

$$\begin{aligned}
\boldsymbol{u}^T\nabla^2\mathcal{L}(\boldsymbol{\xi})\boldsymbol{u} &= \boldsymbol{u}^T\left[\frac{1}{n}\sum_{j\in\mathcal{A}_k^{iid}}\boldsymbol{z}_j\boldsymbol{z}_j^T\right]\boldsymbol{u} + \boldsymbol{u}^T\left[\frac{1}{n}\sum_{j\in(\mathcal{A}_k^{iid})^c}\boldsymbol{z}_j\boldsymbol{z}_j^T\right] \\
&\geq \frac{|\mathcal{A}_k^{iid}|}{n}^T\left[\frac{1}{|\mathcal{A}_k^{iid}|}\sum_{j\in\mathcal{A}_k^{iid}}\boldsymbol{z}_j\boldsymbol{z}_j^T\right] \\
&\geq \frac{|\mathcal{A}_k^{iid}|}{n}^T\nabla\mathcal{L}_{\mathcal{A}}(\boldsymbol{\xi}) \\
&\geq \frac{|\mathcal{A}_k^{iid}|}{n}\frac{\kappa}{2s}\|\boldsymbol{u}_S\|_1^2 \\
&= \frac{|\mathcal{A}_k^{iid}|\kappa}{2ns}\|\boldsymbol{u}_S\|_1^2.
\end{aligned} \tag{EC.75}$$

LEMMA EC.7. *Let the whole sample size be n and the set for iid random sample in U_k be $\mathcal{A}$, $k \in \mathcal{K}$. If assumptions* **A.4** *and* **A.5** *hold, then the follow result holds.*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{lasso} - \boldsymbol{\beta}^{true}\|_1 \leq \frac{96ns\lambda}{|\mathcal{A}|\kappa}\right) \geq 1 - \exp\left(-C_1|\mathcal{A}|\right) - \exp\left(-\frac{n\lambda^2}{8x_{\max}^2} + \log d\right), \tag{EC.76}$$

where $C_1 = \min\left\{1, \kappa^2/\left(192s\sigma_2 x_{\max}^2(3+2\sqrt{\sigma_2}x_{\max})\right)^2\right\}$.

Proof of lemma EC.7 Let $\mathcal{L}_{\mathcal{A}}(\beta)$ be the loss function only includes samples in $\mathcal{A}$. Under assumption **A.4**, we have

$$\frac{\kappa}{s}\|\boldsymbol{u}_{\mathcal{S}}\|_1^2 \leq \boldsymbol{u}^T\mathbb{E}[\nabla^2\mathcal{L}_{\mathcal{A}}(\boldsymbol{\xi})]\boldsymbol{u}, \tag{EC.77}$$

for all u such that $\|\boldsymbol{u}_{\mathcal{S}^c}\|_1 \leq 3\|\boldsymbol{u}_{\mathcal{S}}\|_1$. The following result follows from (EC.77) and Lemma EC.5:

$$\mathbb{P}\left(\frac{\kappa}{2s}\|\boldsymbol{u}_{\mathcal{S}}\|_1^2 \leq \boldsymbol{u}^T\nabla^2\mathcal{L}_{\mathcal{A}}(\boldsymbol{\xi})\boldsymbol{u}\right) \geq 1 - \exp(-C_1|\mathcal{A}|). \tag{EC.78}$$

Moreover, via Lemma EC.6, for all $\boldsymbol{u}$ such that $\|\boldsymbol{u}_{\mathcal{S}^c}\|_1 \leq 3\|\boldsymbol{u}_{\mathcal{S}}\|_1$ the follow inequality holds.

$$\mathbb{P}\left(\frac{|\mathcal{A}|\kappa}{2ns}\|\boldsymbol{u}_{\mathcal{S}}\|_1^2 \leq \boldsymbol{u}^T\nabla^2\mathcal{L}(\boldsymbol{\xi})\boldsymbol{u}\right) \geq 1 - \exp\left(-C_1|\mathcal{A}|\right). \tag{EC.79}$$

Since $\boldsymbol{\beta}^{lasso}$ is the optimal solution to the Lasso problem, we can ensure the following inequality:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}^{lasso}) + \lambda\|\boldsymbol{\beta}^{lasso}\|_1 &\leq \mathcal{L}(\boldsymbol{\beta}^{true}) + \lambda\|\boldsymbol{\beta}^{true}\|_1 \\
\mathcal{L}(\boldsymbol{\beta}^{lasso}) - \mathcal{L}(\boldsymbol{\beta}^{true}) + \lambda\|\boldsymbol{\beta}^{lasso}\|_1 &\leq \lambda\|\boldsymbol{\beta}^{true}\|_1
\end{aligned} \tag{EC.80}$$

$$\nabla\mathcal{L}(\boldsymbol{\beta}^{true})^T(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})+\lambda\|\boldsymbol{\beta}^{lasso}\|_1\leq\lambda\|\boldsymbol{\beta}^{true}\|_1 \tag{EC.81}$$

$$-\|\nabla\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty\|\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true}\|_1+\lambda\|\boldsymbol{\beta}^{lasso}\|_1\leq\lambda\|\boldsymbol{\beta}^{true}\|_1, \tag{EC.82}$$

where (EC.81) uses the convexity of $\mathcal{L}(\boldsymbol{\beta}^{lasso})$. Denote event $\mathcal{E}_0$ as follows.

$$\mathcal{E}_0=\left\{\|\nabla\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty<\frac{1}{2}\lambda\right\}. \tag{EC.83}$$

Under $\mathcal{E}_0$, (EC.82) can be further simplified into

$$\begin{aligned}
-\frac{1}{2}\lambda\|\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true}\|_1+\lambda\|\boldsymbol{\beta}^{lasso}\|_1&\leq\lambda\|\boldsymbol{\beta}^{true}\|_1\\
-\frac{1}{2}\|\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true}\|_1+\|\boldsymbol{\beta}^{lasso}\|_1&\leq\|\boldsymbol{\beta}^{true}\|_1\\
-\frac{1}{2}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1-\frac{1}{2}\|\boldsymbol{\beta}_{S^c}^{lasso}-\boldsymbol{\beta}_{S^c}^{true}\|_1+\|\boldsymbol{\beta}_S^{lasso}\|_1+\|\boldsymbol{\beta}_{S^c}^{lasso}\|_1&\leq\|\boldsymbol{\beta}_S^{true}\|_1+\|\boldsymbol{\beta}_{S^c}^{true}\|_1.
\end{aligned} \tag{EC.84}$$

As $\boldsymbol{\beta}_{S^c}^{true}=\mathbf{0}$ by definition, we then have

$$\begin{aligned}
-\frac{1}{2}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1-\frac{1}{2}\|\boldsymbol{\beta}_{S^c}^{lasso}-\boldsymbol{\beta}_{S^c}^{true}\|_1+\|\boldsymbol{\beta}_S^{lasso}\|_1+\|\boldsymbol{\beta}_{S^c}^{lasso}-\mathbf{0}\|_1&\leq\|\boldsymbol{\beta}_S^{true}\|_1+0\\
-\frac{1}{2}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1-\frac{1}{2}\|\boldsymbol{\beta}_{S^c}^{lasso}-\boldsymbol{\beta}_{S^c}^{true}\|_1+\|\boldsymbol{\beta}_S^{lasso}\|_1+\|\boldsymbol{\beta}_{S^c}^{lasso}-\boldsymbol{\beta}_{S^c}^{true}\|_1&\leq\|\boldsymbol{\beta}_S^{true}\|_1+0
\end{aligned} \tag{EC.85}$$

Rearrange (EC.85) and we may have

$$\|\boldsymbol{\beta}_{S^c}^{lasso}-\boldsymbol{\beta}_{S^c}^{true}\|_1\leq 3\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1 \tag{EC.86}$$

Denote $\boldsymbol{u}=\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true}$. Then, we have $\|\boldsymbol{u}_{S^c}\|_1\leq 3\|\boldsymbol{u}_S\|_1$. Connecting (EC.79), we can obtain

$$\mathbb{P}\left((\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})^T\nabla^2\mathcal{L}(\boldsymbol{\xi})(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})\geq\frac{|\mathcal{A}|\kappa}{2ns}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2\right)\geq 1-\exp\left(-C_1|\mathcal{A}|\right). \tag{EC.87}$$

Now, we turn back to (EC.80) and use the Taylor expansion on $\mathcal{L}(\boldsymbol{\beta}^{lasso})$ at $\boldsymbol{\beta}^{true}$ the following inequality holds for some $\boldsymbol{\xi}$.

$$\nabla\mathcal{L}(\boldsymbol{\beta}^{true})^T(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})+\frac{1}{2}(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})^T\nabla^2\mathcal{L}(\boldsymbol{\xi})(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})+\lambda\|\boldsymbol{\beta}^{lasso}\|_1\leq\lambda\|\boldsymbol{\beta}^{true}\|_1. \tag{EC.88}$$

Combining (EC.82) and (EC.88), we know that with probability $1-\exp(-C_1n)$, the follow results hold.

$$\begin{aligned}
&-\|\nabla\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty\|(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})\|_1+\frac{|\mathcal{A}|\kappa}{4ns}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2+\lambda\|\boldsymbol{\beta}^{lasso}\|_1\leq\lambda\|\boldsymbol{\beta}^{true}\|_1\\
\Rightarrow&-\|\nabla\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty\|(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})\|_1+\frac{|\mathcal{A}|\kappa}{4ns}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2\leq\lambda(\|\boldsymbol{\beta}^{true}\|_1-\|\boldsymbol{\beta}^{lasso}\|_1)\\
\Rightarrow&-\|\nabla\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty\|(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})\|_1+\frac{|\mathcal{A}|\kappa}{4ns}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2\leq\lambda\|\boldsymbol{\beta}^{true}-\boldsymbol{\beta}^{lasso}\|_1
\end{aligned} \tag{EC.89}$$

Under event $\mathcal{E}_0$, we have

$$\begin{aligned}
&-\frac{1}{2}\lambda\|(\boldsymbol{\beta}^{lasso}-\boldsymbol{\beta}^{true})\|_1+\frac{|\mathcal{A}|\kappa}{4ns}\|\boldsymbol{\beta}_{\mathcal{S}}^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2\leq\lambda\|\boldsymbol{\beta}^{true}-\boldsymbol{\beta}^{lasso}\|_1\\
\Rightarrow&\frac{|\mathcal{A}|\kappa}{4ns}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2\leq\frac{3}{2}\lambda\|\boldsymbol{\beta}^{true}-\boldsymbol{\beta}^{lasso}\|_1\\
\Rightarrow&\frac{|\mathcal{A}|\kappa}{4ns}\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1^2\leq 6\lambda\|\boldsymbol{\beta}_{\mathcal{S}}^{true}-\boldsymbol{\beta}_{\mathcal{S}}^{lasso}\|_1\\
\Rightarrow&\|\boldsymbol{\beta}_S^{lasso}-\boldsymbol{\beta}_S^{true}\|_1\leq\frac{24ns}{|\mathcal{A}|\kappa}\lambda
\end{aligned} \tag{EC.90}$$

$$\Rightarrow \|\boldsymbol{\beta}^{lasso} - \boldsymbol{\beta}^{true}\|_1 \leq \frac{96ns}{|\mathcal{A}|\kappa}\lambda, \tag{EC.91}$$

where (EC.90) and (EC.91) use $\|\boldsymbol{\beta}^{lasso}_{S^c} - \boldsymbol{\beta}^{true}_{S^c}\|_1 \leq 3\|\boldsymbol{\beta}^{lasso}_{S} - \boldsymbol{\beta}^{true}_{S}\|_1$ in (EC.86).

Now, we assess the probability of event $\mathcal{E}_0$. The i-th element of $\nabla\mathcal{L}(\boldsymbol{\beta}^{ture})$ is $\frac{1}{n}\sum_{i=1}^{n} x_{ji} f^{'}(r_i|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})$. Denote $X_{ji} = x_{ji} f^{'}(r_i|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})$ for $j = 1, 2, ...n$. Under assumptions **A.1** and **A.5** , X_{ji} are $x_{\max}\sigma$-subguassian random variables with mean 0. We can use Hoeffding inequality to build the following probability bound.

$$\begin{aligned}
&\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} x_{ji} f^{'}(r_i|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})| \geq t) \leq \exp\left(-\frac{nt^2}{2\sigma^2 x_{\max}^2}\right) \\
\Rightarrow &\mathbb{P}\left(\max_j |\frac{1}{n}\sum_{i=1}^{n} x_{ji} f^{'}(r_i|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})| \leq t\right) \geq 1 - \sum_{j=1}^{p}\mathbb{P}\left(|\frac{1}{n}\sum_{i=1}^{n} x_{ji} f^{'}(r_i|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})| \geq t\right) \\
&\geq 1 - d\exp\left(-\frac{nt^2}{2\sigma^2 x_{\max}^2}\right)
\end{aligned} \tag{EC.92}$$

Set $t = \frac{1}{2}\lambda$, and we will have event $\mathcal{E}_0$ defined in (EC.83) holds with at least probability $1 - \exp(-\frac{n\lambda^2}{8x_{\max}^2} + \log d)$. The desirable result follows by (EC.87) and (EC.92).

LEMMA EC.8. *Let $t_0 = 2C_0|\mathcal{K}|$, $C_0 = \max\{10, 16/p^*\}$, and $T \geq \max\{(t_0+1)^2/e^2 - 1, e\}$. Under assumptions* **A.3** *and* **A.4**, *the following statements hold.*

1. $\mathbb{P}\left\{n < \frac{1}{2}C_0(T+1) \text{ or } n > 6C_0\log(T+1)\right\} \leq \frac{2}{T+1}$
2. $\mathbb{P}\left\{|\mathcal{A}| < \frac{1}{4}p^*C_0\log(T+1)\right\} \leq \frac{1}{T+1}$
3. $\mathbb{P}\left\{|\mathcal{A}|/n < \frac{1}{24}p^*\right\} \leq \frac{3}{T+1}$

Proof of EC.8 ***To show statement 1.*** From Proposition 2, we have

$$\mathbb{P}\left(C_0(1+\log(T+1)-\log(t_0)) \leq n \leq 3C_0(1+\log(T)-\log(t_0))\right) \geq 1 - \frac{2}{T+1}. \tag{EC.93}$$

As we have $T \geq e$, the following result holds.

$$3C_0(1+\log(T)-\log(t_0)) \leq 3C_0(\log(T)+\log(T)-0) \leq 6C_0\log(T) < 6C_0\log(T+1). \tag{EC.94}$$

From $T \geq (t_0+1)^2/e^2 + 1 \Rightarrow \frac{1}{2}\log(T+1) - \log(t_0+1) \geq -1$, we have

$$\begin{aligned}
C_0(1+\log(T+1)-\log(t_0+1)) &= C_0(1+\frac{1}{2}\log(T+1)+\frac{1}{2}\log(T+1)-\log(t_0+1)) \\
&\geq C_0(1+\frac{1}{2}\log(T+1)-1) \\
&= \frac{1}{2}C_0\log(T+1).
\end{aligned} \tag{EC.95}$$

The statement 1 is obtained by combining (EC.94),(EC.95) and (EC.93).

To show statement 2. In assumption **A.4**, we assume that for $\boldsymbol{x} \in U_k,\ k \in \mathcal{K}$, the restricted eigenvalue condition is held. And under Assumption **A.3**, we have $\mathbb{P}(\boldsymbol{x} \in U_k) \geq p^*$. Thus, among all n samples, the expected number of samples belong to U_k will be lower bounded by:

$$\mathbb{E}[\mathbb{1}(\boldsymbol{x} \in U_k)] \geq p^*C_0(1+\log(T+1)-\log(t_0+1)). \tag{EC.96}$$

Since $T > (t_0+1)^2/e^2 - 1$ implies $\frac{1}{2}\log(T+1) > \log(t_0+1) - 1$. (EC.96) can be simplified into the following inequality.

$$\mathbb{E}[\sum_{i=1}^{n} \mathbb{1}(x_i \in U_k)] \geq \frac{1}{2} p^* C_0 \log(T+1). \tag{EC.97}$$

We apply the Chernoff inequality on $\sum_{i=1}^{n} \mathbb{1}(x_i \in U)$:

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}(\boldsymbol{x}_i \in U_k) < \frac{1}{2}\mathbb{E}[\sum_{i=1}^{n} \mathbb{1}(\boldsymbol{x}_i \in U_k)]\right) \leq \exp\left(-\frac{1}{8}\mathbb{E}[\sum_{i=1}^{n} \mathbb{1}(\boldsymbol{x}_i \in U_k)]\right) \\
\Rightarrow &\mathbb{P}\left(\sum_{i=1}^{n} \mathbb{1}(\boldsymbol{x}_i \in U_k) < \frac{1}{4} p^* C_0 \log(T+1)\right) \leq \exp\left(-\frac{1}{16} p^* C_0 \log(T+1)\right),
\end{aligned} \tag{EC.98}$$

where (EC.98) uses (EC.97). The statement 2 of Lemma EC.8 can be proved by (EC.98) with $C_0 \geq 16/p^*$.

To show statement 3. Notice that the follow result hold.

$$\begin{aligned}
\left\{|\mathcal{A}|/n \geq \frac{1}{24}p^*\right\} &\supseteq \left\{|\mathcal{A}| \geq \frac{1}{4}C_0 p^* \log(T+1)\right\} \cap \{n \leq 6C_0 \log(T+1)\} \\
&= \left(\left\{|\mathcal{A}| < \frac{1}{4}C_0 p^* \log(T+1)\right\} \cup \{n > 6C_0 \log(T+1)\}\right)^c.
\end{aligned} \tag{EC.99}$$

Hence we can obtain

$$\begin{aligned}
\mathbb{P}\left\{|\mathcal{A}|/n \geq \frac{1}{24}p^*\right\} &\geq \mathbb{P}\left\{\left(\left\{|\mathcal{A}_k| < \frac{1}{4}C_0 p^* \log(T+1)\right\} \cup \{n > 6C_0 \log(T+1)\}\right)^c\right\} \\
&= 1 - \mathbb{P}\left\{\left\{|\mathcal{A}| < \frac{1}{4}C_0 p^* \log(T+1)\right\} \cup \{n > 6C_0 \log(T+1)\}\right\} \\
&= 1 - \mathbb{P}\left\{|\mathcal{A}| < \frac{1}{4}C_0 p^* \log(T+1)\right\} - \mathbb{P}\left\{n > 6C_0 \log(T+1)\right\}.
\end{aligned} \tag{EC.100}$$

The remaining part follows by combining the statement 1 and statement 2 with (EC.100).

LEMMA EC.9. *Let* $t_0 = 2C_0|\mathcal{K}|$, $T \geq \max\{(t_0+1)^2/e^2 - 1, e\}$, $\lambda = C_5\sqrt{1 + \frac{\log d}{\log(T+1)}}$, *and* $a > \frac{2304s}{p^*\kappa}$. *If assumptions* **A.1**,**A.3**,**A.4** *and* **A.5** *hold, we have*

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_1 \leq \min\left\{\frac{1}{\sigma x_{\max}}, \frac{h}{4e\sigma R_{\max} x_{\max}}\right\}\right) \geq 1 - \frac{7}{T+1}, \tag{EC.101}$$

where

$$C_0 = \max\left\{10, \frac{16}{p^*}, \frac{4}{p^* C_1}, \frac{4x_{\max}^2}{C_5^2}\left((\frac{1}{4} - \frac{576s}{p^*\kappa a})\min\left\{1, \frac{\mu_0 p^*}{192 s x_{\max}^2}\right\}\right)^{-2}, \frac{32\sigma_2 s x_{\max}^2(1+\log s)}{p^*\mu_0}, \frac{4\sigma^2 x_{\max}^2(1+\log s)}{t^2}\right\},$$

$t \leq \min\left\{\frac{\mu_0 p^*\sqrt{\tilde{C}_2\lambda}}{48}, \frac{p^*\mu_0}{48\sigma\sqrt{s}x_{\max}}, \frac{hp^*\mu_0}{192e\sigma\sqrt{s}R_{\max}x_{\max}}\right\}$, $\tilde{C}_2 = \frac{\mu_0 p^*}{2\sigma_3 s x_{\max}^3(\mu_0 p^* + 48 s x_{\max}^2)}$ *and* $C_5 = \frac{\boldsymbol{\beta}_{\min} p^* \kappa}{(2304s + a p^* \kappa)\sqrt{1+\log d}}$

Proof of Lemma EC.9 Using Lemma EC.8, $t_0 = 2C_0|\mathcal{K}|$, $T \geq \max\{(t_0+1)^2/e^2 - 1, e\}$, and $C_0 \geq \max\{10, 16/p^*\}$, we have

$$\mathbb{P}\left\{n \geq \frac{1}{2}C_0 \log(T+1)\right\} \leq 1 - \frac{2}{T+1} \tag{EC.102}$$

$$\mathbb{P}\left\{|\mathcal{A}| \geq \frac{1}{4}p^* C_0 \log(T+1)\right\} \geq 1 - \frac{1}{T+1} \tag{EC.103}$$

$$\mathbb{P}\left\{\frac{|\mathcal{A}|}{n} \geq \frac{1}{24}p^*\right\} \geq \frac{3}{T+1}. \tag{EC.104}$$

Thus with probability $1-\frac{3}{T+1}$ we have

$$\begin{aligned}
\beta_{\min} &= (\frac{2304s}{p^*\kappa} + a)C_5\sqrt{1+\log d} \geq (\frac{2304s}{p^*\kappa} + a)\lambda \geq (\frac{96ns}{\kappa|\mathcal{A}|} + a)\lambda \\
a &> \frac{2304s}{p^*\kappa} \geq \frac{96ns}{\kappa|\mathcal{A}|} \\
\tilde{C}_2 &= \frac{\mu_0 p^*}{2\sigma_3 s x_{\max}^3(\mu_0 p^* + 48 s x_{\max}^2)} \leq \frac{\mu_0|\mathcal{A}|}{2\sigma_3 s x_{\max}^3(\mu_0|\mathcal{A}| + n2s x_{\max}^2)} = C_2
\end{aligned} \tag{EC.105}$$

If we require $t \leq \frac{\mu_0|\mathcal{A}|\sqrt{\tilde{C}_2\lambda}}{2n} \leq \frac{\mu_0|\mathcal{A}|\sqrt{C_2\lambda}}{2n}$, from (6) in Proposition 3 , we can obtain the following inequality.

$$\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \geq \frac{2nt}{|\mathcal{A}|\mu_0}\right) \leq \delta_2(n,|\mathcal{A}|,\lambda) + \delta_3(|\mathcal{A}|) + \delta_4(n,|\mathcal{A}|,t). \tag{EC.106}$$

Since $\delta_2(n,|\mathcal{A}|,\lambda), \delta_3(|\mathcal{A}|)$ and $\delta_4(n,|\mathcal{A}|,t)$ decrease when we have larger $|\mathcal{A}|$ and n, we may pick proper C_0 such that at given time T we will have enough $|\mathcal{A}|$ and n according to (EC.102)-(EC.104). As we require $C_0 = \max\left\{\frac{4}{p^*C_1}, \frac{4x_{\max}^2}{C_5^2}\left((\frac{1}{4} - \frac{576s}{p^*\kappa a})\min\left\{1, \frac{\mu_0 p^*}{192 s x_{\max}^2}\right\}\right)^{-2}, \frac{32\sigma_2 s x_{\max}^2(1+\log s)}{p^*\mu_0}, \frac{4\sigma^2 x_{\max}^2(1+\log s)}{t^2}\right\}$ and $\lambda = C_5\sqrt{1+\log d/\log(T+1)}$, one may verify the the follow result hold with probability $1-\frac{3}{T+1}$.

$$\delta_2(n,|\mathcal{A}|,\lambda) + \delta_3(|\mathcal{A}|) + \delta_4(n,|\mathcal{A}|,t) \leq \frac{4}{T+1}. \tag{EC.107}$$

Hence, we have

$$\begin{aligned}
&\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_2 \leq \frac{2nt}{|\mathcal{A}|\mu_0}\right) \geq 1 - \frac{7}{T+1} \\
\Rightarrow &\mathbb{P}\left(\|\boldsymbol{\beta}^{MCP} - \boldsymbol{\beta}^{true}\|_1 \leq \frac{2nt\sqrt{s}}{|\mathcal{A}|\mu_0}\right) \geq 1 - \frac{7}{T+1},
\end{aligned} \tag{EC.108}$$

where (EC.108) uses $\boldsymbol{\beta}^{MCP}$ being the oracle solution with $\boldsymbol{\beta}^{MCP}_{S^c} = \boldsymbol{\beta}^{true}_{S^c} = \mathbf{0}$. Moreover, combine $t \leq \min\left\{\frac{p^*\mu_0}{48\sigma\sqrt{s}x_{\max}}, \frac{hp^*\mu_0}{192e\sigma\sqrt{s}R_{\max}x_{\max}}\right\}$, (EC.104) and we have the following results.

$$\frac{2nt\sqrt{s}}{|\mathcal{A}|\mu_0} \leq \frac{2nhp^*\mu_0\sqrt{s}}{192e\sigma\sqrt{s}R_{\max}x_{\max}|\mathcal{A}|\mu_0} = \frac{h}{4e\sigma R_{\max}x_{\max}} \cdot \frac{n}{|\mathcal{A}|} \cdot \frac{p^*}{24} \leq \frac{h}{4e\sigma R_{\max}x_{\max}} \tag{EC.109}$$

$$\frac{2nt\sqrt{s}}{|\mathcal{A}|\mu_0} \leq \frac{p^*\mu_0\sqrt{s}}{48\sigma\sqrt{s}x_{\max}|\mathcal{A}|\mu_0} = \frac{1}{\sigma_2 x_{\max}} \cdot \frac{n}{|\mathcal{A}|} \cdot \frac{p^*}{24} \leq \frac{1}{\sigma x_{\max}} \tag{EC.110}$$

Desirable result follows immediately.

LEMMA EC.10. *Under assumptions* **A.3** *and* **A.5**, *for any* $\boldsymbol{x} \in U_k, i \in \mathcal{K}$, *the following two statements hold.*

1. $|\mathbb{E}(R_i|x,\boldsymbol{\beta}_i^{true}) - \mathbb{E}(R_i|x,\boldsymbol{\beta}_i^{MCP})| \leq R_{\max}e^{\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP} - \boldsymbol{\beta}_i^{true}\|_1}\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP} - \boldsymbol{\beta}_i^{true}\|_1$
2. *Moreover, if* $\|\boldsymbol{\beta}_i^{MCP} - \boldsymbol{\beta}_i^{true}\|_1 \leq \min\left\{\frac{1}{\sigma x_{\max}}, \frac{h}{4e\sigma R_{\max}x_{\max}}\right\}$, $k \in \mathcal{K}$, *we have* $\mathbb{E}(R_i|\boldsymbol{x},\beta_i^{MCP}) \geq \max_{j\neq i}\mathbb{E}(R_j|\boldsymbol{x},\beta_j^{MCP}) + \frac{h}{2}$.

Proof of Lemma EC.10 ***To show the part 1.*** We first expand the left-hand-side as follows.

$$
\begin{aligned}
&\left|\mathbb{E}(R_i|\boldsymbol{x},\boldsymbol{\beta}_i^{true})-\mathbb{E}(R_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP})\right| \\
=&\left|\int_{-\infty}^{+\infty} r_i dF(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{true})-\int_{-\infty}^{+\infty} r_i dF(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{MCP})\right| \\
=&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{true})}dr_i-\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{MCP})}dr_i\right| \qquad \text{(EC.111)} \\
=&\left|\int_{-\infty}^{+\infty} r_i\left(e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{true})}-e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{MCP})}\right)dr_i\right|. \\
=&\left|\int_{-\infty}^{+\infty} -r_i\left(e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i)}\right)'\Big|_{\boldsymbol{\beta}_i=\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}}\boldsymbol{x}^T(\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true})dr_i\right|, \qquad \text{(EC.112)}
\end{aligned}
$$

where (EC.111) uses f being the negative log density function and $\boldsymbol{\delta}$ is between $\mathbf{0}$ and $\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}$. We then pull $\boldsymbol{x}^T(\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true})$ out of the integral.

$$
\begin{aligned}
&\left|\int_{-\infty}^{+\infty} -r_i\left(e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i)}\right)'\Big|_{\boldsymbol{\beta}_i=\boldsymbol{\beta}_i^{true}+\delta}\boldsymbol{x}^T(\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true})dr_i\right| \\
=&\left|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true})\int_{-\infty}^{+\infty} -r_i\left(e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i)}\right)'\Big|_{\boldsymbol{\beta}_i=\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}}dr_i\right| \\
\le&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))}f'(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))dr_i\right| x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1. \qquad \text{(EC.113)}
\end{aligned}
$$

As we assume $|f'(\cdot|\cdot)|$ is bounded by σ in assumption **A.5**, (EC.113) is upper bounded by

$$
\begin{aligned}
&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))}f'(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))dr_i\right| x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \\
\le&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))}dr_i\right|\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1. \qquad \text{(EC.114)}
\end{aligned}
$$

We then expand term $f(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))$ in (EC.114), and there exists a $\boldsymbol{\xi}$ between $\mathbf{0}$ and $\boldsymbol{\beta}^{true}+\boldsymbol{\delta}$ such that

$$
\begin{aligned}
&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T(\boldsymbol{\beta}_i^{true}+\boldsymbol{\delta}))}dr_i\right|\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \\
=&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{true})-f'(r_i|\boldsymbol{x}^T\boldsymbol{\xi})\boldsymbol{x}^T\boldsymbol{\delta}}dr_i\right|\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_j^{true}\|_1 \\
\le&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{true})+|f'(r_i|\boldsymbol{x}^T\boldsymbol{\xi})|\|\boldsymbol{x}\|_\infty\|\boldsymbol{\delta}\|_1}dr_i\right|\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \\
\le&\left|\int_{-\infty}^{+\infty} r_i e^{-f(r_i|\boldsymbol{x}^T\boldsymbol{\beta}_i^{true})}dr_i\right| e^{\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1}\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \qquad \text{(EC.115)} \\
=&|\mathbb{E}(R_i|x,\boldsymbol{\beta}_i^{true})|e^{\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1}\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \qquad \text{(EC.116)}
\end{aligned}
$$

where (EC.115) uses that $\boldsymbol{\delta}$ is between $\mathbf{0}$ and $\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}$, which implies $\|\boldsymbol{\delta}\|_1\le\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1$, and (EC.116) comes from the definition of $\mathbb{E}(R_i|\boldsymbol{x},\boldsymbol{\beta}_i^{true})$. Combining $|r_i|\le R_{\max}$, (EC.116), and (EC.112), we have:

$$
\left|\mathbb{E}(R_i|x,\boldsymbol{\beta}_i^{true})-\mathbb{E}(R_i|x,\boldsymbol{\beta}_i^{MCP})\right|\le R_{\max}e^{\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1}\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1. \qquad \text{(EC.117)}
$$

To show the part 2. Note that the assumption $\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \le \frac{1}{\sigma x_{\max}},\ k\in\mathcal{K}$ implies the following inequality:

$$\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \le \frac{1}{\sigma x_{\max}} \Rightarrow e^{\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1} \le e \tag{EC.118}$$

Combining (EC.118) and (EC.117), we obtain

$$\begin{aligned}\left|\mathbb{E}(r_i|x,\boldsymbol{\beta}_i^{true})-\mathbb{E}(r_i|x,\boldsymbol{\beta}_i^{MCP})\right| \le & R_{\max}e^{\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1}\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \\ \le & R_{\max}e\sigma x_{\max}\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \end{aligned}\tag{EC.119}$$

Under assumption **A.3**, for any $x\in U_k$, the following inequalities hold:

$$\begin{aligned}&\mathbb{E}(R_i|\boldsymbol{x},\boldsymbol{\beta}_i^{true}) \ge \max_{j\neq i}\mathbb{E}(R_j|\boldsymbol{x},\boldsymbol{\beta}_j^{true})+h \\ \Rightarrow\ &\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{true})-\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP}) \ge \max_{j\neq i}\left[\mathbb{E}(r_j|x,\boldsymbol{\beta}_j^{true})-\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{MCP})\right] \\ &\qquad +\max_{j\neq i}\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{MCP})-\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP})+h \\ \Rightarrow\ &\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP})-\max_{j\neq i}\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{MCP}) \ge -\left|\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP})-\mathbb{E}(r_i|x,\boldsymbol{\beta}_i^{true}))\right| \\ &\qquad -\max_{j\neq i}\left|\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{true})-\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{MCP})\right|+h.\end{aligned}\tag{EC.120}$$

As we assume $\|\boldsymbol{\beta}_k^{MCP}-\boldsymbol{\beta}_k^{true}\|_1 \le \frac{h}{4e\sigma R_{\max}x_{\max}},\ k\in\mathcal{K}$, we have

$$\|\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true}\|_1 \le \frac{h}{4e\sigma R_{\max}x_{\max}} \Rightarrow \|R_{\max}e\sigma x_{\max}(\boldsymbol{\beta}_i^{MCP}-\boldsymbol{\beta}_i^{true})\|_1 \le \frac{h}{4}\tag{EC.121}$$

Combining (EC.119),(EC.121) and (EC.120), we will have

$$\begin{aligned}&\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP})-\max_{j\neq i}\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{MCP}) \ge -\frac{h}{4}-\frac{h}{4}+h \\ \Rightarrow &\mathbb{E}(r_i|\boldsymbol{x},\boldsymbol{\beta}_i^{MCP}) \ge \max_{j\neq i}\mathbb{E}(r_j|\boldsymbol{x},\boldsymbol{\beta}_j^{MCP})+\frac{h}{2}.\end{aligned}\tag{EC.122}$$

LEMMA EC.11. *Denote events $\mathcal{E}_3,\mathcal{E}_4$, and $\mathcal{E}_5$ as follows*

$$\mathcal{E}_3=\left\{\|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty \le \left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{4}\right\}\tag{EC.123}$$

$$\mathcal{E}_4=\left\{\|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty \le \left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\mu_0|\mathcal{A}|\lambda}{8snx_{\max}^2}\right\}\tag{EC.124}$$

$$\mathcal{E}_5=\left\{\|\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true}\|_2 \le \sqrt{C_2\lambda}\right\},\tag{EC.125}$$

where $C_2 \doteq \frac{\mu_0|\mathcal{A}|}{2\sigma_3 sx_{\max}^3(\mu_0|\mathcal{A}|+2snx_{\max}^2)}$. Under assumption **A.1** *and* **A.5**, *events $\mathcal{E}_3,\mathcal{E}_4$ and $\mathcal{E}_5$ implies $\mathcal{E}_2$ defined in* (EC.10).

Proof of Lemma EC.11 We first expend $\nabla\mathcal{L}(\boldsymbol{\beta}^{oracle})$ at $\boldsymbol{\beta}^{true}$.

$$\nabla\mathcal{L}(\boldsymbol{\beta}^{oracle}) = \nabla\mathcal{L}(\boldsymbol{\beta}^{true})+\nabla^2\mathcal{L}(\boldsymbol{\xi})(\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true})\tag{EC.126}$$

$$= \nabla\mathcal{L}(\boldsymbol{\beta}^{true})+\nabla^2\mathcal{L}(\boldsymbol{\beta}^{true})(\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true})+(\nabla^2\mathcal{L}(\boldsymbol{\xi})-\nabla^2\mathcal{L}(\boldsymbol{\beta}^{true}))(\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true}),\tag{EC.127}$$

where $\boldsymbol{\xi} = \tau\boldsymbol{\beta}^{true} + (1-\tau)\boldsymbol{\beta}^{oracle}$, $\tau \in [0,1]$. The last term in (EC.127) can be further expanded as follows

$$\begin{aligned}
&(\nabla^2\mathcal{L}(\boldsymbol{\xi}) - \nabla^2\mathcal{L}(\boldsymbol{\beta}^{true}))(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}) \\
=&\frac{1}{n}\sum_{j=1}^{n}\left[f^{''}(r_j|\boldsymbol{x}_j^T\boldsymbol{\xi}) - f^{''}(r_j|\boldsymbol{x}_j^T\boldsymbol{\beta}^{true})\right]\boldsymbol{x}_j\boldsymbol{x}_j^T(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}) \\
=&\frac{1}{n}\sum_{j=1}^{n}\left[-f^{'''}(r_j|\boldsymbol{x}_j^T\eta)\boldsymbol{x}_j^T(\boldsymbol{\xi} - \boldsymbol{\beta}^{true})\right]\boldsymbol{x}_j\boldsymbol{x}_j^T(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}), \qquad \text{(EC.128)}
\end{aligned}$$

where (EC.128) comes from the mean value theorem and the fact that η is on the line of of $\boldsymbol{\xi}$ and $\boldsymbol{\beta}^{true}$. Hence, assumption **A.5** and (EC.128) imply

$$\begin{aligned}
&\|(\nabla^2\mathcal{L}(\boldsymbol{\xi}) - \nabla^2\mathcal{L}(\boldsymbol{\beta}^{true}))(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\|_\infty \\
=&\left\|\frac{1}{n}\sum_{j=1}^{n}\left[-f^{'''}(r_j|\boldsymbol{x}_j^T\eta)\boldsymbol{x}_j^T(\boldsymbol{\xi} - \boldsymbol{\beta}^{true})\right]\boldsymbol{x}_j\boldsymbol{x}_j^T(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\right\|_\infty \\
\le&\left\|\frac{1}{n}\sum_{j=1}^{n}\sigma_3 x_{\max}(\boldsymbol{\xi} - \boldsymbol{\beta}^{true})\boldsymbol{x}_j\boldsymbol{x}_j^T(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\right\|_\infty \\
\le&\left\|\frac{1}{n}\sum_{j=1}^{n}\sigma_3 x_{\max}(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})^T\boldsymbol{x}_j\boldsymbol{x}_j^T(\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true})\right\|_\infty \\
\le&\sigma_3 x_{\max}\lambda_{\max}(\frac{1}{n}X_{\mathcal{S}}X_{\mathcal{S}}^T)\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2 \\
\le&\sigma_3 s x_{\max}^3\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2. \qquad \text{(EC.129)}
\end{aligned}$$

Combining (EC.127), (EC.129), and the fact $\boldsymbol{\beta}_{\mathcal{S}^c}^{oracle} = \boldsymbol{\beta}_{\mathcal{S}^c}^{true} = 0$, we have

$$\|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{oracle})\|_\infty \le \|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty + \|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\boldsymbol{\beta}_{\mathcal{S}}^{oracle} - \boldsymbol{\beta}_{\mathcal{S}}^{true})\|_\infty + \sigma_3 s x_{\max}^3\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2. \qquad \text{(EC.130)}$$

In addition, from $\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{oracle}) = 0$ and (EC.127), we have

$$(\boldsymbol{\beta}_{\mathcal{S}}^{oracle} - \boldsymbol{\beta}_{\mathcal{S}}^{true}) = -(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}(\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}) + (\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\xi}) - \nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))(\boldsymbol{\beta}_{\mathcal{S}}^{oracle} - \boldsymbol{\beta}_{\mathcal{S}}^{true})). \qquad \text{(EC.131)}$$

Under events $\mathcal{E}_3$, $\mathcal{E}_4$, and (EC.131), the inequality (EC.130) can be upper bounded as follows.

$$\begin{aligned}
&\|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{oracle})\|_\infty \le \left(1 - \frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{4} + \sigma_3 x_{\max}\lambda_{\max}(\frac{1}{n}X_{\mathcal{S}}X_{\mathcal{S}}^T)\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2 \\
&\quad + \|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}(\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}) + (\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\xi}) - \nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))(\boldsymbol{\beta}_{\mathcal{S}}^{oracle} - \boldsymbol{\beta}_{\mathcal{S}}^{true}))\|_\infty \\
&\le \left(1 - \frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{4} + \sigma_3 s x_{\max}^3\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2 \\
&\quad + \left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}\right\|\left(\|\nabla_{\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})\|_\infty + \sigma_3 s x_{\max}^3\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2\right) \\
&\le \left(1 - \frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{4} + \sigma_3 s x_{\max}^3\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|_2^2 \\
&\quad + \left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}\right\|\left(\left(1 - \frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\mu_0|\mathcal{A}|\lambda}{8snx_{\max}^2} + \sigma_3 s x_{\max}^3\|\boldsymbol{\beta}^{oracle} - \boldsymbol{\beta}^{true}\|^2\right). \qquad \text{(EC.132)}
\end{aligned}$$

Note that the maximum value of $\left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}\right\|$ can be bounded.

$$\left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}\right\| \le \max_{\|v\|=1}\left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}v\right\|. \qquad \text{(EC.133)}$$

From (EC.133) and Lemma EC.1, the following inequality holds with probability $1-2s\exp\left(-\frac{|\mathcal{A}|\mu_0}{4s\sigma_2 x_{\max}^2}\right)$.

$$\begin{aligned}\max_{\|v\|=1}\left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})(\nabla^2_{\mathcal{S},\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true}))^{-1}v\right\| &\le \frac{2n}{\mu_0|\mathcal{A}|}\max_{\|v\|=1}\left\|\nabla^2_{\mathcal{S}^c,\mathcal{S}}\mathcal{L}(\boldsymbol{\beta}^{true})v\right\| \\ &\le \frac{2n}{\mu_0|\mathcal{A}|}\cdot sx_{\max}^2 = \frac{2snx_{\max}^2}{\mu_0|\mathcal{A}|}. \end{aligned} \tag{EC.134}$$

Thus, (EC.132) can be simplified to:

$$\begin{aligned}\|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{oracle})\|_\infty &\le \left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{4}+\sigma_3 sx_{\max}^3\|\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true}\|^2 \\ &\quad+\frac{2snx_{\max}^2}{\mu_0|\mathcal{A}|}\left(\left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\mu_0|\mathcal{A}|\lambda}{8snx_{\max}^2}+\sigma_3 sx_{\max}^3\|\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true}\|^2\right) \\ &= \left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{2}+\frac{\sigma_3 sx_{\max}^3(\mu_0|\mathcal{A}|+2snx_{\max}^2}{\mu_0|\mathcal{A}|}\|\boldsymbol{\beta}^{oracle}-\boldsymbol{\beta}^{true}\|_2^2. \end{aligned} \tag{EC.135}$$

Further, conditioning on event $\mathcal{E}_5$ defined in (EC.125), we have:

$$\begin{aligned}\|\nabla_{\mathcal{S}^c}\mathcal{L}(\boldsymbol{\beta}^{oracle})\|_\infty &\le \left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\frac{\lambda}{2}+\frac{\sigma_3 sx_{\max}^3(\mu_0|\mathcal{A}|+2snx_{\max}^2}{\mu_0|\mathcal{A}|}\left(\sqrt{C_2\lambda}\right)^2 \\ &\le \left(1-\frac{96ns}{|\mathcal{A}|\kappa a}\right)\lambda, \end{aligned} \tag{EC.136}$$

where (EC.136) uses $C_2=\frac{\mu_0|\mathcal{A}|}{2\sigma_3 sx_{\max}^3(\mu_0|\mathcal{A}|+2snx_{\max}^2)}$. The inequality (EC.136) directly implies event $\mathcal{E}_2$.