Online Assortment Optimization 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

In this research, we consider an online assortment optimization problem, where a decision-maker sequentially offers assortments to users instantaneously upon their arrivals and users select products from offered assortments according to the contextual multinomial logit choice model. We propose a computationally efficient Lasso-RP-MNL algorithm for the online assortment optimization problem under the cardinality constraint in high-dimensional settings. The Lasso-RP-MNL algorithm combines the Lasso and random projection as dimension reduction techniques to alleviate the computational complexity and improve the learning and estimation accuracy under high-dimensional data. For each arriving user, the Lasso-RP-MNL algorithm constructs an upper-confidence bound for each individual product's attraction parameter, based on which the optimistic assortment can be identified by solving a reformulated linear programming problem. We demonstrate that for the significant feature dimension s, the total feature dimension d, and the sample size dimension T, the expected cumulative regret under the Lasso-RP-MNL algorithm is upper bounded by $\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{2}{3}})$, where $\tilde{\mathcal{O}}$ suppresses the logarithmic dependence on T. Such a regret upper bound matches the informational theoretical regret lower bound under high-dimensional setting with limited samples, i.e., $\Omega(T^{\frac{2}{3}})$. Furthermore, as the sample size increases, we can further improve the Lasso-RP-MNL algorithm's regret upper bound to $\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{1}{2}})$, which also matches the regret lower bound in data-rich regimes, i.e., $\Omega(T^{\frac{1}{2}})$. Finally, through synthetic-data-based experiments and a high-dimensional XianYu assortment recommendation experiment, we show that the Lasso-RP-MNL algorithm is computationally efficient and outperforms other benchmarks in terms of the expected cumulative regret.

Key words: Online assortment optimization, contextual information, high-dimensional data, Lasso, random projection, multinomial logit model, upper-confidence bound.

1. Introduction

Online assortment optimization problems have recently emerged in many internet applications, such as e-tailing, digital advertising, recommendation, etc., where a decision-maker sequentially offers assortments of substitutable products to users instantaneously upon their arrivals. For example, an internet retailer offers an assortment of customized items to an arriving consumer; a search engine displays an assortment of profitable advertisements in response to a user's search query; an online marketplace recommends an assortment of personalized products based on a user's browsing

and purchasing history. To select reward-maximizing assortments, the decision-maker first needs to accurately assess users' utilities and choices, which are typically unknown a priori but can be gradually learned through observing users' responses to various assortments (Mahajan and van Ryzin 1999). Hence, online assortment optimization problems require a judicious balance between exploring different assortments to learn users' choices and simultaneously exploiting assortments that maximize immediate rewards. In the big data era, the growing availability of granular data and high-dimensional contextual information for users and products has presented both promising opportunities and vexing challenges for these online assortment optimization problems.

Rich contextual information provides the decision-maker with unprecedented opportunities to improve his learning capability and prediction accuracy regarding users' utilities and choices. Consider the assortment recommendation practice at XianYu, a leading consumer-to-consumer online marketplace for new and preowned products in China. Upon clicking a product link, the user is redirected to the product information page, on which, along with typical product specifics and transaction details, a "Guess What You Like" section displays a personalized assortment consisting up to 20 suggested products. To optimally recommend 20 products, XianYu relies on contextual information about the user and products to learn and predict the user's utility and choice concerning any given assortment. Realizing that more information leads to better learning and prediction, XianYu has dramatically expanded the extent of contextual information and accelerated its data collection efforts. Currently, the available contextual information at XianYu is extremely high-dimensional: It contains more than 2 billion features, including user information (e.g., demographics, geographics, browsing/clicking history, etc.), product information (e.g., brand, color, size, condition, etc.), and information about possible interactions between these two (e.g., the physical distance between the user and the product, whether the user's searching history matches the product, etc.). In practice, using this high-dimensional contextual information has significantly improved XianYu's learning and prediction accuracy regarding users' choices, which in turn enables better assortment decisions.

Yet, the decision-maker's ability to use all available contextual information and effectively learn the influences of all features on users' utilities and choices is often impaired by the fact that there are limited samples in practice. Specifically, to accurately estimate the influences of more than 2 billion features via traditional statistical methods (e.g., maximum likelihood estimation), XianYu will need billions or even trillions random samples. However, with approximately 4 million daily "Guess What You Like" exposures, among which a very small percentage can be chosen to perform costly learning experiments (Bastani and Bayati 2020), it may take decades before XianYu can identify the influences of all features with reasonable accuracy. Moreover, due to the intrinsically ever-changing nature of users' tastes, these estimations are typically time sensitive:

Estimations based on historical data stretching more than a year, or a few months for fashionable products, will be less relevant for predicting users' choices tomorrow. Hence, compared to the scale of high-dimensional contextual information, available samples are extremely limited and therefore constrain the decision-maker's ability to fully utilize all available features to learn and update his estimations.

Furthermore, even with sufficient samples to support effective learning, the decision-maker still has to ensure that the online assortment optimization algorithm is computationally efficient. In XianYu's example, the average time that elapses between a user clicking a product link and the web page displaying the recommended assortment is expected to be less than a half-second, which includes time needed for learning/updating the estimations and optimizing recommended assortments. However, a single estimation update for XianYu's high-dimensional features can easily take hours with high-performance computing technologies and state-of-the-art techniques, especially when the sample size is not too small; combined with the time needed for optimizing assortments, which is a nonlinear combinatorial optimization problem, the total computational time may far exceed the half-second target mark.

To address these challenges, we propose a computationally efficient Lasso-RP-MNL algorithm for online assortment optimization problems under the cardinality constraint in high-dimensional settings. This algorithm combines both the Lasso (Tibshirani 1996) and random projection (Johnson and Lindenstrauss 1984) to improve the learning and estimation accuracy for high-dimensional features with limited samples and follows the idea of upper-confidence bound (UCB) approach (Auer 2002) to identify the optimistic assortment under the multinomial logit (MNL) choice model. In particular, with a period length that exponentially increases in time, we periodically threshold the Lasso estimator to identify and update significant features that have strong influences on users' utilities and choices; then, for each arriving user, we adopt random projection to reduce the high-dimensional contextual information, excluding significant features already identified by thresholding the Lasso, to a low-dimensional space and then estimate coefficients for both original features thresholded by Lasso and projected features by random projection. Through this process, the learning and parameter estimation can be performed in a low-dimensional fashion to significantly trim down the computational time, while maintaining high accuracy in predicting users' choices. Furthermore, we show that thresholding the Lasso for feature selection will limit the long-term negative influence of the information loss that is intrinsic to random projection and that random projection can in turn alleviate the negative influence of possible model misspecification in the Lasso due to limited samples. Next, to further reduce the computational complexity, instead of constructing upper confidence bound for all assortments, we establish the upper-confidence bound

for each individual product and then identify the optimistic assortment. Note that under the cardinality constrain, the optimal optimistic assortment is a combinatorial optimization problem and can not be identified by the revenue-ordered sets (Rusmevichientong et al. 2010). Hence, we follow Davis et al. (2013) to reformulate the optimal optimistic assortment problem into a linear programming problem, which can be solved by various efficient solution algorithms.

Main Contributions:

We first establish the information theoretical regret lower bound for online assortment optimization problem and show that under high-dimensional settings with limited samples, the $\Omega(T^{\frac{2}{3}})$ regret lower bound is inevitable. Only when there are sufficient samples, the regret lower bound can be transited to the standard $\Omega(T^{\frac{1}{2}})$ regret lower bound.

Next, we demonstrate that the Lasso-RP-MNL algorithm can match the regret lower bound on T and achieve a sub-logarithmic dependence on the feature dimension d. Specifically, we show that the expected cumulative regret of the Lasso-RP-MNL algorithm is upper-bounded by $\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{2}{3}})$, where s is significant feature dimension and d is the total feature dimension. Furthermore, as the sample size increases, we can further improve the Lasso-RP-MNL algorithm's regret upper bound to $\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{1}{2}})$. We believe that the Lasso-RP-MNL algorithm is the first assortment algorithm in s-sparse high-dimensional settings to attain a sub-logarithmic dependence on the feature dimension.

Finally, we benchmark the Lasso-RP-MNL algorithm to existing state-of-the-art algorithms in the literature and industrial practice through both synthetic experiments and a real-life experiment based on XianYu's high-dimensional assortment recommendation dataset. We show that the Lasso-RP-MNL algorithm is computationally efficient and can significantly improve the decision-maker's regret and revenue performance.

2. Related Literature

Our work is related to the dynamic assortment optimization literature, where users' utilities (i.e., model parameters) are unknown to the decision-maker at the beginning but can be gradually learned over multiple periods. Various models have been used in the assortment optimization literature, such as the MNL model (e.g., Ryzin and Mahajan 1999, Mahajan and Van Ryzin 2001), nested logit (e.g., McFadden 1980, Gallego and Wang 2014, Li et al. 2015), exogenous demand model (e.g., Smith and Agrawal 2000, Netessine and Rudi 2003), Markov chain model (e.g., Blanchet et al. 2016, Feldman and Topaloglu 2017), and non-parametric models (e.g., Rusmevichientong et al. 2006, Farias et al. 2013). Among these models, the MNL model, which is adopted in this work, is the most commonly used choice model in Economics, Marketing, and Operations Management literature (Kök and Fisher 2007), mainly by virtue of its tractability in estimating unknown parameters and identifying optimal assortments. For extensive literature review on the MNL model and

other assortment optimization models, we refer to Mahajan and van Ryzin (1999) and Kök et al. (2015).

When there are limited number of products repetitively offered to incoming users, it is natural to consider the setting where the utility of each product is represented by an unknown attraction parameter in the MNL model. Under this setting, Rusmevichientong et al. (2010) and Sauré and Zeevi (2013) propose two explore-then-exploit algorithms, where the decision-maker first offers pre-selected assortments in the exploration phase to attend desired estimation accuracy for these unknown parameters, and then goes to the exploitation phase to maximize his expected reward. Rusmevichientong et al. (2010) show that their Adaptive Assortment algorithm can attain $\mathcal{O}(N^2 \log^2 T)$ cumulative regret bound, and Sauré and Zeevi (2013) demonstrate that their separation-based policy can achieve $\mathcal{O}(N \log T)$ regret ratio bound, where N is the number of candidate products. Kallus and Udell (2016) consider a personalized assortment model to extend the homogeneous users case to the heterogeneous case, and Bernstein et al. (2018) adopt a Bayesian semi-parametric framework to propose a dynamic clustering policy to map users' profiles to groups/clusters. It is worth noting that these works require certain a prior knowledge of the separation gap parameter, which gauges the reward difference between the optimal and the second-best assortment to regulate the exploration phase, without which these algorithms can perform quite poorly (Agrawal et al. 2019). Without assuming any prior knowledge of the separation gap parameter, Agrawal et al. (2019) propose a UCB-type MNL-Bandit algorithm. Under a mild assumption, the authors establish the regret upper bound $\tilde{\mathcal{O}}(\sqrt{NT})$. Agrawal et al. (2017) further propose another Thompson-Sampling based algorithm that can attend a similar regret bound with improved empirical performance. Cheung and Simchi-Levi (2017a) propose a UCB-type policy under resource constrain and show that this policy can also attain $\tilde{\mathcal{O}}(\sqrt{T})$ regret upper bound.

Note that all of previous works assume that unknown parameters are associated with products themselves (i.e., each product has an unique unknown attraction parameter). In practice, however, the number of products can be enormous, and available products change from user to user, both of which lead to an unnecessarily large number of unknown parameters needed to be learned. Therefore, recognizing the facts that the difference among products can be represented by their intrinsic features and that a smaller number of features are sufficient to identify a large number of products in practice (Agrawal et al. 2019), the contextual MNL model assigns unknown parameters to every unique feature and estimates these feature parameters separately. As features can be shared among multiple products, the learning can now cross products (Oh and Iyengar 2021), which suggests that the regret bound for the contextual MNL model can be independent of the number of candidate products N.

Chen et al. (2018) consider the contextual MNL model in which the feature information of products can change over time (i.e., the underlying choice model is non-stationary) and develop an explore-then-exploit UCB-based policy with $\tilde{\mathcal{O}}(d\sqrt{T})$ regret bound. Following a similar setting, Oh and Iyengar (2021) propose another two explore-then-exploit UCB-based algorithms: The first computationally efficient algorithm attain the regret bound of $\tilde{\mathcal{O}}(d\sqrt{T})$, and the second algorithm reduces the regret bound to $\tilde{\mathcal{O}}(\sqrt{dT})$ under the Relaxed Symmetry assumption. Oh and Iyengar (2019) develop two Thompson sampling algorithms and achieve $\tilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ and $\tilde{\mathcal{O}}(d\sqrt{T})$ Bayesian regret, respectively. Ou et al. (2018) consider a linear utility MNL model, where item utilities are represented by linear functions of d-dimension features, and propose the LUMB algorithm, which achieves $\tilde{\mathcal{O}}(dK\sqrt{T})$ regret bound. Different from previous papers that study homogeneous users under the stochastic arrival setting, Cheung and Simchi-Levi (2017b) study heterogeneous users under the adversarial user arrival. The authors propose a Thompson Sampling based Pao-Ts policy whose Bayesian regret upper bound satisfies $\tilde{\mathcal{O}}(N\sqrt{T})$.

Yet, when the contextual information is high-dimensional, a polynomial, linear, or sublinear dependence on the feature dimension d often hinders these algorithms from practically implementing for online assortment optimization problems, mainly due to dissatisfied regret performance and the excessive computational burden. In this work, we consider a s-sparse contextual MNL model under high-dimensional setting and combine both the Lasso and random projection to develop a simultaneously-explore-and-exploit Lasso-RP-MNL algorithm that is computationally efficient. We show that Lasso-RP-MNL algorithm improves the regret bound to $\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{2}{3}})$ in data-poor regimes, where there are limited samples, and to $\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{1}{2}})$ in data-rich regimes, where samples are large related to data dimensions. In addition, we prove that these two regret upper bounds on T for both regimes match the regret lower bound up to a logarithmic factor. We believe that the Lasso-RP-MNL algorithm is the first online assortment optimization algorithm in high-dimensional settings to attain sub-logarithmic dependence on the feature dimension. We summarize the theoretical bounds comparisons in Table 1.

At last, as our algorithm combines the Lasso and random projection to handle the high-dimensional data challenges, this paper is also related to these two streams of literature. In high-dimensional statistics, Lasso-type methods (Tibshirani 1996) have been proposed to explore the high-dimensional data's underlying latent sparse structure and become a standard approach for high-dimensional feature selection and learning (Fan and Li 2001, Meinshausen et al. 2006, 2009, Zhang et al. 2010, Loh and Wainwright 2013). For example, Belloni et al. (2013) study the OLS post-Lasso estimator that first uses the Lasso for feature selection and then applies OLS for parameter estimation. The authors show that it performs strictly better than the Lasso and has the advantage of a smaller bias, even when the feature selection misses some parameters of the true

Table 1 Regret comparisons under data-rich regimes for the MNL model with the cardinality constrain. T is the number of time/periods, d is the total number of features, s is the number of significant features, N is the number of products, and K is the maximum assortment size.

Non-contextual MNL	Upper bound	Lower bound
Cheung and Simchi-Levi (2017a)	$\tilde{\mathcal{O}}(N^{1/3}\sqrt{KT})$	
Agrawal et al. (2017)	$\tilde{\mathcal{O}}(K\sqrt{NT})$	
Chen and Wang (2018)		$\Omega(\sqrt{NT/K})$
Agrawal et al. (2019)	$\tilde{\mathcal{O}}(\sqrt{NT})$	$\Omega(\sqrt{NT/K})$
Contextual MNL		
Cheung and Simchi-Levi (2017b)	$\tilde{\mathcal{O}}(dN\sqrt{KT})$	
Chen et al. (2018)	$\tilde{\mathcal{O}}(d\sqrt{T})$	$\Omega(d\sqrt{T}/K)$
Ou et al. (2018)	$\tilde{\mathcal{O}}(dK\sqrt{T})$	
Oh and Iyengar (2019)	$\tilde{\mathcal{O}}(d\sqrt{T})$	
Oh and Iyengar (2021)	$\tilde{\mathcal{O}}(d\sqrt{T})$ for UCB-MNL	
High-dimensional contextual MNL (this paper)		
Data-poor regime	$\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot T^{\frac{2}{3}})$ by Theorem 5	$\Omega(s^{\frac{1}{3}}T^{\frac{2}{3}})$ by Theorem 1
Data-rich regime	$\tilde{\mathcal{O}}(s\sqrt{\log d}\cdot\sqrt{T})$ by Corollary 2	$\Omega(\sqrt{dT})$ by Theorem 1

model (i.e., model misspecification). Lee et al. (2016) propose a general approach to valid confidence intervals after model selection via the Lasso. Recently, Lasso-type methods have been introduced to Bandit models and shown promising results (Bastani and Bayati 2020, Wang et al. 2018a, Kim and Paik 2019, Hao et al. 2020, Oh et al. 2020). Our algorithm also periodically adopts the Lasso for feature selection. Yet, the Lasso may suffer from model misspecification, especially under limited samples, and can be computationally challenging, therefore restraining these algorithms from being implemented directly in online settings. Random projection (Johnson and Lindenstrauss 1984) has been proposed as a computationally efficient method to deal with high-dimensional data (Fern and Brodley 2003, Pilanci and Wainwright 2015). Specifically, random projection is one of matrix sketching methods (Matoušek 2008, Luo et al. 2016, Ghashami et al. 2016, Clarkson and Woodruff 2017) that approximate a high-dimensional matrix by a more compact low-dimension one with certain approximation guarantee. Therefore, the estimation for unknown parameters and assessment of utilities can be completed in a low-dimensional fashion to significantly reduce the computational complexity (Vershynin 2010) with acceptable accuracy loss. Yet, the distortion and information loss intrinsic to random projection may lead to significant regret loss. In this work, we combine random projection and the Lasso to limit the information loss and to curb model misspecification, while maintaining the computational efficiency.

3. Problem Statement and Regret Lower Bounds

Consider a sequential decision-making process: At each time $t \in \{1, 2, ..., T\}$, a single user/consumer arrives, and the decision-maker then offers this user an assortment A_t from a candidate set containing N_t products indexed by $1, 2, ..., N_t$. The number of available products and the product candidate set may change frequently over time, because some products may be sold out and unavailable in the future, some new products can be added to the candidate set, or some products should be excluded from certain user groups according to legal/managerial policies. Therefore, at different times, the same product index may refer to different products. At each time, due to the cardinality constrain (e.g., limited display capacity), the decision-maker can offer at most K products to the user, $|A_t| \leq K$.

Users are heterogeneous, and the contextual information that characterizes N_t products and the user arriving at time t – the user-product pair – are prescribed by feature vectors $x_{t,1}, x_{t,2}, ..., x_{t,N_t} \in \mathbb{R}^d$, which are drawn i.i.d. from some unknown distributions (Chen et al. 2018, Oh and Iyengar 2021). For simplicity, below, we suppress the time index t, as long as doing so does not cause any misinterpretation. These feature vectors are high-dimensional and include rich contextual information about the user, the product, and possible interactions between these two. The user's utility from choosing product $i \in A$ with a feature vector x_i is stochastic and follows the following linear form:

$$U_i = x_i^T \beta^* + \zeta_i, \ i \in A \cup \{0\}, \tag{1}$$

where $\beta^* \in \mathbb{R}^d$ is the unknown true parameter/coefficient vector for contextual information and the unknown error term ζ_i follows a Gumbel distribution with location parameter 0 and scale parameter 1. Note that we define U_0 in Eq. (1) as the user's utility from the *no-purchase option*, which means that the user chooses not to pick any product from the assortment A and has a zero feature vector (i.e., $x_0 = \mathbf{0}$).

We consider the unknown true parameter vector β^* to be *s-sparse*:

$$\|\beta^*\|_0 = \sum_{i=1}^{d} \mathbb{1}\{\beta_i^* \neq 0\} = s.$$

This is because that the high-dimensional feature vector x_i includes all the information available to the decision-maker, but not all available features are equally valuable for predicting the user's utility and choice. For example, the user's age, the product's brand name, and the name of the product's distributor may all be available to the decision-maker and are included in the feature vector; among these three features, the first two are typically more informative for assessing this user's utility and choice than the last one. Hence, in practice, the unknown true coefficient vector

β^* naturally exhibits a latent sparse structure. Let $\mathcal{S}^* = \{j : \beta_j^* \neq 0\}$ denote the true index set for significant features (e.g., the user's age and the product's brand name), which have nonzero coefficient values and are therefore important bases for the decision-maker's predictions. Note that the true index set $\mathcal{S}^*$ and its cardinality s are unknown to the decision-maker at the beginning.

The decision-maker will earn non-negative rewards, depending on the user's selection choice. In particular, if the user chooses a product i from the offered assortment A, then the decision-maker will collect a reward r_i, which may take the form of click-through, gross merchandise volume, commission revenue, etc. Without loss of generality, we normalize the decision-maker's rewards from the no-purchase option to be zero (i.e., $r_0 = 0$). For an arbitrary assortment policy $\pi = \{A_t\}_{t\geq 1}$, where A_t is the assortment prescribed by policy π at time t, the decision-maker's the expected cumulative reward over T periods can be presented as follows:

$$\sum_{t=1}^{T} \sum_{i \in A_t} \mathbb{E}_{\boldsymbol{\zeta}} \left[r_{t,i} \cdot \mathbb{1}(U_i > U_j \text{ for } j \in A_t \cup \{0\} \setminus \{i\}) \right],$$

where $\mathbb{1}(\cdot)$ is the standard indicator function. Note that given the stochastic linear utility function in Eq. (1), the probability that the user will choose product i from the given assortment A to maximize her own utility, $p_{\beta^*,A}(i)$, can be derived (see Anderson et al. 1992) and written as

$$p_{\beta^*,A}(i) := \mathbb{E}_{\boldsymbol{\zeta}} \left[\mathbb{1}(U_i > U_j \text{ for } j \in A \cup \{0\} \setminus \{i\}) \right] = e^{x_i^T \beta^*} / (1 + \textstyle\sum_{j \in A} e^{x_j^T \beta^*}).$$

Following the MNL literature (McFadden et al. 1973), we will refer $e^{x_i^T \beta^*}$ to as the attraction parameter for product i. Note that the true coefficient vector β^* is unknown to the decision-maker at the beginning, so it is generally intractable to directly analyze the expected cumulative reward equation. Instead, we benchmark the policy π to an oracle policy, where the decision-maker knows the true coefficient vector β^* and always picks the assortment that generates the highest expected reward. Specifically, we define the decision-maker's expected cumulative regret up to time T under the policy π as

$$\text{Regret}(T) = \sum_{t=1}^{T} \left\{ \max_{\substack{\tilde{A}_t \subseteq \{1,2,\ldots,N_t\} \\ |\tilde{A}_t| \leq K}} \left[\sum_{j \in \tilde{A}_t} r_{t,j} p_{\beta^*,\tilde{A}_t}(j) \right] - \sum_{i \in A_t} r_{t,i} p_{\beta^*,A_t}(i) \right\}, \tag{2}$$

which is the difference between the expected reward under the oracle policy[1] and that under the current policy π. The decision-maker will explore to select the optimal policy π to minimize the expected cumulative regret.

[1] Since feature vectors and the candidate set can change over time, there may not exist a best set A^* that maximizes the expected reward in all rounds. Therefore, we use the round-wise optimal assortment (i.e., $\tilde{A}_t$) to benchmark the regret performance instead.

Finally, to avoid trivial assortment decisions, we assume that the feature vector, the coefficient vector, and the rewards are bounded so that the maximum reward is also upper-bounded. Formally, we make the following technical assumption:

Assumption A.1: There exist positive constants b, $x_{\max}$, and $R_{\max} \geq 1$ such that $\|\beta\| \leq b$, $\|x_i\|_1 \leq x_{\max}$, $r_i \in (0, R_{\max}]$ for any product i.

3.1. Regret Lower Bound for High-Dimensional Online Assortment Optimization Problems

Before presenting the Lasso-RP-MNL algorithm, we first establish the information theoretical regret lower bound, which applies to any possible dynamic assortment strategies, for the online assortment optimization problems under high-dimensional data, stated in Eq. (2), as follows:

THEOREM 1. *For any policy π for the high-dimensional online assortment optimization problem described in Eq.* (2)*, there exists a product set and s-sparse coefficient vector such that*

$$\textit{Regret}(T) \geq \min\left\{C_{l1} \cdot (s-1)^{\frac{1}{3}} \cdot T^{\frac{2}{3}}, C_{l2} \cdot d^{\frac{1}{2}} \cdot T^{\frac{1}{2}}\right\},$$

where $C_{l1} = \min\left\{\frac{\exp(-12)}{96\left(\cosh\left(\frac{1}{3}(s-1)^{\frac{1}{3}}T^{-\frac{1}{3}}\right)+1\right)}, \frac{1}{288\kappa^2}\right\}$ *and* $C_{l2} = \min\left\{\frac{1}{96\left(\cosh\left(\frac{1}{3}d^{\frac{1}{2}}T^{-\frac{1}{2}}\right)+1\right)}, \frac{(1-\kappa)+d^{-\frac{1}{2}}T^{\frac{1}{2}}}{288d(s-1)^{-1}\kappa^2}\right\}$.

Note that conditioning on the sample availability and dimensions, the regret lower bound can be further simplified. In particular, when $d \geq (s-1)^{2/3}C_{l1}^2 T^{1/3}/C_{l2}^2$, or equivalently $T \leq C_{l2}^6 d^3/(C_{l1}^6(s-1)^2)$, the regret is lower bounded by $\Omega(T^{\frac{2}{3}})$, which suggests that under high-dimensional settings with limit samples, the regret lower bound is inevitably higher than the conventional $\Omega(T^{\frac{1}{2}})$ regret lower bound derived in the literature for low-dimensional problems. Only as the number of samples T increases, the regret lower bound can be improved to recover the standard $\Omega(T^{\frac{1}{2}})$. In the next section, we will show that the proposed Lasso-RP-MNL algorithm matches the $\Omega(T^{\frac{2}{3}})$ lower bound, and then by considering the minimum signal strength on $\boldsymbol{\beta}^*$, the regret upper bound can be improved to match the $\Omega(T^{\frac{1}{2}})$ lower bound in data-rich regimes.

4. The Lasso-RP-MNL Algorithm

In the section, we describe the Lasso-RP-MNL algorithm and establish its regret performance. We start with the learning and estimation of the unknown coefficient β^* using the Lasso and random projection. Specifically, §4.1 discusses the process of thresholding the Lasso estimator to learn the significant feature set $\mathcal{S}^*$ and demonstrates that this method can asymptotically recover significant features with high probability. §4.2 constructs the permutation matrix and the projection matrix to reduce the high-dimensional estimation problem into a low-dimensional space and shows that the coefficient vector under the proposed permutation and projection is nearly invariant. Moreover, we demonstrate that using the Lasso for feature selection will limit the negative influence of the

information loss that is intrinsic to random projection and that random projection can in turn alleviate the negative influence of possible model misspecification in the Lasso due to limited samples. Next, in §4.3, we construct the upper-confidence bound for each individual product's utility, identify the optimistic assortment by solving a reformulated linear programming problem, and establish the single-period regret upper bound for this optimistic assortment. Finally, in §4.4 and §4.5, we formally present the Lasso-RP-MNL algorithm and derive its expected cumulative regret upper bounds.

4.1. The Lasso and Feature Selection

Denote the observed users' choices, in response to assortments $\{A_1, A_2, ..., A_T\}$ up to time T, as $\{c_1, c_2, ..., c_T\}$. We denote the index set for *whole samples* as $\mathcal{W}$. Further, we use $\mathcal{W}_R$ to denote the index set for *random samples*; that is, for $t \in \mathcal{W}_R$, the decision-maker randomly selects K products from the product candidate set and offers to the user arriving at time t. Note that besides random samples, $\mathcal{W}$ also includes non-random samples, in which the decision-maker selects assortments to maximize his revenue performance, and therefore we have $\mathcal{W}_R \subset \mathcal{W}$. In §4.4, we detail the mechanics of how these random samples are generated via the random decay sampling schedule. Let n_T denote the size of the nonempty index set $\mathcal{W}_R$, i.e., $n_T = |\mathcal{W}_R| > 0$. The Lasso estimator for the unknown coefficient vector β^* can be defined as follows:

$$\hat{\beta} = \arg\min_{\beta} L(\beta) + \lambda \|\beta\|_1, \text{ where } L(\beta) := \frac{1}{n_T} \sum_{t \in \mathcal{W}_R} \log\left(p_{\beta, A_t}(c_t)\right). \tag{3}$$

The λ in Eq. (3) is a positive regularization parameter and decreases in the random sample size n_T.

Compared to the standard maximum likelihood estimator, the Lasso estimator in Eq. (3) introduces a ℓ_1 penalty term, $\lambda\|\beta\|_1$, to retain significant features with nonzero coefficients while pushing coefficients of insignificant features towards zero. Note that the Lasso estimator is identified in Eq. (3) merely by using *random samples* in the index set $\mathcal{W}_R$, but not by using *all samples* $\mathcal{W}$ observed up to time T. This is because that these random samples preserve the iid property necessary for the desired asymptotic performance of the Lasso estimator.

To ensure the identifiability of the Lasso estimator in Eq. (3), we need the following compatibility condition for $L(\beta)$ that is constructed on the random samples set $\mathcal{W}_R$:

Assumption A.2: There exists a $\kappa > 0$ such that for all vector u with $3\|u_S\|_1 \geq \|u_{S^c}\|_1$ and $|S| \leq s$, we have $\mathbb{E}\left[u^T \nabla^2 L(\xi) u\right] \geq \frac{\kappa}{s}\|u_S\|_1^2$, where ξ is any feasible solution and $L(\xi)$ is defined in Eq. (3) by using only random samples collected in $\mathcal{W}_R$.

The Assumption A.2 is analogy to the standard technical assumption in the Lasso literature (Candes et al. 2007, Bickel et al. 2009, Bühlmann and Van De Geer 2011) and high-dimensional

bandit literature (Bastani and Bayati 2020, Wang et al. 2018a, Kim and Paik 2019). This assumption regulates the covariance matrix's behavior in a restricted region and is necessary to ensure that the Lasso estimator asymptotically converges to its true value with high probability. It is worth noting that $L(\beta)$ is constructed by using only random samples in $\mathcal{W}_R$, but not on all samples $\mathcal{W}$ up to time T. Therefore, this Assumption A.2 merely asks random samples to be diverse, but not necessarily for all assortments.

Now, we can show that the Lasso estimator defined in Eq. (3) satisfies the following inequality:

LEMMA 1. *Set the parameter* $\lambda = 2\sqrt{2x_{max}^2(\log d + \log T)/n_T}$. *Under Assumptions A.1-A.2, when* $n_T \geq \mathcal{O}(s^2 \log T)$, *the event* $\mathcal{E}_{lasso}(T) := \left\{ \|\hat{\beta} - \beta^*\|_1 \leq C_{lasso} \cdot s\sqrt{\frac{\log d + \log T}{n_T}} := \mathcal{G}_0(T, s) \right\}$ *holds with probability* $1 - \mathcal{O}(1/T)$, *where* $C_{lasso} = \frac{48\sqrt{2}x_{\max}}{K(K-1)\kappa}$.

In a nutshell, Lemma 1 demonstrates that when the random sample size n_T is large enough, the Lasso estimator $\hat{\beta}$ will be close to the true feature coefficient β^* with high probability. Moreover, it is directly to show that as the random sample size n_T increases, $\mathcal{G}_0(T, s)$ decreases towards 0, which suggests that the Lasso estimator asymptotically converges to its true value.

Recall that through introducing the ℓ_1 penalty term, the Lasso method can perform feature selection by identifying potentially significant features. In particular, we can threshold the Lasso estimate by only keeping dimensions whose estimated coefficient values $|\hat{\beta}_j|$ exceed the threshold value $h(T)$:

THEOREM 2. *Let* $h(T)$ *be a non-negative function of* T *and the thresholded index set* $\mathcal{S} := \{j : |\hat{\beta}_j| \geq h(T)\}$. *Under the event* $\mathcal{E}_{lasso}(T)$, *we have (i)* $|\beta_j^*| \leq h(T) + \mathcal{G}_0(T, s)$ *for all* $j \notin \mathcal{S}$, *and (ii)* $|\mathcal{S}| \leq s + \mathcal{G}_0(T)/h(T)$.

The first part of Theorem 2 demonstrates that when the feature j is not in the thresholded index set for significant features (i.e., $j \notin \mathcal{S}$), then its underlying true coefficient value β_j^* will be small. In other words, features outside of the thresholded index set $\mathcal{S}$ will have little influence on the user's utility and choice probabilities, regardless of whether this feature is actually a significant feature, $j \in \mathcal{S}^*$, or an insignificant feature, $j \notin \mathcal{S}^*$. Further, note that $\mathcal{G}_0(T, s)$ decreases in the random sample size n_T. Hence, if we choose $h(T)$ to also be decreasing function of n_T, then the influence of unselected features can be controlled by the random sample size.

The second part of Theorem 2 reveals that the thresholded index set for significant features identified by thresholding the Lasso estimator can be upper bounded. In fact, we can view $h(T)$ as the controlling parameter to balance the trade-off between the computation efficiency and the selection bias. By setting a large $h(T)$, we lower the selected dimension $\mathcal{S}$, which reduces the computational cost. However, lowering the selected dimension will hurts the regret performance, as more significant dimensions might be dropped from the thresholded index set $\mathcal{S}$.

4.2. Random Projection and Coefficient Estimation

If the decision-maker relies merely on features in the thresholded index set to assess users' utilities and choices, then he can reduce the original high-dimensional parameter estimation problem to a low-dimensional one by ignoring all features outside of the thresholded index set $\mathcal{S}$. Yet, without sufficient random samples, the Lasso may erroneously include insignificant features and exclude some significant features in the underlying true model, which causes the *model misspecification* problem. As many significant features/information will be hidden outside of the thresholded index set $\mathcal{S}$, ignoring these details will lead to a suboptimal assortment selection, lowering the decision-maker's reward. However, estimating coefficients for all features outside of the thresholded index set $\mathcal{S}$ will still be time-consuming, because these features remain high-dimensional. Therefore, to recycle information contained in these features, we propose reducing the dimensionality of these features to a low-dimensional space via random projection and then estimating coefficients for features in both the index set $\mathcal{S}$ and the projected low-dimensional space.

In a nutshell, random projection achieves dimension reduction by multiplying the original high-dimensional matrix by a random projection matrix, resulting a low-dimensional subspace with the same number of samples but fewer projected features. The Johnson-Lindenstrauss Lemma (Johnson and Lindenstrauss 1984) shows that the distance among points under the original high-dimensional space can be largely preserved under the projected low-dimensional space with high probability, and many theoretical studies and empirical applications have demonstrated the value of random projection as a computationally efficient method for dimension reduction (Pilanci and Wainwright 2015). In this study, we will project high-dimensional $(d-|\mathcal{S}|)$ features outside of the thresholded index set $\mathcal{S}$ into a low-dimensional m projected features by multiplying a random projection matrix $P \in \mathbb{R}^{m\times(d-|\mathcal{S}|)}$.

There are two popular choices for the random projection matrix P in the literature: Gaussian random projection matrix and sparse random projection matrix. In Gaussian random projection matrix, each entry $P_{i,j}$ is i.i.d. distributed and follows Gaussian distribution $N(0,1/m)$, whereas in spare random projection matrix, entries take values $\{-\sqrt{v},0,\sqrt{v}\}$ with probabilities $\{1/(2v),1-1/v,1/(2v)\}$, where $v>1$ is a parameter selected by the decision-maker. Clearly, increasing v decreases the number of nonzero elements in spare random projection matrix – the projection matrix becomes sparser. Therefore, spare random projection matrix is faster to generate, manipulate, and store than Gaussian random projection matrix. Yet, projecting high-dimensional data into a low-dimensional space will inevitably result in *information loss* (e.g., the Euclidean distance under the original high-dimensional space may not be precisely preserved under the projected low-dimensional space), so the cost of choosing sparse random projection is additional information loss

in preserving the pairwise distances (Li et al. 2006). In this research, we focus on Gaussian random projection matrix for a tighter theoretical regret bound.

First, we can bound the distance between the original high-dimensional vector and the projected low-dimensional vector with a certain probability guarantee.

LEMMA 2. *[Norm preservation] Let $P = (p_{ij})$ be a random $d \times m$ matrix such that each entry p_{ij} is chosen independently according to $N(0, 1/m)$. For every vector $u \in \mathbb{R}^d$ and $\epsilon \in (0, 1/2]$, the event $\mathcal{E}_{rp}(m, d, \epsilon) := \{|\|Pu\|_2^2 - \|u\|_2^2| \leq \epsilon \|u\|_2^2\}$ holds with probability $1 - 2\exp(-\epsilon^2 m/8)$.*

Lemma 2 demonstrates that Gaussian random projection can largely preserve the geometry structure of the original vector with reasonable distortions with high probability. Hence, by adopting random projection techniques, the decision-maker can significantly reduce the computational time without much sacrifice to the accuracy of parameter estimation. Yet, distortions or information loss in the process of projecting high-dimensional data to a low-dimensional space could lead to a worse regret performance, because such distortions do not vanish over time (Kuzborskij et al. 2018). To limit the negative influence of information loss in random projection, we propose combining random projection with the Lasso.

Recall that as the random sample size increases, the Lasso can learn and gradually identify significant features by the thresholded index set $\mathcal{S}$ (Theorem 2). When the random sample size is large enough, most useful information will already be contained within these features identified by thresholding the Lasso. Therefore, the Lasso-RP-MNL algorithm will only project high-dimensional features *outside* of the thresholded index set $\mathcal{S}$ to a low-dimensional space and then estimate coefficients for features in both the thresholded index set $\mathcal{S}$ and the projected space. Therefore, the long-term information loss (due to random projection) will be limited by the Lasso, and the negative influence of model misspecification (due to the Lasso under limited samples) can be mitigated by recycling features outside of the thresholded index set $\mathcal{S}$ via random projection.

Now, given an thresholded index set $\mathcal{S}$, we describe the process of constructing the projection matrix P_0 and the permutation matrix Q. Through these two matrices, the decision-maker can keep features in the index set $\mathcal{S}$ unchanged while randomly projecting the remaining $(d - |\mathcal{S}|)$ features to a lower m dimensions. To this end, we first need to generate a random projection matrix $P \in \mathbb{R}^{m \times (d - |\mathcal{S}|)}$, where $P_{i,j}$ follows Gaussian distribution $N(0, 1/m)$. Then, combining the random projection matrix P with an identity matrix $I \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}|}$, we can construct the projection matrix $P_0 = \begin{pmatrix} I & 0 \\ 0 & P \end{pmatrix} \in \mathbb{R}^{(|\mathcal{S}|+m) \times d}$. Next, we use $Q \in \mathbb{R}^{d \times d}$ to denote the permutation matrix, which moves the original feature vector x's significant features in the index set $\mathcal{S}$ to the top $|\mathcal{S}|$ places in the permuted feature vector Qx. Hence, by multiplying the projection matrix P_0 by the permuted feature vector Qx, we project the original d dimensional vector to a low-dimensional $(|\mathcal{S}| + m)$

vector, in which the first $|\mathcal{S}|$ elements are original features in the thresholded index set $\mathcal{S}$ identified by the Lasso and the remaining m elements are the projected features by projecting the original features not in the thresholded index set $\mathcal{S}$ via the random projection matrix P. For simplicity of notation, we define z as the projected feature vector, $z := P_0 Q x$. Similarly, the following notations are used throughout this paper: $\theta^* := P_0 Q \beta^*$ and $\Sigma := Q^T P_0^T P_0 Q$.

THEOREM 3. *Let matrix P_0 and Q be constructed by the thresholded index set $\mathcal{S} := \{j : |\hat{\beta}_j| \geq h(T)\}$. When event $\mathcal{E}_{lasso}(T)$ holds, the event $\mathcal{E}_2(m, T, \epsilon) := \{\|\beta^* - \Sigma\beta^*\|_2 \leq \epsilon\sqrt{s} \cdot (h(T) + \mathcal{G}_0(T, s)) := \mathcal{G}_1(m, T, \epsilon)\}$ holds with probability $1 - 4\exp\left(-\frac{m}{8}\epsilon^2\right)$.*

Theorem 3 demonstrates that the true feature coefficient vector β^* is nearly invariant under the projection Σ, which directly suggests that our proposed projection scheme is nearly optimal in the sense that it will not introduce estimation error when predicting users' utilities and choices asymptotically. Specifically, consider projecting both the feature vector x and the coefficient vector β^* by using P_0 and Q. Then, the projected utility can be written as $(P_0 Q x)^T P_0 Q \beta^* = x^T \Sigma \beta^*$. Note that as illustrated in Theorem 3, the time dependence of the term $\|(I - \Sigma)\beta^*\|$ is on the order of $n_T^{-1/2} \log^{3/2} T$. Therefore, if we can ensure that the random sample size n_T is on the order of at least $\mathcal{O}(T^c)$ up to time T (to be detailed in §4.4), where c is an arbitrary positive constant, then $\|(I - \Sigma)\beta^*\|$ will converge to 0 with high probability.

By combining the Lasso and random projection, the decision-maker can project the original high-dimensional d features into a low-dimensional $(|\mathcal{S}| + m)$ space so that the parameter estimation can be performed in a low-dimensional fashion. Specifically, we estimate the coefficients for the projected feature vector $z = P_0 Q x$ as follows:

$$\hat{\theta} = \arg \min_{\|\theta - \theta_0\| \leq \tau} L_z(\theta), \text{ where } L_z(\theta) := \frac{1}{T} \sum_{t=1}^{T} \log \left(e^{z_{c_t}^T \theta} / \left(1 + \sum_{i \in A_t} e^{z_i^T \theta} \right) \right). \tag{4}$$

The τ in Eq. 4 is a positive constant selected by the decision-maker and $\theta_0 = \arg\min_\theta \|\theta - P_0 Q \hat{\beta}\|$. The $\|\theta - \theta_0\| \leq \tau$ is a local constraint added in Eq. (4) to prevent over-fitting, and we solve $\hat{\theta}$ only in the local space around θ_0. Because from Lemma 1, we know that $\hat{\boldsymbol{\beta}}$ will not be far away from $\boldsymbol{\beta}^*$, it implies that $\hat{\theta}$ is also close to $P_0 Q \beta^*$ with high probability.

Similarly to Assumption A.2, which ensures the identifiability of the Lasso estimator in Eq. (3) under the original high-dimensional space, we need the last technical assumption, which requires L_z to be strongly convex under the projected space for random samples, to achieve the identifiability of the estimator $\hat{\theta}$ in Eq. (4) under the projected space:

Assumption A.3: When all samples in $L_z(\theta)$ are i.i.d. random samples, there exists a $\mu > 0$ such that for any v and feasible solution ξ in the projected space, we have $\mathbb{E}\left[v^T \nabla^2 L_z(\xi) v\right] \geq \mu \|v\|^2$.

4.3. Assortment Selection

In this subsection, using the estimated coefficient vector in the projected space, we construct the upper-confidence bound for each individual product's attraction parameter, identify the optimistic assortment, and establishes the single-period regret upper bound for the optimistic assortment.

Given an arbitrary assortment $\mathcal{A}$, we denote the decision-maker's expected reward for a coefficient vector θ under the projected space as

$$\mathcal{R}_{\mathcal{A}}(\theta) = \sum_{i\in\mathcal{A}} \frac{r_i e^{z_i^T\theta}}{1+\sum_{j\in\mathcal{A}} e^{z_j^T\theta}}.$$

The following Lemma establishes an upper bound on the expected reward difference between the estimator $\hat{\theta}$ in Eq. (4) under the projected space and the projected true coefficient vector $P_0Q\beta^*$:

LEMMA 3. *Denote* $f_{\mathcal{A}}(\theta) = \mathbb{E}[p_{\Sigma\beta^*,\mathcal{A}}(i)\log(p_{Q^TP_0^T\theta,\mathcal{A}}(i)/p_{\Sigma\beta^*,\mathcal{A}}(i))]$ *and* $\delta = \|\hat{\theta} - P_0Q\beta^*\|$. *Let* L_3, λ_{max}, *and* ρ *be positive constants such that for all* $t>0$ *and feasible* θ_2, θ_1, ξ, *we have* $\|\nabla^2 f_{\mathcal{A}_t}(\theta_1) - \nabla^2 f_{\mathcal{A}_t}(\theta_2)\|_{op} \le L_3\|\theta_1-\theta_2\|$, $\|\nabla^2\mathcal{R}_{\mathcal{A}_t}(\xi)\|_{op} \le \lambda_{\max}$, *and* $\min_{\mathcal{A},i\in\mathcal{A}} p_{\beta^*,\mathcal{A}}(i) \ge \rho$. *Under Assumption A.1 and A.3, if* $\delta \le \min\{\frac{n_T\mu}{4TL_3}, \frac{\rho}{8Kx_{\max}}\}$, $\mathcal{G}_1(m,T,\epsilon) \le \frac{\rho}{8Kx_{\max}}$, *and events* $\mathcal{E}_2(m,T,\epsilon)$ *and* $\mathcal{E}_{rp}(m,d,1/2)$ *hold for all products, then the following inequality holds for all assortment* $\mathcal{A}$ *with probability* $1-4\exp\left(-\frac{m}{8}\epsilon^2\right) - \mathcal{O}(1/T)$*:*

$$|\mathcal{R}_{\mathcal{A}}(\hat{\theta}) - \mathcal{R}_{\mathcal{A}}(P_0Q\beta^*)| \le \sqrt{2}R_{\max}\omega_T \sqrt{\left\|\left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-\frac{1}{2}} \nabla^2 f_{\mathcal{A}}(\hat{\theta}) \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-\frac{1}{2}}\right\|_{op}} + \frac{1}{2}\lambda_{\max}\delta^2, \tag{5}$$

where $\omega_T = 4\sqrt{4x_{\max}^2 T\mathcal{G}_1(m,T,\epsilon)\delta + 32(2(|\mathcal{S}|+m)+1)\log(T) + 2\Gamma_T}$ *and* $\Gamma_T = \max\{0, TL_z(\hat{\theta})\}$.

The upper bound established in Lemma 3 has two components. The first term in the right-hand-side of Eq. (5) is the typical UCB-type upper-confidence bound, whereas the additional second term, $\lambda_{\max}\delta^2/2$, accounts for the possible influence of model misspecification and information loss. In the classic setting, we assume that the true model always falls in the solution space so that the estimator enjoys asymptotical unbiasedness. However, this assumption no longer holds when the Lasso fails to identify all significant features and random projection is used to compress all remaining high-dimensional features, potentially including true significant features. In such a scenario, the decision-maker will face an upper bound worse than the typical UCB-type bound. Yet, such negative influences of model misspecification and information loss should not be of much concern for a large sample size. In particular, note that in Lemma 3, we require that δ converges to zero at the rate of $\mathcal{O}(n_T/T)$, and therefore the additional second term $\lambda_{\max}\delta^2/2$ diminishes at the rate of $\mathcal{O}(T^{-1/2})$, if we ensure $n_T = \tilde{\mathcal{O}}(T^{1/2})$ (see §4.4 for details), under which case the upper bound established in Lemma 3 converges to the typical UCB-type bound.

When the given assortment includes merely a single item, we can establish an upper bound for each individual product's attraction parameter. Specifically, consider a single-item assortment $\mathcal{A}$ that contains a single product with the feature vector x and reward $r = 1$. The decision-maker's expected reward can be simplified to $(e^{x^T\beta^*}/(1+e^{x^T\beta^*}))$, and the attraction parameter can be upper bounded as in the following corollary.

COROLLARY 1. *Let $\mathcal{A}$ be the assortment with a single item characterized by the feature vector x. If the same conditions stated in Lemma 3 hold, then with probability $1-4\exp\left(-\frac{m}{8}\epsilon^2\right)-\mathcal{O}(1/T)$, we have $e^{x^T\beta^*} \le v^{ucb}$, where $\eta = \exp(x_{\max}b)$ and*

$$v^{ucb} = \exp(x^T Q^T P_0^T \hat{\theta}) + \eta x_{\max}\mathcal{G}_1(m,T,\epsilon) + \frac{\lambda_{\max}}{2+2\eta^2}\delta^2 + \frac{\sqrt{2}R_{\max}\omega_t}{1+\eta^2}\sqrt{\left\|\left(\sum_{i=1}^{T}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\nabla^2 f_{\mathcal{A}}(\hat{\theta})\left(\sum_{i=1}^{T}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}}. \quad (6)$$

This upper bound for the single product's attraction parameter will facilitate the analysis of the regret upper bound for the decision-maker's "optimal" static assortment. In particular, given the current estimator $\hat{\theta}$, in order to maximize his single-period expected reward, the decision-maker needs to offer at most K products out of N candidates. Equivalently, the decision-maker solves the following static assortment optimization problem under the cardinality constraint:

$$\max_{\substack{\mathcal{A}\subseteq\{1,2,\ldots,N\}\\ |\mathcal{A}|\le K}} \left\{R_{\mathcal{A}}\left(\hat{\theta}\right)\right\}.$$

The key challenge in identifying the optimal assortment is that the decision-maker must search through a combinatorial space of N products. In practice, the number of products can easily exceed hundreds or thousands, which makes the problem computationally intractable for online assortment optimization problems. Therefore, following Davis et al. (2013), we reformulate this combinatorial optimization problem as a linear programming problem:

$$\max_{\boldsymbol{w}} \sum_{i\in N} r_i w_i, \text{ s.t. } \sum_{i\in N} w_i + w_0 = 1;\ \sum_{i\in N}\frac{w_i}{v_i} \le K w_0;\ 0 \le w_i \le w_0 v_i \text{ for } i \in N. \quad (7)$$

As at most K decision variables will be none-zero under the optimal solution, various efficient solution algorithms, such as column-generation techniques, can be adopted to expedite the computation for this LP problem. Now, we replace the product i's attraction parameter v_i in Eq. (7) by v_i^{ucb} and denote the optimal solution to the resulting problem as $\boldsymbol{w}^*$. Then, we refer to the assortment $\mathcal{A}^{SRP} = \{i \in N : w_i^* > 0\}$ as the static assortment under random projection. It is also worth noting that that the static assortment under random projection $\mathcal{A}^{SRP}$ may not be the true optimal assortment under the original high-dimensional space for two reasons: The projected space

may not contain the true coefficients, where the systemic bias may appear, and the estimator $\hat{\theta}$ may not match the best possible candidate of the true coefficients in the projected space.

Now, we denote the decision-maker's expected reward for a given assortment $\mathcal{A}$ under the true coefficient β^* in the original high-dimensional space as

$$\mathcal{R}_{\beta^*}(\mathcal{A}) = \sum_{i \in \mathcal{A}} \frac{r_i e^{x_i^T \beta^*}}{1 + \sum_{j \in \mathcal{A}} e^{x_j^T \beta^*}}.$$

Further, we use $\mathcal{A}^*$ to denote the optimal assortment, which can be identified by searching the combinatorial space of all products to maximize $\mathcal{R}_{\beta^*}(\mathcal{A})$, i.e., $\mathcal{A}^* = \arg\max_{\mathcal{A}, |\mathcal{A}| \leq K} \mathcal{R}_{\beta^*}(\mathcal{A})$. Next, we bound the expected reward difference between the static assortment under random projection $\mathcal{A}^{SRP}$ and the optimal assortment $\mathcal{A}^*$ in the following theorem:

THEOREM 4. *Let $\mathcal{A}^{SRP}$ be the static assortment under random projection and $\mathcal{A}^*$ be the optimal assortment. Under the same conditions as in Lemma 3, the following inequality holds with probability $1 - \mathcal{O}(1/T)$:*

$$\begin{aligned}\mathcal{R}_{\beta^*}(\mathcal{A}^*) - \mathcal{R}_{\beta^*}(\mathcal{A}^{SRP}) \leq\ & R_{\max} K \eta x_{\max} (2\delta + 2\mathcal{G}_1(m, T, \epsilon)) + \frac{R_{\max} K \lambda_{\max}}{2 + 2\eta^2} \delta^2 \\ & + \frac{R_{\max}^2 K \eta^{3/2} (1 + K\eta) \sqrt{2} \omega_t}{(1 + \eta^2)(1 + \eta)} \sqrt{\left\| \left(\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta}) \right)^{-1/2} \nabla^2 f_{\mathcal{A}^{\mathcal{SRP}}}(\hat{\theta}) \left(\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta}) \right)^{-1/2} \right\|_{op}},\end{aligned}$$

The upper regret bound for the static assortment under random projection $\mathcal{A}^{SRP}$ can be divided into two parts. On the one hand, the first part, $R_{\max} K \eta x_{\max} (2\delta + 2\mathcal{G}_1(m, T, \epsilon)) + \frac{R_{\max} K \lambda_{\max}}{2 + 2\eta^2} \delta^2$, comes from the single product utility decomposition. The magnitude of the first part of the upper-confidence bound can be regulated by a judiciously designed random sample schedule. In particular, for a small sample size, it is straightforward to show that $\mathcal{G}_1(m, T, \epsilon) = \mathcal{O}(\sqrt{\log T / n_T})$ and $\delta = \mathcal{O}(n_T / T)$. Therefore, if we can design the random sample size n_T to be on the order of at least $\tilde{\mathcal{O}}(T^{2/3})$, then the first part will be upper-bounded at most by the order $\tilde{\mathcal{O}}(T^{-1/3})$. On the other hand, the second part is the typical UCB-type upper-confidence bound, which shares the typical quadratic upper bound as in UCB-type algorithms and can be further bounded by the elliptical potential lemma (see Dani et al. 2008, Rusmevichientong et al. 2010, Filippi et al. 2010, Li et al. 2017).

It is worth mentioning that the common K parameter in the right-hand-side stems from the fact that we construct the product-based bound instead of the assortment-based bound. In particular, instead of enumerating all possible combinations (choose K out of N products) and building confidence bounds for each combination (as in Chen et al. 2018), we construct confidence bounds for each product (see Corollary 1) so that the estimation error will accumulate with respect to

the assortment size K. Although constructing the product dependent bound instead of assortment dependent bound results in additional cost, the decision-maker benefits from the significant reduction in computational time. For example, consider a scenario where the decision-maker selects 20 out of $1,000$ products: Instead of constructing 3.4×10^{41} confidence bounds for assortments, we need to build only 10^3 confidence bounds – one confidence bound for each product. Hence, in practice, where the decision-maker typically faces a large product candidate set, our algorithm becomes more efficient and pragmatic in online decision-making settings.

4.4. Lasso-RP-MNL Algorithm

Recall that the estimation accuracy and assortments' regret bounds are contingent on the number of random samples n_T (see Theorem 2-4). Therefore, we need to design a sampling schedule that generates sufficient, but not excessive, random samples:

***Random Decay Sampling Schedule*:** At the beginning of each time t, the decision-maker draws a random number r_t that follows Bernoulli distribution with success probability $\mathbb{P}(r_t = 1) = \min\{1, C_0 t^{-c_1}\} := P_{C_0,c_1}(t)$, where C_0 is selected by the decision-maker. If $r_t = 1$, then the decision-maker randomly selects K products as the assortment to the user arriving at time t.

Not that under the random decay sampling schedule, the random sampling probability $P_{C_0,c_1}(t)$ decreases at the rate of t^{-c_1}. Therefore, as time t increases, the probability of random samples decreases towards zero. Hence, by controlling the sampling schedule, the decision-maker can generate enough random samples for proper estimation without sacrificing too much regret.

Now, we are ready to present the proposed Lasso-RP-MNL algorithm as follows:

Algorithm : Lasso-RP-MNL Algorithm

Require: Input $P_{C_0,c_1}(t)$, $h(t)$, m, λ_0, and the Lasso step set T_{lasso}. Initialize $t=1$, $\mathcal{W}_R=\emptyset$, $\mathcal{W}=\emptyset$, $P_0 \in \mathbb{R}^{m\times d}$ with i.i.d $N(0,1/m)$ Gaussian random elements, $Q=I$, $\theta_0=\mathbf{0}$, $\{\omega_t\}$, and $\tau=+\infty$.

for $t=1,2,...$ **do**

Draw a Bernoulli random number b_t with success probability $P_{C_0,c_1}(t)$.

if $b_t=1$ **then**

1. Randomly select K products from the candidate set $\{1,2,...,N_t\}$ as the assortment $\mathcal{A}_t$.

2. Observe the user's choice $c_t \in \mathcal{A}_t \cup \{0\}$ and update $\mathcal{W}_R=\mathcal{W}_R\cup\{t\}$ and $\mathcal{W}=\mathcal{W}\cup\{t\}$.

else

1. Solve Eq. (4) for $\hat{\theta}$ with samples in $\mathcal{W}$ and update attraction parameters' upper bounds $v_i=v_i^{ucb}$ for $i\in\{1,2,...,N_t\}$ according to Eq. (6).

2. Plug the updated v_i back to Eq. (7), solve for $\boldsymbol{w}^*$, and offer $\mathcal{A}_t=\{i\in N_t : w_i^*>0\}$.

3. Observe the user's choice $c_t\in\mathcal{A}_t\cup\{0\}$ and update $\mathcal{W}=\mathcal{W}\cup\{t\}$.

end if

if $t\in T_{lasso}$ **then**

Solve Eq. (3) for $\hat{\beta}$ with samples in $\mathcal{W}_R$ and $\lambda=\lambda_0\sqrt{(\log d+\log t)/|\mathcal{W}_R|}$; re-construct the thresholded index set $\mathcal{S}$, the projection matrix P_0, and the permutation matrix Q; set $\theta_0=\arg\min\|\theta-P_0Q\hat{\beta}\|$ and $\tau=\min\{\frac{\rho}{8Kx_{\max}},\frac{n_t\mu}{4tL_3}\}$.

end if

end for

The Lasso-RP-MNL algorithm starts with assigning values for system parameters and initialing intermediate matrices and variables. For a user arriving at time t, the decision-maker will follow the random decay sampling schedule to draw a Bernoulli random number. If this random number equals 1, then the decision-maker will randomly select K products from the product candidate set, offer the resulting assortment to the user, observe the user's choice, and finally include this sample in both the random sample index set $\mathcal{W}_R$ and the whole sample index set $\mathcal{W}$. Otherwise, if the Bernoulli random number equals 0, then the decision-maker will first estimate the coefficients for the projected low-dimensional feature vector $\hat{\theta}$, based on which the decision-maker will update the attraction parameters' upper-confidence bounds for all products in the candidate set; next, the decision-maker will treat these upper-confidence bounds as new attraction parameters in the MNL model and plug them back into Eq. (7) to identify the assortment that maximizes his expected reward; then, the decision-maker will offer this assortment to the user, observe the user's choice, and include this sample in the whole sample index set $\mathcal{W}$ only.

Finally, before moving to the next user, the decision-maker will check whether the current time t belongs to a predetermined Lasso step set T_{lasso}. If not, then the decision-maker does not need to do anything and will move directly to the next user. Otherwise, if $t \in T_{lasso}$, then the decision-maker will update the thresholded index set $\mathcal{S}$ via the Lasso using only random samples in the index set $\mathcal{W}_R$, reconstruct the projection matrix P_0 and the permutation matrix Q, recalculate the local solution θ_0, and then move to the next user.

The next theorem establishes the expected cumulative regret upper bound for the Lasso-RP-MNL algorithm.

THEOREM 5. *Under Assumptions $A1 - A3$, if we set $T_{lasso} = \{c^i,\ i = 0, 1, 2, ...\}$ with a positive integer $c > 1$, $h(t) = \mathcal{G}_0(T, \hat{s})$ for a positive $\hat{s}$, $C_0 \leq 1$, $c_1 = 1/3$, $\lambda_0 = 2\sqrt{2}x_{\max}$, $m = \max\{8\hat{s}\log T, 32\log(TN)\}$, and $T \geq (2\log T/(CC_0))^{\frac{3}{2}}$. Then, with probability $1 - T^{\hat{s}/s} - \mathcal{O}(T^{-1})$, the expected cumulative regret of the Lasso-RP-MNL algorithm is upper bounded as follows:*

$$\begin{aligned} \textit{Regret}(T) &\leq \left(\tilde{C}_{f,1}\left(3s + \hat{s} + m\right) + \tilde{C}_{f,2} + \tilde{C}_{f,3}\sqrt{(2s+m)} + 2C_0\right) T^{\frac{2}{3}} \\ &\lesssim \tilde{\mathcal{O}}\left(s\sqrt{\log d} \cdot T^{\frac{2}{3}}\right), \end{aligned}$$

where $\tilde{C}_{f,1} = \tilde{\mathcal{O}}(\sqrt{\log d})$, $\tilde{C}_{f,2} = \tilde{\mathcal{O}}(1)$ and $\tilde{C}_{f,3} = \tilde{\mathcal{O}}(1)$.

From the regret lower bound established in Theorem 1, we can directly argue that the Lasso-RP-MNL algorithm matches the regret lower bound in the sample size dimension T, up to a logarithmic factor, in the data-poor regime (i.e., $\Omega(T^{\frac{2}{3}})$). We want to highlight this result by comparing it to the scenario where the decision-maker relies only on the Lasso to perform dimension reduction. In particular, consider an auxiliary Lasso-only algorithm in which, keeping everything else unchanged, the decision-maker estimates coefficients only for features thresholded by the Lasso and ignores all remaining features. Under such an auxiliary Lasso-only algorithm, if the Lasso fails to fully identify significant features in the data-poor regime, which is highly possible due to insufficient random samples, then the decision-maker will rely on the misspecified model to perform coefficient estimation and assortment selection. Under this scenario, the expected single-step regret will be proportional to the strength of the model bias and can lead to a linear cumulative regret on T. The Lasso-RP-MNL algorithm, however, estimates coefficients for two sets of features. The first set is the $|\mathcal{S}|$ features by thresholding the Lasso estimator, and the second set is all remaining $(d - |\mathcal{S}|)$ features projected down to m dimensions by random projection. Therefore, the negative influence of model misspecification can be partially mitigated by recycling features in the second set, so we can improve the cumulative regret's dependence on time T from linear to sublinear, $\tilde{\mathcal{O}}(T^{\frac{2}{3}})$.

Theorem 5 further shows that the Lasso-RP-MNL algorithm can improve the regret upper bound on the feature dimension from a linear dependence in the literature (e.g., $\mathcal{O}(d)$ in Chen et al.

2018, Oh and Iyengar 2019) to a sub-logarithmic dependence $\mathcal{O}(\sqrt{\log d})$. This improvement is of particular importance for regret performance under high-dimensional settings, where the feature dimension d is extremely large, and we believe that the Lasso-RP-MNL algorithm is the first algorithm being able to reach the logarithmic bound on the feature dimension for assortment optimization problems.

REMARK 1. Note that the Lasso-RP-MNL algorithm runs the Lasso to update the index set $\mathcal{S}$ only when the current time t belongs to the set T_{lasso}. This is because solving the Lasso problem can be time-consuming, which makes it impractical to update the Lasso for every arriving user under the online decision-making scenario. Hence, we construct a very sparse Lasso step set such that the number of users between two consecutive Lasso runs increases exponentially by setting $T_{lasso} = \{t : t = c^i,\ i = 0, 1, 2, ...\}$ with a positive integer $c > 1$. Therefore, as time progresses, the frequency of updating the Lasso decreases at an exponential rate, which alleviates the computational burden associated with solving the Lasso under high-dimensional data with large sample sizes, while maintaining proper accuracy for parameter estimation.

REMARK 2. The algorithm does not assume any knowledge of the true value for the significant dimension s. Yet, for better empirical performance, we tend to select $\hat{s}$ in Theorem 5 to be sufficiently large so that it will be larger than s, i.e., $\hat{s} \geq s$, under which case we can get rid of the $T^{\hat{s}/s}$ term and have the probability to be $1 - \mathcal{O}(T^{-1})$. In practice, the decision-maker typically relies on experts' opinions or from previous offline data to have a rough guess on s and then to set $\hat{s}$ to upper bound that value. If, however, $\hat{s}$ is picked to be smaller than s, then in order to get rid of the $T^{\hat{s}/s}$ term, there will be one additional $\mathcal{O}(s^{\frac{1}{2}})$ term in the regret upper bound, which can be derived by setting $\epsilon = O(1)$ in Eq. (EC.65) and used the same proof procedure of Theorem 5. In addition, it is also possible to follow Oh et al. (2020) to design an algorithm without any knowledge of s and without the need for random sampling by introducing the relaxed symmetry assumption, and we will leave such consideration for future exploration.

4.5. Improved Upper Bound

Next, we show that the Lasso-RP-MNL algorithm's regret upper bound can be further sharpen to $\tilde{O}(T^{\frac{1}{2}})$ in the data-rich regime. Let's denote $\beta_{\min} = \min_{j \in \mathcal{S}^*} |\beta_j^*|$ as how the significant signal is bound away from zero (i.e., the minimum signal strength), and intuitively, the feature selection by thresholding the Lasso estimator will be much easier under a large $\beta_{\min}$ value. In theorem 2, we show that for $j \notin \mathcal{S}$, we have $|\beta_j^*| \leq h(T) + \mathcal{G}_0(T, s)$. Hence, if we set $h(T) = \mathcal{G}_0(T, \hat{s})$, then for a large T, the term $h(T) + \mathcal{G}_0(T, s)$ becomes smaller than $\beta_{\min}$, which implies that features outside of the thresholded index set $\mathcal{S}$ are actually insignificant features. In this case, the projection will not introduce any distortion, i.e., $\|\beta^* - \Sigma\beta^*\|_2 = 0$. As a result, the high-dimensional problem can

be reduced to the conventional low dimensional setting, and the regret upper bound derived in Theorem 5 can be improved to $\tilde{O}(T^{\frac{1}{2}})$, which is formally summarized in the following corollary.

COROLLARY 2. *Under the same conditions as in Theorem 5 and set* $c_1 = 1/2$*, when* $T \gtrsim \mathcal{O}((s^2 \log d(\beta_{\min})^{-2})^2)$*, the expected cumulative regret for the Lasso-RP-MNL algorithm is upper bounded as follows:*

$$\textit{Regret}(T) \lesssim \tilde{\mathcal{O}}\left(s\sqrt{\log d} \cdot T^{\frac{1}{2}}\right).$$

5. Empirical Experiments

In this section, we benchmark the Lasso-RP-MNL algorithm to existing state-of-the-art algorithms in the literature and industrial practices. In §5.1, we first explore the benefits of the Lasso-RP-MNL algorithm by comparing it to four benchmarks to illustrate the value of the high-dimensional contextual information and the value of dimension reduction techniques. In §5.2, we simulate the real practice environment, where users are heterogeneous, the product candidate set is large, and the feature vector is high-dimensional, to examine the impacts of the size of the product candidate set N, the feature dimension d, and the projection dimension m on Lasso-RP-MNL's cumulative regret performance and computational time. Finally, in §5.3, we use the high-dimensional XianYu online assortment recommendation dataset to evaluate the Lasso-RP-MNL algorithm's performance in a real practice scenario, where the technical assumptions specified early may not hold.

5.1. The Benefits of Lasso-RP-MNL: A Preliminary Illustration

The benefits of the Lasso-RP-MNL algorithm can be justified by two key factors: incorporating high-dimensional contextual information and combining two dimension reduction techniques (i.e., the Lasso and random projection). Hence, to separately gauge the impacts of these two factors, we consider four benchmark algorithms in the first synthetic experiment, as follows:

- *MNL-Bandit*: Proposed by Agrawal et al. (2019), MNL-Bandit is a UCB-based algorithm without the contextual information.
- *Benchmark 1 (With Features)*: Benchmark 1 follows the same structure as the Lasso-RP-MNL algorithm, but without using the Lasso and random projection; this benchmark estimates the unknown coefficient vector β^* under the original high-dimensional space.
- *Benchmark 2 (RP Only)*: Benchmark 2 follows the same structure as the Lasso-RP-MNL algorithm, but it does not update the thresholded index set via the Lasso (i.e., $\mathcal{S} \equiv \emptyset$); instead, it projects all features into a low m-dimensional space via random projection.
- *Benchmark 3 (Lasso Only)*: Benchmark 3 follows the same structure as the Lasso-RP-MNL algorithm, but it estimates coefficients only for features in the thresholded index set $\mathcal{S}$ and ignores all remaining features outside of $\mathcal{S}$.

Note that MNL-Bandit does not use high-dimensional contextual information to estimate each user-product pair's utility. Instead, it assigns a unique attraction parameter for each product and directly estimates these attraction parameters using the sample average (see Agrawal et al. 2019). Therefore, comparing Benchmark 1 to MNL-Bandit[2] could shed light on the value of incorporating the contextual information. By comparing Benchmark 2 and Benchmark 3 to Benchmark 1, we can separately assess the benefits of the Lasso and random projection under the high-dimensional online assortment optimization setting. Finally, we compare the Lasso-RP-MNL algorithm to Benchmark 2 and Benchmark 3 to gauge the benefits of combining the Lasso and random projection.

5.1.1. Data Generation and Parameter Inputs: In the first experiment, we consider that a decision-maker needs to offer at most 5 products (i.e., $K = 5$) out of a candidate set of 20 products (i.e., $N = 20$) to users. The unknown true coefficient vector β^* is sparse, and only 5 out of a total of 20 coefficients are non-zero (i.e., $d = 20$ and $s = 5$). Without loss of generality, we set the first five features to be significant (i.e., $\boldsymbol{\beta}^* = \{\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, \beta_5^*, 0, 0, ..., 0\}$), and their values are also independently and identically generated by a Gaussian distribution. Finally, the corresponding reward r_i for $i = \{1, 2, ..., 20\}$ is generated from a uniform distribution from 0 to 1. In the experiment, we arbitrarily set the parameters $\lambda_0 = 1$, $C_0 = 3$, $c = 2$ and the projection dimension $m = 3$.

Note that MNL-Bandit can not be directly applied to our setting with heterogeneous users and changing product candidate sets. This is because that the learning in MNL-Bandit is associated with the attraction parameter for each product. Therefore, to learn these attraction parameters, MNL-Bandit requires that the number of products is not too large and the true values of these attraction parameters remain unchanged in the experiment. Hence, to benchmark against MNL-Bandit, we will consider a setting where feature vectors x_i for $i = \{1, 2, ..., N\}$ contain only product features, and the same group of feature vectors is repetitively offered to the decision-maker. In other words, in the first experiment, the decision-maker will repetitively choose from a fix set of 20 products to a group of T homogeneous users. Technically, we generate feature vectors x_i for 20 products once at the beginning of the experiment, which are independently and identically generated from the standard Gaussian distribution, and offer these 20 products repetitively to the decision-maker for every incoming user. This constrain will be relaxed in other synthetic experiments in §5.2 and the XianYu experiment in §5.3.

[2] Besides Benchmark 1, MLE-UCB by Chen et al. (2018) can also be used to illustrate the value of contextual information. Yet, as MLE-UCB does not use dimension reduction techniques for learning and relies on computing individual assortment bounds to identify the optimal assortment, it is highly computational expensive and therefore is unsuitable for online assortment optimization under high-dimensional data.

5.1.2. Results: For each algorithm, we perform 100 trials and report the average cumulative regret in Figure 1 for the first $1,000$ users (i.e., $T = 1,000$), at which time all algorithms, except Benchmark 2, seem to have converged. The average computational time (in seconds) for one trial is 11 for MNL-Bandit, 53 for Benchmark 1, 26 for Benchmark 2, 25 for Benchmark 3, and 33 for Lasso-RP-MNL. Recall that MNL-Bandit uses the sample mean to update its estimators for unknown attraction parameters instead of adopting MLE in all other algorithms. Therefore, MNL-Bandit tends to have better computational time performance, when the number of products in the candidate set N is not large. Without using any dimension reduction techniques, Benchmark 1 requires the longest computational time among all algorithms.

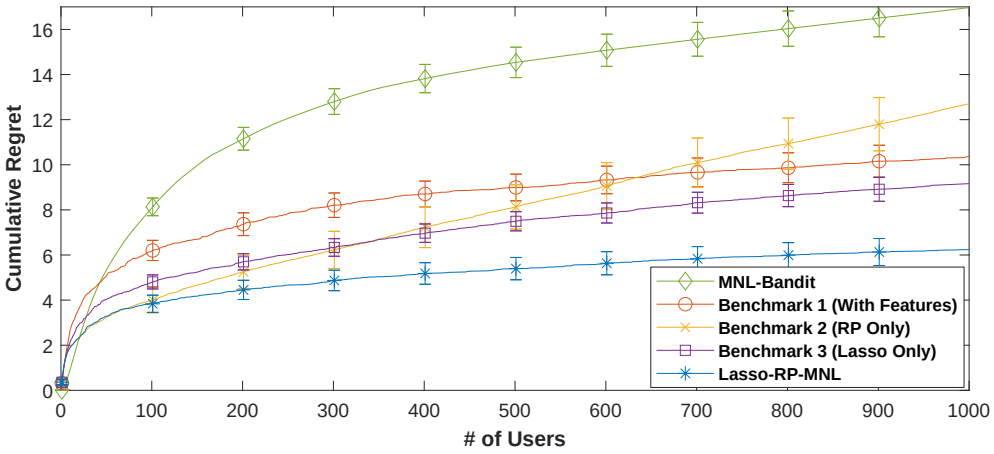


Figure 1 **The impact of T on the cumulative regret, where $d = 20$, $s = 5$, $N = 20$, $K = 5$, and $m = 3$.**

Figure 1 illustrates the cumulative regret performance (with 90% confidence interval error bars) for all algorithms. First, when comparing Benchmark 1 to MNL-Bandit, we observe that algorithms using available contextual information could significantly improve the decision-maker's regret performance. Second, note that as the underlying data possess a sparse structure, the decision-maker could use the Lasso (in Benchmark 3) to perform feature selection, identify the underlying sparse data structure, and improve the accuracy of parameter estimation. Indeed, we observe that with the Lasso, Benchmark 3 further reduces the decision-maker's cumulative regret from Benchmark 1. Third, the Lasso may suffer from model misspecification, especially with limited samples, leading to inaccurate assessment of users' utilities and suboptimal assortment recommendations. Hence, by adding random projection to Benchmark 3 so that features outside the thresholded feature set can be recycled and reused to improve users' utility assessment, Lasso-RP-MNL performs the best among all algorithms.

Finally, it is worth mentioning that Benchmark 2 seems to perform well at the beginning but fails to converge in the experiment. Specifically, we first observe that Benchmark 2 performs exceptionally well under very limited samples. To explain, note that with limited samples, estimating a large number of parameters will inevitably lead to high variances and poor estimates. Hence, by projecting high-dimensional data into a lower dimension, Benchmark 2 could significantly reduce the number of parameters that are needed to be estimated and improve the estimation accuracy, which in turn enables better assortment recommendations. However, Benchmark 2 suffers from information loss in the process of projecting high-dimensional data into a low-dimensional space, which cannot be corrected asymptotically. Hence, as the sample size T increases, the cumulative regret of Benchmark 2 will eventually exceed Benchmark 3, Benchmark 1, and MNL-Bandit sequentially.

5.2. The Impacts of N, d, and m on Lasso-RP-MNL

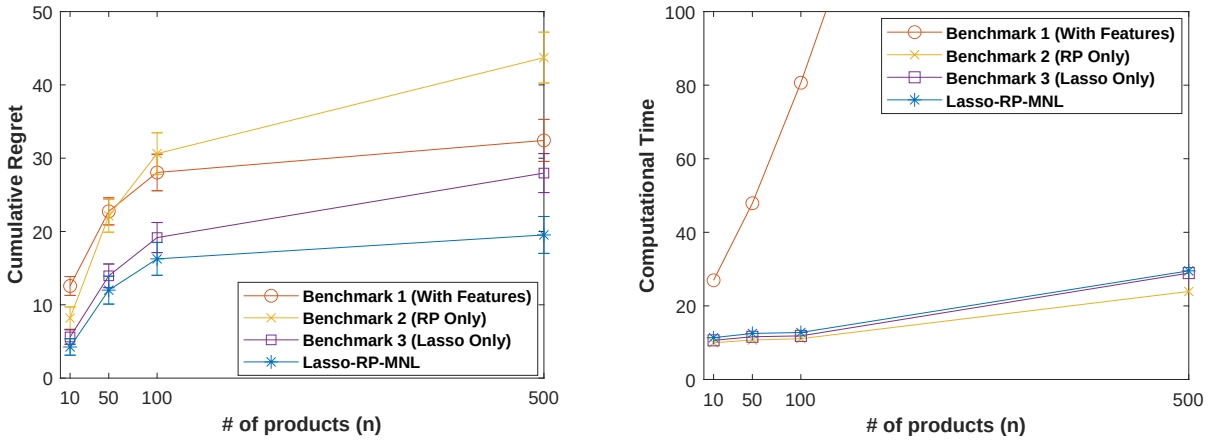
In this subsection, we examine the influences of the number of products in the candidate set N, the feature dimension d, and the projection dimension m. Recall that in the first experiment in §5.1, we consider homogeneous users and restrict the product candidate set to be small and remain unchanged for all users so that MNL-Bandit can be included as a benchmark to illustrate the value of contextual information. In this subsection, however, we simulate a real practice environment, where the decision-maker faces heterogeneous users and available products can be innumerable and change from user to user.

To this end, we largely follow the data generation and parameter inputs discussed in §5.1, except for the process of generating the feature vectors for user-product pairs. In particular, for each arriving user, we will regenerate the feature vectors for N user-product pairs (i.e., x_i for $i = \{1, 2, ..., N\}$) by independently and identically drawing them from a Gaussian distribution. Here, the changes in user-product pairs reflect the changes in both users' features (i.e., heterogeneous users) and products' features (i.e., a different product candidate set). With heterogeneous users and changing product candidate sets, we will benchmark the Lasso-RP-MNL algorithm against Benchmarks 1, 2, and 3 in the following three synthetic experiments.

5.2.1. Impact of the size of the product candidate set N: To examine the impact of the number of products in the candidate set, we vary $N = \{10, 50, 100, 500\}$ while keeping parameters $s = 5$, $d = 50$, $K = 5$, and $m = 5$ unchanged. For different values of N, the Lasso-RP-MNL algorithm always converges before 500 users, so we present the cumulative regret performance and the computational time for Lasso-RP-MNL and Benchmarks 1, 2, and 3 at time $T = 500$.

Figure 2(a) suggests that the cumulative regret for all algorithms seems to increase in the number of products in N. But, among all algorithms, Lasso-RP-MNL has the the lowest cumulative

Figure 2 **The impact of N on the cumulative regret and the computational time, where $T = 500$ $s = 5$, $d = 50$, $K = 5$, and $m = 5$.**



(a) Cumulative regret

(b) Average computational time per trial

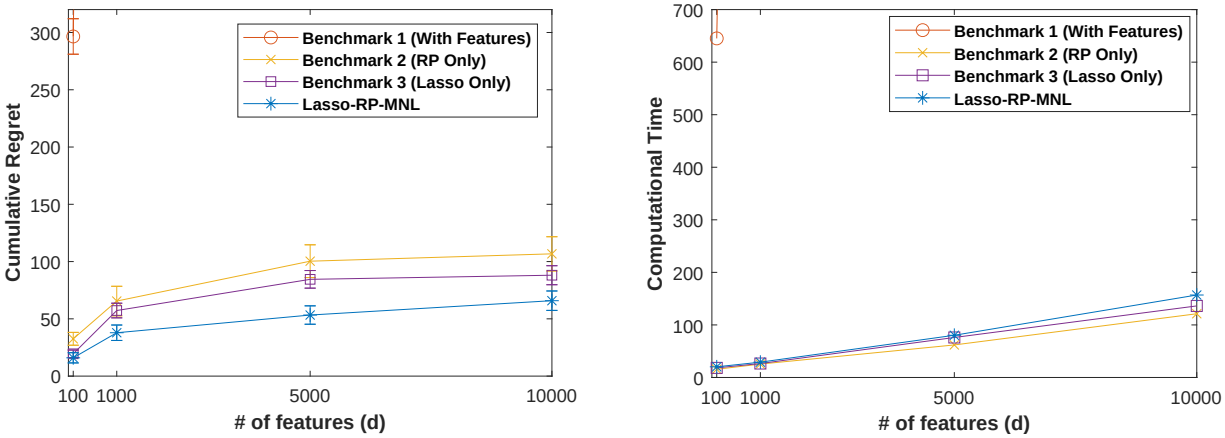
regret, which also tends to grow at the slowest pace. We also observe that algorithms with dimension reduction techniques (i.e., Benchmark 2, Benchmark 3, and Lasso-RP-MNL) are much more computationally efficient compared to the algorithm without such techniques (i.e., Benchmark 1). Specifically, Figure 2(b) presents the influence of N on the average computational time per trial in seconds. Using the Lasso and/or random projection as dimension reduction techniques could significantly scales down the computational time for estimating parameters and calculating users' utilities for all products, and therefore Benchmark 2, Benchmark 3, and Lasso-RP-MNL perform much better than Benchmark 1, regardless of the size of the candidate set.

5.2.2. Impact of the feature dimension d: Next, we examine the influence of the feature dimension by varying $d = \{100, 1000, 5000, 10000\}$ while keeping parameters $s = 5$, $N = 100$, $K = 5$, and $m = 5$ unchanged. Similarly to the second synthetic experiment, we report the cumulative regret and computational time for Lasso-RP-MNL, Benchmark 1, 2, and 3 at $T = 500$ in Figure 3.

We first observe that compared to other algorithms, Benchmark 1's cumulative regret and computational time grow dramatically and rapidly out of the chart, as the feature dimension d exceeds 100 in the experiment[3]. This is as expected. When we expand the feature dimension without using any dimension reduction techniques, Benchmark 1 will require a larger number of available samples to achieve reasonable estimation accuracy. Yet, as we increase the feature dimension while keeping the sample size unchanged at $T = 500$, the regret performance of Benchmark 1 inevitably suffers. Further, note that Benchmark 1 will need to estimate coefficients for all features. Consequently, as the feature dimension increases, its computational time surges.

[3] In our experiments, when the feature dimension d equals $1,000$, the computational time for a single trial of Benchmark 1 will easily exceed one hour. We therefore only plot Benchmark 1's results for $d = 100$ in Figure 3.

Figure 3 **The impact of d on the cumulative regret and the computational time, where $T = 500$ $s = 5$, $N = 100$, $K = 5$, and $m = 5$.**



(a) Cumulative Regret (b) Average Computational Time per Trial

Adopting dimension reduction techniques, Lasso-RP-MNL and the remaining two benchmarks have much lower cumulative regret performance, compared to Benchmark 1. Furthermore, as the feature dimension d increases, the cumulative regret for these three algorithms grows. Similarly to previous experiments, Lasso-RP-MNL continues to have the lowest regret performance, and Benchmark 3 performs better than Benchmark 2. From the computational time's perspective, Benchmark 2, Benchmark 3, and Lasso-RP-MNL all seem to be computationally efficient, while Benchmark 2 maintains a slight advantage over Benchmark 3 and Lasso-RP-MNL.

5.2.3. Impact of the Projection Dimension m: In the final synthetic data experiment, we explore the influences of the projection dimension by varying $m = \{1, 2, 3, 4, 5, 10, 20, 50, 100, 200\}$ and keeping parameters $d = 200$, $s = 5$, $N = 30$, and $K = 5$ unchanged. The cumulative regret and computational time for Lasso-RP-MNL at $T = 500$ are presented in Figure 4. As expected, we first observe that the computational time for Lasso-RP-MNL increases in the projection dimension m.

The cumulative regret of Lasso-RP-MNL, however, exhibits an unimodal property with respect to the projection dimension m in our experiment. In all experiments, we observe that the projection dimension that minimizes the cumulative regret is typically fairly small compared to the feature dimension. In fact, we can show that if the projection dimension m is chosen to be on the order of $\mathcal{O}(\log d)$, then the Gaussian random projection can be confined to a fixed distortion (see the proof of Lemma 2). In Figure 4, for the feature dimension of 200, we merely need to set the projection dimension to be 3 to minimize Lasso-RP-MNL's cumulative regret performance.

5.3. XianYu Assortment Recommendation Experiment

In the last experiment, we consider a high-dimensional assortment recommendation problem faced by XianYu in practice. To ensure that our experiment is manageable for a single PC, we trimmed

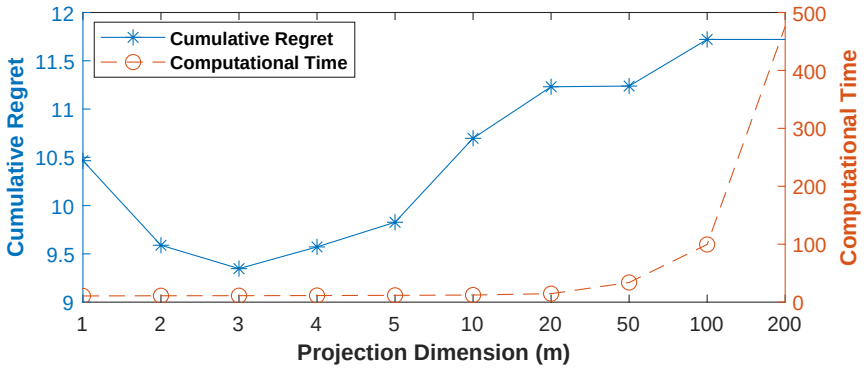


Figure 4 **The impact of m on the cumulative regret and the computational time, where $d = 200$, $s = 5$, $N = 30$, $K = 5$, and $T = 500$.**

the original XianYu dataset, which consists of more than 2 billion features for the user-product pair from 4 million samples (e.g., assortments offered to users and their corresponding responses), to include only 10 thousands features[4] that appear with the highest frequency in these samples.

In practice, for each arriving user, XianYu will first pre-select a personalized product candidate set. Typically, to offer an assortment for an arriving user, there are approximately more than 1 billion available products for XianYu to choose from. However, assessing the user's utilities for all available products to identify the optimal assortment is computationally infeasible under the online setting, where the user expects less than half a second delay. Therefore, depending on the arriving user's specific characteristics (such as searching keywords, current browsing page, demographics, etc.), XianYu will first pre-select $1,000$ highly correlated products from all available products using its efficient recall mechanisms. Then, XianYu's assortment optimization algorithm – the Top-K algorithm – will pick 20 products from the pre-selected $1,000$ products and offer them to the user.

In this experiment, we include XianYu's Top-K algorithm as another benchmark algorithm. The Top-K algorithm is a hybrid online-offline algorithm: The assessment for each user's choice probability and the assortment optimization are performed online, but the parameter estimation and updating are done offline. In the Top-K algorithm, XianYu treats each product in an assortment separately and uses logistic regression to individually assess the user's selection probability for each product. Specifically, for each arriving user, XianYu will first assess this user's selection probabilities, based on the user-product feature pair via logistic regression, for all products in the pre-selected candidate set of $1,000$ products, then calculate the user's expected reward for these products separately, and finally offer the top $K = 20$ products with the highest expected reward as the assortment to this user. In practice, XianYu periodically (typically every couple of hours)

[4] We could extend the experiment to include more features, but doing so would not qualitatively change our results and insights but would considerably increase the computational burden.

updates its estimates for unknown coefficients in the logistic regression via the maximum likelihood estimation. In our experiment, we allow XianYu to update its coefficients at a more frequent rate (i.e., at the same frequency as the Lasso updates in the Lasso-RP-MNL algorithm).

At XianYu, the $1,000$ pre-selected products vary significantly from user to user. Therefore, to simulate such a dynamic environment in the experiment, for each arriving user, we randomly select $1,000$ products from the candidate set of $20,000$ high-frequency products in the dataset and then use different assortment optimization algorithms to select 20 products for the user. Finally, we use the actual asking prices for these products as the reward that XianYu will receive when a user clicks/buys a recommended product. The underlying user's choice model is estimated using the original untrimmed dataset.

In this experiment, we assess the lost revenue under Benchmark 2, Benchmark 3, Lasso-RP-MNL, and the Top-K algorithm by comparing them to an oracle policy. Note that we are unable to include Benchmark 1 in this experiment, because a single trial of Benchmark 1 would take more than 24 hours to finish. It is worth mentioning that as the true coefficient vector is unknown, the "true" oracle policy is impossible to implement in our experiment. Therefore, the oracle policy represents the scenario in which XianYu already has access to all sample data in the original untrimmed dataset to estimate the unknown coefficient vector and identify the optimal assortment accordingly. For each algorithm, we perform 60 trials and report the average loss of revenue for the first $5,000$ users. The computational time for each algorithm per trial is as follows: 259 seconds for the Top-K algorithm[5], 349 seconds for Benchmark 2, 362 seconds for Benchmark 3, and 678 seconds for Lasso-RP-MNL.

Figure 5 plots the cumulative revenue loss (compared to the oracle policy) under Benchmark 2, Benchmark 3, Lasso-RP-MNL, and the Top-K algorithm. In this experiment, the Top-K algorithm and Benchmark 2 seem to fail to converge with $5,000$ users and lead to significant revenue loss. In contrast, Benchmark 3 and Lasso-RP-MNL algorithms have much lower revenue loss and are able to converge with less than 2,000 users. Among all algorithms, the Lasso-RP-MNL algorithm performs the best in terms of cumulative revenue loss.

6. Conclusion

In this paper, we propose a computationally efficient Lasso-RP-MNL algorithm for online assortment optimization problems under high-dimensional settings. This algorithm periodically thresholds the Lasso estimator to identify significant features that strongly influence users' choices

[5] In practice, the Top-K algorithm updates its coefficient estimation in an offline fashion, so the computational time reported for the Top-K algorithm excludes the coefficient estimation time to reflect such a practice.

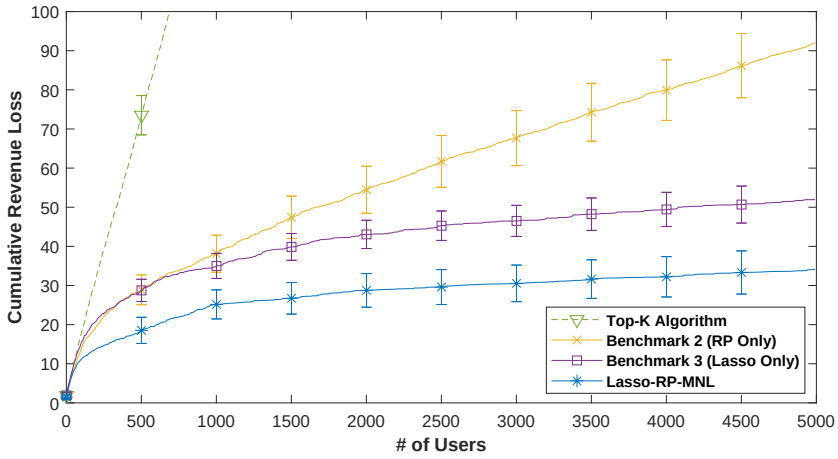


Figure 5 **The cumulative revenue loss for XianYu assortment recommendation experiment where** $T = 5,000$**,** $d = 10,000$ $N = 1,000$**,** $K = 20$**, and** $m = 10$**.**

and adopts random projection to reduce the high-dimensional contextual information to a low-dimensional space. Therefore, the learning and parameter estimation can be performed in a low-dimensional fashion to significantly trim down the computational time while maintaining high accuracy in predicting users' utilities and choices. For each arriving user, the Lasso-RP-MNL algorithm constructs an upper-confidence bound for every product's attraction parameter, based on which the optimistic assortment can be identified through solving a reformulated linear programming problem.

We demonstrate that the Lasso-RP-MNL algorithm's regret upper bound matches the regret lower bound on T and achieves a sub-logarithmic dependence on the feature dimension d. Specifically, we show that the expected cumulative regret of the Lasso-RP-MNL algorithm is upper-bounded by $\tilde{\mathcal{O}}(s\sqrt{\log T} \cdot T^{\frac{2}{3}})$. Furthermore, when the sample size is large, we can further improve the Lasso-RP-MNL algorithm's regret upper bound to $\tilde{\mathcal{O}}(s\sqrt{\log T} \cdot T^{\frac{1}{2}})$. Finally, through synthetic-data-based experiments and a high-dimensional XianYu assortment recommendation experiment, we show that compared to existing state-of-the-art algorithms in the literature and industrial practices, the Lasso-RP-MNL algorithm is computationally efficient and can significantly improve the decision-maker's regret performance.

References

Agrawal S, Avadhanula V, Goyal V, Zeevi A (2017) Thompson sampling for the mnl-bandit. *Conference on Learning Theory*, 76–78 (PMLR).

Agrawal S, Avadhanula V, Goyal V, Zeevi A (2019) Mnl-bandit: A dynamic learning approach to assortment selection. *Operations Research* 67(5):1453–1485.

Anderson SP, De Palma A, Thisse JF (1992) *Discrete choice theory of product differentiation* (MIT press).

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

Bastani H, Bayati M (2020) Online decision making with high-dimensional covariates. *Operations Research* 68(1):276–294.

Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* 2(1):183–202.

Belloni A, Chernozhukov V, et al. (2013) Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19(2):521–547.

Bernstein F, Modaresi S, Sauré D (2018) A dynamic clustering approach to data-driven assortment personalization. *Management Science* 65(5):2095–2115.

Bickel PJ, Ritov Y, Tsybakov AB, et al. (2009) Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics* 37(4):1705–1732.

Blanchet J, Gallego G, Goyal V (2016) A markov chain approximation to choice modeling. *Operations Research* 64(4):886–905.

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

Candes E, Tao T, et al. (2007) The dantzig selector: Statistical estimation when p is much larger than n. *The annals of Statistics* 35(6):2313–2351.

Chen X, Shi C, Wang Y, Zhou Y (2018) Dynamic assortment selection under the nested logit models. *arXiv preprint arXiv:1806.10410* .

Chen X, Wang Y (2018) A note on a tight lower bound for capacitated mnl-bandit assortment selection models. *Operations Research Letters* 46(5):534–537.

Cheung WC, Simchi-Levi D (2017a) Assortment optimization under unknown multinomial logit choice models. *arXiv preprint arXiv:1704.00108* .

Cheung WC, Simchi-Levi D (2017b) Thompson sampling for online personalized assortment optimization problems with multinomial logit choice models. *Available at SSRN 3075658* .

Clarkson KL, Woodruff DP (2017) Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)* 63(6):54.

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

Davis J, Gallego G, Topaloglu H (2013) Assortment planning under the multinomial logit model with totally unimodular constraint structures. *Work in Progress* .

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.

Farias VF, Jagabathula S, Shah D (2013) A nonparametric approach to modeling choice with limited data. *Management science* 59(2):305–322.

Feldman JB, Topaloglu H (2017) Revenue management under the markov chain choice model. *Operations Research* 65(5):1322–1342.

Fern XZ, Brodley CE (2003) Random projection for high dimensional data clustering: A cluster ensemble approach. *Proceedings of the 20th international conference on machine learning (ICML-03)*, 186–193.

Filippi S, Cappe O, Garivier A, Szepesvári C (2010) Parametric bandits: The generalized linear case. *Advances in Neural Information Processing Systems*, 586–594.

Gallego G, Wang R (2014) Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* 62(2):450–461.

Ghashami M, Liberty E, Phillips JM, Woodruff DP (2016) Frequent directions: Simple and deterministic matrix sketching. *SIAM Journal on Computing* 45(5):1762–1792.

Hao B, Lattimore T, Wang M (2020) High-dimensional sparse linear bandits. *arXiv preprint arXiv:2011.04020* .

Jaggi M (2013) Revisiting frank-wolfe: Projection-free sparse convex optimization. *International Conference on Machine Learning*, 427–435 (PMLR).

Johnson WB, Lindenstrauss J (1984) Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics* 26(189-206):1.

Kallus N, Udell M (2016) Dynamic assortment personalization in high dimensions. *arXiv preprint arXiv:1610.05604* .

Kim GS, Paik MC (2019) Doubly-robust lasso bandit. *Advances in Neural Information Processing Systems*, 5869–5879.

Kök AG, Fisher ML (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* 55(6):1001–1021.

Kök AG, Fisher ML, Vaidyanathan R (2015) Assortment planning: Review of literature and industry practice. *Retail supply chain management*, 175–236 (Springer).

Kuzborskij I, Cella L, Cesa-Bianchi N (2018) Efficient linear bandits through matrix sketching. *arXiv preprint arXiv:1809.11033* .

Lattimore T, Szepesvári C (2020) *Bandit algorithms* (Cambridge University Press).

Lee JD, Sun DL, Sun Y, Taylor JE, et al. (2016) Exact post-selection inference, with application to the lasso. *The Annals of Statistics* 44(3):907–927.

Li G, Rusmevichientong P, Topaloglu H (2015) The d-level nested logit model: Assortment and price optimization problems. *Operations Research* 63(2):325–342.

Li L, Lu Y, Zhou D (2017) Provably optimal algorithms for generalized linear contextual bandits. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2071–2080 (JMLR. org).

Li P, Hastie TJ, Church KW (2006) Very sparse random projections. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 287–296 (ACM).

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.

Luo H, Agarwal A, Cesa-Bianchi N, Langford J (2016) Efficient second order online learning by sketching. *Advances in Neural Information Processing Systems*, 902–910.

Mahajan S, Van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Operations Research* 49(3):334–351.

Mahajan S, van Ryzin GJ (1999) Retail inventories and consumer choice. *Quantitative models for supply chain management*, 491–551 (Springer).

Matoušek J (2008) On variants of the johnson–lindenstrauss lemma. *Random Structures & Algorithms* 33(2):142–156.

McFadden D (1980) Econometric models for probabilistic choice among products. *Journal of Business* S13–S29.

McFadden D, et al. (1973) Conditional logit analysis of qualitative choice behavior .

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.

Netessine S, Rudi N (2003) Centralized and competitive inventory models with demand substitution. *Operations research* 51(2):329–335.

Oh Mh, Iyengar G (2019) Thompson sampling for multinomial logit contextual bandits. *NeurIPS*, 3145–3155.

Oh Mh, Iyengar G (2021) Multinomial logit contextual bandits: Provable optimality and practicality. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 9205–9213.

Oh Mh, Iyengar G, Zeevi A (2020) Sparsity-agnostic lasso bandit. *arXiv preprint arXiv:2007.08477* .

Ou M, Li N, Zhu S, Jin R (2018) Multinomial logit bandit with linear utility functions. *arXiv preprint arXiv:1805.02971* .

Pilanci M, Wainwright MJ (2015) Randomized sketches of convex programs with sharp guarantees. *IEEE Transactions on Information Theory* 61(9):5096–5115.

Rusmevichientong P, Shen ZJM, Shmoys DB (2010) Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations research* 58(6):1666–1680.

Rusmevichientong P, Van Roy B, Glynn PW (2006) A nonparametric approach to multiproduct pricing. *Operations Research* 54(1):82–98.

Ryzin Gv, Mahajan S (1999) On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* 45(11):1496–1509.

Sauré D, Zeevi A (2013) Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management* 15(3):387–404.

Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. *Operations Research* 48(1):50–64.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1):267–288.

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

Vershynin R (2010) Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027* .

Wang X, Wei MM, Yao T (2018a) Minimax concave penalized multi-armed bandit model with high-dimensional convariates. *International Conference on Machine Learning*, 5187–5195.

Wang X, Wei MM, Yao T (2018b) Online learning and decision-making under generalized linear model with high-dimensional data. *Available at SSRN 3294832* .

Ye Y, Todd MJ, Mizuno S (1994) An o(nl)-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of operations research* 19(1):53–67.

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

Electronic Companion to "Online Assortment Optimization with High-Dimensional Data"

Parameters	Explanation
t, T	Time indexes.
$\{1,2,3,...,N_t\}$	The candidate set of available products at time t.
A, A_t	Assortment at time t, i.e., $A_t \subseteq \{1,2,3,...,N_t\}$.
K	The cardinality constrain, i.e., $\|A\| \leq K$.
$x_i, x_{i,t}$	The feature vector characterize product i and user arriving at time t.
β^*	The true parameter vector for features.
d, s	Total/significant feature dimension.
$\mathcal{S}^*$	The true index set for significant features, i.e., $\mathcal{S}^* = \{j : \beta_j^* \neq 0\}$.
$r_i, r_{i,t}$	The reward for product i at time t.
$b, x_{\max}, R_{\max}$	Upper-bound parameters for β, x_i, and r_i defined in Assumption A.1.
C_{l1}, C_{l2}, κ	Constants defined in the proof of Theorem 1; $\kappa \in (0,1)$.
$\mathcal{W}$	The index set for all samples collected up to time T.
$\mathcal{W}_R$	The index set for random samples collected up to time T by random decay sampling schedule and $\mathcal{W}_R \subset \mathcal{W}$.
n_T	The size of the index set for random samples $\mathcal{W}_R$, i.e., $n_T = \|\mathcal{W}_R\|$.
λ	A positive regularization parameter for Lasso.
$L(\beta)$	Log-likelihood function: $L(\beta) = \frac{1}{n_T}\sum_{t\in\mathcal{W}_R} \log\left(p_{\beta,A_t}(c_t)\right)$.
C_{lasso}, C	Constants defined in the proofs of Lemma 1: $C_{lasso} = \frac{48\sqrt{2}x_{\max}}{K(K-1)\kappa}$, $C = \frac{1}{2}K(K-1)\left(\kappa/256sx_{\max}^2(3+2\sqrt{2}(1+2x_{\max})\right)^2$.
$\mathcal{G}_0(T,s)$	A function defined in Lemma 1: $\mathcal{G}_0(T,s) = C_{lasso}\cdot s\sqrt{\frac{\log d+\log T}{n_T}}$.
$h(T)$	A thresholding function defined in Theorem 2.
$\mathcal{S}$	The thresholded index set for features selected by thresholding the Lasso estimator: $\mathcal{S} = \{j : \|\hat{\beta}_j\| \geq h(T)\}$.
P, P_0, Q	The random projection matrix, the projection matrix, and the permutation matrix; $P \in \mathbb{R}^{m\times(d-\|\mathcal{S}\|)}$, $P_0 = \begin{pmatrix} I & 0 \\ 0 & P \end{pmatrix} \in \mathbb{R}^{(\|\mathcal{S}\|+m)\times d}$, and $Q \in \mathbb{R}^{d\times d}$.
z, θ	The projected feature vector and the projected coefficient vector; $z := P_0Qx$ and $\theta := P_0Q\beta$.
$\mathcal{G}_1(m,T,\epsilon)$	A function defined in Theorem 3: $\mathcal{G}_1(m,T,\epsilon) = \epsilon\sqrt{s}\cdot(h(T)+\mathcal{G}_0(T,s))$.
τ	A positive constant in Eq. (4).
$\delta, L_3, \lambda_{max}, \rho, \omega_T, \Gamma_T$	Constants defined in Lemma 3.
η	A constant defined in Corollary 1.
C_0, c_1	Constants in the random decay sampling schedule.
T_{lasso}	Lasso step set for the Lasso-RP-MNL algorithm.
$\tilde{C}_{f,1}, \tilde{C}_{f,2}, \tilde{C}_{f,3}$	Constants defined in the proofs of Theorem 5: $\tilde{C}_{f,1} = c\|1-c^{\frac{2}{3}}\|^{-1}\max\{\tilde{C}_{tmp,1}c^{\frac{1}{3}}, \tilde{C}_{tmp,4}c^{\frac{1}{6}}, 32\log^{\frac{1}{2}}T\}$, $\tilde{C}_{f,2} = \tilde{C}_{tmp,2}c\|1-c^{\frac{2}{3}}\|^{-1}$, $\tilde{C}_{f,3} = 4\tilde{C}_{tmp,3}\sqrt{2\Gamma_T}c\|1-c^{\frac{2}{3}}\|^{-1}$, where $\tilde{C}_{tmp,1} = 4C_0^{-\frac{1}{2}}KR_{\max}\eta x_{\max}\cdot C_{lasso}\sqrt{\log d+\log T}$, $\tilde{C}_{tmp,2} = 6\max\left\{\frac{\mu^2KR_{\max}\lambda_{\max}}{32L_3^2(1+\eta^2)}, \frac{\mu KR_{\max}\eta x_{\max}}{2L_3}\right\}\max\{C_0^2, C_0\}$, $\tilde{C}_{tmp,3} = 2\frac{R_{\max}^3K\eta^{3/2}(1+K\eta)\sqrt{2}}{(1+\eta^2)(1+\eta)}\cdot\max\left\{4\log\left(\frac{8(T-1)K^2x_{\max}^2}{\mu}\right), \sum_{t=1}^{T}\frac{1}{t}\right\}$, $\tilde{C}_{tmp,4} = 8\sqrt{2C_0}x_{\max}\sqrt{(2L_3)^{-1}C_0C_{lasso}\mu\sqrt{2(\log d+\log T)C_0^{-1}}}$.

EC.1. Appendix: Main Proofs

EC.1.1. Proof of Theorem 1

First, we state three basic results, which will be used frequently in the proof, as follows:

$$\frac{1}{2}\tanh\left(\frac{x}{2}\right) = \frac{1}{1+\exp(-x)} - \frac{1}{2} = \frac{1}{2}\cdot\frac{1-\exp(-x)}{1+\exp(-x)}$$

$$\left(\tanh\left(\frac{x}{2}\right)\right)' = \frac{1}{\cosh(x)+1}$$

$$\tanh(x) = \tanh(0) + \frac{1}{\cosh(2\xi x)+1}\cdot x \geq \frac{x}{\cosh(2x)+1}, \text{ where } \xi \in [0,1].$$

To prove the regret lower bound, our approach is standard and relies on information theory (e.g., Part IV in Lattimore and Szepesvári 2020). We start proving this lower bound theorem by considering a special single-item assortment problem, where the decision-maker offers an assortment with a single item, the candidate set remains the same for every arriving consumer, and the prices for all products in the candidate set are equal to 1. Now, let's construct two sets as follows:

$$\mathcal{S} \doteq \left\{x \in \mathbb{R}^d \,|x_j \in \{-1,0,1\} \text{ for } j \in \{1,2,...,d-1\}, \|x\|_1 = s-1, \text{ and } x_d = 0\right\},$$

$$\mathcal{H} \doteq \left\{x \in \mathbb{R}^d \,|x_j \in \{-\kappa,\kappa\} \text{ for } j \in \{1,2,...,d-1\}, \text{ and } x_d = 1\right\},$$

where $\kappa \in (0,1)$. Now, we define the candidate set $\mathcal{A}$ as the union of these two sets, i.e., $\mathcal{A} = \mathcal{S} \cup \mathcal{H}$. Next, let define a coefficient vector β as follows:

$$\beta = (\underbrace{\epsilon,...,\epsilon}_{s-1}, \underbrace{0,...,0}_{d-s}, -1),$$

where $\epsilon > 0$. Under the assortment problem parameterized by the coefficient vector β, if we pick the item x_h in $\mathcal{H}$, then the corresponding reward equals to the probability that the item x_h will be chosen:

$$\frac{1}{1+\exp(-\beta^T x_h)} = \frac{1}{1+\exp(-\epsilon\sum_{i=1}^{s-1} x_{h,i}+1)}.$$

Similarly for x_s in $\mathcal{S}$, the reward will be

$$\frac{1}{1+\exp(-\beta^T x_s)} = \frac{1}{1+\exp(-\epsilon\sum_{i=1}^{s-1} x_{s,i})}.$$

When ϵ is small enough, the reward associated with x_h will be smaller than that of x_s, which implies that the x_h will lead to a higher regret. On the other hand, x_h is highly informative, which hints that we need to address the tradeoff between the regret and the information. Clearly, the optimal single-item assortment in $\mathcal{A}$, denoted as x^*, is in $\mathcal{S}$ and given as follows:

$$x^* = \arg\max_{x\in\mathcal{A}} \frac{1}{1+\exp(-\beta^T x)} = \arg\max_{x\in\mathcal{A}} \beta^T x = (\underbrace{1,...,1}_{s-1}, 0, ..., 0).$$

Next, we will construct an alternative assortment problem parameterized by a coefficient vector $\tilde{\beta}$, under which it is hard to distinguish $\tilde{\beta}$ and β and the optimal assortment for β is suboptimal for $\tilde{\beta}$ and vice versa. We denote $\mathbb{P}_\beta$ and $\mathbb{P}_{\tilde{\beta}}$ as measures on the sequence of assortments and their corresponding rewards $(x_1, y_1, ..., x_n, y_n)$, where x_t is the single-item assortment offered at time t and y_t is the corresponding reward,

by β and $\tilde{\beta}$ respectively, and $\mathbb{E}_\beta$ and $\mathbb{E}_{\tilde{\beta}}$ as the corresponding expectation operators. Then, we construct a new set $\mathcal{S}'$ as follows:

$$\mathcal{S}' \doteq \left\{x \in \mathbb{R}^d | x_j = 0 \text{ for } j = \{1, 2, ..., s-1, d\}, x_j \in \{-1, 0, 1\} \text{ for } j \in \{s, s+1, ..., d-1\}, \|x\|_1 = s-1\right\}.$$

Note that $\mathcal{S}'$ is a subset of $\mathcal{S}$, i.e., $\mathcal{S}' \subset \mathcal{S}$. Further, we denote

$$\tilde{x} = \arg\min_{z \in \mathcal{S}'} \mathbb{E}_\beta \left[\sum_{t=1}^{n} \left(x_t^T z\right)^2\right]$$

and construct the alternative coefficient vector $\tilde{\beta}$ as follows:

$$\tilde{\beta} = \beta - 2\epsilon\tilde{x}.$$

Finally, we define an event

$$\mathcal{E} = \left\{\frac{1}{3}\sum_{t=1}^{n} \mathbb{1}(x_t \in \mathcal{S}) \cdot \beta^T x_t \leq n \tanh\left(\frac{(s-1)\epsilon}{6}\right)\right\} \tag{EC.1}$$

and will show that when the event $\mathcal{E}$ occurs, the expected cumulative regret is large under the assortment problem parameterized by β; when it doesn't occur, then the expected cumulative regret is large under the alternative assortment problem parameterized by $\tilde{\beta}$. Specifically, we have the following technical lemma (the proofs for all technical lemmas are deferred to the second section of this Electronic Companion):

LEMMA EC.1. *For small enough ϵ, the expected cumulative regret lower bounds with respect to the event $\mathcal{E}$ are $R_\beta(n) \geq \frac{n}{4}\tanh\left(\frac{(s-1)\epsilon}{6}\right)\mathbb{P}_\beta(\mathcal{E})$ and $R_{\tilde{\beta}}(n) \geq \frac{n}{4}\tanh\left(\frac{(s-1)\epsilon}{6}\right)\mathbb{P}_{\tilde{\beta}}(\mathcal{E}^c)$.*

Combining Lemma EC.1 and Bretagnolle-Huber Inequality, we can directly show that

$$\begin{aligned} R_\beta(n) + R_{\tilde{\beta}}(n) &\geq \frac{n}{4}\tanh\left(\frac{(s-1)\epsilon}{6}\right)\left(\mathbb{P}_\beta(\mathcal{E}) + \mathbb{P}_{\tilde{\beta}}(\mathcal{E}^c)\right) \\ &\geq \frac{n}{8}\tanh\left(\frac{(s-1)\epsilon}{6}\right)\exp\left(-\mathrm{KL}\left(\mathbb{P}_\beta, \mathbb{P}_{\tilde{\beta}}\right)\right), \end{aligned} \tag{EC.2}$$

in which the KL divergence between $\mathbb{P}_\beta$ and $\mathbb{P}_{\tilde{\beta}}$, i.e., $\mathrm{KL}\left(\mathbb{P}_\beta, \mathbb{P}_{\tilde{\beta}}\right)$, can be further upper bounded by the following lemma:

LEMMA EC.2. $KL\left(\mathbb{P}_\beta, \mathbb{P}_{\tilde{\beta}}\right) \leq 12\epsilon^2\left(\frac{n(s-1)^2}{d-s} + \sum_{t=1}^{n}\mathbb{1}(x_t \in \mathcal{H})(s-1)\kappa^2\right)$.

Combining (EC.2) with Lemma EC.2, we can show that when $\sum_{t=1}^{n}\mathbb{1}(x_t \in \mathcal{H}) \leq \frac{1}{12\epsilon^2\kappa^2(s-1)}$, the following inequality holds:

$$\begin{aligned} R_\beta(n) + R_{\tilde{\beta}}(n) &\geq \frac{n}{8}\tanh\left(\frac{(s-1)\epsilon}{6}\right)\exp\left(-\frac{12\epsilon^2(s-1)^2}{d-s}\cdot n\right)\cdot\exp(-1) \\ &= \frac{n}{8e}\tanh\left(\frac{(s-1)\epsilon}{6}\right)\exp\left(-\frac{12\epsilon^2(s-1)^2}{d-s}\cdot n\right). \end{aligned}$$

On the other hand, if $\sum_{t=1}^{n}\mathbb{1}(x_t \in \mathcal{H}) \geq \frac{1}{12\epsilon^2\kappa^2(s-1)}$, then we have the following inequality:

$$R_\beta(n) \geq \mathbb{E}_\beta\left[\sum_{t=1}^{n}\mathbb{1}(x_t \in \mathcal{H})\min_{x \in \mathcal{H}}\left(\frac{1}{1+\exp(-\beta^T x^*)} - \frac{1}{1+\exp(-\beta^T x_t)}\right)\right]$$

$$
\begin{aligned}
&= \mathbb{E}_\beta \left[\sum_{t=1} \mathbb{1}(x_t \in \mathcal{H}) \min_{x \in \mathcal{H}} \left(\frac{1}{1+\exp(-\beta^T x^*)} - \frac{1}{2} + \frac{1}{2} - \frac{1}{1+\exp(-\beta^T x_t)} \right) \right] \\
&= \frac{1}{2} \mathbb{E}_\beta \left[\sum_{t=1} \mathbb{1}(x_t \in \mathcal{H}) \min_{x_t \in \mathcal{H}} \left(\tanh\left(\frac{\beta^T x^*}{2} \right) - \tanh\left(\frac{\beta^T x_t}{2} \right) \right) \right] \\
&= \frac{1}{2} \mathbb{E}_\beta \left[\sum_{t=1} \mathbb{1}(x_t \in \mathcal{H}) \min_{x_t \in \mathcal{H}} \left(\frac{\beta^T (x^* - x_t)}{2(\cosh(\xi_t)+1)} \right) \right], \qquad \text{(EC.3)}
\end{aligned}
$$

where $x_t = (\overbrace{\kappa, ..., \kappa}^{s-1}, 0, ..., 0, 1)$ and ξ_t is between $\epsilon(s-1)\kappa - 1$ and $\epsilon(s-1)$. When ϵ is small enough, we have $\cosh(\xi_t) \leq 2$. Accordingly, we can simplify (EC.3) as follows:

$$
\begin{aligned}
R_\beta(n) &\geq \frac{1}{2} \mathbb{E}_\beta \left[\sum_{t=1} \mathbb{1}(x_t \in \mathcal{H}) \frac{\epsilon(s-1) - \epsilon(s-1)\kappa + 1}{6} \right] \\
&= \frac{1}{12} \mathbb{E}_\beta \left[\sum_{t=1} \mathbb{1}(x_t \in \mathcal{H}) (\epsilon(s-1)(1-\kappa) + 1) \right] \\
&\geq \frac{1}{12} \cdot \frac{\epsilon(s-1)(1-\kappa)+1}{12\epsilon^2\kappa^2(s-1)}.
\end{aligned}
$$

Combining these two cases, we can show that

$$
R_\beta(n) + R_{\tilde{\beta}}(n) \geq \min \left\{ \underbrace{\frac{n}{8e} \tanh\left(\frac{(s-1)\epsilon}{6} \right) \exp\left(-\frac{12\epsilon^2(s-1)^2}{d-s} \cdot n \right)}_{(a)}, \underbrace{\frac{\epsilon(s-1)(1-\kappa)+1}{144\epsilon^2\kappa^2(s-1)}}_{(b)} \right\}. \qquad \text{(EC.4)}
$$

For small sample size (e.g., $n \leq (d-s)^3/(s-1)^2$), we can set $\epsilon = (s-1)^{-2/3} n^{-1/3}$. Next, we consider term (a) and term (b) in (EC.4) separately. For term (a), we can show that

$$
\begin{aligned}
(a) \geq & \frac{n}{8e} \tanh\left(\frac{(s-1)^{1/3} n^{-1/3}}{6} \right) \exp(-12) \\
= & \frac{n}{8e} \left[0 + \frac{(s-1)^{1/3} n^{-1/3}}{6(\cosh(\xi(s-1)^{1/3} n^{-1/3}/3) + 1)} \right] \exp(-12) \\
\geq & \frac{n}{8e} \left[0 + \frac{(s-1)^{1/3} n^{-1/3}}{6(\cosh((s-1)^{1/3} n^{-1/3}/3) + 1)} \right] \exp(-12) \\
= & \frac{\exp(-12)}{48e(\cosh((s-1)^{1/3} n^{-1/3}/3) + 1)} \cdot (s-1)^{1/3} n^{2/3},
\end{aligned}
$$

where $\xi \in [0, 1]$. For term (b), we can show that

$$
\begin{aligned}
(b) = \frac{\epsilon(s-1)(1-\kappa)+1}{144\epsilon^2\kappa^2(s-1)} &= \frac{(s-1)^{1/3} n^{-1/3}(1-\kappa)+1}{144(s-1)^{-1/3} n^{-2/3} \kappa^2} \\
&= \frac{(1-\kappa)(s-1)^{2/3} n^{1/3}}{144\kappa^2} + \frac{(s-1)^{1/3} n^{2/3}}{144\kappa^2} \\
&= \left(\frac{(1-\kappa)(s-1)^{1/3} n^{-1/3} + 1}{144\kappa^2} \right) (s-1)^{1/3} n^{2/3} \\
&\geq \left(\frac{1}{144\kappa^2} \right) (s-1)^{1/3} n^{2/3}.
\end{aligned}
$$

Therefore, combining these two results, we have

$$
\max\left(R_\beta(n), R_{\tilde{\beta}}(n) \right) \geq \frac{1}{2} (R_\beta(n) + R_{\tilde{\beta}}(n)) \geq C_{l1} (s-1)^{1/3} n^{2/3}, \qquad \text{(EC.5)}
$$

where

$$C_{l1} = \min\left\{\frac{\exp(-12)}{96e(\cosh((s-1)^{1/3}n^{-1/3}/3)+1)}, \frac{1}{288\kappa^2}\right\}.$$

On the other hand, for large samples (e.g., $n \geq (d-s)^3/(s-1)^2$), we can choose $\epsilon = \frac{1}{s-1}\sqrt{\frac{d}{n}}$ so that

$$\begin{aligned}(a) &\geq \frac{n}{8e}\tanh\left(\frac{d^{1/2}n^{-1/2}}{6}\right)\exp(-12)\\ &\geq \frac{1}{48e(\cosh(d^{1/2}n^{-1/2}/3)+1)}\cdot d^{1/2}n^{1/2},\end{aligned}$$

and

$$(b) \geq \frac{d^{1/2}n^{1/2}(1-\kappa)+n}{144d(s-1)^{-1}\kappa^2} = \frac{(1-\kappa)+d^{-1/2}n^{1/2}}{144d(s-1)^{-1}\kappa^2}\cdot d^{1/2}n^{1/2},$$

combining which two, we can show that

$$\max\left(R_\beta(n), R_{\tilde{\beta}}(n)\right) \geq C_{l2}d^{1/2}n^{1/2}, \tag{EC.6}$$

where

$$C_{l2} = \min\left\{\frac{1}{96e(\cosh(d^{1/2}n^{-1/2}/3)+1)}, \frac{(1-\kappa)+d^{-1/2}n^{1/2}}{288d(s-1)^{-1}\kappa^2}\right\}$$

Finally, combining (EC.5) and (EC.6), we have

$$\max\left(R_\beta(n), R_{\tilde{\beta}}(n)\right) \geq \min\left\{C_{l1}(s-1)^{1/3}n^{2/3}, C_{l2}d^{1/2}n^{1/2}\right\}.$$

EC.1.2. Proof of Lemma 1

Since $\hat{\beta}$ is the optimal solution for the Lasso problem, we have

$$\nabla L(\hat{\boldsymbol{\beta}}) + \lambda\partial\|\hat{\boldsymbol{\beta}}\|_1 = 0,$$

where $\partial(\cdot)$ denotes the subgradient. As the negative log-likelihood function L is twice differentiable, there exists a $\boldsymbol{\xi}$ such that

$$\begin{aligned}&\nabla L(\boldsymbol{\beta}^*) - \nabla^2 L(\boldsymbol{\xi})(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) = \nabla L(\hat{\boldsymbol{\beta}})\\ \Rightarrow\ &\nabla L(\boldsymbol{\beta}^*) - \nabla^2 L(\boldsymbol{\xi})(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) + \lambda\partial\|\hat{\boldsymbol{\beta}}\|_1 = 0\\ &\Rightarrow (\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})^T\nabla^2 L(\boldsymbol{\xi})(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) = (\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})^T\left(\nabla L(\boldsymbol{\beta}^*) + \lambda\partial\|\hat{\boldsymbol{\beta}}\|_1\right)\\ &\Rightarrow (\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})^T\nabla^2 L(\boldsymbol{\xi})(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}) \leq \|\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}\|_1\left(\|\nabla L(\boldsymbol{\beta}^*)\|_\infty + \lambda\right).\end{aligned} \tag{EC.7}$$

Next, we will build the lower bound for the left-hand-side of (EC.7) by the following technical lemma.

LEMMA EC.3. *Denote n_T as the random sample size up to time T. If Assumption A.2 holds, then the following inequality holds with probability $1-\exp(-Cn_T)$:*

$$\boldsymbol{u}^T\nabla^2 L(\boldsymbol{\xi}|\boldsymbol{x}, \mathcal{A})\boldsymbol{u} \geq \frac{K(K-1)\kappa}{4s}\|\boldsymbol{u}_S\|_1^2, \tag{EC.8}$$

where $\boldsymbol{u}$ *satisfy* $\|\boldsymbol{u}_{S^c}\|_1 \leq 3\|\boldsymbol{u}_S\|_1$ *and* $C = \frac{1}{2}K(K-1)\left(\kappa/256sx_{\max}^2(3+2\sqrt{2}(1+2x_{\max}))\right)^2$. *Moreover, when* $n_T > \log T/C$, *we have*

$$\mathbb{P}\left(\boldsymbol{u}^T\nabla^2(\boldsymbol{\xi}|\boldsymbol{x},\mathcal{A})\boldsymbol{u} \geq \frac{K(K-1)\kappa}{4s}\|\boldsymbol{u}_S\|_1^2\right) \geq 1-\frac{1}{T} \tag{EC.9}$$

Note that based on Lemma EC.3, to show that the left-hand-side of (EC.7) is lower bounded by $K(K-1)\kappa/(4s)\cdot\|\boldsymbol{\beta}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1^2$, we only need to prove $\|\boldsymbol{\beta}^*_{(\mathcal{S}^*)^c}-\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}\|_1 \leq 3\|\boldsymbol{\beta}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1$. As $\hat{\boldsymbol{\beta}}$ is the optimal solution for the Lasso problem, we have

$$\begin{aligned}
&L(\hat{\boldsymbol{\beta}})+\lambda\|\hat{\boldsymbol{\beta}}\|_1 \leq L(\boldsymbol{\beta}^*)+\lambda\|\boldsymbol{\beta}^*\|_1\\
\Rightarrow\ &L(\hat{\boldsymbol{\beta}})-L(\boldsymbol{\beta}^*) \leq \lambda(\|\boldsymbol{\beta}^*\|_1-\|\hat{\boldsymbol{\beta}}\|_1)\\
\Rightarrow\ &\nabla L(\boldsymbol{\beta}^*)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*) \leq \lambda\left(\|\boldsymbol{\beta}^*\|_1-\|\hat{\boldsymbol{\beta}}\|_1\right)\\
\Rightarrow\ &-\|\nabla L(\boldsymbol{\beta}^*)\|_\infty\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_1 \leq \lambda\left(\|\boldsymbol{\beta}^*\|_1-\|\hat{\boldsymbol{\beta}}\|_1\right)\\
\Rightarrow\ &-\|\nabla L(\boldsymbol{\beta}^*)\|_\infty(\|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}-\boldsymbol{\beta}^*_{(\mathcal{S}^*)^c}\|_1+\|\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}-\boldsymbol{\beta}^*_{\mathcal{S}^*}\|_1) \leq \lambda(\|\hat{\boldsymbol{\beta}}^*_{\mathcal{S}^*}\|_1+0-\|\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1-\|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}\|_1)\\
\Rightarrow\ &-\|\nabla L(\boldsymbol{\beta}^*)\|_\infty(\|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}-\boldsymbol{\beta}^*_{(\mathcal{S}^*)^c}\|_1+\|\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}-\boldsymbol{\beta}^*_{\mathcal{S}^*}\|_1) \leq \lambda(\|\hat{\boldsymbol{\beta}}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1-\|\hat{\boldsymbol{\beta}}^*_{(\mathcal{S}^*)^c}-\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}\|_1)\\
\Rightarrow\ &(\lambda-\|\nabla L(\boldsymbol{\beta}^*)\|_\infty)\,\|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}-\boldsymbol{\beta}^*_{(\mathcal{S}^*)^c}\|_1 \leq (\lambda+\|\nabla L(\boldsymbol{\beta}^*)\|_\infty)\,\|\hat{\boldsymbol{\beta}}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1.
\end{aligned} \tag{EC.10}$$

Therefore, if we have $\|\nabla L(\boldsymbol{\beta}^*)\|_\infty \leq \frac{1}{2}\lambda$, then (EC.10) directly implies $\|\boldsymbol{\beta}^*_{(\mathcal{S}^*)^c}-\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}\|_1 \leq 3\|\boldsymbol{\beta}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1$. Such a condition can be shown in the following technical lemma.

LEMMA EC.4. *Let* n *denote the random sample size. If Assumption A.1 holds, then for any* $T>0$, *we have*

$$\mathbb{P}\left(\|\nabla L(\boldsymbol{\beta}^*)\|_\infty \geq \sqrt{\frac{2x_{\max}^2(\log d+\log T)}{n}}\right) \leq \frac{2}{T}. \tag{EC.11}$$

If we set $\lambda = 2\sqrt{\frac{2x_{\max}^2(\log d+\log T)}{n}}$, then Lemma EC.4 suggests that with probability $1-O(T^{-1})$, we have

$$\|\nabla L(\boldsymbol{\beta}^*)\|_\infty \leq \frac{1}{2}\lambda. \tag{EC.12}$$

Combining (EC.10) and (EC.12), we have

$$\|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c}-\boldsymbol{\beta}^*_{(\mathcal{S}^*)^c}\|_1 \leq 3\|\hat{\boldsymbol{\beta}}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1. \tag{EC.13}$$

Hence, We can use Lemma EC.3 to show that with high probability, the following inequality holds:

$$(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}})^T\nabla^2 L(\boldsymbol{\xi})(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}) \geq \frac{K(K-1)\kappa}{4s}\|\boldsymbol{\beta}^*_{\mathcal{S}^*}-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1^2. \tag{EC.14}$$

Using (EC.13) and (EC.14) we can show

$$(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}})^T\nabla^2 L(\xi)(\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}) \geq \frac{K(K-1)\kappa}{16s}\|\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}\|_1^2. \tag{EC.15}$$

Moreover, combine (EC.15), (EC.12) and (EC.7) and we reach

$$\begin{aligned}
&\frac{K(K-1)\kappa}{16s}\|\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}\|_1^2 \leq \frac{3}{2}\lambda\|\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}\|_1\\
&\Rightarrow \|\boldsymbol{\beta}^*-\hat{\boldsymbol{\beta}}\|_1 \leq \frac{24s}{K(K-1)\kappa}\lambda.
\end{aligned} \tag{EC.16}$$

The remaining part of this Lemma follows directly by setting $C_{lasso} = \frac{48\sqrt{2}x_{\max}}{K(K-1)\kappa}$ and $n \geq \log T/C = \mathcal{O}(s^2 \log T)$.

EC.1.3. Proof of Theorem 2

When event $\mathcal{E}_{lasso}(T)$ holds, we have

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 \leq \mathcal{G}_0(T,s),$$

which implies that $|\hat{\beta}_j| \leq |\beta_j^*| + \mathcal{G}_0(T,s)$ and $|\hat{\beta}_j| \geq |\beta_j^*| - \mathcal{G}_0(T,s)$ for any j.

Combining $|\hat{\beta}_j| \leq |\beta_j^*| + \mathcal{G}_0(T,s)$ with the definition of the index set $\mathcal{S} = \{j : |\hat{\beta}_j| \geq h(T)\}$, we can show that

$$j \notin \mathcal{S} \Rightarrow |\hat{\beta}_j| < h(T) \Rightarrow |\beta_j^*| < h(T) + \mathcal{G}_0(T,s). \tag{EC.17}$$

Next, we consider the upper bound for $|\mathcal{S}|$:

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 \geq \|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c} - \mathbf{0}\|_1 \geq \|\hat{\boldsymbol{\beta}}_{(\mathcal{S}^*)^c \cap \mathcal{S}}\|_1 \geq |\mathcal{S} - \mathcal{S}^*| \min_{j \in \mathcal{S}} |\hat{\beta}_j| \geq |\mathcal{S} - \mathcal{S}^*| h(T). \tag{EC.18}$$

Combining (EC.18) with $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 \leq \mathcal{G}_0(T,s)$, we will have

$$|\mathcal{S} - \mathcal{S}^*| h(T) \leq \mathcal{G}_0(T) \Rightarrow |\mathcal{S}| \leq \frac{\mathcal{G}_0(T)}{h(T)} + |\mathcal{S}^*| \leq \frac{\mathcal{G}_0(T)}{h(T)} + s. \tag{EC.19}$$

EC.1.4. Proof of Lemma 2

According to the choice of $\boldsymbol{P}$, for and $\boldsymbol{x} \in \mathbb{R}^d$ and j-th row of matrix $\boldsymbol{P}$ with $j \leq m$, we have

$$\mathbb{E}\left[\boldsymbol{P}_j \boldsymbol{x}\right] = 0 \text{ and } \mathbb{E}\left[(\boldsymbol{P}_j \boldsymbol{x})^2\right] = \frac{1}{m}\|\boldsymbol{x}\|_2^2, \tag{EC.20}$$

Due to the gaussian choice, above inequality implies $\tilde{z}_j = \sqrt{m}\boldsymbol{P}_j \boldsymbol{x}/\|\boldsymbol{x}\|_2$ is distributed as $\mathcal{N}(0,1)$, and $\tilde{z}_j$ are independent among $j = 1, 2, ..., m$. Then,

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{x}\|_2^2 > (1+\epsilon)\|\boldsymbol{x}\|_2^2\right) = \mathbb{P}\left(\sum_{j=1}^{m} \tilde{z}_j^2 > (1+\epsilon)m\right) = \mathbb{P}\left(\chi_m^2 < (1+\epsilon)m\right), \tag{EC.21}$$

where χ_m^2 is the chi-squared distribution with m degrees of freedom. Via standard tail-inequality for chi-squared distribution, for $\epsilon \in (0, 1/2)$,

$$\mathbb{p}\left(\chi_m^2 > (1+\epsilon)m\right) \leq \exp\left(-\frac{m}{8}\epsilon^2\right). \tag{EC.22}$$

Combining (EC.21) and (EC.22), we have

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{x}\|_2^2 > (1+\epsilon)\|\boldsymbol{x}\|_2^2\right) \leq \exp\left(-\frac{m}{8}\epsilon^2\right). \tag{EC.23}$$

Similarly, one can verify

$$\mathbb{P}\left(\|\boldsymbol{P}\boldsymbol{x}\|_2^2 < (1-\epsilon)\|\boldsymbol{x}\|_2^2\right) \leq \exp\left(-\frac{m}{8}\epsilon^2\right). \tag{EC.24}$$

The desirable result is reached.

EC.1.5. Proof of Theorem 3

As Q is the permutation matrix, we have $Q^TQ = I$. Hence, we can show that:

$$\|(\boldsymbol{I}-\boldsymbol{\Sigma})\boldsymbol{\beta}^*\|_2 \le \max_{\boldsymbol{x}\in\mathcal{S}^{d-1}} \left|\boldsymbol{x}^T(\boldsymbol{I}-\boldsymbol{\Sigma})\boldsymbol{\beta}^*\right|, \tag{EC.25}$$

where $\mathcal{S}^{d-1}$ is the unit d dimension sphere.

$$\begin{aligned}
\left|\boldsymbol{x}^T(\boldsymbol{I}-\boldsymbol{\Sigma})\boldsymbol{\beta}^*\right| &= \left|\boldsymbol{x}^T\left(\boldsymbol{I}-\boldsymbol{Q}^T\boldsymbol{P}_0^T\boldsymbol{P}_0\boldsymbol{Q}\right)\boldsymbol{\beta}^*\right| \\
&= \left|(\boldsymbol{Q}\boldsymbol{x})^T\left(\boldsymbol{I}-\boldsymbol{P}_0^T\boldsymbol{P}_0\right)\boldsymbol{Q}\boldsymbol{\beta}^*\right| \\
&= \left|\left(\boldsymbol{x}_S^T\ \boldsymbol{x}_{S^c}^T\right)\left[\begin{pmatrix}\boldsymbol{I} & \\ & \boldsymbol{I}\end{pmatrix}-\begin{pmatrix}\boldsymbol{I} & \\ & \boldsymbol{P}^T\boldsymbol{P}\end{pmatrix}\right]\begin{pmatrix}\boldsymbol{\beta}_S^* \\ \boldsymbol{\beta}_{S^c}^*\end{pmatrix}\right| \qquad \text{(EC.26)} \\
&= \left|\boldsymbol{x}_{S^c}^T(\boldsymbol{I}-\boldsymbol{P}^T\boldsymbol{P})\boldsymbol{\beta}_{S^c}^*\right| \\
&= \left|\boldsymbol{x}_{S^c}^T\boldsymbol{\beta}_{S^c}^* - (\boldsymbol{P}\boldsymbol{x}_{S^c})^T\boldsymbol{P}\boldsymbol{\beta}_{S^c}^*\right|, \qquad \text{(EC.27)}
\end{aligned}$$

where (EC.26) comes from the definition and construction of the permutation matrix Q and the projection matrix P_0.

We then apply the technical Lemma EC.11 to show that with probability $1-4\exp\left(-\frac{m}{8}\epsilon^2\right)$, we have

$$\left|\boldsymbol{x}_{S^c}^T\boldsymbol{\beta}_{S^c}^* - (\boldsymbol{P}\boldsymbol{x}_{S^c})^T\boldsymbol{P}\boldsymbol{\beta}_{S^c}^*\right| \le \epsilon\|\boldsymbol{x}_{S^c}^T\|_2\|\boldsymbol{\beta}_{S^c}^*\|_2 \le \epsilon\|\boldsymbol{\beta}_{S^c}^*\|_2, \tag{EC.28}$$

where last inequality uses $\boldsymbol{x}\in\mathcal{S}^{d-1}$. Thus via EC.25, EC.27 and EC.28, we have

$$\|(\boldsymbol{I}-\boldsymbol{\Sigma})\boldsymbol{\beta}^*\|_2 \le \epsilon\|\boldsymbol{\beta}_{S^c}^*\|_2 \tag{EC.29}$$

holds with probability $1-4\exp\left(-\frac{m}{8}\epsilon^2\right)$. Finally, by the fact that $\|\boldsymbol{\beta}_{S^c}^*\|_2 \le \sqrt{s}\|\boldsymbol{\beta}_{S^c}^*\|_\infty \le \sqrt{s}(h(T)+\mathcal{G}_0(T,s))$, we have

$$\mathbb{P}\left(\|(\boldsymbol{I}-\boldsymbol{\Sigma})\boldsymbol{\beta}^*\|_2 \le \epsilon\sqrt{s}\cdot(h(T)+\mathcal{G}_0(T,s))\right) \le 1-4\exp\left(-\frac{m}{8}\epsilon^2\right). \tag{EC.30}$$

EC.1.6. Proof of Lemma 3

To simplify the notations in this proof, we will ignore the $\mathcal{A}$ subscript in probability term $p_{\cdot,\mathcal{A}}(\cdot)$ and re-define $\mathcal{A}$ as the assortment including both the original assortment $\mathcal{A}$ and the no-purchase option (i.e., $\mathcal{A} := \mathcal{A}\cup\{0\}$), as long as doing so does not cause any misinterpretation. Accordingly, the decision-maker's expected reward for a coefficient vector θ under this re-defined assortment $\mathcal{A}$ can be simplified into $R_{\mathcal{A}}(\theta) = \frac{\sum_{i\in\mathcal{A}} r_i\exp((P_0Qx_i)^T\theta)}{\sum_{i\in\mathcal{A}}\exp((P_0Qx_i)^T\theta)}$.

Using Taylor expansion, we can show that there exists a ξ such that

$$\begin{aligned}
|R_{\mathcal{A}}(\hat\theta) - R_{\mathcal{A}}(P_0Q\beta^*)| &= \left\|\nabla R_{\mathcal{A}}(\hat\theta)^T(\hat\theta-P_0Q\beta^*) - \frac{1}{2}(\hat\theta-P_0Q\beta^*)^T\nabla^2R_{\mathcal{A}}(\xi)(\hat\theta-P_0Q\beta^*)\right\| \\
&\le \|\nabla R_{\mathcal{A}}(\hat\theta)^T(\hat\theta-P_0Q\beta^*)\| + \left\|\frac{1}{2}(\hat\theta-P_0Q\beta^*)^T\nabla^2R_{\mathcal{A}}(\xi)(\hat\theta-P_0Q\beta^*)\right\| \\
&\le \sqrt{\|\nabla R_{\mathcal{A}}(\hat\theta)^T(\hat\theta-P_0Q\beta^*)\|^2} + \frac{1}{2}\left\|\nabla^2\mathcal{R}_{\mathcal{A}}(\xi)\right\|_{op}\|\hat\theta-P_0Q\beta^*\|^2 \\
&= \overbrace{\sqrt{(\hat\theta-P_0Q\beta^*)^T\nabla R_{\mathcal{A}}(\hat\theta)\nabla R_{\mathcal{A}}(\hat\theta)^T(\hat\theta-P_0Q\beta^*)}}^{\text{ⓐ}} + \overbrace{\frac{1}{2}\left\|\nabla^2\mathcal{R}_{\mathcal{A}}(\xi)\right\|_{op}\|\hat\theta-P_0Q\beta^*\|^2}^{\text{ⓑ}}.
\end{aligned}$$

Now, we will separately build upper bounds for ⓐ and ⓑ.

Analysis for ⓐ: From the definition of $R_{\mathcal{A}}(\hat{\theta})$ and using $p_{Q^T P_0^T \theta}(i)$ to represent the probability of selecting product i from the assortment $\mathcal{A}$ (i.e., $p_{Q^T P_0^T \theta}(i) = 1/(\sum_{i\in\mathcal{A}} \exp((P_0 Q x_i)^T \theta))$), we can show that

$$\begin{aligned}
\nabla R_{\mathcal{A}}(\hat{\theta}) &= \frac{\sum_{i\in\mathcal{A}} r_i \exp(x_i^T Q^T P_0^T \hat{\theta}) P_0 Q x_i \cdot (\sum_{j\in\mathcal{A}} \exp(x_j^T Q^T P_0^T \hat{\theta}))}{(\sum_{j\in\mathcal{A}} \exp(x_j^T Q^T P_0^T \hat{\theta}))^2} \\
&\quad - \frac{(\sum_{i\in\mathcal{A}} r_i \exp(x_i^T Q^T P_0^T \hat{\theta})) \cdot (\sum_{j\in\mathcal{A}} \exp(x_j^T Q^T P_0^T \hat{\theta}) P_0 Q x_j)}{(\sum_{j\in\mathcal{A}} \exp(x_j^T Q^T P_0^T \hat{\theta}))^2} \\
&= \sum_{i\in\mathcal{A}} P_0 Q x_i r_i p_{Q^T P_0^T \hat{\theta}}(i) - \sum_{i\in\mathcal{A}} r_i p_{Q^T P_0^T \hat{\theta}}(i) \cdot \sum_{j\in\mathcal{A}} P_0 Q x_j p_{Q^T P_0^T \hat{\theta}}(j) \\
&= \sum_{i\in\mathcal{A}} r_i p_{Q^T P_0^T \hat{\theta}}(i) \left[P_0 Q x_i - \sum_{j\in\mathcal{A}} P_0 Q x_j p_{Q^T P_0^T \hat{\theta}}(j) \right] \\
&= \sum_{i\in\mathcal{A}} p_{Q^T P_0^T \hat{\theta}}(i) \left[r_i - \sum_{j\in\mathcal{A}} p_{Q^T P_0^T \hat{\theta}}(j) r_i \right] \left[P_0 Q x_i - \sum_{j\in\mathcal{A}} p_{Q^T P_0^T \hat{\theta}}(j) P_0 Q x_j \right]. \qquad \text{(EC.31)}
\end{aligned}$$

We write (EC.31) in short-hand notation as follows:

$$\text{(EC.31)} := \tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right) \right], \qquad \text{(EC.32)}$$

where $\tilde{\mathbb{E}}$ represents the expectation w.r.t. probability distribution $\{p_{Q^T P_0^T \hat{\theta}}(i)\}$ and $z_i = P_0 Q x_i$ for all $i \in \mathcal{A}$. Then we have

$$\begin{aligned}
\nabla R_S(\hat{\theta}) \nabla R_S(\hat{\theta})^T &= \tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right) \right] \cdot \tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right) \right]^T \\
&= \tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right) \cdot \overbrace{\tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right) \right]^T}^{*} \right] \qquad \text{(EC.33)} \\
&\preceq \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right) \cdot \left(r - \tilde{\mathbb{E}}[r] \right) \left(z - \tilde{\mathbb{E}}[z] \right)^T \right] \right] \qquad \text{(EC.34)} \\
&= \tilde{\mathbb{E}} \left[\left(r - \tilde{\mathbb{E}}[r] \right)^2 \left(z - \tilde{\mathbb{E}}[z] \right) \left(z - \tilde{\mathbb{E}}[z] \right)^T \right] \\
&\preceq R_{\max}^2 \tilde{\mathbb{E}} \left[\left(z - \tilde{\mathbb{E}}[z] \right) \left(z - \tilde{\mathbb{E}}[z] \right)^T \right],
\end{aligned}$$

where in (EC.34) we uses the Jensen's inequality (e.g., equation 2.2.2 in Tropp et al. 2015) on the * term in (EC.33).

To simplify notation, we can show that $\nabla^2 f_{\mathcal{A}}(\hat{\theta}) = \tilde{\mathbb{E}} \left[\left(z - \tilde{\mathbb{E}}[z] \right) \left(z - \tilde{\mathbb{E}}[z] \right)^T \right]$ (see Lemma EC.10). Then, we have

$$\sqrt{(\hat{\theta} - P_0 Q \beta^*)^T \nabla R_{\mathcal{A}}(\hat{\theta}) \nabla R_{\mathcal{A}}(\hat{\theta})^T (\hat{\theta} - P_0 Q \beta^*)} \le R_{\max} \sqrt{(\hat{\theta} - P_0 Q \beta^*)^T \nabla^2 f_{\mathcal{A}}(\hat{\theta}) (\hat{\theta} - P_0 Q \beta^*)}. \qquad \text{(EC.35)}$$

Analysis for ⓑ: As we assume $\|\nabla^2 \mathcal{R}_{\mathcal{A}}(\xi)\|_{op} \le \lambda_{\max}$, we can bound ⓑ as follows:

$$\frac{1}{2} \left\| \nabla^2 \mathcal{R}_{\mathcal{A}}(\xi) \right\|_{op} \|\hat{\theta} - P_0 Q \beta^*\|^2 \le \frac{1}{2} \lambda_{\max} \delta^2.$$

Combining the upper bounds for the part ⓐ and part ⓑ, we have

$$|R_{\mathcal{A}}(\hat{\theta}) - R_{\mathcal{A}}(P_0Q\beta^*)| \le R_{\max}\sqrt{(\hat{\theta} - P_0Q\beta^*)^T\nabla^2 f_{\mathcal{A}}(\hat{\theta})(\hat{\theta} - P_0Q\beta^*)} + \frac{1}{2}\lambda_{\max}\delta^2$$
$$\Rightarrow |R_{\mathcal{A}}(\hat{\theta}) - R_{\mathcal{A}}(P_0Q\beta^*)| - \frac{1}{2}\lambda_{\max}\delta^2 \le R_{\max}\sqrt{(\hat{\theta} - P_0Q\beta^*)^T\nabla^2 f_{\mathcal{A}}(\hat{\theta})(\hat{\theta} - P_0Q\beta^*)} \tag{EC.36}$$

Denote $H^* = \left(\sum_i^T \nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)$. We show that H^* is positive definite with high probability in Lemma EC.8. Therefore, (EC.36) implies

$$\begin{aligned}
&|R_{\mathcal{A}}(\hat{\theta}) - R_{\mathcal{A}}(P_0Q\beta^*)| - \frac{1}{2}\lambda_{\max}\delta^2 \\
&\le R_{\max}\sqrt{(\hat{\theta} - P_0Q\beta^*)^T(H^*)^{1/2}(H^*)^{-1/2}\nabla^2 f_{\mathcal{A}}(\hat{\theta})(H^*)^{-1/2}(H^*)^{1/2}(\hat{\theta} - P_0Q\beta^*)} \\
&\le R_{\max}\left\|(\hat{\theta} - P_0Q\beta^*)^T(H^*)^{1/2}\right\|\sqrt{\left\|(H^*)^{-1/2}\nabla^2 f_{\mathcal{A}}(\hat{\theta})(H^*)^{-1/2}\right\|_{op}} \\
&= R_{\max}\sqrt{(\hat{\theta} - P_0Q\beta^*)^T H^*(\hat{\theta} - P_0Q\beta^*)}\cdot\sqrt{\left\|(H^*)^{-1/2}\nabla^2 f_{\mathcal{A}}(\hat{\theta})(H^*)^{-1/2}\right\|_{op}}.
\end{aligned} \tag{EC.37}$$

Then, we use the following Lemma:

LEMMA EC.5. *Let $\delta = \|\theta - P_0Q\beta^*\|$. Under Assumptions A.1 and A.3, events $\mathcal{E}_2(m,T,\epsilon)$ and $\mathcal{E}_{rp}(m,d,1/2)$, if $\delta \le \min\{\frac{3}{4}\frac{n_T\mu}{TL_3}, \frac{\rho}{8Kx_{\max}}\}$ and $\mathcal{G}_1(m,T,\epsilon) \le \frac{\rho}{8Kx_{\max}}$, then the following inequality holds for $T \ge 2$ with probability $1 - 4\exp(-\frac{m}{8}\epsilon^2) - \mathcal{O}(1/T)$:*

$$(\theta - P_0Q\beta^*)^T\left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*) \le 16x_{\max}^2 T\mathcal{G}_1(m,T,\epsilon)\delta + 128(2(|\mathcal{S}|+m)+1)\log(T) + 8\Gamma_T,$$

where $\Gamma_T := \max\left\{0, \sum_t \hat{f}_t(\hat{\theta})\right\}$.

Combining Lemma EC.5, (EC.37), $H^* = \left(\sum_i^T \nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)$, $\sum_{t=1}^T \hat{f}_t(\hat{\theta}) = TL_z(\hat{\theta})$ and the definition of ω_t, we can show that

$$\begin{aligned}
&|R_{\mathcal{A}}(\hat{\theta}) - R_{\mathcal{A}}(P_0Q\beta^*)| - \frac{1}{2}\lambda_{\max}\delta^2 \\
&\le R_{\max}\omega_t\sqrt{\left\|\left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\nabla^2 f_{\mathcal{A}}(\hat{\theta})\left(\sum_{i=1}^T \nabla^2 f(P_0Q\beta^*)\right)^{-1/2}\right\|_{op}}.
\end{aligned}$$

Furthermore, for all $i \le T$, we can show that

$$\begin{aligned}
&\left\|\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*) - \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right\|_{op} \le L_3\|P_0Q\beta^* - \hat{\theta}\| = L_3\delta \\
\Rightarrow &\nabla^2 f_{\mathcal{A}_i}(\hat{\theta}) - L_3\delta I \preceq \nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*).
\end{aligned}$$

The lower bound for $\nabla^2 f(\hat{\theta})$ can be established by using Lemma EC.8, which shows that the following inequality holds with probability $1 - \mathcal{O}(1/T)$:

$$\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta}) \succeq \frac{1}{2}\mu n_T I.$$

When $\delta \le \frac{\mu n_T}{4L_3 T}$, we have $\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta}) - L_3\delta I \succeq \frac{1}{2}\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})$, which leads to

$$\sqrt{2}\left(\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \succeq \left(\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(P_0 Q\beta^*)\right)^{-1/2}.$$

Therefore, with probability $1-4\exp\left(-\frac{m}{8}\epsilon^2\right) - O(T^{-1})$ the following result holds:

$$|R_{\mathcal{A}}(\hat{\theta}) - R_{\mathcal{A}}(P_0 Q\beta^*)| \le \sqrt{2} R_{\max}\omega_T \sqrt{\left\|\left(\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \nabla^2 f_{\mathcal{A}}(\hat{\theta}) \left(\sum_{i=1}^{T} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}} + \frac{1}{2}\lambda_{\max}\delta^2.$$

EC.1.7. Proof of Corollary 1

If we choose $\mathcal{A}$ to be the assortment with a single item with vector x and $r=1$, we then have

$$R_{\mathcal{A}}(\theta) = \frac{\exp(x^T Q^T P_0^T \theta)}{1+\exp(x^T Q^T P_0^T \theta)} \tag{EC.38}$$

Consider the function $\phi(x) = x/(1+x)$, which monotonically increases in x for $x \in (0, x_0)$. Therefore, we have

$$x_1 - x_2 \le \frac{|\phi(x_1)-\phi(x_2)|}{\phi'(x_0)} = \frac{|\phi(x_1)-\phi(x_2)|}{(1+x_0)^2} \le \frac{|\phi(x_1)-\phi(x_2)|}{1+x_0^2}, \tag{EC.39}$$

where last inequality uses $(1+x)^2 \ge 1+x^2$ for $x \ge 0$. Thus, applying Lemma 2, we will have

$$\begin{aligned}
&\exp(x^T\Sigma\beta^*) - \exp(x^T Q^T P_0^T\theta) \le \frac{|R_s(P_0Q\beta^*) - R_s(\theta)|}{1+\exp(2x_{\max}b)} \\
&\le \frac{\sqrt{2}\omega_t}{1+\exp(2x_{\max}b)} \sqrt{\left\|\left(\sum_{i=1}^{t-1} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \nabla^2 f_{\mathcal{A}}(\hat{\theta}) \left(\sum_{i=1}^{t-1} \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}} \\
&+ \frac{\lambda_{\max}}{2+2\exp(2x_{\max}b)}\delta^2.
\end{aligned} \tag{EC.40}$$

We then bound the difference between $\exp(x^T\beta^*)$ and $\exp(x^T\Sigma\beta^*)$. Via Taylor expansion, there exists a ξ such that

$$\exp(x^T\beta^*) - \exp(x^T\Sigma\beta^*) = \exp(x^T\xi)x^T(\beta^* - \Sigma\beta^*) \le \exp(x_{\max}b)x_{\max}\mathcal{G}_1(m,T,\epsilon), \tag{EC.41}$$

where the last inequality uses Assumption A.1 and Theorem 2. Combining (EC.40), (EC.41) and $\eta = \exp(x_{\max}b)$, the desirable result follows.

EC.1.8. Proof of Theorem 4

For simplification, we will ignore the subscript t in this proof. In Corollary 1, we show that for any x we have

$$\exp(x^T\beta^*) \le v^{ucb}. \tag{EC.42}$$

Let $R^{ucb}_{\Sigma\beta^*}(\mathcal{A}) = \frac{\sum_{i\in\mathcal{A}} r_i v_i^{ucb}}{\sum_{i\in\mathcal{A}} v_i^{ucb}}$. Combining (EC.42) with Lemma A.3 in Agrawal et al. (2019), we can directly show that

$$R^{ucb}_{\Sigma\beta^*}(\mathcal{A}^*) \ge R_{\beta^*}(\mathcal{A}^*).$$

Using the fact that $\mathcal{A}^{SRP}$ is the optimal assortment under utilities $\{v_i^{ucb}\}$, we can further show that

$$R^{ucb}_{\Sigma\beta^*}(\mathcal{A}^{SRP}) \geq R_{\beta^*}(\mathcal{A}^*). \tag{EC.43}$$

In addition, we can show that

$$R^{ucb}_{\Sigma\beta^*}(\mathcal{A}) - R_{\beta^*}(\mathcal{A}) \leq \frac{\sum_{i\in\mathcal{A}} r_i(v_i^{ucb} - \exp(x_i^T\beta^*))}{\sum_{i\in\mathcal{A}} v_i^{ucb}} \leq \sum_{i\in\mathcal{A}} r_i(v_i^{ucb} - \exp(x_i^T\beta^*)), \tag{EC.44}$$

where we use $v^{ucb} \geq \exp(x\beta^*)$ and $\sum_{i\in\mathcal{A}} v_i^{ucb} \geq 1$. Therefore, we can show that

$$\begin{aligned}
R_{\beta^*}(\mathcal{A}^*) - R_{\beta^*}(\mathcal{A}^{SRP}) &\leq R^{ucb}_{\Sigma\beta^*}(\mathcal{A}^{SRP}) - R_{\beta^*}(\mathcal{A}^{SRP}) \\
&\leq R_{\max} \sum_{i\in\mathcal{A}^{SRP}} (v_i^{ucb} - \exp(x_i^T\beta^*)) \\
&= R_{\max} \sum_{i\in\mathcal{A}^{SRP}} \left(\exp(x^TQ^TP_0^T\hat{\theta}) - \exp(x_i^T\beta^*)\right) \\
&+ R_{\max} \sum_{i\in\mathcal{A}^{SRP}} \left(\frac{\lambda_{\max}}{2\phi'\left(e^{x_{\max}b}\right)}\delta^2 + e^{x_{\max}b}x_{\max}\mathcal{G}_1(m,T,\epsilon)\right) \\
&+ R_{\max} \sum_{i\in\mathcal{A}^{SRP}} \left(\frac{\sqrt{2}R_{\max}\omega_t}{\phi'\left(e^{x_{\max}b}\right)} \sqrt{\left\| \left(\sum_{i=1}^t \nabla^2 f_o(\hat{\theta})\right)^{-1/2} \nabla^2 f_{\mathcal{A}^i}(\hat{\theta}) \left(\sum_{i=1}^t \nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\right)^{-1/2} \right\|_{op}}\right),
\end{aligned} \tag{EC.45}$$

where the second inequality uses (EC.44). In (EC.45) we denote the assortment with single item $i \in \mathcal{A}$ as $\mathcal{A}^i$ and use the definition of v^{ucb}.

Next, we need to upper bound the term $\sum_{i\in\mathcal{A}^{SRP}} \left(\exp(x_i^TQ^TP_0^T\hat{\theta}) - \exp(x_i^T\beta^*)\right)$. By Taylor expansion, there exists a set of $\{\xi_i : \xi_i \text{ is between } x^TQ^TP_0^T\hat{\theta} \text{ and } x_i^T\beta^*\}$ such that

$$\begin{aligned}
\sum_{i\in\mathcal{A}^{SRP}} \left(\exp(x^TQ^TP_0^T\hat{\theta}) - \exp(x_i^T\beta^*)\right) &= \sum_{i\in\mathcal{A}^{SRP}} \exp(\xi_i)x_i^T(Q^TP_0^T\hat{\theta} - \beta^*) \\
&= \sum_{i\in\mathcal{A}^{SRP}} \exp(\xi_i)x_i^T(Q^TP_0^T\hat{\theta} - \Sigma\beta^* + \Sigma\beta^* - \beta^*) \\
&= \sum_{i\in\mathcal{A}^{SRP}} \exp(\xi_i)\left[x_i^T(Q^TP_0^T\hat{\theta} - \Sigma\beta^*) + x_i^T(\Sigma\beta^* - \beta^*)\right] \\
&= \sum_{i\in\mathcal{A}^{SRP}} \exp(\xi_i)\left[z_i^T(\hat{\theta} - P_0Q\beta^*) + x_i^T(\Sigma - I)\beta^*\right] \\
&\leq \sum_{i\in\mathcal{A}^{SRP}} \exp(\xi_i)\left[\max_i \|z_i\|\delta + x_{\max}\mathcal{G}_1(m,T,\epsilon)\right],
\end{aligned} \tag{EC.46}$$

where last inequality uses $\|x_i\| \leq x_{\max}$ and $\delta = \|\hat{\theta} - P_0Q\beta^*\|$ in Assumption A.1 and event $\mathcal{E}_2(m,T)$.

Under the event $\mathcal{E}_{rp}(m,d,1/2)$, we have $\|z_i\| \leq 2\|x_i\| \leq 2x_{\max}$. Combining this result with (EC.46), we have

$$\sum_{i\in\mathcal{A}^{SRP}} \left(\exp(z_i^T\hat{\theta}) - \exp(x_i^T\beta^*)\right) \leq \sum_{i\in\mathcal{A}^{SRP}} \exp(\xi_i)\cdot x_{\max}(2\delta + \mathcal{G}_1(m,T,\epsilon)) \leq K\exp(x_{\max}b)x_{\max}(2\delta + \mathcal{G}_1(m,T,\epsilon)). \tag{EC.47}$$

Then, we will upper bound the term $\sum_{i\in\mathcal{A}^{SRP}}\sqrt{\left\|\left(\sum_{i=1}^{T}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\nabla^2 f_{\mathcal{A}^i}(\hat{\theta})\left(\sum_{i=1}^{T}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}}$. First, using Lemma EC.10, we can show that

$$\begin{aligned}
\nabla^2 f_{\mathcal{A}}(\hat{\theta}) &= \tilde{\mathbb{E}}\left[\left(z-\tilde{\mathbb{E}}[z]\right)\left(z-\tilde{\mathbb{E}}[z]\right)^T\right]\\
&= \tilde{\mathbb{E}}\left[zz^T\right]-\tilde{\mathbb{E}}\left[z\right]\tilde{\mathbb{E}}\left[z^T\right]\\
&= \tilde{\mathbb{E}}\left[z\left(z^T-\tilde{\mathbb{E}}\left[z\right]\right)^T\right]\\
&= \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\hat{\theta}}(i)\left[z_i\left(\sum_{j\in\mathcal{A}_t} p_{Q^T P_0^T\hat{\theta}}(j)(z_i-z_j)\right)^T\right]\\
&= \sum_{i,j\in\mathcal{A}} p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(j)\left[z_i\left(z_i-z_j\right)^T\right]\\
&= \sum_{i>j\in\mathcal{A}} p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(j)\left[z_i\left(z_i-z_j\right)^T+z_j\left(z_j-z_i\right)^T\right]\\
&= \sum_{i>j\in\mathcal{A}} p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(j)\left[\left(z_i-z_j\right)\left(z_i-z_j\right)^T\right]\\
&\succeq \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(0)\left[z_i z_i^T\right]
\end{aligned}$$

Moreover, by the definition of single item assortment $\mathcal{A}^i$, we have

$$\begin{aligned}
\nabla^2 f_{\mathcal{A}^i}(\hat{\theta}) &= \tilde{\mathbb{E}}\left[\left(z-\tilde{\mathbb{E}}[z]\right)\left(z-\tilde{\mathbb{E}}[z]\right)^T\right]\\
&= p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)z_i z_i^T - p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)\cdot(z_i)\left(p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)\cdot(z_i)\right)^T\\
&= p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(0)z_i z_i^T\\
\Rightarrow z_i z_i^T &= \frac{\nabla^2 f_{\mathcal{A}^i}(\hat{\theta})}{p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(0)}
\end{aligned}$$

Therefore

$$\begin{aligned}
\nabla^2 f_{\mathcal{A}}(\hat{\theta}) &\succeq \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(0)\left[z_i z_i^T\right]\\
&= \sum_{i\in\mathcal{A}} \frac{p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(0)}{p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(0)}\nabla^2 f_{\mathcal{A}^i}(\hat{\theta})\\
\Rightarrow \nabla^2 f_{\mathcal{A}}(\hat{\theta}) &\succeq \min_i\left\{\frac{p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(0)}{p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(0)}\right\}\sum_{i\in\mathcal{A}}\nabla^2 f_{\mathcal{A}^i}(\hat{\theta}),
\end{aligned}$$

which implies that

$$\sum_{i\in\mathcal{A}}\nabla^2 f_{\mathcal{A}^i}(\hat{\theta}) \preceq \max_i\left\{\frac{p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(0)}{p_{Q^T P_0^T\hat{\theta}}(i)p_{Q^T P_0^T\hat{\theta}}(0)}\right\}\nabla^2 f_{\mathcal{A}}(\hat{\theta}).$$

As x and β have upper bounds $x_{\max}$, b and $\Sigma\hat{\beta}$ is feasible, we know that

$$\exp(x^T\Sigma\hat{\beta})\in[1/\exp(x_{\max}b),\exp(x_{\max}b)].$$

As $\eta := \exp(x_{\max}b)$, then we have

$$p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(i)\le\eta/(1+\eta),\ p_{Q^T P_0^T\hat{\theta},\mathcal{A}^i}(0)\le\eta/(1+\eta)$$

$$p_{Q^T P_0^T \hat{\theta}}(i) \geq (\eta + K\eta^2)^{-1},\ p_{Q^T P_0^T \hat{\theta}}(0) \geq (1+K\eta)^{-1}$$

Thus we have

$$\begin{aligned}
&\sum_i \nabla^2 f_{\mathcal{A}^i}(\hat{\theta}) \preceq \frac{\eta^2(\eta+K\eta^2)(1+K\eta)}{(1+\eta)^2}\nabla^2 f_{\mathcal{A}}(\hat{\theta}) = \frac{\eta^3(1+K\eta)^2}{(1+\eta)^2}\nabla^2 f_{\mathcal{A}}(\hat{\theta}) \\
\Rightarrow &\nabla^2 f_{\mathcal{A}^i}(\hat{\theta}) \preceq \frac{\eta^3(1+K\eta)^2}{(1+\eta)^2}\nabla^2 f_{\mathcal{A}}(\hat{\theta}) \\
\Rightarrow &\sum_{i\in\mathcal{A}^{SRP}} \sqrt{\left\| \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \nabla^2 f_{\mathcal{A}^i}(\hat{\theta}) \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \right\|_{op}} \\
&\leq K\sqrt{\left\| \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \frac{\eta^3(1+K\eta)^2}{(1+\eta)^2}\nabla^2 f_{\mathcal{A}}(\hat{\theta}) \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \right\|_{op}} \\
&= \frac{K\eta^{3/2}(1+K\eta)}{(1+\eta)}\sqrt{\left\| \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \nabla^2 f_{\mathcal{A}}(\hat{\theta}) \left(\sum_{i=1}^T \nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2} \right\|_{op}}. \qquad \text{(EC.48)}
\end{aligned}$$

Finally, the theorem follows directly by combining (EC.45), (EC.47), and (EC.48).

EC.1.9. Proof of Theorem 5

Let both events $\mathcal{E}_{lasso}(T)$ and $\mathcal{E}_{rp}(m,d,1/2)$ hold. We separate the expected cumulative regret under random samples from that without random samples:

$$\text{REGRET}(T) \leq \overbrace{\sum_{t\in random}^{T} \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)]}^{\text{(Random samples)}} + \overbrace{\sum_{t\notin random}^{T} \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)]}^{\text{(Non-random samples)}}.$$

Non-random samples part: Recall that we periodically update the projection matrix P_0 based on the Lasso problem. We start with considering the cumulative regret for a arbitrary single period. Without loss of generality, we consider the period starting from T_a and ending at T_b.

By Theorem 4, with high probability we have

$$\begin{aligned}
&\sum_{t\notin random, t\in[T_a,T_b)} R_{t,\beta^*}(\mathcal{A}^*) - R_{t,\beta^*}(\mathcal{A}_t) \\
&\leq \overbrace{\sum_{t\notin random, t\in[T_a,T_b)} 2R_{\max}K\eta x_{\max}\mathcal{G}_1(m,T,\epsilon)}^{\text{ⓐ}} \\
&+ \overbrace{\sum_{t\notin random, t\in[T_a,T_b)} \frac{R_{\max}K\lambda_{\max}}{2+2\eta^2}\delta^2 + 2KR_{\max}\eta x_{\max}\delta}^{\text{ⓑ}}
\end{aligned}$$

$$+\overbrace{\sum_{t\notin random, t\in[T_a,T_b)}\min\left\{R_{\max},\frac{K\eta^{3/2}(1+K\eta)\sqrt{2}\omega_t}{(1+\eta^2)(1+\eta)}\sqrt{\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\nabla^2 f_{\mathcal{A}^{\mathcal{SRP}}}(\hat{\theta})\left(\sum_{i\in\mathcal{W}}\nabla^2 f(\hat{\theta})\right)^{-1/2}\right\|_{op}}\right\}}^{\text{ⓒ}}. \tag{EC.49}$$

The bound for part ⓐ:

$$\begin{aligned}
\text{ⓐ}\leq&\sum_{t=T_a}^{T_b-1}2KR_{\max}\eta x_{\max}\cdot\epsilon\sqrt{s}\cdot C_{lasso}(s+\hat{s})\sqrt{\frac{\log d+\log t}{n_{T_a}}}\\
\leq&\sum_{t=T_a}^{T_b-1}2KR_{\max}\eta x_{\max}\cdot\epsilon\sqrt{s}\cdot C_{lasso}(s+\hat{s})\sqrt{\frac{\log d+\log T}{n_{T_a}}}\\
\leq&(T_b-T_a-1)\cdot 2KR_{\max}\eta x_{\max}\cdot\epsilon\sqrt{s}\cdot C_{lasso}(s+\hat{s})\sqrt{\frac{\log d+\log T}{n_{T_a}}}.
\end{aligned}\tag{EC.50}$$

The bound for part ⓑ: As $\delta_t\leq\frac{n_t\mu}{4tL_3}$, we can show that

$$\begin{aligned}
\text{ⓑ}\leq&\frac{KR_{\max}\lambda_{\max}}{2+2\eta^2}\sum_{t=T_a}^{T_b-1}\delta^2+2KR_{\max}\eta x_{\max}\sum_{t=T_a}^{T_b-1}\delta\\
\leq&\frac{\mu^2KR_{\max}\lambda_{\max}}{32L_3^2(1+\eta^2)}\sum_{t=T_a}^{T_b-1}\frac{n_t^2}{t^2}+\frac{\mu KR_{\max}\eta x_{\max}}{2L_3}\sum_{t=T_a}^{T_b-1}\frac{n_t}{t}.
\end{aligned}\tag{EC.51}$$

The bound for part ⓒ:

Note for $R_{\max}\geq 1$ and $\frac{KR^2\eta^{3/2}(1+K\eta)\sqrt{2}\omega_t}{(1+\eta^2)(1+\eta)}\geq 1$, we can show that

$$\begin{aligned}
\text{ⓒ}\leq&\frac{R_{\max}^3K\eta^{3/2}(1+K\eta)\sqrt{2}\omega_t}{(1+\eta^2)(1+\eta)}\sum_{t=T_a}^{T_b-1}\min\left\{1,\sqrt{\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}}\right\}\\
\leq&\frac{R_{\max}^3K\eta^{3/2}(1+K\eta)\sqrt{2}\omega_t}{(1+\eta^2)(1+\eta)}\sqrt{T_b-T_a-1}\sqrt{\sum_{t=T_a}^{T_b-1}\min\left\{1,\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}\right\}},
\end{aligned}$$

where the last inequality uses the fact that $\sum_{i=1}^b\sqrt{c_i}\leq\sqrt{b}\sqrt{\sum_{i=1}^b c_i}$ holds for all $b,c_i>0$. As $f(\cdot)$ has Lipschitz hessian, we have

$$\left\|\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})-\nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right\|_{op}\leq L_3\|\hat{\theta}-P_0Q\beta^*\|\leq L_3\delta_t. \tag{EC.52}$$

When $\delta_t\leq\frac{n_t\mu}{4tL_3}$, using Lemma EC.8, we can verify that with high probability

$$\begin{aligned}
&\sum_t\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\succeq\frac{1}{2}\sum_t\nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\\
\Rightarrow&\left(\sum_t\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\right)^{-1/2}\preceq\sqrt{2}\left(\sum_t\nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)^{-1/2}.
\end{aligned}$$

Therefore,

$$\sum_{t=T_a}^{T_b-1}\min\left\{1,\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(\hat{\theta})\right)^{-1/2}\right\|_{op}\right\}$$

$$
\begin{aligned}
&\leq 2\sum_{t=T_a}^{T_b-1}\min\left\{1,\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\right\|_{op}\right\}\\
&\leq 2\sum_{t=T_a}^{T_b-1}\min\left\{1,\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\right\|_{op}\right\}\\
&+2\sum_{t=T_a}^{T_b-1}\left\|\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\left(\nabla^2 f_{\mathcal{A}_t}(\hat{\theta})-\nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)\left(\sum_{i\in\mathcal{W}}\nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*)\right)^{-1/2}\right\|_{op}\\
&\leq 4(|\mathcal{S}|+m)\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu n_{T_a}}\right)+2\sum_{t=T_a}^{T_b-1}L_3\delta_t\left\|\left(\sum_{t\in\mathcal{W}}\nabla^2 f(P_0Q\beta^*)\right)^{-1}\right\|_{op} \qquad \text{(EC.53)}\\
&\leq 4(|\mathcal{S}|+m)\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu}\right)+\sum_{t=T_a}^{T_b-1}\frac{4L_3\delta_t}{\mu n_t} \qquad \text{(EC.54)}\\
&\leq 4(|\mathcal{S}|+m)\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu}\right)+\sum_{t=T_a}^{T_b-1}\frac{4L_3}{\mu n_t}\frac{n_t\mu}{4tL_3} \qquad \text{(EC.55)}\\
&\leq 4(|\mathcal{S}|+m)\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu}\right)+\sum_{t=T_a}^{T_b-1}\frac{1}{t},
\end{aligned}
$$

where (EC.53) uses Lemma EC.9 and (EC.52), (EC.54) uses Lemma EC.8, and (EC.55) uses $\delta_t \leq \frac{n_t\mu}{4tL_3}$. Therefore, we can upper bound part ⓒ as follow:

$$
ⓒ \leq \frac{R_{\max}^3K\eta^{3/2}(1+K\eta)\sqrt{2}\omega_t}{(1+\eta^2)(1+\eta)}\sqrt{T_b-T_a-1}\sqrt{4(|\mathcal{S}_i|+m)\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu}\right)+\sum_{t=T_a}^{T_b-1}\frac{1}{t}}. \qquad \text{(EC.56)}
$$

Hence, we can show that the upper bound for non-random part is

$$
\begin{aligned}
&\sum_{t\notin random, t\in[T_a,T_b)}^{T}\mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*)-R_{t,\beta^*}(\mathcal{A}_t)]\\
&\leq (T_b-T_a-1)\cdot 2KR_{\max}\eta x_{\max}\cdot\epsilon\sqrt{s}\cdot C_{lasso}(s+\hat{s})\sqrt{\frac{\log d+\log T_b}{n_{T_a}}}\\
&+\frac{\mu^2KR_{\max}\lambda_{\max}}{32L_3^2(1+\eta^2)}\sum_{t=T_a}^{T_b-1}\frac{n_t^2}{t^2}+\frac{\mu KR_{\max}\eta x_{\max}}{2L_3}\sum_{t=T_a}^{T_b-1}\frac{n_t}{t}\\
&+\frac{R_{\max}^3K\eta^{3/2}(1+K\eta)\sqrt{2}\omega_t}{(1+\eta^2)(1+\eta)}\sqrt{T_b-T_a-1}\sqrt{4(|\mathcal{S}_i|+m)\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu}\right)+\sum_{t=T_a}^{T_b-1}\frac{1}{t}}. \qquad \text{(EC.57)}\\
&\leq C_{tmp,1}(T_b-T_a)\cdot\epsilon\sqrt{s}(s+\hat{s})n_{T_a}^{-1/2}+C_{tmp,2}\sum_{t=T_a}^{T_b-1}\left(\frac{n_t^2}{t^2}+\frac{n_t}{t}\right)+C_{tmp,3}\omega_{T_b}\sqrt{T_b-T_a}\sqrt{|\mathcal{S}_i|+m+1},
\end{aligned}
$$

where we set

$$
\begin{aligned}
C_{tmp,1}&=2KR_{\max}\eta x_{\max}\cdot C_{lasso}\sqrt{\log d+\log T_b}\\
C_{tmp,2}&=\max\left\{\frac{\mu^2KR_{\max}\lambda_{\max}}{32L_3^2(1+\eta^2)},\frac{\mu KR_{\max}\eta x_{\max}}{2L_3}\right\}\\
C_{tmp,3}&=\frac{R_{\max}^3K\eta^{3/2}(1+K\eta)\sqrt{2}}{(1+\eta^2)(1+\eta)}\cdot\max\left\{4\log\left(\frac{8(T_b-T_a)K^2x_{\max}^2}{\mu}\right),\sum_{t=T_a}^{T_b-1}\frac{1}{t}\right\}
\end{aligned}
$$

By Chernoff bound, we can show that with probability $1-\mathcal{O}(T^{-1})$, we have

$$\frac{1}{2}C_0 t^{2/3} \le n_t \le 2C_0 t^{2/3}. \tag{EC.58}$$

Combining (EC.58) and (EC.57), we can further simplify as follow:

$$\begin{aligned}
&\sum_{t\notin random, t\in[T_a,T_b)}^{T} \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)] \\
&\le 2C_{tmp,1}C_0^{-\frac{1}{2}}(T_b-T_a)\cdot\epsilon\sqrt{s}(s+\hat{s})T_a^{-\frac{1}{3}} + C_{tmp,2}\sum_{t=T_a}^{T_b-1}\left(4C_0^2 t^{-\frac{2}{3}} + 2C_0 t^{-\frac{1}{3}}\right) + C_{tmp,3}\omega_{T_b}\sqrt{T_b-T_a}\sqrt{|\mathcal{S}_i|+m+1} \\
&\le \tilde{C}_{tmp,1}\cdot\epsilon\sqrt{s}(s+\hat{s})\cdot T_b T_a^{-\frac{1}{3}} + \tilde{C}_{tmp,2}T_b^{\frac{2}{3}} + \tilde{C}_{tmp,3}\omega_{T_b}T_b^{\frac{1}{2}}\sqrt{|\mathcal{S}_i|+m},
\end{aligned} \tag{EC.59}$$

where

$$\begin{aligned}
\tilde{C}_{tmp,1} &= 2C_{tmp,1}C_0^{-1/2} \\
\tilde{C}_{tmp,2} &= 6C_{tmp,2}\max\{C_0^2, C_0\} \\
\tilde{C}_{tmp,3} &= 2C_{tmp,3}.
\end{aligned}$$

As we use $T_{lasso} = \{c^i, i = 0,1,2,...\}$ random sampling schedule, we have $T_b = c^i$ and $T_a = c^{i-1}$, and we can further simplify (EC.59) as follows:

$$\begin{aligned}
&\sum_{t\notin random, t\in\text{Period } i-1} \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)] \\
&\le \tilde{C}_{tmp,1}\cdot\epsilon\sqrt{s}(s+\hat{s})\cdot c^i(c^{i-1})^{-\frac{1}{3}} + \tilde{C}_{tmp,2}6(c^i)^{\frac{2}{3}} + \tilde{C}_{tmp,3}\omega_{T_b}(c^i)^{\frac{1}{2}}\sqrt{|\mathcal{S}_i|+m} \\
&\le \tilde{C}_{tmp,1}\cdot\epsilon\sqrt{s}(s+\hat{s})\cdot c^{i-\frac{i-1}{3}} + \tilde{C}_{tmp,2}c^{\frac{2i}{3}} + \tilde{C}_{tmp,3}\omega_{T_b}c^{\frac{i}{2}}\sqrt{|\mathcal{S}_i|+m} \\
&\le \left(\tilde{C}_{tmp,1}c^{\frac{1}{3}}\cdot\epsilon\sqrt{s}(s+\hat{s}) + \tilde{C}_{tmp,2} + \tilde{C}_{tmp,3}\omega_{T_b}\sqrt{|\mathcal{S}_i|+m}\right)c^{\frac{2i}{3}}.
\end{aligned} \tag{EC.60}$$

Moreover, as $\omega_{T_b} = 4\sqrt{4x_{\max}^2 T_b\mathcal{G}_1(m,T_b,\epsilon)\delta + 64(|\mathcal{S}_i|+m))\log(T_b) + 2\Gamma_{T_b}}$, $\mathcal{G}_1(m,T_b,\epsilon) = \epsilon\sqrt{s}\cdot(h(T_b) + \mathcal{G}_0(T,s))$, and $\delta_{T_b} \le \frac{n_t\mu}{4tL_3}$. We can upper bound ω_{T_b} as follow

$$\begin{aligned}
\omega_{T_b} &\le 8x_{\max}\sqrt{T_b\epsilon\sqrt{s}\cdot(\mathcal{G}_0(T_b,\hat{s}) + \mathcal{G}_0(T_b,s))\frac{n_{T_b}\mu}{4T_bL_3}} + 32\sqrt{|\mathcal{S}_i|+m}\log^{\frac{1}{2}}(T_b) + 4\sqrt{2\Gamma_{T_b}} \qquad \text{(EC.61)} \\
&\le 8x_{\max}\cdot\sqrt{2}C_0^{\frac{1}{2}}c^{\frac{i}{3}}\cdot\sqrt{\epsilon\sqrt{s}}\cdot\sqrt{C_{lasso}(s+\hat{s})\sqrt{\frac{\log d+\log T_b}{\frac{1}{2}C_0c^{\frac{2(i-1)}{3}}}\cdot\frac{2C_0c^{\frac{2i}{3}}\mu}{4c^iL_3}}} + 32\sqrt{|\mathcal{S}_i|+m}\log^{\frac{1}{2}}(T_b) + 4\sqrt{2\Gamma_{T_b}} \\
&= C_{tmp,4}c^{\frac{1}{6}}\epsilon^{\frac{1}{2}}s^{\frac{1}{4}}\cdot\sqrt{s+\hat{s}} + 32\sqrt{|\mathcal{S}_i|+m}\log^{\frac{1}{2}}(T_b) + 4\sqrt{2\Gamma_{T_b}}, \qquad \text{(EC.62)}
\end{aligned}$$

where (EC.61) uses the fact that $\sqrt{a+b+c} \le \sqrt{a}+\sqrt{b}+\sqrt{c}$ for all $a,b,c \ge 0$ and

$$C_{tmp,4} = 8x_{\max}\cdot\sqrt{2}C_0^{\frac{1}{2}}\cdot\sqrt{(2L_3)^{-1}C_0C_{lasso}\mu\sqrt{2(\log d+\log T_b)C_0^{-1}}}.$$

Directly combining (EC.62) with (EC.60), we can show that

$$\sum_{t\notin random} \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)]$$

$$
\begin{aligned}
&\leq \left(\tilde{C}_{tmp,1} c^{\frac{1}{3}} \cdot \epsilon\sqrt{s}(s+\hat{s}) + \tilde{C}_{tmp,2}\right) \sum_i c^{\frac{2i}{3}} \\
&+ \left(\tilde{C}_{tmp,3}\left(C_{tmp,4} c^{\frac{1}{6}} \epsilon^{\frac{1}{2}} s^{\frac{1}{4}} \sqrt{s+\hat{s}} + 32\sqrt{\max_i |\mathcal{S}_i| + m}\log^{\frac{1}{2}}(T) + 4\sqrt{2\Gamma_T}\right)\sqrt{\max_i |\mathcal{S}_i| + m}\right) \sum_i c^{\frac{2i}{3}} \\
&= \left(\tilde{C}_{tmp,1} c^{\frac{1}{3}} \epsilon\sqrt{s}(s+\hat{s}) + \tilde{C}_{tmp,4} c^{\frac{1}{6}} \epsilon^{\frac{1}{2}} s^{\frac{1}{4}} \sqrt{s+\hat{s}}\sqrt{\max_i |\mathcal{S}_i| + m} + 32\tilde{C}_{tmp,3}\log^{\frac{1}{2}}(T)(\max_i |\mathcal{S}_i| + m)\right) \sum_i c^{\frac{2i}{3}} \\
&+ \left(\tilde{C}_{tmp,2} + 4\tilde{C}_{tmp,3}\sqrt{2\Gamma_T(\max_i |\mathcal{S}_i| + m)}\right) \sum_i c^{\frac{2i}{3}} \\
&\leq \left(\tilde{C}_{tmp,1} c^{\frac{1}{3}} \epsilon\sqrt{s}(s+\hat{s}) + \tilde{C}_{tmp,4} c^{\frac{1}{6}} \epsilon^{\frac{1}{2}} s^{\frac{1}{4}} (s+\hat{s}+\max_i |\mathcal{S}_i| + m) + 32\tilde{C}_{tmp,3}\log^{\frac{1}{2}}(T)(\max_i |\mathcal{S}| + m)\right) \sum_i c^{\frac{2i}{3}} \\
&+ \left(\tilde{C}_{tmp,2} + 4\tilde{C}_{tmp,3}\sqrt{2\Gamma_T(\max_i |\mathcal{S}_i| + m)}\right) \sum_i c^{\frac{2i}{3}} \\
&\leq \left(C_{tmp,5}\max\{\epsilon\sqrt{s}, \epsilon^{\frac{1}{2}} s^{\frac{1}{4}}\}\left(s+\hat{s}+\max_i |\mathcal{S}| + m\right) + \tilde{C}_{tmp,2} + 4\tilde{C}_{tmp,3}\sqrt{2\Gamma_T(\max_i |\mathcal{S}_i| + m)}\right) c|1-c^{\frac{2}{3}}|^{-1} T^{\frac{2}{3}},
\end{aligned}
\tag{EC.63}
$$

where in last inequality we uses the formula of exponential series summation.

Random samples part: Since we use random decay sampling schedule, with high probability we have

$$
\sum_{t\in random}^{T} \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)] \leq R_{\max} n_T \leq 2C_0 T^{\frac{2}{3}}. \tag{EC.64}
$$

Combining both non-random samples part and random samples part, i.e., (EC.64) and (EC.63), we can upper bound the cumulative regret up to time T as follow:

$$
\begin{aligned}
&\sum_t \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)] \\
&\leq \left(C_{tmp,5}\max\{\epsilon\sqrt{s}, \epsilon^{\frac{1}{2}} s^{\frac{1}{4}}\}\left(s+\hat{s}+\max_i |\mathcal{S}_i| + m\right) + \tilde{C}_{tmp,2} + 4\tilde{C}_{tmp,3}\sqrt{2\Gamma_T(\max_i |\mathcal{S}_i| + m)}\right) c|1-c^{\frac{2}{3}}|^{-1} T^{\frac{2}{3}} + 2C_0 T^{\frac{2}{3}} \\
&\leq \mathcal{O}\left((s+\hat{s}+\max\{|\mathcal{S}|\}+m)\log^{\frac{1}{2}} d\log^{\frac{5}{2}} T \cdot T^{\frac{2}{3}}\right),
\end{aligned}
\tag{EC.65}
$$

where we choose $\epsilon = s^{-1/2}$ in the last inequality. Via Theorem 2, we know that $\max_i |\mathcal{S}_i| \leq s + \mathcal{G}_0(T,s)/h(T) = s + \frac{s}{\hat{s}} \leq 2s$, and then above result can be further simplified as

$$
\begin{aligned}
&\sum_t \mathbb{E}[R_{t,\beta^*}(\mathcal{A}_t^*) - R_{t,\beta^*}(\mathcal{A}_t)] \\
&\leq \left(\tilde{C}_{f,1}(3s+\hat{s}+m) + \tilde{C}_{f,2} + \tilde{C}_{f,3}\sqrt{2s+m} + 2C_0\right) T^{\frac{2}{3}} \\
&= \mathcal{O}\left(sm\sqrt{\log d}\cdot T^{\frac{2}{3}}\log^{\frac{5}{2}} T\right),
\end{aligned}
\tag{EC.66}
$$

where $\tilde{C}_{f,1} = \frac{cC_{tmp,5}}{|1-c^{\frac{2}{3}}|}$, $\tilde{C}_{f,2} = \frac{c\tilde{C}_{tmp,2}}{|1-c^{\frac{2}{3}}|}$, and $\tilde{C}_{f,3} = \frac{4c\tilde{C}_{tmp,3}\sqrt{2\Gamma_T}}{|1-c^{\frac{2}{3}}|}$.

We then consider the probability part. Note that in previous proofs, we assume events $\mathcal{E}_{lasso}(T)$ for T and $\mathcal{E}_{rp}(m,d,1/2)$ hold for all products. The remaining task is to bound the probability that those two events happen simultaneously. As we require $C_0 > 0$, then when $T \geq = (2\log T/(CC_0))^{\frac{3}{2}}$, we have $n_T \geq \mathcal{O}(s^2\log T)$.

Using Lemma 1, we have

$$\mathbb{P}(\mathcal{E}_{lasso}(T)) \geq 1 - \mathcal{O}(T^{-1}). \tag{EC.67}$$

Via Lemma 2 and union bound, we have

$$\mathbb{P}(\mathcal{E}_{rp}(m,d,1/2)) \geq 1 - 2(NT)^2 \exp\left(-\frac{m}{32}\right) \geq 1 - \exp\left(-\frac{m}{32} + 2\log(2NT)\right). \tag{EC.68}$$

We then use the union bound over (EC.67) and (EC.68) and the desirable result follows:

$$\mathbb{P}(\mathcal{E}_{lasso}(T) \cap \mathcal{E}_{rp}(m,d,1/2)) \geq 1 - \exp\left(-\frac{m}{32} + 2\log(2NT)\right) - \exp\left(-\frac{m}{8s} + \log(4\log T)\right) - \mathcal{O}(T^{-1}).$$

At last, the probability part follows directly by setting $m = \mathcal{O}\left(\hat{s}\log(NT)\right)$.

EC.1.10. Proof of Corollary 2

The proof procedure is analogy to the proof of Theorem 5. We omit it for brevity.

EC.2. Appendix: Technical Lemmas

EC.2.1. Proof of Lemma EC.1

We start with proving the first part of this lemma. By using the definition of x^* and β, we can show that

$$\begin{aligned}
R_\beta(n) &= \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\exp(\beta^T x^*)}{1+\exp(\beta^T x^*)}\right] - \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\exp(\beta^T x_t)}{1+\exp(\beta^T x_t)}\right] \\
&= \mathbb{E}_\beta\left[\frac{n}{1+\exp(-(s-1)\epsilon)} - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{1+\exp(-\beta^T x_t)} - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{1+\exp(-\beta^T x_t)}\right] \\
&= \mathbb{E}_\beta\left[\frac{n}{1+\exp(-(s-1)\epsilon)} - \frac{n}{2} - \sum_{t=1}^{n}\left(\frac{\mathbb{1}(x_t\in\mathcal{H})}{1+\exp(-\beta^T x_t)} - \frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\right) - \sum_{t=1}^{n}\left(\frac{\mathbb{1}(x_t\in\mathcal{S})}{1+\exp(-\beta^T x_t)} - \frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\right)\right] \\
&= \mathbb{E}_\beta\left[\frac{n}{2}\left(\frac{1-\exp(-(s-1)\epsilon)}{1+\exp(-(s-1)\epsilon)}\right) - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\frac{1-\exp(-\beta^T x_t)}{1+\exp(-\beta^T x_t)}\cdot - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\frac{1-\exp(-\beta^T x_t)}{1+\exp(-\beta^T x_t)}\right] \\
&= \mathbb{E}_\beta\left[\frac{n}{2}\tanh\left(\frac{(s-1)\epsilon}{2}\right) - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{\beta^T x_t}{2}\right) - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\tanh\left(\frac{\beta^T x_t}{2}\right)\right].
\end{aligned}$$

Note that when ϵ is small enough, we have

$$\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{\beta^T x_t}{2}\right) < 0.$$

Therefore, we can further show that

$$\begin{aligned}
R_\beta(n) &\geq \mathbb{E}_\beta\left[\frac{n}{2}\tanh\left(\frac{(s-1)\epsilon}{2}\right) - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\tanh\left(\frac{\beta^T x_t}{2}\right)\right] \\
&= \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{(s-1)\epsilon}{2}\right) + \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\left(\tanh\left(\frac{(s-1)\epsilon}{2}\right) - \tanh\left(\frac{\beta^T x_t}{2}\right)\right)\right] \\
&\geq \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{(s-1)\epsilon}{6}\right) + \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\underbrace{\left(\tanh\left(\frac{(s-1)\epsilon}{2}\right) - \tanh\left(\frac{\beta^T x_t}{2}\right)\right)}_{(*)}\right],
\end{aligned}$$

where the last inequality uses the fact that the $\tanh(\cdot)$ function is monotonically increasing and $\tanh(0)=0$. Applying the Taylor expansion on the $(*)$ term, we can show that

$$R_\beta(n) \geq \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{(s-1)\epsilon}{6}\right) + \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\frac{(s-1)\epsilon-\beta^T x_t}{2(\cosh(\xi_t)+1)}\right],$$

where ξ_t is between $\beta^T x_t$ and $(s-1)\epsilon$. By the construction of β and $x_t\in\mathcal{S}$, we can directly show that $(s-1)\epsilon/2 - \beta^T x_t/2 \geq 0$. By picking ϵ to be small enough, we have $\cosh(\xi_t)\in[1,2]$, which implies

$$\begin{aligned}
R_\beta(n) &\geq \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{(s-1)\epsilon}{6}\right) + \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\frac{(s-1)\epsilon-\beta^T x_t}{6}\right] \\
&\geq \mathbb{E}_\beta\left[\sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{H})}{2}\tanh\left(\frac{(s-1)\epsilon}{6}\right) + \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\left(\tanh\left(\frac{(s-1)\epsilon}{6}\right) - \frac{\beta^T x_t}{6}\right)\right] \\
&= \mathbb{E}_\beta\left[\frac{n}{2}\tanh\left(\frac{(s-1)\epsilon}{6}\right) - \sum_{t=1}^{n}\frac{\mathbb{1}(x_t\in\mathcal{S})}{2}\frac{\beta^T x_t}{6}\right],
\end{aligned}$$

where the last inequality uses the fact that $x > \tanh(x)$ for all $x \geq 0$. Conditioning on the event $\mathcal{E}$, we can show that

$$\begin{aligned} R_\beta(n) &\geq \mathbb{E}_\beta \left[\frac{n}{2} \tanh\left(\frac{(s-1)\epsilon}{6}\right) - \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \frac{\beta^T x_t}{6} \right] \\ &\geq \mathbb{E}_\beta \left[\frac{n}{2} \tanh\left(\frac{(s-1)\epsilon}{6}\right) - \frac{n}{4} \tanh\left(\frac{(s-1)\epsilon}{6}\right) \right] \mathbb{P}_\beta(\mathcal{E}) \\ &\geq \frac{n}{4} \tanh\left(\frac{(s-1)\epsilon}{6}\right) \mathbb{P}_\beta(\mathcal{E}). \end{aligned} \tag{EC.69}$$

Next, we will derive the second part of this Lemma. Denote $\tilde{x}^*$ as the optimal single-item assortment for the assortment problem parameterized by the coefficient vector $\tilde{\beta}$. By similar analysis, we can show that

$$\begin{aligned} R_{\tilde{\beta}}(n) &= \mathbb{E}_{\tilde{\beta}} \left[\frac{n}{2} \tanh\left(\frac{2(s-1)\epsilon}{2}\right) - \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{H}) + \mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{\tilde{\beta}^T x_t}{2}\right) \right] \\ &\geq \mathbb{E}_{\tilde{\beta}} \left[\frac{n}{2} \tanh\left(\frac{2(s-1)\epsilon}{2}\right) - \underbrace{\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{\tilde{\beta}^T x_t}{2}\right)}_{(**)} \right]. \end{aligned}$$

Now, we will analyze the upper bound for the $(**)$ term:

$$\begin{aligned} (**) &= \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{(\beta - 2\epsilon\tilde{x})^T x_t}{2}\right) \\ &\leq \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{\beta^T x_t + 2\epsilon \sum_{j \in \text{supp}(\tilde{x})} |x_{t,j}|}{2}\right) \\ &\leq \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{\sum_{j=1}^{s-1} |x_{t,j}|\epsilon + 2\epsilon \sum_{j \in \text{supp}(\tilde{x})} |x_{t,j}|}{2}\right), \end{aligned}$$

where these two inequalities use $\tilde{x} \in \mathcal{S}'$ with $\|\tilde{x}\|_\infty = 1$, the definition of β and $x_t \in \mathcal{S}$. Note that by the construction of $\mathcal{S}'$, the first $(s-1)$ elements in $\tilde{x}$ are zero, which implies that

$$\begin{aligned} &\sum_{j=1}^{s-1} |x_{t,j}| + \sum_{j \in \text{supp}(\tilde{x})} |x_{t,j}| \leq \sum_{j=1}^{d} |x_{t,j}| = \|x_t\|_1 = s-1 \\ \Rightarrow &\sum_{j=1}^{s-1} |x_{t,j}|\epsilon + 2\epsilon \sum_{j \in \text{supp}(\tilde{x})} |x_{t,j}| \leq 2(s-1)\epsilon - \sum_{j=1}^{s-1} |x_{t,j}|\epsilon. \end{aligned}$$

Therefore, we have

$$(**) \leq \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{2(s-1)\epsilon - \sum_{j=1}^{s-1} |x_{t,j}|\epsilon}{2}\right),$$

which leads to

$$\begin{aligned} R_{\tilde{\beta}}(n) &\geq \mathbb{E}_{\tilde{\beta}} \left[\frac{n}{2} \tanh\left(\frac{2(s-1)\epsilon}{2}\right) - \sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \tanh\left(\frac{2(s-1)\epsilon - \sum_{j=1}^{s-1} |x_{t,j}|\epsilon}{2}\right) \right] \\ &\geq \mathbb{E}_{\tilde{\beta}} \left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \left(\tanh\left(\frac{2(s-1)\epsilon}{2}\right) - \tanh\left(\frac{2(s-1)\epsilon - \sum_{j=1}^{s-1} |x_{t,j}|\epsilon}{2}\right) \right) \right] \end{aligned}$$

$$= \mathbb{E}_{\tilde{\beta}}\left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \frac{\sum_{j=1}^{s-1} |x_{t,j}|\epsilon}{2(\cosh(\xi_t)+1)}\right], \tag{EC.70}$$

where ξ_t is between $\frac{2(s-1)\epsilon}{2}$ and $\frac{2(s-1)\epsilon - \sum_{j=1}^{s-1}|x_{t,j}|\epsilon}{2} \geq \frac{(s-1)\epsilon}{2} > 0$. Based on the monotonicity of the $\cosh(\cdot)$ function, we can show that

$$\begin{aligned}
R_{\tilde{\beta}}(n) \geq & \mathbb{E}_{\tilde{\beta}}\left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \frac{(\sum_{j=1}^{s-1} |x_{t,j}| + 1)\epsilon}{2(\cosh(2(s-1)\epsilon)+1)}\right] \\
\geq & \mathbb{E}_{\beta}\left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S})}{2} \frac{\sum_{j=1}^{s-1} |x_{t,j}|\epsilon}{2(\cosh(2(s-1)\epsilon)+1)}\right] \\
\geq & \mathbb{E}_{\beta}\left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S}) \sum_{j=1}^{s-1} |x_{t,j}|\epsilon}{12}\right],
\end{aligned}$$

where the last inequality holds when ϵ is small enough so that $\cosh(2(s-1)\epsilon) \in [1,2]$. Conditioning on the event $\mathcal{E}^c$, we can show that

$$\begin{aligned}
R_{\tilde{\beta}}(n) \geq & \mathbb{E}_{\tilde{\beta}}\left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S}) \sum_{j=1}^{s-1} x_{t,j}\epsilon}{12}\right] \\
\geq & \mathbb{E}_{\tilde{\beta}}\left[\sum_{t=1}^{n} \frac{\mathbb{1}(x_t \in \mathcal{S}) \sum_{j=1}^{s-1} x_{t,j}\epsilon}{12}\right] \mathbb{P}_{\tilde{\beta}}(\mathcal{E}^c) \\
\geq & \frac{n}{4} \tanh\left(\frac{(s-1)\epsilon}{6}\right) \mathbb{P}_{\tilde{\beta}}(\mathcal{E}^c).
\end{aligned} \tag{EC.71}$$

EC.2.2. Proof of Lemma EC.2

Via Lemma 15.1 in Lattimore and Szepesvári (2020), we can directly show that

$$\mathrm{KL}\left(\mathbb{P}_{\beta}, \mathbb{P}_{\tilde{\beta}}\right) = \sum_{t=1}^{n} \mathrm{KL}\left(P(x_t), P'(x_t)\right), \tag{EC.72}$$

where $P(a)$ and $P'(a)$ denote the Bernoulli distributions with success probability $1/(1+\exp(-\beta^T a))$ and $1/(1+\exp(-\tilde{\beta}^T a))$ respectively. We first analyze $\mathrm{KL}\left(P(x_t), P'(x_t)\right)$:

$$\begin{aligned}
& \mathrm{KL}\left(P(x_t), P'(x_t)\right) \\
= & \frac{1}{1+\exp(-x_t^T\beta)} \log\left(\frac{1+\exp(-x_t^T\tilde{\beta})}{1+\exp(-x_t^T\beta)}\right) + \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \log\left(\frac{\exp(-x_t^T\beta)(1+\exp(-x_t^T\tilde{\beta}))}{\exp(-x_t^T\tilde{\beta})(1+\exp(-x_t^T\beta))}\right) \\
= & \frac{1}{1+\exp(-x_t^T\beta)} \log\left(\frac{1+\exp(-x_t^T\tilde{\beta})}{1+\exp(-x_t^T\beta)}\right) + \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \log\left(\frac{1+\exp(-x_t^T\tilde{\beta})}{1+\exp(-x_t^T\beta)}\right) + \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \log\left(\frac{\exp(-x_t^T\beta)}{\exp(-x_t^T\tilde{\beta})}\right) \\
= & \log\left(\frac{1+\exp(-x_t^T\tilde{\beta})}{1+\exp(-x_t^T\beta)}\right) - \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} x_t^T(\beta - \tilde{\beta}) \\
= & \log\left(\frac{1+\overbrace{\exp(-x_t^T(\beta - 2\epsilon\tilde{x}))}^{(***)}}{1+\exp(-x_t^T\beta)}\right) - \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \cdot 2x_t^T\tilde{x}\epsilon.
\end{aligned}$$

By using Taylor expansion on the $(***)$ term, we can show that there exists a constant $\delta_x \in [0,1]$ such that

$$\mathrm{KL}\left(P(x_t), P(x_t)'\right) = \log\left(\frac{1+\exp(-x_t^T\beta) + \exp(-x_t^T(\beta - \delta_x 2\epsilon\tilde{x})) \cdot 2\epsilon x_t^T\tilde{x}}{1+\exp(-x_t^T\beta)}\right) - \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \cdot 2x_t^T\tilde{x}\epsilon$$

$$= \log\left(1 + \frac{\exp(-x_t^T(\beta - \delta_x 2\epsilon\tilde{x}))2\epsilon x_t^T\tilde{x}}{1+\exp(-x_t^T\beta)}\right) - \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \cdot 2x_t^T\tilde{x}\epsilon.$$

Then, by using the fact that $\log(1+x) \le x$ for $x \ge 0$, we have

$$\begin{aligned}
\mathrm{KL}\left(P(x_t), P'(x_t)\right) \le & \frac{\exp(-x_t^T(\beta - \delta_x 2\epsilon x^T\tilde{x}))2\epsilon x_t^T\tilde{x}}{1+\exp(-x_t^T\beta)} - \frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} \cdot 2x_t^T\tilde{x}\epsilon \\
= & \frac{\exp(-x_t^T(\beta - \delta_x 2\epsilon\tilde{x})) - \exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)} 2\epsilon x_t^T\tilde{x} \\
= & \frac{\exp(-x_t^T(\beta - \tilde{\delta}_x 2\epsilon\tilde{x})) \cdot 2\delta_x\epsilon x_t^T\tilde{x}}{1+\exp(-x_t^T\beta)} 2\epsilon x_t^T\tilde{x},
\end{aligned}$$

where $\tilde{\delta}_x \in [0, \delta_x] \le 1$. We can further show that

$$\begin{aligned}
\mathrm{KL}\left(P(x_t), P(x_t)'\right) \le & \delta_x \frac{\exp(-x_t^T(\beta - \tilde{\delta}_x 2\epsilon\tilde{x}))}{1+\exp(-x_t^T\beta)} \left(2\epsilon x_t^T\tilde{x}\right)^2 \\
= & \delta_x \frac{\exp(-x_t^T\beta + 2\tilde{\delta}_x x_t^T\epsilon\tilde{x})}{1+\exp(-x_t^T\beta)} \left(2\epsilon x_t^T\tilde{x}\right)^2 \\
= & \overbrace{\delta_x}^{\le 1} \exp(2\epsilon \overbrace{\tilde{\delta}_x}^{\le 1} \overbrace{x_t^T\tilde{x}}^{\le s-1}) \overbrace{\frac{\exp(-x_t^T\beta)}{1+\exp(-x_t^T\beta)}}^{\le 1} \left(2\epsilon x_t^T\tilde{x}\right)^2 \\
\le & \exp(2\epsilon(s-1)) \left(2\epsilon x_t^T\tilde{x}\right)^2 \\
\le & 12\epsilon^2 \left(x_t^T\tilde{x}\right)^2, \qquad \text{(EC.73)}
\end{aligned}$$

where the last inequality holds for a small enough ϵ. Combining (EC.72), (EC.73) and the definition of $\tilde{\beta}$, we have

$$\mathrm{KL}\left(\mathbb{P}_\beta, \mathbb{P}_{\tilde{\beta}}\right) \le 12\epsilon^2 \sum_{t=1}^n \left(x_t^T\tilde{x}\right)^2 = 12\epsilon^2 \min_{z\in\mathcal{S}'} \sum_{t=1}^n \left(x_t^T z\right)^2 \le 12\epsilon^2 \frac{1}{|\mathcal{S}'|} \underbrace{\sum_{z\in\mathcal{S}'} \sum_{t=1}^n \left(x_t^T z\right)^2}_{(****)}, \qquad \text{(EC.74)}$$

where the last inequality uses the fact that the minimum value is always no larger than its corresponding average value. Next, we analyze the upper bound for the $(****)$ term:

$$\begin{aligned}
(****) &= \sum_{z\in\mathcal{S}'} \sum_{t=1}^n \left(\sum_{j=1}^d x_{t,j} z_j\right)^2 \\
&= \sum_{z\in\mathcal{S}'} \sum_{t=1}^n \left[\sum_{j=1}^d \left(x_{t,j} z_j\right)^2 + 2\sum_{i<j} x_{t,j} x_{t,i} z_i z_j\right].
\end{aligned}$$

We can bound the above two terms separately. Specifically, to bound the first term, we observe that

$$\begin{aligned}
& \sum_{z\in\mathcal{S}'} \sum_{t=1}^n \left[\sum_{j=1}^d \left(x_{t,j} z_j\right)^2\right] \\
= & \sum_{z\in\mathcal{S}'} \sum_{t=1}^n \mathbb{1}(x_t \in \mathcal{S}) \sum_{j=1}^d \left(x_{t,j} z_j\right)^2 + \sum_{z\in\mathcal{S}'} \sum_{t=1}^n \mathbb{1}(x_t \in \mathcal{H}) \sum_{j=1}^d \left(x_{t,j} z_j\right)^2 \\
\le & \sum_{t=1}^n \mathbb{1}(x_t \in \mathcal{S}) \sum_{z\in\mathcal{S}'} \sum_{j=1}^d \left(x_{t,j} z_j\right)^2 + \sum_{z\in\mathcal{S}'} \sum_{t=1}^n \mathbb{1}(x_t \in \mathcal{H}) \sum_{j=1}^d \left(x_{t,j} z_j\right)^2.
\end{aligned}$$

Since x_t and z are $(s-1)$-sparse, we have $\sum_{j=1}^{d}(x_{t,j}z_j)^2 \le s-1$. Note that only when x_t and β are overlapping on at least one dimension, $\sum_{j=1}^{d}(x_{t,j}z_j)^2$ will be nonzero. Therefore, we have

$$\sum_{z\in\mathcal{S}'}\sum_{j=1}^{d}(x_{t,j}z_j)^2 \le \binom{d-s-1}{s-2}(s-1)$$
$$\Rightarrow \sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{S})\sum_{z\in\mathcal{S}'}\sum_{j=1}^{d}(x_{t,j}z_j)^2 \le \binom{d-s-1}{s-2}\cdot n(s-1).$$

In addition, since β is $(s-1)$-sparse with 0 on its last dimension, we have

$$\sum_{z\in\mathcal{S}'}\sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{H})\sum_{j=1}^{d}(x_{t,j}z_j)^2 \le \binom{d-s}{s-1}\sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{H})(s-1)\kappa^2.$$

Therefore, we have

$$\sum_{z\in\mathcal{S}'}\sum_{t=1}^{n}\left[\sum_{j=1}^{d}(x_{t,j}z_j)^2\right] \le \binom{d-s-1}{s-2}\cdot n(s-1) + \binom{d-s}{s-1}\cdot\sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{H})(s-1)\kappa^2.$$

Now, we bound the second term. Based on the symmetry construction of $\mathcal{S}^{'}$, we observe that

$$\sum_{z\in\mathcal{S}'}x_{t,j}x_{t,i}z_iz_j = 0$$
$$\Rightarrow \sum_{z\in\mathcal{S}'}\sum_{t=1}^{n}2\sum_{i<j}x_{t,j}x_{t,i}z_iz_j = 2\sum_{i<j}\sum_{t=1}^{n}\sum_{z\in\mathcal{S}'}x_{t,j}x_{t,i}z_iz_j = 0.$$

Thus, we can show that

$$\sum_{z\in\mathcal{S}'}\sum_{t=1}^{n}(x_t^Tz)^2 \le \binom{d-s-1}{s-2}\cdot n(s-1) + \binom{d-s}{s-1}\cdot\sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{H})(s-1)\kappa^2.$$
$$\Rightarrow \frac{1}{|\mathcal{S}'|}\sum_{z\in\mathcal{S}'}\sum_{t=1}^{n}(x_t^Tz)^2 \le \frac{n(s-1)^2}{d-s} + \sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{H})(s-1)\kappa^2. \tag{EC.75}$$

Combining (EC.75) and (EC.74), we have

$$\mathrm{KL}\left(\mathbb{P}_\beta,\mathbb{P}_{\tilde{\beta}}\right) \le 12\epsilon^2\left(\frac{n(s-1)^2}{d-s} + \sum_{t=1}^{n}\mathbb{1}(x_t\in\mathcal{H})(s-1)\kappa^2\right).$$

EC.2.3. Proof of Lemma EC.3

To simplify the notation in this proof, we use $\nabla^2 L(\xi)$ to denote $\nabla^2 L(\xi|x,\mathcal{A})$, which can be re-written as follows:

$$\nabla^2 L(\xi) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi)x_{k,i})(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi)x_{k,i}^T)}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi)^2} - \frac{\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi)x_{k,i}x_{k,i}^T}{\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi)}\right)$$
$$= -\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{k_1\in\mathcal{A}_i}\sum_{k_2\in\mathcal{A}_i}\exp(x_{k_1,i}^T\xi)\exp(x_{k_2,i}^T\xi)x_{k_1,i}x_{k_2,i}^T}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi)^2}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\frac{\sum_{k_1\in\mathcal{A}_i}\sum_{k_2\in\mathcal{A}_i}\exp(x_{k_2,i}^T\xi)\exp(x_{k_1,i}^T\xi)x_{k_1,i}x_{k_1,i}^T}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi))^2}$$
$$= -\frac{1}{n}\sum_{i=1}^{n}\sum_{k1,k2}\frac{\exp(x_{k_2,i}^T\xi)\exp(x_{k_1,i}^T\xi)x_{k_1,i}(x_{k_2,i}-x_{k_1,i})^T}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi))^2}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{k_1,k_2}\Phi_i(k_1,k_2),$$

where we use $\Phi_i(k_1,k_2)$ to denote the following shorthand:

$$\Phi_i(k_1,k_2) := -\frac{\exp(x_{k_2,i}^T\xi)\exp(x_{k_1,i}^T\xi)x_{k_1,i}(x_{k_2,i}-x_{k_1,i})^T}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi))^2}$$

Further, we denote $\phi_i(k_1,k_2) := \frac{\exp(x_{k_2,i}^T\xi)\exp(x_{k_1,i}^T\xi)}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi))^2}$. Hence, when $k_1 \neq k_2$, we have

$$\begin{aligned}\Phi_i(k_1,k_2)+\Phi_i(k_2,k_1) =& -\frac{\exp(x_{k_2,i}^T\xi)\exp(x_{k_1,i}^T\xi)(x_{k_1,i}(x_{k_2,i}-x_{k_1,i})^T + x_{k_2,i}(x_{k_1,i}-x_{k_2,i})^T)}{(\sum_{k\in\mathcal{A}_i}\exp(x_{k,i}^T\xi))^2}\\ =& -\phi_i(k_1,k_2)(x_{k_1,i}(x_{k_2,i}-x_{k_1,i})^T + x_{k_2,i}(x_{k_1,i}-x_{k_2,i})^T),\end{aligned}$$

Then, for any $z\in\mathbb{R}^d$, we have

$$\begin{aligned}\frac{z^T(\Phi_i(k_1,k_2)+\Phi_i(k_2,k_1))z}{\phi_i(k_1,k_2)} &= -z^T(x_{k_1,i}(x_{k_2,i}-x_{k_1,i})^T + x_{k_2,i}(x_{k_1,i}-x_{k_2,i})^T)z\\ &= -\left(z^Tx_{k_1,i}x_{k_2,i}^Tz - z^Tx_{k_1,i}x_{k_1,i}^Tz + z^Tx_{k_2,i}x_{k_1,i}^Tz - z^Tx_{k_2,i}x_{k_2,i}^Tz\right)\\ &= -\left(-\|z^Tx_{k_1,i}\|^2 - \|z^Tx_{k_2,i}\|^2 + 2\langle z^Tx_{k_1,i}, z^Tx_{k_2,i}\rangle\right)\\ &= \|z^T(x_{k_1,i}-x_{k_2,i})\|^2 \geq 0\end{aligned}$$

Further, we can show that

$$\begin{aligned}z^T\nabla^2L(\xi)z &= \frac{1}{n}\sum_i\sum_{k_1,k_2}z^T\Phi_i(k_1,k_2)z\\ &= \frac{1}{n}\sum_i\sum_{k_1<k_2}\phi_i(k_1,k_2)\|z^T(x_{k_1,i}-x_{k_2,i})\|^2\\ \Rightarrow \nabla^2L(\xi) &= \frac{1}{n}\sum_i\sum_{k_1<k_2}\phi_i(k_1,k_2)(x_{k_2,i}-x_{k_2,i})(x_{k_2,i}-x_{k_2,i})^T\\ \Rightarrow \nabla^2L(\xi) &= \frac{1}{n}\sum_i\sum_{k_1<k_2}y_{k_1,k_2,i}y_{k_1,k_2,i}^T, \qquad \text{(EC.76)}\end{aligned}$$

where we set $y_{k_1,k_2,i} = \sqrt{\phi_i(k_1,k_2)}(x_{k_1,i}-x_{k_2,1})$. By the definition of $\phi_i(k_1,k_2)$, we know that $\|y_{k_1,k_2,i}\|_\infty \leq 2x_{\max} := y_{\max}$. Let $K = y_{\max}$, $\sigma_0 = \sqrt{2}y_{\max}$, we can verify the follow inequalitys hold

$$K^2\left(\mathbb{E}[\exp(y_{k1,k2,i,j}^2/K^2)-1]\right) \leq y_{\max}^2(e-1) \leq \sigma_0^2, \qquad \text{(EC.77)}$$

where $y_{k1,k2,i,j}$ is the j-th element of $y_{k1,k2,i}$. Then, via the exercise 14.3 in Bühlmann and Van De Geer (2011), we have the following inequality for $t>0$:

$$\begin{aligned}&\mathbb{P}\left\{\left\|\frac{2}{nK(K-1)}\sum_i\sum_{k_1<k_2}y_{k_1,k_2,i}y_{k_1,k_2,i}^T - \mathbb{E}[y_{k_1,k_2,i}y_{k_1,k_2,i}^T]\right\|_\infty\right.\\ &\left.\geq 2y_{\max}^2t + 4z_{\max}^2\sqrt{t} + \sqrt{8}y_{\max}^2\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right)\right\} \leq \exp\left(-\frac{1}{2}nK(K-1)t\right), \qquad \text{(EC.78)}\end{aligned}$$

where

$$\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right) = \sqrt{\frac{2\log(d(d-1))}{n}} + \frac{y_{\max}\log(d(d-1))}{n}$$

Note that when $t<1$ and $n\geq \log d/t$, we will have the following inequalities:

$$2y_{\max}^2 t + 4y_{\max}^2\sqrt{t} \leq 6y_{\max}^2\sqrt{t}$$
$$\sqrt{8}y_{\max}^2\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right) \leq \sqrt{8}y_{\max}^2\left(\sqrt{\frac{4\log d}{n}} + \frac{2y_{\max}\log d}{n}\right) \leq 4\sqrt{2}y_{\max}^2(1+y_{\max})\sqrt{t}.$$

Combining these two inequalities, we have

$$2y_{\max}^2 t + 4y_{\max}^2\sqrt{t} + \sqrt{8}y_{\max}^2\lambda\left(\frac{\sqrt{2}}{2}, n, \binom{d}{2}\right) \leq 2y_{\max}^2(3+2\sqrt{2}(1+y_{\max}))\sqrt{t}. \tag{EC.79}$$

When $t = \left(\frac{\kappa}{64sy_{\max}^2(3+2\sqrt{2}(1+y_{\max}))}\right)^2 = \left(\kappa/256sx_{\max}^2(3+2\sqrt{2}(1+2x_{\max})\right)^2$, we have

$$2y_{\max}^2(3+2\sqrt{2}(1+y_{\max}))\sqrt{t} = \frac{\kappa}{32s}. \tag{EC.80}$$

Via (EC.78), (EC.79) and (EC.80), with probability $1-\exp(\frac{1}{2}nK(K-1)t)$ we have

$$\left\|\frac{2}{nK(K-1)}\sum_i\sum_{k_1<k_2} y_{k_1,k_2,i}y_{k_1,k_2,i}^T - \mathbb{E}[y_{k_1,k_2,i}y_{k_1,k_2,i}^T]\right\|_\infty \leq \frac{\kappa}{23s}. \tag{EC.81}$$

Then, via the Corollary 6.8 in Bühlmann and Van De Geer (2011), when Assumption A.2 holds, then (EC.81) leads the following result

$$\begin{aligned}\|u_S\|_1^2 &\leq \frac{s}{\kappa/2}u^T\left[\frac{2}{nK(K-1)}\sum_i\sum_{k_1<k_2} y_{k_1,k_2,i}y_{k_1,k_2,i}^T\right]u \\ \Rightarrow \|u_S\|_1^2 &\leq \frac{4s}{K(K-1)\kappa}u^T\nabla^2 L(\xi)u,\end{aligned} \tag{EC.82}$$

where last inequality uses (EC.76).

Hence, when choosing $C = \frac{1}{2}K(K-1)\left(\kappa/256sx_{\max}^2(3+2\sqrt{2}(1+2x_{\max})\right)^2$, then with probability $1-\exp(-Cn)$, we have

$$u^T\nabla^2(\xi|x,\mathcal{A})u \geq \frac{K(K-1)\kappa}{4s}\|u_S\|_1^2,$$

where $\|u_{S^C}\|_1 \leq 3\|u_S\|$. The remaining of this lemma follows directly by using $n>\log T/C$.

EC.2.4. Proof of Lemma EC.4

We start with showing that the expectation of $\nabla\log(p_{\beta^*,\mathcal{A}}(j))$ among all possible choice $j\in\mathcal{A}$ is 0, i.e., $\mathbb{E}_{j\in\mathcal{A}}[\nabla\log(p_{\beta^*,\mathcal{A}}(j))]=0$.

$$\begin{aligned}\mathbb{E}_{j\mathcal{A}}[\nabla\log(p_{\beta^*,\mathcal{A}}(j))] &= \frac{1}{n}\sum_{j\in\mathcal{A}_t} p_{\beta^*,\mathcal{A}}(j)\nabla\log(p_{\beta^*,\mathcal{A}}(j)) \\ &= \sum_{j\in\mathcal{A}_t} p_{\beta^*,\mathcal{A}}(j)\cdot\frac{1}{p_{\beta^*,\mathcal{A}}(j)}\nabla p_{\beta^*,\mathcal{A}}(j) \\ &= \sum_{j\in\mathcal{A}_t}\nabla p_{\beta^*,\mathcal{A}}(j) \\ &= \sum_{j\in\mathcal{A}_t}\nabla\left(\frac{\exp(x_j^T\beta^*)}{\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)}\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\in\mathcal{A}_t}\left(\frac{\exp(x_j^T\beta^*)x_j}{\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)} - \frac{\exp(x_j^T\beta^*)\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)x_i}{(\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*))^2}\right)\\
&= \frac{\sum_{j\in\mathcal{A}_t}\exp(x_j^T\beta^*)x_j}{\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)} - \frac{\sum_{j\in\mathcal{A}_t}\exp(x_j^T\beta^*)\cdot\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)x_i}{(\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*))^2}\\
&= \frac{\sum_{j\in\mathcal{A}_t}\exp(x_j^T\beta^*)x_j}{\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)} - \frac{\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)x_i}{\sum_{i\in\mathcal{A}_t}\exp(x_i^T\beta^*)}\\
&= 0
\end{aligned}
$$

Combining with the fact that $\nabla\log(p_{\beta^*,\mathcal{A}}(j))$ is element-wise bounded by $x_{\max}$, we can conclude that every dimension of $\nabla\log(p_{\beta^*,\mathcal{A}}(j))$ is a zero mean $x_{\max}^2$-subgaussian random variable and that $\nabla L(\beta)$ is the finite average of zero mean i.i.d subgaussian random vector, i.e.,$\nabla L(\beta^*) = \frac{1}{n}\sum_t \nabla\log(p_{\beta^*,\mathcal{A}}(c_t))$.

Via Hoeffding inequality, for any $\epsilon > 0$, we have

$$
\mathbb{P}\left(\left|\sum_t \nabla_i\log(p_{\beta^*,\mathcal{A}}(c_t))\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2nx_{\max}^2}\right), \tag{EC.83}
$$

where $\nabla_i\log(q_{\beta^*}(c_t))$ is the i-th dimension of $\nabla\log(q_{\beta^*}(c_t))$. Hence, via union bound, we have

$$
\begin{aligned}
&\mathbb{P}\left(\left\|\sum_t \nabla\log(p_{\beta^*,\mathcal{A}}(c_t))\right\|_\infty \geq \epsilon\right) \leq 2d\exp\left(-\frac{\epsilon^2}{2nx_{\max}^2}\right)\\
\Rightarrow &\mathbb{P}\left(\|\nabla L(\beta^*|x,\mathcal{A})\|_\infty \geq \epsilon/n\right) \leq 2\exp\left(-\frac{\epsilon^2}{2nx_{\max}^2} + \log(d)\right).
\end{aligned}
$$

If we set $\epsilon = \sqrt{2nx_{\max}^2(\log d + \log T)}$, then

$$
\begin{aligned}
&\Rightarrow\mathbb{P}\left(\|\nabla L(\beta^*|x,\mathcal{A})\|_\infty \geq \frac{\sqrt{2nx_{\max}^2(\log d+\log T)}}{n}\right) \leq \frac{2}{T}\\
&\Rightarrow\mathbb{P}\left(\|\nabla L(\beta^*|x,\mathcal{A})\|_\infty \geq \sqrt{\frac{2x_{\max}^2(\log d+\log T)}{n}}\right) \leq \frac{2}{T}.
\end{aligned}
$$

EC.2.5. Proof of Lemma EC.5

By standard covering number arguments (e.g., van de Geer, 2000), an ϵ covering set $\mathcal{H}(\epsilon)$ for $\|\theta-\theta_0\| \leq \delta$ has a finite elements upper bounded by $\exp((|\mathcal{S}|+m)\log(3\delta/\epsilon))$. As we require event $\mathcal{E}_2(m,T)$, $\delta \leq \frac{\rho}{8Kx_{\max}}$ and $\mathcal{G}_1(m,T,\epsilon) \leq \frac{\rho}{8Kx_{\max}}$, via Lemma EC.6 and the union bound, any $\theta\in\mathcal{H}(\epsilon)$, we can show that the following inequality holds with probability $1-\delta_4\cdot\exp((|\mathcal{S}|+m)\log(3\delta/\epsilon))$:

$$
\left|\sum_{t=1}^T[\hat{f}_t(\theta) - f_{\mathcal{A}_t}(\theta)]\right| \leq \frac{2}{3}\log(1/\delta_4) + 3\sqrt{\log(1/\delta_4)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)} \tag{EC.84}
$$

If we set $\delta_4 = \exp(-(|\mathcal{S}|+m)\log(3\delta/\epsilon) - \log T)$ and $\epsilon = \frac{1}{2}\delta$, the above inequality directly suggests that the following result holds with probability $1-T^{-1}$:

$$
\begin{aligned}
\left|\sum_{t=1}^T[\hat{f}_t(\theta) - f_{\mathcal{A}_t}(\theta)]\right| &\leq \frac{2}{3}((|\mathcal{S}|+m)\log(6)+\log(T)) + 3\sqrt{((|\mathcal{S}|+m)\log(6)+\log(T))\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)}\\
&\leq \frac{2}{3}(2(|\mathcal{S}|+m)+1)\log(T) + 3\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)},
\end{aligned} \tag{EC.85}
$$

where last inequality we uses $T \geq 2$ in Lemma statement.

Next, we will bound the term $|\sum_t f_{\mathcal{A}_t}(\hat{\theta})|$:

$$\begin{aligned}
|\sum_t f_{\mathcal{A}_t}(\hat{\theta})| &= \sum_t f_{\mathcal{A}_t}(\hat{\theta}) \\
&= \sum_t f_{\mathcal{A}_t}(\hat{\theta}) - \sum_t \hat{f}_t(\hat{\theta}) + \sum_t \hat{f}_t(\hat{\theta}) \\
&\leq |\sum_t \hat{f}_t(\hat{\theta}) - \sum_t f_{\mathcal{A}_t}(\hat{\theta})| + \max\left\{0, \sum_t \hat{f}_t(\hat{\theta})\right\}, \qquad \text{(EC.86)}
\end{aligned}$$

where the first equality uses the fact that $f_{\mathcal{A}_t}(\cdot) \geq 0$.

Let $\Gamma_T := \max\left\{0, \sum_t \hat{f}_t(\hat{\theta})\right\}$ and $x = \sqrt{\sum_t f_{\mathcal{A}_t}(\hat{\theta})}$. We combine (EC.85) and (EC.86) to show that x^2, or equivalently $|\sum_t f_{\mathcal{A}_t}(\hat{\theta})|$, can be bounded as follows with probability $1 - \mathcal{O}(T^{-1})$:

$$\begin{aligned}
x^2 &\leq \frac{2}{3}(2(|\mathcal{S}|+m)+1)\log(T) + 3\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} \cdot x + \Gamma_T \\
\Rightarrow x^2 - 3\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} \cdot x &- (\Gamma_T + \frac{2}{3}(2(|\mathcal{S}|+m)+1)\log(T)) \leq 0. \qquad \text{(EC.87)}
\end{aligned}$$

Note that the inequality (EC.87) can be viewed as a quadratic function in x. Hence, we can solve for the upper bound of x:

$$\begin{aligned}
x &\leq \frac{3\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} + \sqrt{9(2(|\mathcal{S}|+m)+1)\log(T) + 4(\Gamma_T + \frac{2}{3}(2(|\mathcal{S}|+m)+1)\log(T))}}{2} \\
&= \frac{3\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} + \sqrt{(9+8/3)(2(|\mathcal{S}|+m)+1)\log(T) + 4\Gamma_T}}{2} \\
&< \frac{3\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} + 4\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} + 2\sqrt{\Gamma_T}}{2} \qquad \text{(EC.88)} \\
&\leq \frac{7}{2}\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} + \sqrt{\Gamma_T} \\
\Rightarrow \sqrt{\sum_t f_{\mathcal{A}_t}(\hat{\theta})} &\leq 4\sqrt{(2(|\mathcal{S}|+m)+1)\log(T)} + \sqrt{\Gamma_T} \\
\Rightarrow \sum_t f_{\mathcal{A}_t}(\hat{\theta}) &\leq 32(2(|\mathcal{S}|+m)+1)\log(T) + 2\Gamma_T, \qquad \text{(EC.89)}
\end{aligned}$$

where in (EC.88) we first enlarge $(9+8/3)$ to 16 and then uses the fact that $\sqrt{a^2+b^2} \leq a+b$ for $a, b \geq 0$, and in (EC.89) we uses the fact that $(a+b)^2 \leq 2a^2 + 2b^2$ for all $a, b \in \mathbb{R}$. The remaining part follows by combing above results with Lemma EC.7.

EC.2.6. Lemma EC.6

LEMMA EC.6. *Denote the empirical version* $\hat{f}_t(\theta) = \log(p_{Q^T P_0^T \theta}(c_t)/p_{\beta^*}(c_t))$ *for* $t > 0$. *If event* $\mathcal{E}_2(m,T)$ *holds and* $\max\{\|\theta - P_0 Q\beta^*\|, \mathcal{G}_1(m,T,\epsilon)\} \leq \frac{\rho}{8K x_{\max}}$, *then with probability* $1-\delta_4$ *we have*

$$\left|\sum_t [\hat{f}_t(\theta) - f_{\mathcal{A}_t}(\theta)]\right| \leq \frac{2}{3}\log(1/\delta_4) + 3\sqrt{\log(1/\delta_4)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)}. \qquad \text{(EC.90)}$$

Proof. We can construct a Doob's martingale $\{M(i), i = 0, 1, 2, ..., T\}$ as follow

$$M(i) = \mathbb{E}\left[\sum_t \hat{f}_t(\theta)|\mathcal{H}_i\right], i = 1, 2, ...T \qquad \text{(EC.91)}$$

Using Bernstein's inequality, we can show that for $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(|M(T) - M(0)| \geq t\right) \leq \exp\left(-\frac{t^2}{2k + 2t/3}\right) \\
\Rightarrow &\mathbb{P}\left(\left|\sum_t [\hat{f}_t(\theta) - f_{\mathcal{A}_t}(\theta)]\right| \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2k + 2\epsilon/3}\right),
\end{aligned} \tag{EC.92}$$

where $k \geq \sum_{i=1}^{T} Var[M(i) - M(i-1)|\mathcal{H}_{i-1}]$. Next, we will upper bound k.
First, we show that the mean different is zero:

$$\begin{aligned}
&\mathbb{E}[M(i) - M(i-1)|\mathcal{H}_{i-1}] \\
&= \mathbb{E}\left[\mathbb{E}[\sum_t \hat{f}_t(\theta)|\mathcal{H}_i] - \mathbb{E}[\sum_t \hat{f}_t(\theta)|\mathcal{H}_{i-1}]\right] \\
&= \mathbb{E}[\sum_t \hat{f}_t(\theta)] - \mathbb{E}[\sum_t \hat{f}_t(\theta)] = 0.
\end{aligned}$$

As $\mathbb{E}[M(i) - M(i-1)|\mathcal{H}_{i-1}] = 0$, we can show that

$$\begin{aligned}
&Var[M(i) - M(i-1)|\mathcal{H}_{i-1}] \\
&= \mathbb{E}[(M(i) - M(i-1)|\mathcal{H}_{i-1})^2] \\
&= \mathbb{E}[(\hat{f}_i(\theta) - f_{\mathcal{A}_i}(\theta))^2] \\
&= \mathbb{E}[\hat{f}_i(\theta)^2] - f_{\mathcal{A}_i}(\theta)^2,
\end{aligned}$$

where the second-to-last equality follows from the fact that $\mathbb{E}[\hat{f}_i(\theta)] = f_{\mathcal{A}_t}(\theta)$. As we have $2x^2 > \log^2(1+x)$ for all $x > -1/2$. Then, when $(p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j))/(p_{\beta^*}(j)) \geq -1/2$ holds for all $j \in \mathcal{A}_t$, we have

$$\begin{aligned}
\mathbb{E}[\hat{f}_t(\theta)^2] &= \sum_{j \in \mathcal{A}_t} p_{\beta^*}(j) \left(\log\left(\frac{p_{Q^T P_0^T \theta}(j)}{p_{\beta^*}(j)}\right)\right)^2 \\
&= \sum_{j \in \mathcal{A}_t} p_{\beta^*}(j) \left(\log\left(1 + \frac{p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)}{p_{\beta^*}(j)}\right)\right)^2 \\
&\leq 2 \sum_{j \in \mathcal{A}_t} \frac{(p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j))^2}{p_{\beta^*}(j)}
\end{aligned}$$

In addition, we can also show that

$$\begin{aligned}
f_{\mathcal{A}_t}(\theta) &= \sum_{j \in \mathcal{A}_t} p_{\beta^*}(j) \log\left(\frac{p_{Q^T P_0^T \theta}(j)}{p_{\beta^*}(j)}\right) \\
&= \sum_{j \in \mathcal{A}_t} p_{\beta^*}(j) \left(\frac{p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)}{p_{\beta^*}(j)} - \frac{1}{2(\xi_j)^2}\left(\frac{p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)}{p_{\beta^*}(j)}\right)^2\right) &\text{(EC.93)} \\
&= \sum_{j \in \mathcal{A}_t} (p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)) + \sum_j \frac{1}{2(\xi_j)^2} \frac{(p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j))^2}{p_{\beta^*}(j)} &\text{(EC.94)} \\
&= \sum_{j \in \mathcal{A}_t} \frac{1}{2(\xi_j)^2} \frac{(p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j))^2}{p_{\beta^*}(j)^2} \\
&\geq \frac{1}{2\max_j(\xi_j)^2} \sum_{j \in \mathcal{A}_t} p_{\beta^*}(j) \frac{(p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j))^2}{p_{\beta^*}(j)^2}
\end{aligned}$$

$$\geq \frac{1}{2\max_j(\xi_j)^2}\mathbb{E}[\hat{f}_t(\theta)^2], \tag{EC.95}$$

where (EC.93) uses the Taylor's expansion of $\log(a)$ at $a=1$ and (EC.94) uses $\sum_j p_{\beta^*}(j) = \sum_j p_{Q^T P_0^T \theta}(j) = 1$. We then analyze the value of ξ_j. Since we expend the $\log(a)$ function at $a=1$, there exists an α_j such that

$$\begin{aligned}\xi_j &= \alpha_j + (1-\alpha_j)\frac{p_{Q^T P_0^T \theta}(j)}{p_{\beta^*}(j)} \\ &= 1 + (1-\alpha)\frac{p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)}{p_{\beta^*}(j)}.\end{aligned}$$

Claim: If $\max\{\|\theta - P_0 Q\beta^*\|, \mathcal{G}_1(m,T,\epsilon)\} \leq \frac{\rho}{8Kx_{\max}}$, then $-\frac{1}{2} \leq \frac{p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)}{p_{\beta^*}(j)} \leq \frac{1}{2}$ for all j.

$$\begin{aligned}|p_{\beta^*}(j) - p_{Q^T P_0^T \theta}(j)| &= |p_{\beta^*}(j) - p_{\Sigma\beta^*}(j) + p_{\Sigma\beta^*}(j) - p_{Q^T P_0^T \theta}(j)| \\ &\leq |p_{\Sigma\beta^*}(j) - p_{Q^T P_0^T \theta}(j)| + |p_{\beta^*}(j) - p_{\Sigma\beta^*}(j)| \\ &\leq \|\nabla p_{\nu_1}(j)\|\|Q^T P_0\theta - \Sigma\beta^*\| + \|\nabla p_{\nu_2}(j)\|\|\beta^* - \Sigma\beta^*\|, \end{aligned}\tag{EC.96}$$

where ν_1 and ν_2 are in between $\{Q^T P_0\theta, \Sigma\beta^*\}$ and $\{\beta^*, \Sigma\beta^*\}$ respectively. We now analyze the upper bound of $\|\nabla p_\xi(j)\|$ as follow

$$\begin{aligned}\nabla p_\nu(j) &= \frac{\exp(x_j^T\nu)x_j}{\sum\exp(x^T\nu)} - \frac{\exp(x_j^T\nu)\sum\exp(x^T\nu)x}{(\sum\exp(x^T\nu))^2} \\ &= p_\nu(j)\sum_{i\in\mathcal{A}_t}(x_j - p_\nu(i)x_i) \\ \Rightarrow \|\nabla p_\nu(j)\| &\leq 2Kx_{\max}, \end{aligned}\tag{EC.97}$$

Combining (EC.96), (EC.97), event $\mathcal{E}_2(m,T)$ and $\max\{\|\theta - P_0Q\beta^*\|, \mathcal{G}_1(m,T,\epsilon)\} \leq \frac{\rho}{8Kx_{\max}}$, then we have

$$\begin{aligned}|p_{\beta^*}(j) - p_{Q^T P_0^T \theta}(j)| &= | \leq 2Kx_{\max}\|\theta - P_0Q\beta^*\| + 2Kx_{\max}\|(\Sigma - I)\beta^*\| \\ &\leq 2Kx_{\max}\cdot\frac{\rho}{8Kx_{\max}} + 2Kx_{\max}\mathcal{G}_1(m,T,\epsilon) \\ &\leq \frac{\rho}{4} + 2Kx_{\max}\frac{\rho}{8Kx_{\max}} \\ &\leq \frac{\rho}{2} \leq \frac{1}{2}p_{\beta^*}(j) \\ \Rightarrow \frac{|p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)|}{p_{\beta^*}(j)} &\leq \frac{1}{2} \\ \Rightarrow -\frac{1}{2} \leq \frac{p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j)}{p_{\beta^*}(j)} &\leq \frac{1}{2}, \end{aligned}$$

which proves the Claim. Hence, if we have $(p_{Q^T P_0^T \theta}(j) - p_{\beta^*}(j))/p_{\beta^*}(j) \leq 1/2$, then we can show that

$$\xi_j \leq 1 + \frac{1}{2} = \frac{3}{2} \Rightarrow 2\max_j(\xi)^2 \leq \frac{9}{2}.$$

Therefore, if we set k as follows:

$$k = \sum_{i=1}^{T}\frac{9}{2}f_{\mathcal{A}_t}(\theta) \geq \sum_{i=1}^{T}\mathbb{E}[\hat{f}_i(\theta)] \geq \sum_{i=1}^{T}Var[M(i) - M(i-1)|\mathcal{H}_{i-1}]$$

then from EC.92, we have

$$\mathbb{P}\left(\left|\sum_t [\hat{f}_t(\theta) - f_{\mathcal{A}_t}(\theta)]\right| \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{9\sum_{i=1}^T f_{\mathcal{A}_t}(\theta) + 2\epsilon/3}\right). \tag{EC.98}$$

Finally, to ensure $\mathbb{P}\left(\left|\sum_t [\hat{f}_t(\theta) - f_{\mathcal{A}_t}(\theta)]\right| \geq \epsilon\right) \leq \delta$, we can set $\delta_4 = \exp\left(-\frac{\epsilon^2}{9\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)+2\epsilon/3}\right)$. We then solve for the ϵ.

$$\begin{aligned}
\delta_4 &= \exp\left(-\frac{\epsilon^2}{9\sum_{i=1}^T f_{\mathcal{A}_t}(\theta) + 2\epsilon/3}\right) \\
\log(1/\delta_4) &= \frac{\epsilon^2}{9\sum_{i=1}^T f_{\mathcal{A}_t}(\theta) + 2\epsilon/3} \\
\log(1/\delta_4)(9\sum_{i=1}^T f_{\mathcal{A}_t}(\theta) + 2\epsilon/3) &= \epsilon^2 \\
\epsilon^2 - \frac{2}{3}\log(1/\delta_4)\epsilon - 9\log(1/\delta_4)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta) &= 0 \\
\Rightarrow \frac{\frac{2}{3}\log(1/\delta_4) + \sqrt{(\frac{2}{3}\log(1/\delta_4))^2 + 36\log(1/\delta_4)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)}}{2} &= \epsilon \\
\Rightarrow \frac{\frac{2}{3}\log(1/\delta_4) + \frac{2}{3}\log(1/\delta_4) + 6\sqrt{\log(1/\delta_4)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)}}{2} &\geq \epsilon \\
\Rightarrow \frac{2}{3}\log(1/\delta_4) + 3\sqrt{\log(1/\delta_4)\sum_{i=1}^T f_{\mathcal{A}_t}(\theta)} &\geq \epsilon.
\end{aligned}$$

The Lemma follows directly by plugging the last inequality back to EC.98.

EC.2.7. Lemma EC.7

LEMMA EC.7. *Let $\delta = \|\theta - P_0 Q\beta^*\|$ and 2 be a positive constant. If events $\mathcal{E}_2(m,T)$ and $\mathcal{E}_{rp}(m,d,1/2)$ hold and $\delta \leq \frac{3}{4}\frac{n_T\mu}{TL_3}$, then the following inequality holds*

$$\begin{aligned}
\sum_t f_{\mathcal{A}_t}(\theta) \geq \frac{1}{4}(\theta - P_0Q\beta^*)^T \left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*) \\
- 4x_{\max}^2 T\mathcal{G}_1(m,T,\epsilon)\delta.
\end{aligned} \tag{EC.99}$$

Proof. We first expend $\sum_t f_{\mathcal{A}_t}(\theta)$ at $P_0Q\beta^*$.

$$\begin{aligned}
\sum_t f_{\mathcal{A}_t}(\theta) \geq \overbrace{\sum_t f_{\mathcal{A}_t}(P_0Q\beta^*)}^{ⓐ} + \overbrace{\sum_t \nabla f_{\mathcal{A}_t}(P_0Q\beta^*)^T(\theta - P_0Q\beta^*)}^{ⓑ} \\
+ \overbrace{\frac{1}{2}(\theta - P_0Q\beta^*)^T\left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*) - \frac{1}{6}\sum_t L_3\|\theta - P_0Q\beta^*\|^3}^{ⓒ}.
\end{aligned} \tag{EC.100}$$

We will derive the lower bounds for these three parts in equation (EC.100).

Lower bound for ⓐ. Since $\sum_t f_{\mathcal{A}_t}(P_0Q\beta^*)$ can be viewed as KL-Divergence, we have

$$f_{\mathcal{A}_t}(P_0Q\beta^*) \geq 0. \tag{EC.101}$$

Lower bound for ⓑ. Via Cauchy inequality we have:

$$\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)^T(\theta - P_0Q\beta^*) \geq -\|\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\|\|\theta - P_0Q\beta^*\| = -\|\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\|\delta. \tag{EC.102}$$

The remaining task is to find the bound for $\|\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\|$.

Claim: $\|\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\| \leq 6K^2x_{\max}^2\mathcal{G}_1(m,T,\epsilon)$ holds for all t.

$$\begin{aligned}
\nabla f_{\mathcal{A}_t}(P_0Q\beta^*) &= \sum_{j\in\mathcal{A}_t} \frac{p_{\beta^*(j)}}{p_{\Sigma\beta^*}(j)} \nabla p_{\Sigma\beta^*}(j) \\
&= \sum_{j\in\mathcal{A}_t} \frac{p_{\beta^*(j)}}{p_{\Sigma\beta^*}(j)} \frac{(\sum \exp(x^T\Sigma\beta^*))\exp(x_j^T\Sigma\beta^*)P_0Qx_j}{(\sum\exp(x^T\Sigma\beta^*))^2} \\
&- \sum_{j\in\mathcal{A}_t} \frac{p_{\beta^*(j)}}{p_{\Sigma\beta^*}(j)} \frac{\exp(x_j^T\Sigma\beta^*)\sum\exp(x^T\Sigma\beta^*)P_0Qx}{(\sum\exp(x^T\Sigma\beta^*))^2} \\
&= \sum_{j\in\mathcal{A}_t} \frac{p_{\beta^*(j)}}{p_{\Sigma\beta^*}(j)} p_{\Sigma\beta^*}(j)P_0Qx_j \\
&- \sum_{j\in\mathcal{A}_t} \frac{p_{\beta^*(j)}}{p_{\Sigma\beta^*}(j)} \frac{p_{\Sigma\beta^*}(j)\sum\exp(x^T\Sigma\beta^*)P_0Qx}{\sum\exp(x^T\Sigma\beta^*)} \\
&= \sum_{j\in\mathcal{A}_t} p_{\beta^*(j)}P_0Qx_j - \sum_{j\in\mathcal{A}_t} p_{\beta^*(j)} \sum_{j\in\mathcal{A}_t} p_{\Sigma\beta^*}(j)P_0Qx_j \\
&= \sum_{j\in\mathcal{A}_t} p_{\beta^*(j)} \left(P_0Qx_j - \sum_{i\in\mathcal{A}_t} p_{\Sigma\beta^*}(i)P_0Qx_i\right) \\
&= \sum_{j\in\mathcal{A}_t} (p_{\beta^*}(j) - p_{\Sigma\beta^*}(j))P_0Qx_j \\
&= \sum_{j\in\mathcal{A}_t} \nabla p_\xi(j)^T(\beta^* - \Sigma\beta^*)P_0Qx_j,
\end{aligned} \tag{EC.103}$$

where ξ is on the line between β^* and $\Sigma\beta^*$.

We can show that

$$\begin{aligned}
\nabla p_\xi(j) &= \frac{\exp(x_j^T\xi)x_j}{\sum\exp(x^T\xi)} - \frac{\exp(x_j^T\xi)\sum\exp(x^T\xi)x}{(\sum\exp(x^T\xi))^2} \\
&= p_\xi(j)\sum_{i\in\mathcal{A}_t}(x_j - p_\xi(i)x_i) \\
\Rightarrow \|\nabla p_\xi(j)\| &\leq 2Kx_{\max},
\end{aligned}$$

where we use the fact that $0 \leq p_\xi(j) \leq 1$ for all $j \in \mathcal{A}_t$. Via event $\mathcal{E}_2(m,T)$, we have

$$\begin{aligned}
\|\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\| = \|\sum_{j\in\mathcal{A}_t} \nabla p_\xi(j)\,(\beta^* - \Sigma\beta^*)\,P_0Qx_j\| &\leq K\cdot 2Kx_{\max}\cdot\|\beta^* - \Sigma\beta^*\| \max_{j\in\mathcal{A}_t}\|P_0Qx_j\| \\
&= 2K^2x_{\max}\mathcal{G}_1(m,T,\epsilon)\max_{j\in\mathcal{A}_t}\|P_0Qx_j\|.
\end{aligned}$$

To prove the Claim, we finally need to bound $\|P_0Qx_j\|$ with $\|x_j\|$. Let $\tilde{x}_j = Qx_j$.

$$\begin{aligned}
\|P_0Qx_j\| &= \|P_0\tilde{x}_j\| \\
&= \left\| \begin{pmatrix} I & \\ & P \end{pmatrix} \begin{pmatrix} \tilde{x}_{j1} \\ \tilde{x}_{j2} \end{pmatrix} \right\| \\
&= \left\| \begin{pmatrix} \tilde{x}_{j1} \\ P\tilde{x}_{j2} \end{pmatrix} \right\| \leq \|\tilde{x}_{j1}\| + \|P\tilde{x}_{j2}\|.
\end{aligned}$$

As event $\mathcal{E}_{rp}(m,d,1/2)$ holds, we have $\|P\tilde{x}_{j2}\| \le (1+1/2)\|\tilde{x}_{j2}\| \le 2\|\tilde{x}_{j2}\|$. Combining this result with the fact that Q is a permutation matrix that won't change the scale of x_j, we have

$$\|P_0 Q x_j\| \le 3x_{\max}$$

holds with high probability. Therefore, we have

$$\|\nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\| \le 2Kx_{\max}\mathcal{G}_1(m,T,\epsilon)\max_{j\in\mathcal{A}_t}\|P_0Qx_j\| \le 6K^2x_{\max}^2\mathcal{G}_1(m,T,\epsilon),$$

which proves the Claim.

Thus, applying this Claim for all t, we have

$$\begin{aligned}
&\|\sum_t \nabla f_{\mathcal{A}_t}(P_0Q\beta^*)\| \le 6TK^2x_{\max}^2\mathcal{G}(T)\\
\Rightarrow &\sum_t \nabla f_{\mathcal{A}_t}(P_0Q\beta^*)^T(\theta - P_0Q\beta^*) \ge -6TK^2x_{\max}^2\mathcal{G}(T)\delta. \qquad \text{(EC.104)}
\end{aligned}$$

Bound for © By Lemma EC.8 and Assumption A.3, with probability $1-\mathcal{O}(1/T)$ we have

$$\begin{aligned}
&\frac{1}{2}(\theta - P_0Q\beta^*)^T\left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*)\\
&\ge \frac{1}{2}(\theta - P_0Q\beta^*)^T\left(\sum_{t\in\mathcal{W}_R} \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*)\\
&\ge \frac{1}{4}\mu n_T\|\theta - P_0Q\beta^*\|^2. \qquad \text{(EC.105)}
\end{aligned}$$

Since we require $\delta \le \frac{3}{4}\frac{n_T\mu}{TL_3}$, we can further show that

$$\begin{aligned}
&\frac{1}{6}L_3\|\theta - P_0Q\beta^*\|^3 \le \frac{n_T\mu}{8T}\|\theta - P_0Q\beta^*\|^2\\
&\le \frac{1}{4T}(\theta - P_0Q\beta^*)^T\left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*)\\
\Rightarrow &\frac{1}{2}(\theta - P_0Q\beta^*)^T\left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*) - \frac{1}{6}\sum_t L_3\|\theta - P_0Q\beta^*\|^3\\
&\ge \frac{1}{4}(\theta - P_0Q\beta^*)^T\left(\sum_t \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)\right)(\theta - P_0Q\beta^*). \qquad \text{(EC.106)}
\end{aligned}$$

EC.2.8. Lemma EC.8

LEMMA EC.8. *Under Assumption A.3, for all feasible θ in the projected space, if $n_T = \mathcal{O}(2(|\mathcal{S}|+m)^2(\log T - \log(|\mathcal{S}|+m)))$, then with probability at least $1-\mathcal{O}(1/T)$, we have*

$$\sum_{t=1}^T \nabla^2 f(\theta) \succeq \frac{1}{2}\mu n_T I.$$

Proof. Since $\nabla^2 f(\hat{\theta})$ is always positive semidefinite, we will have

$$\sum_{t=1}^T \nabla^2 f(\hat{\theta}) \succeq \sum_{i\in\mathcal{W}_R} \nabla^2 f(\hat{\theta}).$$

From Lemma EC.10, we have

$$\nabla^2 f(\hat{\theta}) = \tilde{\mathbb{E}}\left[\left(z-\tilde{\mathbb{E}}[z]\right)\left(z-\tilde{\mathbb{E}}[z]\right)^T\right] \preceq \tilde{\mathbb{E}}[zz^T] \preceq (|\mathcal{S}|+m)z_{\max}^2 I. \tag{EC.107}$$

Combining this inequality with the fact that 1) $i \in \mathcal{W}_R$ are i.i.d random sample and 2) $\nabla^2 f(\hat{\theta})$ is always positive semidefinite, we can use the Matrix Chernoff inequalities to show that

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{i\in\mathcal{W}_R}^{T}\nabla^2 f(\hat{\theta})\right) \le (1-\delta)\mu_{\min}\right) \le (|\mathcal{S}|+m)\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}/R}, \tag{EC.108}$$

where $\mu_{\min} \le \lambda_{\min}\left(\sum_{i\in\mathcal{W}_R}^{T}\mathbb{E}[\nabla^2 f(\hat{\theta})]\right)$ and $R \ge \lambda_{\max}(\nabla^2 f(\hat{\theta}))$. Under assumption A.3, we can show that $\lambda_{\min}\left(\sum_{i\in\mathcal{W}_R}^{T}\mathbb{E}[\nabla^2 f(\hat{\theta})]\right) \ge n_T\mu$. Based on (EC.107), we can show that $\nabla^2 f(\hat{\theta}) \preceq (|\mathcal{S}|+m)z_{\max}^2 I$. Hence, we set $\mu_{\min} = n_T\mu$ and $R = (|\mathcal{S}|+m)z_{\max}^2$.

Moreover, if we pick $\delta = 1/2$, we then have

$$\begin{aligned}
\mathbb{P}\left(\lambda_{\min}\left(\sum_{i\in\mathcal{W}_R}^{T}\nabla^2 f(\hat{\theta})\right) \le \frac{1}{2}n_T\mu\right) &\le (|\mathcal{S}|+m)\left(\frac{e^{-1/2}}{(1/2)^{1/2}}\right)^{n_T\mu/(|\mathcal{S}|+m)z_{\max}^2} \\
&= (|\mathcal{S}|+m)\left(\frac{e}{2}\right)^{-n_T\mu/2(|\mathcal{S}|+m)z_{\max}^2} \\
&= (|\mathcal{S}|+m)\left(\frac{e}{2}\right)^{-\frac{n_T\mu}{2(|\mathcal{S}|+m)z_{\max}^2}}.
\end{aligned}$$

The remaining part of this lemma follows directly by using $n_T \gtrsim \frac{2(|\mathcal{S}|+m)z_{\max}^2(\log T - \log(|\mathcal{S}|+m))}{\mu\log(e/2)}$.

EC.2.9. Lemma EC.9

LEMMA EC.9. *Let* $B_t = (\sum_i^t \nabla^2 f_{\mathcal{A}_i}(P_0Q\beta^*))^{-1/2}\nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*)(\sum_t \nabla^2 f(P_0Q\beta^*))^{-1/2}$. *Under Assumptions A.1, A.2, and A.3, with probability* $1-\mathcal{O}(1/T)$, *we have*

$$\sum_{t=T_0+1}^{T}\min\{1, \|B_t\|_{op}\} \le 2(|\mathcal{S}|+m)\log\left(\frac{8TK^2x_{\max}^2}{\mu n_{T_0}}\right). \tag{EC.109}$$

Proof. Denote the eigenvalue as $\sigma_1(B_t) \ge \sigma_2(B_T) \ge ... \ge 0$ and we can show that

$$\begin{aligned}
\sum_{t=T_0+1}^{T}\nabla^2 f(P_0Q\beta^*) &= \sum_{t=T_0+1}^{T-1}\nabla^2 f(P_0Q\beta^*) + \nabla^2 f_{\mathcal{A}_t}(P_0Q\beta^*) \\
&= (\sum_{t=T_0+1}^{T-1}\nabla^2 f(P_0Q\beta^*))^{1/2}(I+B_T)(\sum_{t=T_0+1}^{T-1}\nabla^2 f(P_0Q\beta^*))^{1/2} \\
\Rightarrow \log\left(\frac{\det\sum_{t=T_0+1}^{T}\nabla^2 f(P_0Q\beta^*)}{\det\sum_{t=T_0+1}^{T-1}\nabla^2 f(P_0Q\beta^*)}\right) &= \sum_j \log(1+\sigma_j(B_T)) \ge \log(1+\sigma_1(B_T)),
\end{aligned}$$

Together with the observation that $2\log(1+x) \ge x$ for $x \in (0,1]$, we can show the following inequality holds:

$$\begin{aligned}
\min\{1, \|B_T\|_{op}\} &\le 2\log(1+\|B_t\|_{op}) \\
&= 2\log(1+\sigma_1(B_t)) \\
&\le 2\log\left(\frac{\det\sum_{t=T_0+1}^{T}\nabla^2 f(P_0Q\beta^*)}{\det\sum_{t=T_0+1}^{T-1}\nabla^2 f(P_0Q\beta^*)}\right)
\end{aligned} \tag{EC.110}$$

Therefore, we can prove that

$$\sum_{t=T_0+1}^{T} \min\{1, \|B_t\|_{op}\} \le 2\log\left(\frac{\det\sum_{t=1}^{T}\nabla^2 f(P_0Q\beta^*)}{\det\sum_{t=1}^{T_0+1}\nabla^2 f(P_0Q\beta^*)}\right).$$

From Lemma EC.8, we can show with probability $1-\mathcal{O}(1/T)$, the following inequality holds:

$$\sum_{t}^{T}\nabla^2 f(P_0Q\beta^*) \succeq \frac{1}{2}\mu n_T.$$

Then, following the similar procedures as in (EC.76) in Lemma EC.3, we can show that

$$\sum_{t}^{T}\nabla^2 f(P_0Q\beta^*) \preceq 4TK^2x_{\max}^2.$$

Hence,

$$\begin{aligned} &2\log\left(\frac{\det\sum_{t=1}^{T}\nabla^2 f(P_0Q\beta^*)}{\det\sum_{t=1}^{T_0+1}\nabla^2 f(P_0Q\beta^*)}\right) \le 2(|\mathcal{S}|+m)\log\left(\frac{8TK^2x_{\max}^2}{\mu n_{T_0}}\right) \\ \Rightarrow &\sum_{t=T_0+1}^{T}\min\{1,\|B_t\|_{op}\} \le 2(|\mathcal{S}|+m)\log\left(\frac{8TK^2x_{\max}^2}{\mu n_{T_0}}\right). \end{aligned}$$

EC.2.10. Lemma EC.10

LEMMA EC.10. $\nabla^2 f_{\mathcal{A}}(\hat{\theta}) = \tilde{\mathbb{E}}\left[\left(z-\tilde{\mathbb{E}}[z]\right)\left(z-\tilde{\mathbb{E}}[z]\right)^T\right]$.

Proof. We first consider the gradient of $f_{\mathcal{A}}(\theta)$:

$$\nabla f_{\mathcal{A}}(\theta) = \sum_{j\in\mathcal{A}} p_{\Sigma\beta^*,\mathcal{A}}(j)\frac{\nabla p_{Q^TP_0^T\theta,\mathcal{A}}(j)}{p_{Q^TP_0^T\theta,\mathcal{A}}(j)}. \tag{EC.111}$$

Then, we compute the term $\nabla p_{Q^TP_0^T\theta,\mathcal{A}}(j)$:

$$\begin{aligned} \nabla p_{Q^TP_0^T\theta,\mathcal{A}}(j) &= \frac{(\sum_{i\in\mathcal{A}}\exp(z_i^T\theta)\exp(z_j^T\theta)z_j - \exp(z_j^T\theta)\sum_{i\in\mathcal{A}}\exp(z_i^T\theta)z_i}{(\sum_{i\in\mathcal{A}}\exp(z_i^T\theta))^2} \\ &= \frac{\exp(z_j^T\theta)z_j}{\sum_{i\in\mathcal{A}}\exp(z_i^T\theta)} - \frac{\exp(z_j^T\theta)}{\sum_{i\in\mathcal{A}}\exp(z_i^T\theta)}\cdot\sum_{i\in\mathcal{A}}\frac{\exp(z_i^T\theta)z_i}{\sum_{i\in\mathcal{A}}\exp(z_i^T\theta)} \\ &= p_{Q^TP_0^T\theta,\mathcal{A}}(j)z_j - p_{Q^TP_0^T\theta,\mathcal{A}}(j)\sum_{i\in\mathcal{A}}p_{Q^TP_0^T\theta,\mathcal{A}}(i)z_i, \end{aligned}$$

which implies that $\frac{\nabla p_{Q^TP_0^T\theta,\mathcal{A}}(j)}{p_{Q^TP_0^T\theta,\mathcal{A}}(j)} = z_j - \sum_{i\in\mathcal{A}}p_{Q^TP_0^T\theta,\mathcal{A}}z_i$. Combining it with (EC.111), we will have

$$\begin{aligned} \nabla f_{\mathcal{A}}(\theta) &= \sum_{j\in\mathcal{A}}p_{\Sigma\beta^*,\mathcal{A}}(j)\left(z_j - \sum_{i\in\mathcal{A}}p_{Q^TP_0^T\theta,\mathcal{A}}z_i\right) \\ &= \sum_{j\in\mathcal{A}}p_{\Sigma\beta^*,\mathcal{A}}(j)z_j - \sum_{j\in\mathcal{A}}p_{\Sigma\beta^*,\mathcal{A}}(j)\sum_{i\in\mathcal{A}}p_{Q^TP_0^T\theta,\mathcal{A}}z_i \\ &= \sum_{j\in\mathcal{A}}p_{\Sigma\beta^*,\mathcal{A}}(j)z_j - \sum_{i\in\mathcal{A}}p_{Q^TP_0^T\theta,\mathcal{A}}z_i. \end{aligned}$$

Therefore, we can show that

$$\nabla^2 f_{\mathcal{A}}(\theta) = \sum_{i\in\mathcal{A}}\nabla p_{Q^TP_0^T\theta,\mathcal{A}}(i)z_i^T$$

$$
\begin{aligned}
&= \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\theta,\mathcal{A}}(i) z_i z_i^T - \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\theta,\mathcal{A}}(i) \sum_{k\in\mathcal{A}} p_{Q^T P_0^T\theta,\mathcal{A}}(k) z_k z_i^T \\
&= \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\theta,\mathcal{A}}(i) z_i z_i^T - \sum_{k\in\mathcal{A}} p_{Q^T P_0^T\theta,\mathcal{A}}(k) z_k \sum_{i\in\mathcal{A}} p_{Q^T P_0^T\theta,\mathcal{A}}(i) z_i^T \\
&= \tilde{\mathbb{E}}[zz^T] - \tilde{\mathbb{E}}[z]\tilde{\mathbb{E}}[z^T] \\
&= \tilde{\mathbb{E}}\left[zz^T - \tilde{\mathbb{E}}[z]\tilde{\mathbb{E}}[z^T] + z\tilde{\mathbb{E}}[z^T] - \tilde{\mathbb{E}}[z]z^T\right] \\
&= \tilde{\mathbb{E}}\left[\left(z - \tilde{\mathbb{E}}[z]\right)\left(z - \tilde{\mathbb{E}}[z]\right)^T\right],
\end{aligned}
$$

where we use the definition of $\tilde{\mathbb{E}}(\cdot)$ in the last three equations.

EC.2.11. Lemma EC.11

LEMMA EC.11. *Let $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^d$ and $\boldsymbol{P}=(p_{ij})$ be a random $d\times m$ matrix such that entry p_{ij} is chosen independently to $\mathcal{N}(0,1/m)$. Then,*

$$
\mathbb{P}\left(\left|\boldsymbol{u}^T\boldsymbol{v} - (\boldsymbol{P}\boldsymbol{u})^T\boldsymbol{P}\boldsymbol{v}\right| > \epsilon\|\boldsymbol{u}\|\|\boldsymbol{v}\|\right) \le 4\exp\left(-\frac{m}{8}\epsilon^2\right) \tag{EC.112}
$$

proof. We first apply Lemma 2 to the vectors $\alpha\boldsymbol{u}+\alpha^{-1}\boldsymbol{v}$ and $\alpha\boldsymbol{u}-\alpha^{-1}\boldsymbol{v}$ for some $\alpha>0$. Then, with probability $1-4\exp\left(-\frac{m}{8}\epsilon^2\right)$, we have the following two inequalities:

$$
\begin{aligned}
(1-\epsilon)\|\alpha\boldsymbol{u}+\alpha^{-1}\boldsymbol{v}\|_2^2 \le \left\|\boldsymbol{P}(\alpha\boldsymbol{u}+\alpha^{-1}\boldsymbol{v})\right\|_2^2 \le (1+\epsilon)\|\alpha\boldsymbol{u}+\alpha^{-1}\boldsymbol{v}\|_2^2 \\
(1-\epsilon)\|\alpha\boldsymbol{u}-\alpha^{-1}\boldsymbol{v}\|_2^2 \le \left\|\boldsymbol{P}(\alpha\boldsymbol{u}-\alpha^{-1}\boldsymbol{v})\right\|_2^2 \le (1+\epsilon)\|\alpha\boldsymbol{u}-\alpha^{-1}\boldsymbol{v}\|_2^2.
\end{aligned}
$$

Combining them together, we have

$$
\begin{aligned}
4(\boldsymbol{P}\boldsymbol{u})^T\boldsymbol{P}\boldsymbol{v} &= \left\|\boldsymbol{P}(\alpha\boldsymbol{u}+\alpha^{-1}\boldsymbol{v})\right\|_2^2 - \left\|\boldsymbol{P}(\alpha\boldsymbol{u}-\alpha^{-1}\boldsymbol{v})\right\|_2^2 \\
&\ge (1-\epsilon)\|\alpha\boldsymbol{u}+\alpha^{-1}\boldsymbol{v}\|_2^2 - (1+\epsilon)\|\alpha\boldsymbol{u}-\alpha^{-1}\boldsymbol{v}\|_2^2 \\
&= 4\boldsymbol{u}^T\boldsymbol{v} - 2\epsilon\left(\alpha^2\|\boldsymbol{u}\|_2^2 + \alpha^{-2}\|\boldsymbol{v}\|_2^2\right).
\end{aligned}
$$

We then choose $\alpha=\sqrt{\frac{\|\boldsymbol{v}\|}{\|\boldsymbol{u}\|}}$ and above inequality implies

$$
(\boldsymbol{P}\boldsymbol{u})^T\boldsymbol{P}\boldsymbol{v} \ge \boldsymbol{u}^T\boldsymbol{v} - \epsilon\|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2. \tag{EC.113}
$$

Similarly, we can prove the other direction of inequality.

EC.3. Appendix: Computation Complexity for the Lasso-RP-MNL Algorithm

The computational costs of the Lasso-RP-MNL algorithm mainly consist of the following three parts:

- *Solving the Lasso*: The Lasso problem will be solved $O(\log T)$ times by time T. As the Lasso is a convex problem, the computational cost will be $O(Td\epsilon^{-1/2})$ by using a gradient-type algorithm (e.g., FISTA in Beck and Teboulle 2009) as the solution scheme, where ϵ is the optimization cost constant.
- *Estimating upper confidence bound for products*: First, we project all N feature vectors x_i from d dimension to $|\mathcal{S}+m|$ dimension, which costs $O(d(|\mathcal{S}+m|)N)$. After the projection, the problem reduces to a low-dimensional setting. Then, we will need to solve a local regression problem (i.e., Eq. (4)) and compute the upper confidence bound for each product. The local regression problem can be solved by a constraint optimization method (e.g., Frank-Wolfe method in Jaggi 2013), whose computation complexity is $O(T(|\mathcal{S}|+m)\epsilon^{-1})$. The upper confidence bounds for products require the matrix inversion and maximum eigenvalue computation so that the cost will be $O((|\mathcal{S}|+m)^3N)$.
- *Identifying optimal assortment*: We need to solve a linear programming problem to identify the optimal assortment. By using the interior point algorithm (e.g., Ye et al. 1994), we will incur at most the computational cost of $O(N^3\epsilon^{-1/2})$.

Combining the above three parts, we can bound the total computational costs of the Lasso-RP-MNL algorithm by time T as follows:

$$O(\log T \cdot Td\epsilon^{-1/2} + Td(|\mathcal{S}|+m)N + T^2(|\mathcal{S}|+m)\epsilon^{-1} + T(|\mathcal{S}|+m)^3N + TN^3\epsilon^{-1}).$$

We want to highlight three major computational improvements of the Lasso-RP-MNL algorithm from the literature. First, compared to the high-dimensional literature (e.g., Bastani and Bayati 2020, Wang et al. 2018a, Kim and Paik 2019), we significantly reduce the usage of the Lasso solver from $O(T)$ to $O(\log T)$, which reduces the computational cost from $O(T^2d)$ to $O(T\log Td)$. Second, by projecting d high-dimensional data into $(|\mathcal{S}|+m)$ low-dimensional space, the computational costs for updating the upper confidence bound for each product can be reduced from $O(d^3N)$ to $O(d(|\mathcal{S}|+m)N+(|\mathcal{S}|+m)^3N)$. Third, instead of enumerating all possible assortments to establish the UCB bound, we construct the UCB bound for every individual product, which helps us eliminate the possible $O(N^K)$ dependence and reduces the computational cost from $O(d^3N\cdot N^K)$ to $O(d(|\mathcal{S}|+m)N+(|\mathcal{S}|+m)^3N+N^3\epsilon^{-1})$.

EC.4. Appendix: Additional experiments on the comparison of the Lasso-RP-MNL algorithm to generalized linear bandit models with $K = 1$

In this section, we benchmark the Lasso-RP-MNL algorithm with several generalized linear bandit algorithms in the literature. For a fair comparison, we restrict $K = 1$ and $N = 2$, under which the MNL model is reduced to a two-arm contextual logistic bandit problem. In particular, we consider the following three benchmark algorithms:

- *UCB-GLM*: This is the Algorithm 1 in Li et al. (2017), which is a UCB-type algorithm for generalized linear model (without using Lasso and random projection).
- *Doubly-Robust Lasso Bandit*: We use the doubly robust lasso estimator introduced in Kim and Paik (2019) for feature selection (without using random projection).
- *G-MCP-Bandit*: This is the G-MCP-Bandit algorithm proposed in Wang et al. (2018b) (without using random projection).

Figure EC.1 reports an representative regret comparison between the Lasso-RP-MNL algorithm and other three benchmark algorithms (other numerical experiments exhibit similar patterns and are omitted for brevity). In particular, in Figure EC.1, we choose $s = 10$ and $d = 100$ with $\beta^* = (1, 2, 3, 4, 5, 1.1, 2.1, 3.1, 4.1, 5.1, 0, 0, ...)$. The feature vectors x_i are randomly generated from the standard gaussian distribution. For each algorithm, we perform 100 trials and report the cumulative regret with 90% confidence interval error bars.

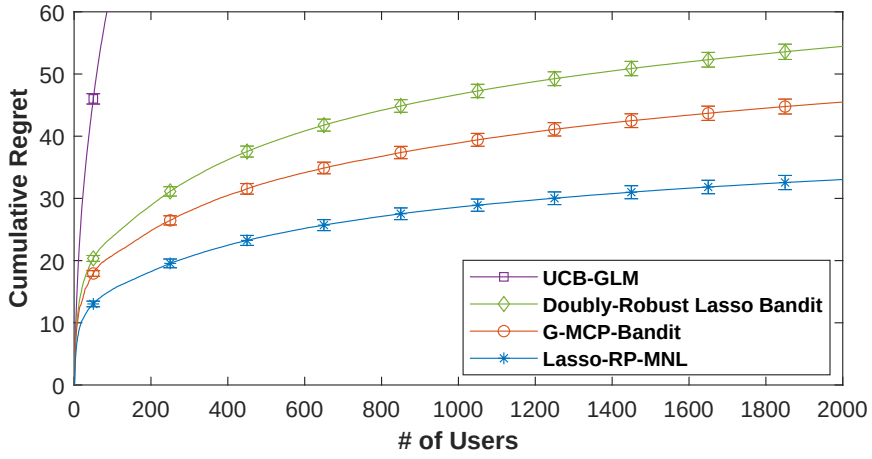


Figure EC.1 **The cumulative regret comparison for a single item assortment, where $d = 100$, $s = 10$, $N = 2$, and $K = 1$.**

We observe that using dimension reduction methods (i.e., Lasso, MCP, or RP), all other three algorithms could significantly improve the regret performance from UCB-GLM and that the Lasso-RP-MNL algorithm continues to outperform other benchmark algorithms and have the least cumulative regret.