Degeneracy is OK: Logarithmic Regret for Network Revenue Management with Indiscrete Distributions

Jiashuo Jiang[†] Will Ma[‡] Jiawei Zhang[§]

[†] Department of Industrial Engineering & Decision Analytics, Hong Kong University of Science and Technology
[‡] Decision, Risk, and Operations Division, Graduate School of Business, Columbia University
[§] Department of Technology, Operations & Statistics, Stern School of Business, New York University

We study the classical Network Revenue Management (NRM) problem with accept/reject decisions and T IID arrivals. We consider a distributional form where each arrival must fall under a finite number of possible categories, each with a deterministic resource consumption vector, but a random value distributed continuously over an interval. We develop an online algorithm that achieves $O(\log^2 T)$ regret under this model, with no further assumptions. We develop another online algorithm that achieves an improved $O(\log T)$ regret, with only a second-order growth assumption. To our knowledge, these are the first results achieving logarithmic-level regret in a continuous-distribution NRM model without further "non-degeneracy" assumptions. Our results are achieved via new techniques including: a new method of bounding myopic regret, a "semi-fluid" relaxation of the offline allocation, and an improved bound on the "dual convergence".

1. Introduction

In the Network Revenue Management (NRM) problem, resources with finite capacities are to be allocated over a finite time horizon of length T. During each time step $t = 1, \ldots, T$, a query t arrives, demanding a vector $\tilde{\boldsymbol{a}}_t$ of resources and providing a reward $\tilde{r}_t$. An irrevocable decision must then be made about whether to serve query t, in which case $\tilde{\boldsymbol{a}}_t$ would be subtracted from the resources and $\tilde{r}_t$ would be collected. Query t is only feasible to serve if the remaining resources exceed $\tilde{\boldsymbol{a}}_t$ component-wise, and a feasible query t can be judiciously rejected, e.g. if $\tilde{r}_t$ is low relative to $\tilde{\boldsymbol{a}}_t$. The goal is to maximize the total reward collected from serving queries, when the values $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$ for each query t are unknown before it arrives but known to be drawn IID across time.

Due to the curse of dimensionality in this problem, a mathematically rich literature has evolved out of developing heuristics and obtaining guarantees on their performance. We consider the stream of literature that analyzes *regret*, which is the additive loss of an online allocation algorithm compared to an optimal offline allocation that knows all values of $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$ beforehand, taken in expectation over the IID query draws (and any further randomness in the algorithm). Generally, regret is larger for longer time horizons T, and this literature is concerned with how the regret grows as a function of T when all other system parameters stay fixed (but the initial resource capacities are also allowed to grow arbitrarily with T).

Two types of assumptions are commonly made in the papers that analyze regret in NRM. The first involves having a *small number of possible realizations* for the values $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$, described by a discrete distribution on N points. As T grows, N stays fixed and is treated as a constant in the analysis. However, such an assumption abandons some natural models, e.g. that of $\tilde{r}_t$ being drawn uniformly from [0,1]. On the other hand, papers that can capture these continuous distributions require a different set of assumptions, which we will call *non-degeneracy*. At a high level, these papers assume that the mathematical program being re-solved by the online algorithm over time to make its decisions is always well-behaved, and has a unique optimal solution. However, such assumptions are difficult to intuit or verify, and appear to be motivated primarily by the analysis.

Our contribution. We establish logarithmic regret in Network Revenue Management with neither the small-N nor non-degeneracy assumptions. We do make some assumptions on the distribution for $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$, but we can show them to be necessary for logarithmic-or-better regret.

To our knowledge, such a result has been missing from the literature, which we now review. Our result builds upon the existing literature that establishes logarithmic-or-better regret in either the small-N or non-degenerate settings, or in the multi-secretary special case.

1.1. Regret Results in Network Revenue Management

We begin with Jasin and Kumar (2012), who establish a constant $O(1)$ regret under *both* the small-N and non-degeneracy assumptions. For NRM literature pre-dating 2012, we refer to the exegesis in Bumpensanti and Wang (2020).

NRM with small-N. Bumpensanti and Wang (2020) and Vera and Banerjee (2021) were the first to establish $O(1)$ regret for a general NRM problem without any non-degeneracy assumptions. We note that Arlotto and Gurvich (2019) first established $O(1)$ regret in the *multi-secretary*[1] special case, where all queries demand one unit of a single resource (i.e. $\tilde{\boldsymbol{a}}_t = (1)$ w.p. 1). These are all surprising results, in that given a fixed discrete distribution for $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$, the regret is upper-bounded by an absolute numerical constant, regardless of how long a time horizon T and how large the resource capacities there are over which regret can be incurred. The intuition is that the further out t is from the end, the harder it is to incur regret, due to the concentration properties of IID draws from a discrete distribution. The total of expected regret over time steps $t = 1, 2, \ldots$ then turns out to be a convergent sum.

[1] Not to be confused with the use of "secretary" to describe a random arrival order (see e.g. Esfandiari et al. 2017).

Multi-secretary with general distributions. A caveat to the aforementioned analysis is that the $O(1)$ constant depends on N, and N can be ∞, e.g. for continuous distributions. To understand continuous distributions better, Bray (2019) studies the example of $\tilde{r}_t$ being drawn uniformly from [0,1] in the multi-secretary special case. He establishes an upper bound on regret that grows logarithmically with T, and importantly, shows this regret rate of $\Theta(\log T)$ to be *tight*—that is, a constant regret is no longer possible. Recently, Besbes et al. (2022) make further progress on the multi-secretary problem by establishing a notion of complexity for general distributions, which affects regret. A corollary relevant to us is that if the reward distribution is supported on an interval (or many disjoint intervals), over which it has a density that is lower-bounded by a positive constant, then $O(\log^2 T)$ regret can be achieved.

NRM with general distributions and non-degeneracy. Although the above papers consider continuous and general reward distributions, it is unclear how they extend beyond the multi-secretary case. Meanwhile, several papers have studied general distributions for NRM under additional assumptions. To be specific, Li and Ye (2021) consider continuous reward distributions and make two additional assumptions, which are: (i) a *non-degeneracy condition* that requires the ex-ante relaxation of the optimal offline allocation to enjoy a unique solution and *strict complementary slackness* being satisfied; and (ii) a *second-order growth condition* that requires the Lagrangian dual function for the ex-ante relaxation to be strongly convex. Though in the literature, these two conditions are always stated together as a whole, we do show that these conditions can indeed be differentiated from each other. In Example 2, we construct an instance where the second-order growth condition is satisfied but the strict complementary slackness of the ex-ante relaxation is violated. Then, equipped with both conditions, Li and Ye (2021) obtain a $O(\log T \log\log T)$ regret bound when the query distribution is unknown. Balseiro et al. (2021) and Bray (2022) assume the query distribution to be known and derive a $O(\log T)$ regret bound. However, the results in Balseiro et al. (2021) and Bray (2022) still require both the non-degeneracy condition and the second-order growth condition to be satisfied, in different forms. We provide a detailed comparison of assumptions in Section 5.

Our model: NRM with discrete demands and continuous rewards. We consider the following structural form of the distribution for $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$. First, the demand vector $\tilde{\boldsymbol{a}}_t$ is drawn from a discrete distribution supported on finite possibilities $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Then, conditional on $\tilde{\boldsymbol{a}}_t = \boldsymbol{a}_j$ for any $j = 1, \ldots, n$, the reward $\tilde{r}_t$ is drawn from a continuous distribution F_j with a lower-bounded density over some interval $[l_j, u_j]$. We develop two algorithms with the following guarantees:

1. $O(\log^2 T)$ regret under our model, with *no further assumptions* (**Section 3**);
2. Improved $O(\log T)$ regret, with the second-order growth assumption (**Section 4**).

In our model, two different indices j, j' can have $\boldsymbol{a}_j = \boldsymbol{a}_{j'}$; moreover, they can have *non-overlapping* intervals with $l_j < u_j < l_{j'} < u_{j'}$. Thus, with a single resource and $\boldsymbol{a}_j = (1)$ for all $j = 1, \dots, n$, we can capture the multi-secretary reward distribution with density lower-bounded over disjoint intervals, and our first result achieves a $O(\log^2 T)$ regret matching the corollary from Besbes et al. (2022). We also emphasize that our base model can exhibit all of: degeneracy, lack of strict complementary slackness, and lack of second-order growth. In Section 4, we present a restricted structural form of the query distribution (Assumption 2) which guarantees second-order growth, but still exhibits degeneracy (Example 2). In both cases, our two algorithms are the only known ways to achieve logarithmic or better regret.

Although our model still imposes a structural form on the distribution for $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$, it has the following justifications. First, both the aspects of n being finite and the density being lower-bounded are *necessary* for logarithmic-or-better regret—see Besbes et al. (2022). Second, because we are restricting the *distribution* instead of the latent mathematical program, our assumption is much easier to intuit—there is a discrete list of n flight itineraries, and a continuous range of customer valuations for each itinerary. Finally, we believe our model is in some sense a natural sandbox for pushing the technical boundaries of the literature, as we now explain in Subsections 1.2 and 1.3.

1.2. $O(\log^2 T)$ Regret with No Further Assumptions

Our $O(\log^2 T)$ result tries to extend the approach of Vera and Banerjee (2021) from discrete distributions, which we now recap. Fix some remaining resource capacities and suppose there are s time steps left. The offline allocation knows for every possible realization $(r, \boldsymbol{a})$, called a "type", the remaining number of queries t with $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t) = (r, \boldsymbol{a})$, denoted by $\tilde{d}_{(r,\boldsymbol{a})}$. Meanwhile, consider an online algorithm that solves for an optimal *fluid* packing, which replaces each $\tilde{d}_{(r,\boldsymbol{a})}$ with its expectation to specify the queries that can be served. The authors compare $\tilde{y}_{(r,\boldsymbol{a})}$, the number of queries of type $(r, \boldsymbol{a})$ accepted by the offline, to $\hat{y}_{(r,\boldsymbol{a})}$, the number of such queries instructed by the fluid packing to accept. Their key argument is that since the number of queries of type $(r, \boldsymbol{a})$ is growing linearly in s and any two of them are interchangeable, the online algorithm only "makes a mistake" if $\tilde{y}_{(r,\boldsymbol{a})}$ and $\hat{y}_{(r,\boldsymbol{a})}$ are distance $\Omega(s)$ apart. This is a highly unlikely event (over the randomness in the offline's draws of $\tilde{d}_{(r,\boldsymbol{a})}$) because $\tilde{y}_{(r,\boldsymbol{a})}$ and $\hat{y}_{(r,\boldsymbol{a})}$ are generally only $O(\sqrt{s})$ apart, specifically an event with probability $O(e^{-s})$ that when summed over s leads to constant regret.

A new, "semi-fluid" relaxation of offline. Comparing $\tilde{y}_{(r,\boldsymbol{a})}$ to $\hat{y}_{(r,\boldsymbol{a})}$ is meaningless under continuous rewards, because there is zero probability of drawing any specific $(r, \boldsymbol{a})$. To cope, we introduce a new *semi-fluid* relaxation that amalgamates decisions over queries with the same demand vector. Specifically, we call each $j = 1, \dots, n$ a *type* in our model of NRM, and let $\tilde{d}_j$ denote the remaining number of queries with $\tilde{\boldsymbol{a}}_t = \boldsymbol{a}_j$ and $\tilde{r}_t$ drawn from F_j. The semi-fluid relaxation knows $\tilde{d}_j$ for all j. However, the semi-fluid relaxation differs from the offline allocation in that it collects exactly the "fluid" value

$$\tilde{d}_j \int_{1-\tilde{y}_j/\tilde{d}_j}^{1} F_j^{-1}(q)dq \tag{1}$$

when it accepts $\tilde{y}_j$ queries of type j. Objective (1) integrates over the $\tilde{y}_j/\tilde{d}_j$ proportion of the $\tilde{d}_j$ type-j queries with the highest rewards, as explained in Section 3. Meanwhile, our algorithm is still based on solving the (fully) fluid packing, whose variables can also be amalgamated into acceptance quantities $\hat{y}_j$ for each type j (and $\tilde{d}_j$ will be replaced $\mathbb{E}[\tilde{d}_j]$, including in the objective (1)). We can then compare the acceptance quantities $\tilde{y}_j$ to $\hat{y}_j$ for the amalgamated types j.

Bounding the myopic regret. Unfortunately, because in our model queries of the same type have different rewards, a mistake can be made without requiring $|\tilde{y}_j - \hat{y}_j| = \Omega(s)$. In fact, a mistake only requires drawing a quantile that is above the acceptance proportion $\tilde{y}_j/\tilde{d}_j$ for the semi-fluid but below $\hat{y}_j/\mathbb{E}[\tilde{d}_j]$ for the fluid (or vice versa), which occurs with probability $|\tilde{y}_j/\tilde{d}_j - \hat{y}_j/\mathbb{E}[\tilde{d}_j]|$. As such, the rough argument that $\tilde{y}_j$ and $\hat{y}_j$ are $O(\sqrt{s})$ apart (and $\mathbb{E}[\tilde{d}_j] = \Omega(s)$) would lead to a mistake probability of $O(1/\sqrt{s})$, and an undesirable overall regret of $O(\sqrt{T})$. Therefore, we instead follow Bray (2019) who argues that for continuous distributions one must quantify the "myopic regret" at each time step (instead of just bounding the probability that it is non-zero). We decompose overall regret in way (see Section 2) such that the regret at a time step can be quanfied as

$$\bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}} + \mathbf{e}_j) - \int_{1-\hat{y}_j/\mathbb{E}[\tilde{d}_j]}^{1} F_j^{-1}(q)dq - \frac{\hat{y}_j}{\mathbb{E}[\tilde{d}_j]} \cdot \bar{V}_{\boldsymbol{c}-\boldsymbol{a}_j}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}}) - \left(1 - \frac{\hat{y}_j}{\mathbb{E}[\tilde{d}_j]}\right) \cdot \bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}}), \tag{2}$$

where $\boldsymbol{c}$ denotes the remaining resource capacities, j denotes the type of the current query, and $\bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\boldsymbol{d})$ denotes the optimal objective value of the semi-fluid relaxation given remaining resources $\boldsymbol{c}$ and a generic vector $\boldsymbol{d} = (d_1, \ldots, d_n)$ counting the remaining queries of each type.

To upper-bound (2), we take an optimal solution $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)$ for $\bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}} + \mathbf{e}_j)$ and modify it into feasible solutions for $\bar{V}_{\boldsymbol{c}-\boldsymbol{a}_j}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}})$ and $\bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}})$, which lower-bounds the latter quantities. As long as these feasible solutions can be constructed by modifying only the j'th coordinate of $\tilde{\boldsymbol{y}}$, we show that (2) is $O((\tilde{y}_j/\tilde{d}_j - \hat{y}_j/\mathbb{E}[\tilde{d}_j])^2)$. Per the earlier discussion, this is roughly $O((1/\sqrt{s})^2) = O(1/s)$, which when summed over s would sufficiently lead to logarithmic regret.

A boundary-attracted algorithm. If the semi-fluid solution $\tilde{y}_j$ is close to 0, however, then $\tilde{\boldsymbol{y}}$ is difficult to modify into a feasible solution for $\bar{V}_{\boldsymbol{c}-\boldsymbol{a}_j}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}})$—one cannot pack into the reduced capacity $\boldsymbol{c} - \boldsymbol{a}_j$ by only reducing coordinate $\tilde{y}_j$. Our strategy is to bypass this boundary situation by tweaking the online algorithm—if there is a risk of this infeasibility, which we show can be identified by the algorithm checking whether $\hat{y}_j = O(\sqrt{s \log s})$, then it *always rejects* the current type-j query. This effectively sets $\hat{y}_j = 0$ in (2) and avoids having to lower-bound the quantity

$\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_j}(\tilde{\boldsymbol{d}})$. On the other extreme, our algorithm *always accepts* if $\hat{y}_j$ is within $O(\sqrt{s \log s})$ of its maximum value $\mathbb{E}[\tilde{d}_j]$. All in all, we use this tweaked version of the fluid re-solving algorithm that is "attracted to boundaries", which is similar in spirit to the thresholding in Bumpensanti and Wang (2020) and the "conservatism with respect to gaps" in Besbes et al. (2022). We provide a new explanation for it based on our analysis, and it allows us to always upper-bound (2) by $O((\tilde{y}_j/\tilde{d}_j - \hat{y}_j/\mathbb{E}[\tilde{d}_j])^2) = O(1/s)$ while sacrificing only a log-factor—ultimately achieving $O(\log^2 T)$ regret.

Lipschitz property for semi-fluid convex program. Finally, it should not be taken for granted that $\tilde{y}_j$ and $\hat{y}_j$ are nearby, an intuition we have been frequently using. Indeed, they correspond to optimal solutions of mathematical programs with different objective functions—the fluid problem replaces $\tilde{d}_j$ with $\mathbb{E}[\tilde{d}_j]$ in (1)—and a simple example (Mangasarian and Shiau 1987, Remark 2.7) reveals that optimal solution sets are highly sensitive to small perturbations in the objective. Nonetheless, we extend (Lemma 3) the Lipschitz analysis of Mangasarian and Shiau (1987) to show that the specific objective function (1) is well-behaved. We also note that due to degeneracy, it is necessary for the Lipschitz property to be of the form "given any optimal solution $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_n)$ to the fluid, *there exists* a nearby optimal solution $\tilde{\boldsymbol{y}}$ to the semi-fluid". General perturbation analysis results for convex programs (Bonnans and Shapiro 2013), which try to argue that *all* optimal solutions $\tilde{\boldsymbol{y}}$ are nearby, do not apply in our setting with degeneracy.

1.3. $O(\log T)$ Regret with Second-order Growth Assumption

Based on the structural form introduced previously, we further assume that the density is also upper bounded, conditional on $\tilde{\boldsymbol{a}}_t = \boldsymbol{a}_j$, and it holds that $l_j = 0$, for any $j = 1, \ldots, n$. Note that the further assumptions we make are to guarantee a second-order growth condition over the Lagrangian dual function of the ex-ante relaxation, which requires the dual function to be strongly convex. As discussed earlier, the second-order growth condition, together with the non-degeneracy assumption, lead to the $O(\log T \log \log T)$ regret bound in Li and Ye (2021) and the $O(\log T)$ regret bound in Balseiro et al. (2021) and Bray (2022). In comparison to the aforementioned papers, our contribution here is to derive a $O(\log T)$ regret bound, without the non-degeneracy assumption, though we still need the second-order growth condition. In what follows, we first illustrate how the two conditions are utilized in the previous literature to derive the logarithmic regret bound, and we then explain how our analysis improves upon them to relax the non-degeneracy assumption.

Analysis in the previous literature. Note that after assuming the strict complementary slackness condition is satisfied by the ex-ante relaxation, there exists a neighborhood of the initial average capacities per period such that the optimal basis (binding constraints) for the ex-ante relaxation remains unchanged even if the average capacities are perturbed, as long as the perturbed average capacities still stay in this neighborhood. Then, one can establish a connection between the average remaining capacities and the optimal dual variables of the ex-ante relaxation, for each period, as long as the average remaining capacities stay in this neighborhood. Such a connection reveals that the dual variables corresponding to the binding constraints will behave as a martingale, while the concentration can be bounded using the second-order growth condition of the dual function. Such an idea is first developed in Jasin and Kumar (2012) with a small-N assumption, and then extended in Li and Ye (2021) and Balseiro et al. (2021) for general distributions. However, we note that all three aforementioned papers use the ex-ante relaxation as the upper bound of the optimal offline allocation to do the regret analysis, which is the key reason why the non-degeneracy condition is inevitable to carry out their analysis.

A tighter relaxation of the offline allocation. Our improvement comes from using a tighter relaxation than the ex-ante relaxation as an upper bound of the offline allocation. Our relaxation is that for each sample path, we relax the integral decision of the offline allocation to be fractional, and we take an expectation over the sample path. Such an LP relaxation of the offline allocation has been derived in Bumpensanti and Wang (2020) and Vera and Banerjee (2021) in the discrete setting and we derive it here for general distributions. Then, by following a myopic regret approach described in Section 2.1, we are able to bound the regret incurred at each period by the variance of the dual variable of the LP relaxation of the offline allocation, no matter what the remaining capacities are. Note that we do not need to consider the optimal basis of the ex-ante relaxation to bound the myopic regret. This is the key distinction between our approach and the martingale-based approach in Li and Ye (2021), Balseiro et al. (2021), Bray (2022), which would require a non-degeneracy assumption to guarantee the optimal basis remains fixed in their analysis.

An improvement on the dual convergence bound. Another improvement we make is regarding the variance of the dual variable of the LP relaxation of the offline allocation, which is referred to as "dual convergence" in Li and Ye (2021) and requires the second-order growth condition of the Lagrangian dual function of the ex-ante relaxation. Note that if one assumes the distribution to be known, then the martingale-based approach in Li and Ye (2021) would not require a bound on the "dual convergence". Instead, as shown in Theorem 1 of Balseiro et al. (2021), the martingale property itself is enough to prove the regret bound. The "dual convergence" is needed in Li and Ye (2021) because they allow the distribution to be unknown and they use the sample average problem to estimate the dual variable, where the data comes from the past periods, and the "dual convergence" is needed to bound the estimation error. To be specific, when there are s data points, Li and Ye (2021) prove the "dual convergence" bound to be at the order of $O(\frac{\log\log s}{s})$, which we improve to $O(\frac{1}{s})$. To obtain this improvement, we utilize both ways of splitting the whole space into a set of small cubes with exponentially increasing edge lengths in Huber (1967) and Li and Ye (2021) to give a probability bound on difference scenarios. To illustrate the main idea, we denote by $\tilde{\boldsymbol{\mu}}$ the dual variable of the sample average problem and $\hat{\boldsymbol{\mu}}$ the dual variable of the ex-ante relaxation. Then, we apply the approach in Huber (1967) to obtain a bound on $P(|\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}| \geq \varepsilon)$, which is derived from Chebyshev's inequality. However, we note that when $\varepsilon > \sqrt{\frac{1}{s}}$, the approach in Li and Ye (2021), which is based on the Hoeffding's inequality, would give us a tighter probability bound. Therefore, by applying different ways to bound $P(|\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}| \geq \varepsilon)$ for different ranges of ε, we get a tighter bound on the "dual convergence" $\mathbb{E}[\|\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}\|_2^2]$.

We do note that the "dual convergence" bound is used in different ways between our analysis vs. Li and Ye (2021). The data points for "dual convergence" in Li and Ye (2021) comes from past periods. By contrast, the "data points" for "dual convergence" in our analysis comes from the future periods. This is because our online decision is made based on the ex-ante relaxation, while our benchmark, the LP relaxation of the offline allocation, makes the decision based on each sample path of future periods. The myopic regret caused by this distinction is shown to be bounded by the variance of the dual variable, where the randomness comes from the sample path of future periods. Though the future sample path is convoluted, we only use its distribution to give a bound and our algorithm does not require any knowledge of the realization.

Discussion on the second-order growth condition. We do acknowledge that the second-order growth condition that is assumed in Section 4 is somewhat stronger than the one in existing literature (Li and Ye 2021, Balseiro et al. 2021, Bray 2022). To be specific, the second-order growth condition assumed in Section 4 holds for the Lagrangian dual function given *any* remaining average capacities per period, while the condition in the existing literature holds given remaining average capacities per period belonging to a neighborhood of the initial average capacities per period. However, the second-order growth condition in Section 4 is a consequence of our problem formulation instead of a primitive assumption. Moreover, since it is assumed that the ex-ante relaxation admits a unique optimal dual variable in Li and Ye (2021), Balseiro et al. (2021), Bray (2022), their second-order growth conditions are stated as the strong-convexity of the Lagrangian dual function. In contrast, the second-order growth condition in our setting is stated as the strong-convexity after projecting every variable into the subspace that is spanned by the set of possible query sizes and we do not require the uniqueness of the optimal dual variable.

1.4. Further Related Work

The network revenue management (NRM) problem has been extensively studied in the literature and one main topic is to develop near-optimal policies with strong theoretical guarantees. One common way is to derive the policy from the optimal solution of the ex-ante relaxation. To be specific, Talluri and Van Ryzin (1998) propose a static bid-price policy based on the optimal dual variable of the ex-ante relaxation and proves that the regret bound is $O(\sqrt{T})$. Then, a dynamic update of the bid-price is considered in the literature. Subsequently, Reiman and Wang (2008) shows that by re-solving the ex-ante relaxation once to update the bid-price, one can obtain an improved regret bound $o(\sqrt{T})$. Then, Jasin and Kumar (2012) shows that under a non-degeneracy condition for the ex-ante relaxation, a policy which re-solves the ex-ante relaxation at each time period will lead to an $O(1)$ regret. The relationship between the performances of the control policies and the number of times of re-solving the ex-ante relaxation is further discussed in their later paper (Jasin and Kumar 2013). More recently, Bumpensanti and Wang (2020) proposes an infrequent re-solving policy and shows a regret bound of $O(1)$ without the "non-degeneracy" assumption. This has been extended by Balseiro and Xia (2022) to fair allocation problems. With a different approach, Vera and Banerjee (2021) proves the same $O(1)$ upper bound for the NRM problem and their approach is further generalized in series of papers (e.g. Freund and Banerjee (2019), Vera et al. (2021), Freund and Zhao (2022)). Recent studies on the NRM problem includes variants such as the reusable resource setting (Baek and Ma 2022), unknown distribution setting (Li et al. 2020, Balseiro et al. 2022) and imperfect distribution knowledge setting under a non-stationary environment (Jiang et al. 2020).

Another problem that is closely related to the NRM problem is called the online packing problem, where a more general formulation is studied and less distribution knowledge is assumed. The packing problem covers a wide range of applications, including secretary problem (Ferguson 1989, Arlotto and Gurvich 2019), online knapsack problem (Arlotto and Xie 2020, Jiang and Zhang 2020), resource allocation problem (Asadpour et al. 2020), network routing problem (Buchbinder and Naor 2009), matching problem (Mehta et al. 2007), etc. The problem is usually studied under either a stochastic model where the reward and size of each query is drawn independently from an unknown distribution $\mathcal{P}$, or a more general the random permutation model where the queries arrive in a random order (Molinaro and Ravi 2014, Agrawal et al. 2014, Kesselheim et al. 2014, Gupta and Molinaro 2014).

2. Problem Formulation and Our Approach

We consider an online resource allocation problem, where there are m resources and each resource $i \in [m]$ has an initial fractional capacity $C_i \in \mathbb{R}_{\geq 0}$. There are T discrete time periods and at each period $t \in [T]$, one query arrives, denoted by query t. Each query t has a *random* size $\tilde{\boldsymbol{a}}_t = (\tilde{a}_{t,1}, \ldots, \tilde{a}_{t,m}) \in \mathbb{R}^m_{\geq 0}$, where $\tilde{a}_{t,i}$ denotes how much resource i will be consumed if query t is served, for all $i \in [m]$, and a *random* reward $\tilde{r}_t \in \mathbb{R}_{\geq 0}$ that denotes how much reward can be collected by serving query t. We assume that the value of $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$ for each $t \in [T]$ is drawn *independently* from an identical distribution denoted by $F(\cdot)$. We suppose that the queries are of finite types, i.e., for each $t \in [T]$, $\tilde{\boldsymbol{a}}_t$ is supported on a finite set $\mathcal{A} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$. We call the situation where $\tilde{\boldsymbol{a}}_t$ is realized as $\boldsymbol{a}_j$ as query t being of type j and we denote by $p_j = P(\tilde{\boldsymbol{a}} = \boldsymbol{a}_j)$, for each $j \in [n]$.

After query t arrives and the value of $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$ is revealed, the decision maker has to decide immediately and irrevocably whether or not to serve query t. Note that query t can only be served if for every resource i its remaining capacity it at least $\tilde{a}_{t,i}$. The goal of the decision maker is to maximize the total collected reward subject to the resource capacity constraint.

Any online policy π for the decision maker is specified by a set of decision variables $\{\tilde{x}_t^\pi\}_{\forall t \in [T]}$, where $\tilde{x}_t^\pi$ is a binary variable and denotes whether query t is served, for all $t \in [T]$. Note that $\tilde{x}_t^\pi$ can be stochastic if π is a randomized policy. Any policy π is feasible if for all $t \in [T]$, $\tilde{x}_t^\pi$ depends only on $F(\cdot)$ and $\{(\tilde{r}_1, \tilde{\boldsymbol{a}}_1), \ldots, (\tilde{r}_t, \tilde{\boldsymbol{a}}_t)\}$, and the following capacity constraint is satisfied:

$$\sum_{t=1}^{T} \tilde{a}_{t,i} \cdot \tilde{x}_t^\pi \le C_i, \quad \forall i \in [m]. \tag{3}$$

The total collected value of policy π is given by $V^\pi(I) = \sum_{t=1}^{T} \tilde{r}_t \cdot \tilde{x}_t^\pi$, where $I = \{(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)\}_{t=1}^T$ denotes the problem instance.

The benchmark is the prophet, which is an offline decision maker that is aware of the value of $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$ for all $t \in [T]$ and always makes the optimal decision in hindsight. We denote by $\{\tilde{x}_t^{\text{off}}\}_{t=1}^T$ the offline decision of the prophet, which is an optimal solution to the following offline problem:

$$\begin{aligned} V^{\text{off}}(I) = \max \ & \sum_{t=1}^{T} \tilde{r}_t \cdot x_t \\ \text{s.t.} \ & \sum_{t=1}^{T} \tilde{a}_{t,i} \cdot x_t \le C_i, \quad \forall i \in [m] \\ & x_t \in \{0,1\} \qquad \forall t \in [T]. \end{aligned} \tag{4}$$

For any feasible online policy π, we use *regret* to measure its performance, which is defined as follows:

$$\text{Regret}(\pi) := \mathbb{E}_{I \sim F}[V^{\text{off}}(I)] - \mathbb{E}_{I \sim F}[V^\pi(I)] \tag{5}$$

where $I = \{(\tilde{r}_t, \tilde{a}_t)\}_{t=1}^T \sim F$ denotes that $(\tilde{r}_t, \tilde{a}_a)$ follows distribution $F(\cdot)$ independently for each $t \in [T]$. In what follows, we describe our general approach to upper bound the regret defined in (5), and discuss how our approach implies online policies under various settings.

2.1. General Description of Our Approach

We now give a general description of our approach. We denote by $\boldsymbol{c} = (c_1, \ldots, c_m) \in \mathbb{R}^m$ any vector of remaining capacities of the resources at the beginning of a period t. Then, on problem instance $I_t = \{(\tilde{r}_t, \tilde{\boldsymbol{a}}_t), \ldots, (\tilde{r}_T, \tilde{\boldsymbol{a}}_T)\}$, we denote by $\bar{V}_{\boldsymbol{c}}(I_t)$ a relaxation of the total reward collected by the prophet from period t up to period T, given the remaining capacity $\boldsymbol{c}$, where the decision variable $x_\tau \in \{0,1\}$

is relaxed into $x_\tau \in [0,1]$ for $\tau = t, \ldots, T$. We specify various formulations of the relaxation $\bar{V}_{\boldsymbol{c}}(I_t)$ to deal with various settings in the following sections. Then, the regret of any online policy π can be upper bounded by the gap between $\mathbb{E}_{I_1 \sim F}[\bar{V}_{\mathbf{C}}(I_1)]$, where $\boldsymbol{C} = (C_1, \ldots, C_m)$ is a vector of initial capacity for all resources, and $\mathbb{E}_{\pi, I_1 \sim F}[V^\pi(I)]$, i.e.,

$$\text{Regret}(\pi) \leq \mathbb{E}_{I_1 \sim \boldsymbol{F}}[\bar{V}_{\mathbf{C}}(I_1)] - \mathbb{E}_{\pi, I_1 \sim \boldsymbol{F}}[V^\pi(I_1)]. \tag{6}$$

Our approach relies on the following decomposition of the upper bound in (6). For each $t \in [T]$, we denote by $\tilde{\boldsymbol{c}}_t^\pi = (\tilde{c}_{t,1}^\pi, \ldots, \tilde{c}_{t,m}^\pi) \in \mathbb{R}^m$ the remaining capacities at the beginning of period t during the execution of the policy π. Note that $\tilde{\boldsymbol{c}}_t^\pi$ is random for each $t \in [T]$, where the randomness comes from the randomness in the problem instance I and any randomness in the policy π. Then, the term $\bar{V}_{\mathbf{C}}(I_1)$ can be telescoped as follows by noting that $\tilde{\boldsymbol{c}}_1^\pi = \boldsymbol{C}$ and $\bar{V}_{\boldsymbol{c}}(I_{T+1}) = 0$ for every $\boldsymbol{c}$:

$$\bar{V}_{\mathbf{C}}(I_1) = \bar{V}_{\tilde{\boldsymbol{c}}_1^\pi}(I_1) - \bar{V}_{\tilde{\boldsymbol{c}}_{T+1}^\pi}(I_{t+1}) = \sum_{t=1}^{T} \left(\bar{V}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) - \bar{V}_{\tilde{\boldsymbol{c}}_{t+1}^\pi}(I_{t+1}) \right). \tag{7}$$

Thus, the regret upper bound (6) can be decomposed as:

$$\begin{aligned}
\mathbb{E}_{I_1 \sim \boldsymbol{F}}[\bar{V}_{\mathbf{C}}(I_1)] - \mathbb{E}_{\pi, I_1 \sim \boldsymbol{F}}[V^\pi(I_1)] &= \mathbb{E}_{\pi, I_t \sim \boldsymbol{F}} \left[\sum_{t=1}^{T} \left(\bar{V}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) - \bar{V}_{\tilde{\boldsymbol{c}}_{t+1}^\pi}(I_{t+1}) - \tilde{r}_t \cdot \tilde{x}_t^\pi \right) \right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{\pi, I_t \sim \boldsymbol{F}} \left[\bar{V}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) - \bar{V}_{\tilde{\boldsymbol{c}}_{t+1}^\pi}(I_{t+1}) - \tilde{r}_t \cdot \tilde{x}_t^\pi \right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{\pi, I_t \sim \boldsymbol{F}} \left[\bar{V}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) - \bar{V}_{\tilde{\boldsymbol{c}}_t^\pi - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^\pi}(I_{t+1}) - \tilde{r}_t \cdot \tilde{x}_t^\pi \right]
\end{aligned}$$

where the third equality follows from the identity that $\tilde{\boldsymbol{c}}_{t+1}^\pi = \tilde{\boldsymbol{c}}_t^\pi - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^\pi$. We proceed to analyze the term for each $t \in [T]$ in the above summation. For each $\boldsymbol{c} \geq 0$, we now denote by

$$\text{Myopic}_t(\pi, \boldsymbol{c}) = \mathbb{E}_{\pi, I_t} \left[\bar{V}_{\boldsymbol{c}}(I_t) - \bar{V}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^\pi}(I_{t+1}) - \tilde{r}_t \cdot \tilde{x}_t^\pi \right]. \tag{8}$$

It is clear that in order to upper bound Regret(π), it is sufficient to upper bound Myopic$_t(\pi, \boldsymbol{c})$ for each $t \in [T]$ and each $\boldsymbol{c} \geq 0$. We summarize the above arguments in the following lemma.

Lemma 1. *For any feasible online policy π, the regret is upper bounded by*

$$\textit{Regret}(\pi) \leq \sum_{t=1}^{T} \mathbb{E}_{\tilde{\boldsymbol{c}}_t^\pi} \left[\textit{Myopic}_t(\pi, \tilde{\boldsymbol{c}}_t^\pi) \right]$$

where the myopic term Myopic$_t(\pi, \tilde{\boldsymbol{c}}_t^\pi)$ is defined in (8).

We now motivate our policy π such that the myopic term Myopic$_t(\pi, \boldsymbol{c})$ can be minimized for each $\boldsymbol{c}$. Now suppose that the online decision maker is allowed to "foresee" the sample path I_t and we denote by

$$M_{\boldsymbol{c}, \tilde{\boldsymbol{a}}_t}(I_{t+1}) = \bar{V}_{\boldsymbol{c}}(I_{t+1}) - \bar{V}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t}(I_{t+1})$$

the marginal increase for the relaxation $\bar{V}$ to have an extra $\tilde{\boldsymbol{a}}_t$ resources from period $t+1$ to T. Clearly, in order to minimize $\mathrm{Myopic}_t(\pi, \boldsymbol{c})$ in (8), we set $\tilde{x}_t^\pi = 1$ if and only if

$$\bar{V}_{\boldsymbol{c}}(I_t) - \bar{V}_{\boldsymbol{c}-\tilde{\boldsymbol{a}}_t}(I_{t+1}) - \tilde{r}_t \leq \bar{V}_{\boldsymbol{c}}(I_t) - \bar{V}_{\boldsymbol{c}}(I_{t+1})$$

which implies that

$$\tilde{x}_t^\pi = \begin{cases} 1, & \text{if } \tilde{r}_t \geq M_{\boldsymbol{c},\tilde{\boldsymbol{a}}_t}(I_{t+1}) \\ 0, & \text{if } \tilde{r}_t < M_{\boldsymbol{c},\tilde{\boldsymbol{a}}_t}(I_{t+1}). \end{cases}$$

However, note that in order for π to be feasible, $\tilde{x}_t^\pi$ must be independent of I_{t+1}. Therefore, instead of comparing $\tilde{r}_t$ to the marginal increase $M_{\boldsymbol{c},\tilde{\boldsymbol{a}}_t}(I_{t+1})$, we compare $\tilde{r}_t$ to an estimator $\hat{M}_{\boldsymbol{c},\tilde{\boldsymbol{a}}_t}$ that is independent of I_{t+1}. Our policy is formalized in Algorithm 1, which takes as an input an exogenous estimator $\hat{M}$ that we further specify in the following sections on different settings.

Algorithm 1 $\hat{M}$-estimator policy $(\pi_{\hat{M}})$

```
Input: an estimator $\hat{M}$.
Initialize the initial capacities $\boldsymbol{c}_1 = \boldsymbol{C}$.
for $t = 1, ..., T$ do
    Observe the value of $(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)$ and obtain the value of $\hat{M}_{\mathbf{c}_t,\tilde{\boldsymbol{a}}_t}$.
    if $\tilde{r}_t \geq \hat{M}_{\mathbf{c}_t,\tilde{\boldsymbol{a}}_t}$ and $\mathbf{c}_t \geq \tilde{\boldsymbol{a}}_t$, then we set $x_t = 1$ and update $\boldsymbol{c}_{t+1} = \boldsymbol{c}_t - \tilde{\boldsymbol{a}}_t$;
    otherwise, set $x_t = 0$ and update $\boldsymbol{c}_{t+1} = \boldsymbol{c}_t$.
end for
Output: online decisions $\boldsymbol{x} = (x_1, ..., x_T)$.
```

3. Policy with Log-squared Regret

In this section, we derive a log-squared bound for (6) under the following assumption over the distribution $F(\cdot)$.

Assumption 1. *We assume that for each $j \in [n]$, conditional on $\tilde{\boldsymbol{a}}_t$ being realized as any $\boldsymbol{a}_j \in \mathcal{A}$, the reward distribution of $\tilde{r}$ is supported on the interval $[l_j, u_j]$ with a density function $f(\cdot|\boldsymbol{a}_j)$, where $u_j \geq l_j \geq 0$, and it satisfies $f(r|\boldsymbol{a}_j) \geq \alpha$, for a constant $\alpha > 0$, for any $r \in [l_j, u_j]$.*

Note that in the above Assumption 1, we allow $l_j = u_j$ for a type j, i.e., the reward distribution for type j query is a point mass. In this case, we let the density be $f_j(r|\boldsymbol{a}_j) = \infty$ for $r = l_j = u_j$ and any constant α would satisfy $f_j(r|\boldsymbol{a}_j) \geq \alpha$ for $r \in [l_j, u_j]$.

In what follows, we first specify the relaxation $\bar{V}$ that will be used in this section, and then we specify the $\hat{M}$-estimator for our algorithm and derive the corresponding regret bound.

3.1. Semi-fluid Relaxation

We now specify the semi-fluid relaxation $\bar{V}$ that will be used in this section. For each $j \in [n]$, we let d_j denote a generic non-negative integer that should be interpreted as the number of type j query arrivals remaining. Meanwhile, we let $\tilde{d}_{j,t}$ be the random variable for the number of type j query arrivals from period t to period T, i.e., the number of times that $\tilde{\boldsymbol{a}}_\tau = \boldsymbol{a}_j$ for $\tau = t, \ldots, T$. We introduce $\bar{V}$ is as follows, for a fixed $\boldsymbol{d} = (d_1, \ldots, d_n) \in \mathbb{R}^n$:

$$\begin{aligned} \bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\boldsymbol{d}) = \max\ & \sum_{j=1}^{n} d_j \cdot \mathbb{E}_{r \sim F(\cdot|\boldsymbol{a}_j)}[r \cdot x_j(r)] \\ \text{s.t.}\ & \sum_{j=1}^{n} d_j \cdot a_{j,i} \cdot \mathbb{E}_{r \sim F(\cdot|\boldsymbol{a}_j)}[x_j(r)] \leq c_i, \quad \forall i \in [m] \\ & x_j(r) \in [0,1] \end{aligned} \tag{9}$$

In the following lemma, we show that the formulation of $\bar{V}$ introduced in (9) implies an upper bound of the offline optimum V^{off} in (4).

Lemma 2. *It holds that $\mathbb{E}_I[\bar{V}_{\boldsymbol{C}}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}}_1)] \geq \mathbb{E}_I[V^{\mathrm{off}}(I)]$, where $\tilde{\boldsymbol{d}} = (\tilde{d}_{1,1}, \ldots, \tilde{d}_{n,1})$ depends on the sample path I.*

In order to see that $\bar{V}_{\boldsymbol{C}}^{\mathrm{Semi}}(\tilde{\boldsymbol{d}}_1)$ is an upper bound, we first fix the type arrivals of the queries that is implied by the sample path I. We then take an ex-ante relaxation over the reward distribution for each type.

Comparison with other relaxations. There are also other relaxations of the prophet (4) existing in the literature and we now compare $\bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\boldsymbol{d})$ (9) with them.

The most natural relaxation of the prophet (4) is an LP relaxation, which is defined for each $t \in [T]$, any $\boldsymbol{c} \geq 0$, and any sample path I.

$$\begin{aligned} \bar{V}_{\boldsymbol{c}}^{\mathrm{Off}}(I_t) = \max\ & \sum_{\tau=t}^{T} \tilde{r}_\tau \cdot x_\tau \\ \text{s.t.}\ & \sum_{\tau=t}^{T} \tilde{a}_{\tau,i} \cdot x_\tau \leq c_i, \quad \forall i \in [m] \\ & x_\tau \in [0,1] \qquad \forall \tau = t, \ldots, T \end{aligned} \tag{10}$$

The only difference between the formulation of $V^{\mathrm{off}}(I)$ and $\bar{V}_{\boldsymbol{C}}^{\mathrm{Off}}(I_1)$ is that the integral decision variables of $V^{\mathrm{off}}(I)$ are relaxed to be fractional in $\bar{V}_{\boldsymbol{C}}^{\mathrm{Off}}(I_1)$.

Another common relaxation in the literature is the so-called ex-ante relaxation, which can be obtained by taking expectation over $\tilde{\boldsymbol{d}}_t$ in the formulation of $\bar{V}_{\boldsymbol{c}}^{\mathrm{Off}}(\tilde{\boldsymbol{d}}_t)$. For any $t \in [T]$ and any $\boldsymbol{c}$, we denote by $\bar{V}_{t,\boldsymbol{c}}^{\mathrm{Fld}}$ the ex-ante relaxation with a formulation given as follows:

$$\bar{V}_{t,\boldsymbol{c}}^{\mathrm{Fld}} = \max \sum_{j=1}^{n} p_j \cdot s \cdot \mathbb{E}_{r \sim F(\cdot|\boldsymbol{a}_j)}[r \cdot x_j(r)] \tag{11}$$

$$\begin{aligned} \text{s.t. } & \sum_{j=1}^{n} p_j \cdot s \cdot a_{j,i} \cdot \mathbb{E}_{r\sim F(\cdot|\boldsymbol{a}_j)}[x_j(r)] \le c_i, \quad \forall i \in [m] \\ & x_j(r) \in [0,1] \end{aligned}$$

where we denote by $s = T - t + 1$ for notation brevity.

3.2. Policy and Regret Analysis

We now develop the estimator that will be used in Algorithm 1 and analyze the regret bound. It is easy to see that the optimal solution of (9) preserves a "threshold" property, i.e. there exists a set of thresholds $\{\kappa_j\}_{j=1}^n$ such that it is optimal to set $x_j^*(r) = 1$ if and only if $r \ge \kappa_j$ and $x_j^*(r) = 0$ if and only if $r < \kappa_j$, for any $j \in [n]$. Therefore, $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d})$ in (9) can be re-written into the following formulation:

$$\begin{aligned} \bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}) = \max & \sum_{j=1}^{n} d_j \cdot \int_{q=1-q_j}^{1} F_j^{-1}(q)dq \\ \text{s.t. } & \sum_{j=1}^{n} d_j \cdot a_{j,i} \cdot q_j \le c_i, \quad \forall i \in [m] \\ & q_j \in [0,1], \quad \forall j \in [n] \end{aligned} \tag{12}$$

where we denote by $F_j(\cdot) = F(\cdot|\boldsymbol{a}_j)$ for notation brevity, for each $j \in [n]$. Here, the decision variable q_j can be interpreted as the probability of serving type j query, for each $j \in [n]$. Denote by $\{\tilde{q}_j^*\}$ one optimal solution to $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\tilde{\boldsymbol{d}}_t)$. Since $\{\tilde{q}_j^*\}$ depends on the sample path I_t, clearly, one cannot directly use $\{\tilde{q}_j^*\}$ to derive a feasible online policy that is "agnostic" about I_{t+1}. Therefore, we will consider using the optimal solution of the ex-ante problem (11) to "approximate" $\{\tilde{q}_j^*\}$. It is clear to see that the optimal solution of the ex-ante problem (11) also preserves a threshold property and $\bar{V}_{t,\boldsymbol{c}}^{\text{Fld}}$ (11) can be re-written into the following formulation:

$$\begin{aligned} \bar{V}_{t,\boldsymbol{c}}^{\text{Fld}} = \max & \sum_{j=1}^{n} p_j \cdot s \cdot \int_{q=1-q_j}^{1} F_j^{-1}(q)dq \\ \text{s.t. } & \sum_{j=1}^{n} p_j \cdot s \cdot a_{j,i} \cdot q_j \le c_i, \quad \forall i \in [m] \\ & q_j \in [0,1], \quad \forall j \in [n] \end{aligned} \tag{13}$$

Note that the formulation of $\bar{V}_{t,\boldsymbol{c}}^{\text{Fld}}$ deviates from the formulation of $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\tilde{\boldsymbol{d}}_t)$ only in that the random variable $\tilde{d}_{j,t}$ is changed into its expectation $p_j \cdot s$. We can bound how this change of parameter would result in a change of the expected reward that we can gain from each type of query in the optimization problem (12) and (13). Our analysis generalizes the Lipschitz analysis in Mangasarian and Shiau (1987) from linear programming to a general convex optimization problem. Our argument is formlized in the following lemma, where the proof is relegated to Appendix B.

Lemma 3. *There exists a constant μ such that for any $\boldsymbol{c} \geq 0$ and any optimal solution $\hat{\boldsymbol{q}}^* = (\hat{q}_j^*)_{j=1}^n$ to (13), it holds that*

$$\|\hat{\boldsymbol{q}}^* - \bar{\boldsymbol{q}}^*\|_\infty \leq \max_{j\in[n]}\left\{\frac{\mu}{p_j}\right\} \cdot \max_{j'\in[n]}\{|p_{j'} - d_{j'}/s|\} + \max_{j\in[n]}\left\{\left|\frac{d_j}{p_j \cdot s} - 1\right|\right\}, \ \forall j \in [n]$$

for any sample path I_t, where $s = T - t + 1$, $\bar{\boldsymbol{q}}^* = (\bar{q}_j^*)_{j=1}^n$ denotes one optimal solution to (12).

Note that the constant μ in Lemma 3 depends only on the matrix $A = (a_{i,j})_{\forall i\in[m],\forall j\in[n]} \in \mathbb{R}^{n\times n}$, and is thus common for any $s, \boldsymbol{c}$ and any parameters $\{l_j, u_j, \alpha\}_{j=1}^n$ given in Assumption 1. The exact formulation for the constant μ is given in (72) in Appendix B. Lemma 3 implies the existence of a constant κ_1 such that

$$\max_{j\in[n]} |\hat{q}_j^* - \tilde{q}_j^*| \leq \kappa_1 \cdot \max_{j\in[n]}\{|d_j/s - p_j|\}$$

which will drive our regret analysis. The formal policy is given in Algorithm 2, which also depends on the constant κ_1 (an upper bound on κ_1 suffices). We now provide the regret analysis.

Algorithm 2 Algorithm achieving $O(\log^2 T)$ Regret

```
Input: the remaining inventory $\boldsymbol{c}$ and the type of query $t$, denoted by $j_t$.
Obtain $\{\hat{q}^*_{j,t}\}$ by solving the optimization problem (13).
if $\hat{q}^*_{j_t,t} \geq 1 - 3\kappa_1 \cdot \sqrt{\frac{\log(T-t+1)}{T-t+1}}$, then we set $\hat{M}_{\boldsymbol{c},\boldsymbol{a}_{j_t}} = l_{j_t}$.
else if $\hat{q}^*_{j_t,t} \leq 3\kappa_1 \cdot \sqrt{\frac{\log(T-t+1)}{T-t+1}}$, then we set $\hat{M}_{\boldsymbol{c},\boldsymbol{a}_{j_t}} = u_{j_t}$.
else if $3\kappa_1 \cdot \sqrt{\frac{\log(T-t+1)}{T-t+1}} \leq \hat{q}^*_{j_t,t} \leq 1 - 3\kappa_1 \cdot \sqrt{\frac{\log(T-t+1)}{T-t+1}}$, then we set $\hat{M}_{\boldsymbol{c},\boldsymbol{a}_{j_t}} = F^{-1}(1 - \hat{q}^*_{j_t,t}|\boldsymbol{a}_{j_t})$.
Output: $\hat{M}_{\boldsymbol{c},\boldsymbol{a}_{j_t}}$
```

Theorem 1. *Suppose that the estimator $\hat{M}$ is given in Algorithm 2. Then, for any $\boldsymbol{c} \geq 0$, as long as $s = T - t + 1 \geq s_0$ for a constant $s_0 \geq 0$ that solely depends on $\{p_j\}_{j=1}^n$, it holds that*

$$\mathit{Myopic}_t(\pi, \boldsymbol{c}) \leq \frac{\kappa_2 \cdot \log s}{s}$$

for a constant $\kappa_2 > 0$.

Proof of Theorem 1:. We have that

$$\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_t) = \sum_{j=1}^n \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}^*_{j,t}}^1 F_j^{-1}(q)dq + \int_{q=1-\tilde{q}^*_{j_t,t}}^1 F_{j_t}^{-1}(q)dq. \tag{14}$$

where we denote by j_t the type of query t in the instance I. Now we plug (14) into the formulation (8). We get

$$\begin{aligned}\text{Myopic}_t(\pi, \boldsymbol{c}) =& \mathbb{E}_{j_t, I_{t+1}}\left[\sum_{j=1}^n \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}^*_{j,t}}^1 F_j^{-1}(q)dq + \int_{q=1-\tilde{q}^*_{j_t,t}}^1 F_{j_t}^{-1}(q)dq\right.\\ &\left. - \int_{q=1-q^\pi_{j_t,t}}^1 F_{j_t}^{-1}(q)dq - \mathbb{E}_{r\sim F_{j_t}(\cdot)}[\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}\cdot \tilde{x}^\pi_t(r)}(I_{t+1})]\right]\end{aligned} \tag{15}$$

where we denote by π our online policy Algorithm 1 with the estimator given in Algorithm 2. Then, $q^{\pi}_{j,t}$ denotes the ex-ante probability that query t will be served by the online policy π. With these notations, (15) can be re-written as

$$\begin{aligned}\text{Myopic}_t(\pi,\boldsymbol{c}) =& \mathbb{E}_{j_t,I_{t+1}}\Bigg[\sum_{j=1}^{n}\tilde{d}_{j,t+1}\cdot\int_{q=1-\tilde{q}^*_{j,t}}^{1}F_j^{-1}(q)dq+\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}}F_{j_t}^{-1}(q)dq\\ &-q^{\pi}_{j_t,t}\cdot\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})-(1-q^{\pi}_{j_t,t})\cdot\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})\Bigg]\end{aligned}$$

We construct feasible solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})$ and $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})$ to upper bound the myopic regret, for all the three cases. We identify a "good" event that $\tilde{q}^*_{j,t}$ is close to $\hat{q}^*_{j,t}$ for each $j\in[n]$. Following Lemma 3, we know that

$$|\tilde{q}^*_{j,t}-\hat{q}^*_{j,t}|\leq\kappa_1\cdot\max_{j'}\{|p_{j'}-\tilde{d}_{j',t+1}/(s-1)|\}. \tag{16}$$

for a constant $\kappa_1>0$. We note that for each $j\in[n]$, $\tilde{d}_{j,t+1}$ is a binomial distribution with mean $p_j\cdot(s-1)$. Then, from Hoeffding's inequality (Lemma 12), we have

$$P(|\tilde{d}_{j,t+1}-p_j(s-1)|\leq\sqrt{4(s-1)\log(s-1)})\geq 1-2\exp(-2\log(s-1))\geq 1-\frac{2}{(s-1)^2}\geq 1-\frac{1}{n(s-1)}$$

as long as $s\geq s_0$ for a constant s_0. We denote by the event

$$\mathcal{G}=\{|\tilde{d}_{j,t+1}-p_j(s-1)|\leq\sqrt{4(s-1)\log(s-1)},\forall j\in[n]\},$$

which is the "good" event that $\tilde{d}_{j,t+1}$ is close to its mean. From union bound, we know that

$$P(\mathcal{G})\geq 1-\sum_{j=1}^{n}(1-P(|\tilde{d}_{j,t+1}-p_j(s-1)|\leq\sqrt{4(s-1)\log(s-1)}))\geq 1-\frac{1}{s-1}.$$

We denote by

$$\begin{aligned}\text{Myopic}_t(\pi,\boldsymbol{c},\mathcal{G}) =& P(\mathcal{G})\cdot\mathbb{E}_{j_t,I_{t+1}}\Bigg[\sum_{j=1}^{n}\tilde{d}_{j,t+1}\cdot\int_{q=1-\tilde{q}^*_{j,t}}^{1}F_j^{-1}(q)dq+\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}}F_{j_t}^{-1}(q)dq\\ &-q^{\pi}_{j_t,t}\cdot\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})-(1-q^{\pi}_{j_t,t})\cdot\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})\mid\mathcal{G}\Bigg]\end{aligned}$$

and

$$\begin{aligned}\text{Myopic}_t(\pi,\boldsymbol{c},\mathcal{G}^c) =& P(\mathcal{G}^c)\cdot\mathbb{E}_{j_t,I_{t+1}}\Bigg[\sum_{j=1}^{n}\tilde{d}_{j,t+1}\cdot\int_{q=1-\tilde{q}^*_{j,t}}^{1}F_j^{-1}(q)dq+\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}}F_{j_t}^{-1}(q)dq\\ &-q^{\pi}_{j_t,t}\cdot\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})-(1-q^{\pi}_{j_t,t})\cdot\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})\mid\mathcal{G}^c\Bigg]\end{aligned}$$

where $\mathcal{G}^c$ is the complement of $\mathcal{G}$. It is clear that

$$\text{Myopic}_t(\pi,\boldsymbol{c})=\text{Myopic}_t(\pi,\boldsymbol{c},\mathcal{G})+\text{Myopic}_t(\pi,\boldsymbol{c},\mathcal{G}^c)$$

We have a direct upper bound

$$\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}^c) \leq \frac{\max_{j\in[n]}\{u_j\}}{s-1}. \tag{17}$$

In what follows, we condition on the event $\mathcal{G}$ happens, and we bound $\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G})$.

Case (i) when $\hat{q}^*_{j_t,t} \geq 1 - 3\kappa_1 \cdot \sqrt{\frac{\log s}{s}}$: in this case, we know that

$$\boldsymbol{c} \geq p_{j_t} \cdot s \cdot \hat{q}^*_{j_t,t} \cdot \boldsymbol{a}_{j_t} \geq \frac{p_{j_t} \cdot s \cdot \boldsymbol{a}_{j_t}}{2} \geq \boldsymbol{a}_{j_t}$$

when $s \geq s_0$ for a constant $s_0 \geq 0$. Therefore, we always have enough remaining capacity to serve query t with type j_t.

Since we have $q^\pi_{j_t,t} = 1$, we only need to construct feasible solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})$. From the feasibility of $\{\tilde{q}^*_{j,t}\}$, we know that

$$\sum_{j=1}^n \tilde{d}_{j,t+1} \cdot a_{j,i} \cdot \tilde{q}^*_{j,t} + a_{j_t,i} \cdot \tilde{q}^*_{j_t,t} \leq c_i, \quad \forall i \in [m]. \tag{18}$$

Note that conditioning on the event $\mathcal{G}$, following (16), we have

$$|\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t}| \leq 2\kappa_1 \cdot \sqrt{\frac{\log(s-1)}{s-1}}$$

which implies $\tilde{q}^*_{j_t,t} \geq 1 - 5\kappa_1 \cdot \sqrt{\frac{\log s}{s}} \geq \frac{1}{2}$ when $s \geq s_0$ for a constant $s_0 > 0$. We construct the following solution $\{\tilde{q}'_{j,t}\}$ for $\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})$ satisfying

$$\tilde{q}'_{j,t} = \tilde{q}^*_{j,t}, \forall j \neq j_t \quad \text{and} \quad \tilde{q}'_{j_t,t} = \tilde{q}^*_{j_t,t} + \frac{\tilde{q}^*_{j_t,t} - 1}{\tilde{d}_{j_t,t+1}}. \tag{19}$$

Since $\sum_{j=1}^n \tilde{d}_{j,t+1} \cdot a_{j,i} \cdot \tilde{q}'_{j,t} \leq c_i - a_{j_t,i}$ for each $i \in [m]$, we know that $\{\tilde{q}'_{j,t}\}$ is a feasible solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(I_{t+1})$, where $\tilde{q}'_{j_t,t} \geq 0$ follows from $\tilde{q}^*_{j_t,t} \geq \frac{1}{2}$ and $\tilde{d}_{j_t,t+1} \geq 1$ conditioning on the event $\mathcal{G}$. Therefore, we have that

$$\begin{aligned} \text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}) \leq & \mathbb{E}_{j_t, I_{t+1}}\left[\sum_{j=1}^n \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}^*_{j,t}}^1 F_j^{-1}(q)dq + \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^\pi_{j_t,t}} F_{j_t}^{-1}(q)dq \right. \\ & \left. - \sum_{j=1}^n \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}'_{j,t}}^1 F_j^{-1}(q)dq \mid \mathcal{G}\right] \\ = & \mathbb{E}_{j_t, I_{t+1}}\left[\tilde{d}_{j_t,t+1} \cdot \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-\tilde{q}'_{j_t,t}} F_{j_t}^{-1}(q)dq + \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^\pi_{j_t,t}} F_{j_t}^{-1}(q)dq \mid \mathcal{G}\right] \end{aligned} \tag{20}$$

We make the following claim.

Claim 1. *For any $q_1, q_2 \in [0,1]$, it holds that*

$$\int_{q=q_1}^{q_2} F_j^{-1}(q)dq \leq F_j^{-1}(q_1) \cdot (q_2 - q_1) + \frac{(q_2 - q_1)^2}{\alpha}$$

for any $j \in [n]$, where α is the lower bound of the density function $f(\cdot|\boldsymbol{a}_j)$ specified in Assumption 1.

The proof of Claim 1 is relegated to Appendix B. Therefore, applying Claim 1, we have

$$\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-\tilde{q}'_{j_t,t}} F^{-1}_{j_t}(q)dq \le F^{-1}_{j_t}(1-\tilde{q}^*_{j_t,t})\cdot \frac{1-\tilde{q}^*_{j_t,t}}{\tilde{d}_{j_t,t+1}} + \frac{(1-\tilde{q}^*_{j_t,t})^2}{\alpha\cdot \tilde{d}^2_{j_t,t+1}} \tag{21}$$

and

$$\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}} F^{-1}_{j_t}(q)dq \le F^{-1}_{j_t}(1-\tilde{q}^*_{j_t,t})\cdot (\tilde{q}^*_{j_t,t}-1) + \frac{(1-\tilde{q}^*_{j_t,t})^2}{\alpha} \tag{22}$$

Plugging (21) and (22) into (20), we have

$$\begin{aligned}\text{Myopic}_t(\pi,\boldsymbol{c},\mathcal{G}) &\le \mathbb{E}_{j_t,I_{t+1}}\left[\frac{(1-\tilde{q}^*_{j_t,t})^2}{\alpha}+\frac{(1-\tilde{q}^*_{j_t,t})^2}{\alpha\cdot\tilde{d}_{j_t,t+1}} \mid \mathcal{G}\right]\cdot P(\mathcal{G}) \\ &\le 2\mathbb{E}_{j_t,I_{t+1}}\left[\frac{(1-\hat{q}^*_{j_t,t})^2}{\alpha}+\frac{(\tilde{q}^*_{j_t,t}-\hat{q}^*_{j_t,t})^2}{\alpha}+\frac{1}{\alpha\cdot\tilde{d}_{j_t,t+1}} \mid \mathcal{G}\right]\cdot P(\mathcal{G}) \\ &\le \frac{2\log s}{\alpha\cdot s}+\frac{2}{\alpha}\cdot\mathbb{E}_{j_t,I_{t+1}}[(\tilde{q}^*_{j_t,t}-\hat{q}^*_{j_t,t})^2]+\frac{2}{\alpha}\cdot\mathbb{E}_{j_t,I_{t+1}}\left[\frac{1}{\tilde{d}_{j_t,I_{t+1}}} \mid \mathcal{G}\right].\end{aligned} \tag{23}$$

Case (ii) when $\hat{q}^*_{j_t,t}\le 3\kappa_1\cdot\sqrt{\frac{\log s}{s}}$: since we have $q^{\pi}_{j_t,t}=0$, we only need to construct feasible solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})$. Note that conditioning on the event $\mathcal{G}$, following (16), we have

$$|\tilde{q}^*_{j_t,t}-\hat{q}^*_{j_t,t}|\le 2\kappa_1\cdot\sqrt{\frac{\log(s-1)}{s-1}}$$

which implies $\tilde{q}^*_{j_t,t}\le 5\kappa_1\cdot\sqrt{\frac{\log s}{s}}\le\frac{1}{2}$ as long as $s\ge s_0$ for a constant $s_0\ge 0$. From the feasibility of $\{\tilde{q}^*_{j,t}\}$ demonstrated in (18), we construct the following solution $\{\tilde{q}''_{j,t}\}$ for $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})$ satisfying

$$\tilde{q}''_{j,t}=\tilde{q}^*_{j,t},\forall j\neq j_t \quad\text{and}\quad \tilde{q}''_{j_t,t}=\tilde{q}^*_{j_t,t}\cdot\frac{\tilde{d}_{j_t,t+1}+1}{\tilde{d}_{j_t,t+1}}. \tag{24}$$

Since $\sum_{j=1}^n \tilde{d}_{j,t+1}\cdot a_{j,i}\cdot\tilde{q}''_{j,t}\le c_i$ for each $i\in[m]$, we know that $\{\tilde{q}''_{j,t}\}$ is a feasible solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})$, where $\tilde{q}''_{j_t,t}\le 1$ follows from $\tilde{q}^*_{j_t,t}\le\frac{1}{2}$ and $\tilde{d}_{j_t,t+1}\ge 1$ conditioning on the event $\mathcal{G}$. Therefore, we have that

$$\begin{aligned}\text{Myopic}_t(\pi,\boldsymbol{c},\mathcal{G}) \le&\mathbb{E}_{j_t,I_{t+1}}\left[\sum_{j=1}^n\tilde{d}_{j,t+1}\cdot\int_{q=1-\tilde{q}^*_{j,t}}^{1}F^{-1}_j(q)dq+\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}}F^{-1}_{j_t}(q)dq-\sum_{j=1}^n\tilde{d}_{j,t+1}\cdot\int_{q=1-\tilde{q}''_{j,t}}^{1}F^{-1}_j(q)dq \mid \mathcal{G}\right] \\ =&\mathbb{E}_{j_t,I_{t+1}}\left[\tilde{d}_{j_t,t+1}\cdot\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-\tilde{q}''_{j_t,t}}F^{-1}_{j_t}(q)dq+\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}}F^{-1}_{j_t}(q)dq \mid \mathcal{G}\right]\end{aligned} \tag{25}$$

Applying Claim 1, we have that

$$\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-\tilde{q}''_{j_t,t}} F^{-1}_{j_t}(q)dq \le -F^{-1}_{j_t}(1-\tilde{q}^*_{j_t,t})\cdot\frac{\tilde{q}^*_{j_t,t}}{\tilde{d}_{j_t,t+1}}+\frac{(\tilde{q}^*_{j_t,t})^2}{\alpha\cdot\tilde{d}^2_{j_t,t+1}} \tag{26}$$

and

$$\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^{\pi}_{j_t,t}} F_{j_t}^{-1}(q)dq \le F_{j_t}^{-1}(1-\tilde{q}^*_{j_t,t}) \cdot \tilde{q}^*_{j_t,t} + \frac{(\tilde{q}^*_{j_t,t})^2}{\alpha} \tag{27}$$

Plugging (26) and (27) into (25), we get

$$\begin{aligned}
\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}) &\le \mathbb{E}_{j_t, I_{t+1}}\left[\frac{(\tilde{q}^*_{j_t,t})^2}{\alpha} + \frac{(\tilde{q}^*_{j_t,t})^2}{\alpha \cdot \tilde{d}_{j_t,t+1}} \mid \mathcal{G}\right] \cdot P(\mathcal{G}) \\
&\le 2\mathbb{E}_{j_t, I_{t+1}}\left[\frac{(\hat{q}^*_{j_t,t})^2}{\alpha} + \frac{(\hat{q}^*_{j_t,t} - \tilde{q}^*_{j_t,t})^2}{\alpha} + \frac{1}{\alpha \cdot \tilde{d}_{j_t,t+1}} \mid \mathcal{G}\right] \cdot P(\mathcal{G}) \\
&\le \frac{2\log s}{\alpha \cdot s} + \frac{2}{\alpha} \cdot \mathbb{E}_{j_t, I_{t+1}}[(\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t})^2] + \frac{2}{\alpha} \cdot \mathbb{E}_{j_t, I_{t+1}}\left[\frac{1}{\tilde{d}_{j_t,t+1}} \mid \mathcal{G}\right].
\end{aligned} \tag{28}$$

Case (iii) when $3\kappa_1 \cdot \sqrt{\frac{\log s}{s}} \le \hat{q}^*_{j_t,t} \le 1 - 3\kappa_1 \cdot \sqrt{\frac{\log s}{s}}$: in this case, we know that

$$\boldsymbol{c} \ge p_{j_t} \cdot s \cdot \hat{q}^*_{j_t,t} \cdot \boldsymbol{a}_{j_t} \ge p_{j_t} \cdot \sqrt{s \log s} \cdot \boldsymbol{a}_{j_t} \ge \boldsymbol{a}_{j_t}$$

when $s \ge s_0$ for a constant $s_0 \ge 0$ that depends solely on $\{p_j\}_{j=1}^n$. Therefore, we always have enough remaining capacity to serve query t with type j_t.

Note that conditioning on the event $\mathcal{G}$, following (16), we have

$$|\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t}| \le 2\kappa_1 \cdot \sqrt{\frac{\log(s-1)}{s-1}}$$

which implies $\kappa_1 \cdot \sqrt{\frac{\log s}{s}} \le \tilde{q}^*_{j_t,t} \le 1 - \kappa_1 \cdot \sqrt{\frac{\log s}{s}}$.

We construct feasible solution $\{\tilde{q}'_{j,t}\}$ for $\bar{V}^{\text{Semi}}_{\boldsymbol{c}-\boldsymbol{a}_{j_t}}(\tilde{\boldsymbol{d}}_{t+1})$ following the definition in (19). Then, $\tilde{q}'_{j_t,t} \ge 0$ follows from the fact that

$$\tilde{q}^*_{j_t,t} \ge \kappa_1 \cdot \sqrt{\frac{\log s}{s}} \ge \frac{2}{p_{j_t} \cdot (s-1)} \ge \frac{1}{\tilde{d}_{j_t,t+1}}$$

conditioning on the event $\mathcal{G}$, as long as $s \ge s_0$ for a constant $s_0 > 0$.

We construct feasible solution $\{\tilde{q}''_{j,t}\}$ for $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\tilde{\boldsymbol{d}}_{t+1})$ following the definition in (24). Then, $\tilde{q}''_{j_t,t} \le 1$ follows from

$$\tilde{q}''_{j_t,t} \le \tilde{q}^*_{j_t,t} + \frac{1}{\tilde{d}_{j_t,t+1}} \le 1 - \kappa_1 \cdot \sqrt{\frac{\log s}{s}} + \frac{3}{2p_{j_t}(s-1)} \le 1$$

conditioning on the event $\mathcal{G}$, as long as $s \ge s_0$ for a constant $s_0 > 0$.

Therefore, we have that

$$\begin{aligned}\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}) \leq & \mathbb{E}_{j_t, I_{t+1}} \left[\sum_{j=1}^{n} \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}^*_{j,t}}^{1} F_j^{-1}(q)dq + \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^\pi_{j_t,t}} F_{j_t}^{-1}(q)dq \right. \\ & \left. - q^\pi_{j_t,t} \cdot \sum_{j=1}^{n} \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}'_{j,t}}^{1} nF_j^{-1}(q)dq - (1-q^\pi_{j_t,t}) \cdot \sum_{j=1}^{n} \tilde{d}_{j,t+1} \cdot \int_{q=1-\tilde{q}''_{j,t}}^{1} F_j^{-1}(q)dq \mid \mathcal{G} \right] \cdot P(\mathcal{G}) \\ = & \mathbb{E}_{j_t, I_{t+1}} \left[q^\pi_{j_t,t} \cdot \tilde{d}_{j_t,t+1} \cdot \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-\tilde{q}'_{j_t,t}} F_{j_t}^{-1}(q)dq + (1-q^\pi_{j_t,t}) \cdot \tilde{d}_{j_t,t+1} \cdot \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-\tilde{q}''_{j_t,t}} F_{j_t}^{-1}(q)dq \right. \\ & \left. + \int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^\pi_{j_t,t}} F_{j_t}^{-1}(q)dq \mid \mathcal{G} \right] \cdot P(\mathcal{G})\end{aligned} \tag{29}$$

Applying Claim 1, we get (21), (26), and

$$\int_{q=1-\tilde{q}^*_{j_t,t}}^{1-q^\pi_{j_t,t}} F_{j_t}^{-1}(q)dq = F_{j_t}^{-1}(1-\tilde{q}^*_{j_t,t}) \cdot (\tilde{q}^*_{j_t,t} - q^\pi_{j_t,t}) + \frac{(\tilde{q}^*_{j_t,t} - q^\pi_{j_t,t})^2}{\alpha}. \tag{30}$$

Therefore, we have

$$\begin{aligned}\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}) & \leq \mathbb{E}_{j_t, I_{t+1}} \left[q^\pi_{j_t,t} \cdot \frac{(1-\tilde{q}^*_{j_t,t})^2}{\alpha \cdot \tilde{d}_{j_t,t+1}} + (1-q^\pi_{j_t,t}) \cdot \frac{(\tilde{q}^*_{j_t,t})^2}{\alpha \cdot \tilde{d}_{j_t,t+1}} + \frac{(\tilde{q}^*_{j_t,t} - q^\pi_{j_t,t})^2}{\alpha} \mid \mathcal{G} \right] \cdot P(\mathcal{G}) \\ & \leq \mathbb{E}_{j_t, I_{t+1}} \left[\frac{1}{\alpha \cdot d_{j_t,t+1}} + \frac{(\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t})^2}{\alpha} \mid \mathcal{G} \right] \cdot P(\mathcal{G}).\end{aligned} \tag{31}$$

From (23), (28) and (31), for all cases, it holds that

$$\begin{aligned}\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}) & \leq \frac{2\log s}{\alpha \cdot s} + \frac{2}{\alpha} \cdot \mathbb{E}_{j_t, I_{t+1}}[(\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t})^2] + \frac{2}{\alpha} \cdot \mathbb{E}_{j_t, I_{t+1}} \left[\frac{1}{\tilde{d}_{j_t, I_{t+1}}} \mid \mathcal{G} \right] \\ & \leq \frac{2\log s}{\alpha \cdot s} + \frac{2}{\alpha} \cdot \mathbb{E}_{j_t, I_{t+1}}[(\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t})^2] + \frac{4}{\alpha(s-1)} \cdot \sum_{j=1}^{n} \frac{1}{p_j}.\end{aligned}$$

From Lemma 3, we know that there exists a constant κ_2 such that

$$\mathbb{E}_{j_t, I_{t+1}}[(\tilde{q}^*_{j_t,t} - \hat{q}^*_{j_t,t})^2] \leq \kappa_2 \cdot \sum_{j=1}^{n} (p_j - \tilde{d}_{j,t}/s)^2 \leq \frac{\kappa_2}{s}$$

Therefore, we know that

$$\text{Myopic}_t(\pi, \boldsymbol{c}, \mathcal{G}) \leq \frac{2\log s}{\alpha \cdot s} + \frac{2\kappa_2}{\alpha \cdot s} + \frac{4}{\alpha(s-1)} \cdot \sum_{j=1}^{n} \frac{1}{p_j}.$$

Together with (17), our proof is completed. □

Following Lemma 1 and Theorem 1, we can see that the regret of the policy π given in Algorithm 2 can be upper bounded by

$$\text{Regret}(\pi) \leq \sum_{s=1}^{T} \frac{\kappa_2 \cdot \log s}{s} \leq \kappa_2 \cdot \log^2 T$$

which is our final regret bound.

4. Policy with Logarithmic Regret under Second-order Growth

In this section, we derive an improved logarithmic regret bound for (6) under the following stronger assumption.

Assumption 2. *We assume that for each $j \in [n]$, conditional on $\tilde{\boldsymbol{a}}_t$ being realized as any $\boldsymbol{a}_j \in \mathcal{A}$, the reward distribution of $\tilde{r}$ is supported on the interval $[0, u_j]$ with a density function $f(\cdot|\boldsymbol{a}_j)$, where $u_j \geq 0$, and it satisfies $\beta \geq f(r|\boldsymbol{a}_j) \geq \alpha$, for a constant $\beta \geq \alpha > 0$, for any $r \in [0, u_j]$.*

Note that Assumption 2 is stronger than Assumption 1 in that we require $l_j = 0$ for each l_j in Assumption 1. Moreover, in Assumption 2, we require not only the lower bound α on the density, but also an upper bound β on the density $f(\cdot|\boldsymbol{a}_j)$ for each $j \in [n]$. Assumption 2 implies the following less restrictive assumption, which will be used in our proofs.

Assumption 3. *There exists a compact convex set $\Omega \subset \mathbb{R}^m_{\geq 0}$ such that for any $t \in [T]$, any $\boldsymbol{c}$ and any problem instance I_t, the relaxed offline optimum $\bar{V}_{t,\boldsymbol{c}}(I)$ (10) possesses one optimal dual solution $\tilde{\boldsymbol{\mu}}$ satisfying $\tilde{\boldsymbol{\mu}} \in \Omega$. Moreover, there exists two positive constants $\underline{\alpha}, \bar{\alpha}$ such that for any $\boldsymbol{\mu}', \boldsymbol{\mu}'' \in \Omega$, it holds that*

$$\underline{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}}\left[(\tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}' - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}'')^2\right] \leq \mathbb{E}_{\tilde{\boldsymbol{a}}}\left[\left(F(\tilde{\boldsymbol{a}}^\top \mu'|\tilde{\boldsymbol{a}}) - F(\tilde{\boldsymbol{a}}^\top \mu''|\tilde{\boldsymbol{a}})\right) \cdot (\tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}' - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}'')\right] \leq \bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}}\left[(\tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}' - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}'')^2\right] \tag{32}$$

where $\tilde{\boldsymbol{a}}$ is a random variable denoting the size of one query and is realized as $\boldsymbol{a}_j$ with probability p_j, for $j \in [n]$.

We now provide an example to illustrate under which condition Assumption 3 holds or not.

Example 1. Consider a special case of our model where there is a single resource, i.e., $m = 1$, and two types of queries, i.e., $n = 2$. For each type $j = 1$ or 2, the size $a_j = 1$ and the reward distribution is a uniform distribution over the interval $[l_j, u_j]$ with $u_1 \leq u_2$. Note that Assumption 3 essentially requires that $u_1 \geq l_2$, i.e., the support for the reward distribution of each type of query overlaps with each other. In order to see this point, the set Ω can be specified as $[l_1, u_2]$ and it is clear to see that (32) will be satisfied. In contrast, if $u_1 < l_2$, then we can see that (32) will be violated by setting $\mu' = l_2$ and $\mu'' = u_1$. However, the results derived in Section 3 will still apply and a $O(\log^2 T)$ regret bound can be obtained.

The only reason why we make Assumption 3 is to establish a second-order growth condition of a dual function (we formalize in Lemma 6), which is a standard condition in the stochastic programming literature. Then, we apply results from the stochastic programming literature to derive our logarithmic regret bound (we formalize in Lemma 9). Note that the second-order growth condition has been assumed frequently in the previous literature (e.g. Bray (2022), Li and Ye (2021), Balseiro et al. (2021)). By deriving our results simply under Assumption 3, our contribution would be to get rid of a so-called "non-degeneracy" assumption, which concerns the "position" of the ex-ante relaxation $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$ and is different from the second-order growth condition. In the following part, we briefly illustrate the "non-degeneracy" and we provide a thorough comparison between our Assumption 3 and the assumptions made in previous literature in Section 5.

Comparison with assumptions made in previous literature. Notably, a common assumption made in the previous literature (Balseiro et al. 2021, Li and Ye 2021, Bray 2022) regarding logariathmic regret with continuous reward distribution is about the "non-degeneracy" of the ex-ante relaxation $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$. The "non-degeneracy" assumption not only requires the optimal solution to $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$ to be unique, but also requires *strict complementary slackness* condition to be satisfied by $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$, i.e., for any binding resource constraint in the optimal solution of $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$, the corresponding optimal dual variable must be strictly positive. In this way, the optimal basis of $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$ will remain unchanged if $\boldsymbol{C}$ is perturbed by a certain amount, which drives all the analysis in Balseiro et al. (2021), Li and Ye (2021), Bray (2022). In contrast, our Assumption 3 simply requires the support of the reward distribution of each type to "overlap" with each other, as shown in Example 1, and our goal of Assumption 3 is to establish the standard second-order growth condition of the dual function. We require nothing over the "position" of the ex-ante relaxation $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$, including unique optimal solution condition, strict complementary slackness condition, and so on.

We now provide an example where the strict complementary slackness condition of the ex-ante relaxation $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$ is violated, while our Assumption 3 is still satisfied.

Example 2. Consider a model with 3 resources, each with an initial capacity of $T\cdot\frac{2\varepsilon(1+\varepsilon)}{1+5\varepsilon}$. There are 3 types of queries, denoted by $j=1,2,3$. The size of each type of query is $\boldsymbol{a}_1=(0,1,1)$, $\boldsymbol{a}_2=(1,0,1)$ and $\boldsymbol{a}_3=(1,1,0)$. The reward distribution of each type of query is a uniform distribution over $[0,1]$ and the arrival probability is $p_1=\frac{2\varepsilon}{1+5\varepsilon}$, $p_2=\frac{1+\varepsilon}{1+5\varepsilon}$ and $p_3=\frac{2\varepsilon}{1+5\varepsilon}$. Clearly, our Assumption 2 is satisfied (thus Assumption 3 satisfied). On the other hand, following (13), the ex-ante relaxation $\bar{V}^{\text{Fld}}/T$ can be formulated as:

$$\begin{aligned} \max\ & \sum_{j=1}^{3} p_j\cdot q_j\cdot(1-\frac{q_j}{2}) \\ \text{s.t.}\ & p_2q_2+p_3q_3\le\frac{2\varepsilon(1+\varepsilon)}{1+5\varepsilon} \\ & p_1q_1+p_3q_3\le\frac{2\varepsilon(1+\varepsilon)}{1+5\varepsilon} \\ & p_1q_1+p_2q_2\le\frac{2\varepsilon(1+\varepsilon)}{1+5\varepsilon} \\ & q_1,q_2,q_3\in[0,1] \end{aligned} \tag{33}$$

Denote by μ_i the dual variable for constraint for resource i. Then, it is clear to see that the primal dual pair $\mu_1^*=\mu_3^*=\frac{1-\varepsilon}{2}$, $\mu_2^*=0$ and $q_1^*=q_3^*=\frac{1+\varepsilon}{2}$, $q_2^*=\varepsilon$ is optimal to (33) by checking that the saddle point condition is satisfied, for any $\varepsilon > 0$. However, while the resource constraint is binding for every $i = 1, 2, 3$, the optimal dual variable $\mu_2^* = 0$, which shows that the *strict complementary slackness* condition is not satisfied for every $\varepsilon > 0$. Therefore, the "non-degeneracy" assumption is violated for (33).

4.1. Decomposition of the Myopic Regret

We now proceed to bound $\mathrm{Myopic}_t(\pi, \tilde{\boldsymbol{c}}_t^\pi)$ in (8) under Assumption 1 and Assumption 3. We use $\bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)$ (10) as the benchmark, which is a LP relaxation of $V^{\mathrm{off}}(I)$ (4) and thus provides more tractablity. We denote by $\{\tilde{x}_t^*\}$ one optimal solution of $\bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)$ (10), where $\tilde{x}_\tau^* \in [0,1]$ for each $\tau = t, \dots, T$. Then, a gap arises from the fact that the online decision for our policy π must be binary as required by problem formulation, while the optimal solution $\tilde{x}_t^*$ in the relaxation (10) can be fractional. In order to deal with this gap, we introduce a "rounded" relaxed offline optimum as an intermediate. To be specific, we denote by

$$\tilde{x}_t^{\mathrm{round}} = \begin{cases} 1, & \text{if } \tilde{r}_t \geq M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1}) = \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi}(I_{t+1}) - \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi - \tilde{\boldsymbol{a}}_t}(I_{t+1}) \text{ and } \tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

as a rounding of $\tilde{x}_t^*$. Then, we have that

$$\bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) = \tilde{x}_t^{\mathrm{round}} \cdot (\tilde{r}_t + \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi - \tilde{\boldsymbol{a}}_t}(I_{t+1})) + (1 - \tilde{x}^{\mathrm{round}}) \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi}(I_{t+1}) + G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) \tag{35}$$

where $G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)$ denotes the gap caused by rounding $\tilde{x}_t^*$ to be $\tilde{x}_t^{\mathrm{round}}$, which can be formulated as follows

$$G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t) = \tilde{r}_t(\tilde{x}_t^* - \tilde{x}_t^{\mathrm{round}}) + \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi - \tilde{\boldsymbol{a}}_t \tilde{x}_t^*}(I_{t+1}) - \left(\tilde{x}_t^{\mathrm{round}} \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi - \tilde{\boldsymbol{a}}_t}(I_{t+1})) + (1 - \tilde{x}_t^{\mathrm{round}}) \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^\pi}(I_{t+1}) \right). \tag{36}$$

Introducing the rounded gap $G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)$ allows us to further decompose the regret of our policy into three terms, as formalized in the following lemma, where the proof is relegated to Appendix C.

Lemma 4. *For any $t \in [T]$, it holds that*

$$\begin{aligned} \mathit{Myopic}_t(\pi, \tilde{\boldsymbol{c}}_t^\pi) \leq & 2\bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \mathrm{Var}(M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1}))] + \mathbb{E}_{I_t}[G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)] \\ & + 2\bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \left(\hat{M}_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t} - \mathbb{E}_{I_{t+1}}[M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1})] \right)^2 \right] \end{aligned} \tag{37}$$

where the variance Var *is taken over the problem instance I_{t+1}.*

A key step in deriving Lemma 4 is to utilize the relationship

$$\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_t) = \tilde{r}_t \cdot \tilde{x}_t^* + \bar{V}^{\mathrm{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^*}(I_{t+1}). \tag{38}$$

We note that the above backward induction holds only on each sample path I, which is the reason why we use a sample-path based relaxed offline optimum $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_t)$ (10) as the benchmark.

In contrast, if we use a benchmark that is *independent* of the sample path I, for example, the deterministic relaxation $\bar{V}^{\mathrm{Fld}}$ (11) as considered in Li and Ye (2021) and Balseiro et al. (2021). Then, we only have the following induction:

$$\bar{V}^{\mathrm{Fld}}_{t,\boldsymbol{c}} = \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r}\cdot x^*(\tilde{r},\tilde{\boldsymbol{a}})] + \bar{V}^{\mathrm{Fld}}_{t+1,\boldsymbol{c}-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{\boldsymbol{a}}\cdot x^*(\tilde{r},\tilde{\boldsymbol{a}})]}. \tag{39}$$

If we plug (39) into the proof of Lemma 4 to derive a similar decomposition of the myopic term $\mathrm{Myopic}_t(\pi,\tilde{\boldsymbol{c}}^\pi_t)$ that is now defined with respect to $\bar{V}^{\mathrm{Fld}}_{t,\boldsymbol{c}}$, then we would have an additional term $\bar{V}^{\mathrm{Fld}}_{t+1,\boldsymbol{c}-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{\boldsymbol{a}}\cdot x^*(\tilde{r},\tilde{\boldsymbol{a}})]} - \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\bar{V}^{\mathrm{Fld}}_{t+1,\boldsymbol{c}-\tilde{\boldsymbol{a}}\cdot x^*(\tilde{r},\tilde{\boldsymbol{a}})}]$ showing up in the decomposition. This additional term would require a "non-degeneracy" assumption (we further discuss in Section 5) to deal with, as shown in Li and Ye (2021) and Balseiro et al. (2021). The above discussion reveals the benefits of considering a sample-path based benchmark $\bar{V}^{\mathrm{Off}}_{t,\boldsymbol{c}}(I_t)$, which enables us to get rid of the "non-degeneracy" assumption at the first hand. The same idea of using the relaxation $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_t)$ and relationship (38) has been developed in Vera and Banerjee (2021) to get rid of the "non-degeneracy" assumption when the reward for each type is deterministic. We generalize this idea to allow the reward for each type having a continuous distribution.

For the RHS of (37), we refer the first term as the variation gap, the second term as the rounding gap, and the third term as the estimator gap. In what follows, we proceed to bound the three terms separately. Our goal is to show that each gap can be bounded at the order of $O(\frac{1}{T-t})$, and these bounds together will imply a $O(\log T)$ bound for Algorithm 1.

Our analysis relies on considering the dual problem of $\bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}^\pi_t}(I_{t+1})$ and $\bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}^\pi_t-\tilde{\boldsymbol{a}}_t}(I_{t+1})$ that define $M_{\tilde{\boldsymbol{c}}^\pi_t,\tilde{\boldsymbol{a}}_t}(I_{t+1})$ in (8), for each fixed $\tilde{\boldsymbol{c}}^\pi_t$ and $\tilde{\boldsymbol{a}}_t$ satisfying $\tilde{\boldsymbol{c}}^\pi_t \ge \tilde{\boldsymbol{a}}_t$. Now for any $\boldsymbol{c}\ge 0$, we introduce a dual variable $\boldsymbol{\mu}$ for the constraints of $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_{t+1})$ (10) and we denote by the function

$$\begin{aligned} L^{\mathrm{Off}}_{\boldsymbol{c},I_{t+1}}(\mu) &:= \max_{\tilde{x}_\tau\in[0,1],\forall\tau=t+1,\dots,T} \left(\frac{\boldsymbol{c}}{s-1}\right)^\top \mu + \frac{1}{s-1}\cdot\sum_{\tau=t+1}^{T}[\tilde{r}_\tau - \tilde{\boldsymbol{a}}_\tau^\top\mu]\cdot\tilde{x}_\tau \\ &= \left(\frac{\boldsymbol{c}}{s-1}\right)^\top \mu + \frac{1}{s-1}\cdot\sum_{\tau=t+1}^{T}[\tilde{r}_\tau - \tilde{\boldsymbol{a}}_\tau^\top\mu]^+. \end{aligned} \tag{40}$$

as the dual function of $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_{t+1})$, scaled by $\frac{1}{s-1}$, with $s=T-t+1$. We now proceed to bound the variation gap, the rounding gap, and the estimator gap. In Section 4.2, we show that bounding the first two gaps can be reduced to bounding a so-called "dual convergence", which concerns the variance of one optimal dual variable for minimizing the dual function $L^{\mathrm{Off}}_{\boldsymbol{c},I_{t+1}}(\mu)$ (40). Then, we propose our $\hat{M}-$estimator and bound the "dual convergence" in Section 4.3 to complete our final bound over the myopic regret.

4.2. Reduction to Dual Convergence

In this section, we show how to reduce bounding each term in (37) to bounding the "dual convergence", and we also propose our $\hat{M}-$estimator.

We first bound the variation gap $\mathrm{Var}(M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}))$ for each fixed $\boldsymbol{c}$ and $\boldsymbol{a}_t$ satisfying $\boldsymbol{c} \geq \boldsymbol{a}_t$. We denote by

$$\tilde{\boldsymbol{\mu}}_1 \in \arg\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Off}}_{\boldsymbol{c},I_{t+1}}(\boldsymbol{\mu}) \quad \text{and} \quad \tilde{\boldsymbol{\mu}}_2 \in \arg\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Off}}_{\boldsymbol{c}-\boldsymbol{a}_t,I_{t+1}}(\boldsymbol{\mu}) \tag{41}$$

Note that $\tilde{\boldsymbol{\mu}}_1$ (resp. $\tilde{\boldsymbol{\mu}}_2$) is one optimal dual variable of $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_{t+1})$ (resp. $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}-\boldsymbol{a}_t}(I_{t+1})$). Also, $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ are random variables, where the randomness comes from the randomness of the problem instance I_{t+1}. The goal of introducing $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ is to lower-bound and upper-bound $M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})$ in (8), as in the following lemma.

Lemma 5. *For any problem instance I_{t+1} and any $\boldsymbol{c} \geq \boldsymbol{a}_t$, it holds that*

$$\boldsymbol{a}_t^\top \tilde{\mu}_1 \leq M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}) = \bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_{t+1}) - \bar{V}^{\mathrm{Off}}_{\boldsymbol{c}-\boldsymbol{a}_t}(I_{t+1}) \leq \boldsymbol{a}_t^\top \tilde{\mu}_2. \tag{42}$$

where $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ are defined in (41).

We proceed to bound $\mathrm{Var}(M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}))$ with the help of Lemma 5. We note that the function $L^{\mathrm{Off}}_{\boldsymbol{c},I_{t+1}}(\boldsymbol{\mu})$ can be regarded as a sample average approximation of the following stochastic optimization problem:

$$\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu}) := \left(\frac{\boldsymbol{c}}{s-1}\right)^\top \boldsymbol{\mu} + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})\sim F}[\tilde{r} - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}]^+ \tag{43}$$

with $s-1 = T-t$ samples. Our analysis relies on showing the second-order growth condition of the "limiting" dual function $L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}$ defined in (43). Denote by $\mathcal{S}$ the subspace spanned by the resource consumption vector $\boldsymbol{a}$ of each query:

$$\mathcal{S} = \{\sum_{j\in[n]} \alpha_j \cdot \boldsymbol{a}_j : \forall \alpha_j \in \mathbb{R}\} \tag{44}$$

As a result of Assumption 3, we have the following second-order growth condition of the limiting dual function $L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}$ for any $\boldsymbol{c} \geq 0$.

Lemma 6. *For any $\boldsymbol{c} \geq 0$, we denote by $\boldsymbol{\mu}^* \in \arg\min\ L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu})$ such that $\boldsymbol{\mu}^* \in \Omega$ with Ω specified in Assumption 3. We also denote by $\mathcal{P}_S$ the projection of any vector to the subspace $\mathcal{S}$ (defined in (44)) spanned by the resource consumption vector $\boldsymbol{a}$ of each query. Then for any $\boldsymbol{\mu} \in \Omega$, it holds that*

$$L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu}) - L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu}^*) \geq \frac{\alpha\underline{\beta}}{2} \cdot \|\mathcal{P}_S(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\|_2^2$$

where $\underline{\beta}$ denotes the smallest positive eigenvalue of $\mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}} \cdot \tilde{\boldsymbol{a}}^\top]$.

The proof is relegated to Appendix C. Denote by

$$\hat{\boldsymbol{\mu}}_1 \in \arg\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu}) \quad \text{and} \quad \hat{\boldsymbol{\mu}}_2 \in \arg\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\boldsymbol{\mu}). \tag{45}$$

Following classical results from sample average approximation for stochastic programming (Shapiro 1993), we know that $\tilde{\boldsymbol{\mu}}_1$ (resp. $\tilde{\boldsymbol{\mu}}_2$) converges to $\hat{\boldsymbol{\mu}}_1$ (resp. $\hat{\boldsymbol{\mu}}_2$) in probability, as $s-1=T-t\to\infty$. Thus, we can use $\hat{\boldsymbol{\mu}}_1$ (resp. $\hat{\boldsymbol{\mu}}_2$) as an approximation of $\mathbb{E}[\tilde{\boldsymbol{\mu}}_1]$ (resp. $\mathbb{E}[\tilde{\boldsymbol{\mu}}_2]$) and obtain a bound over $\mathrm{Var}(\tilde{\boldsymbol{\mu}}_1)$ and $\mathrm{Var}(\tilde{\boldsymbol{\mu}}_2)$, which finally implies an upper bound of $\mathrm{Var}(M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}))$. We summarize the above arguments in the following lemma, which shows that we can reduce bounding the variation gap to bounding the "dual convergence" terms $\mathbb{E}_{I_{t+1}}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2]$ and $\mathbb{E}_{I_{t+1}}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2]$. The proof is relegated to Appendix C.

LEMMA 7. *There exists a constant $\kappa_1>0$ such that for any $\boldsymbol{c}\geq 0$ and any $\boldsymbol{a}_t$ satisfying $\boldsymbol{c}\geq\boldsymbol{a}_t$, it holds that*

$$\mathrm{Var}(M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}))\leq\frac{\kappa_1}{s-1}+\kappa_1\cdot\mathbb{E}_{I_{t+1}}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2]+\kappa_1\cdot\mathbb{E}_{I_{t+1}}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2]$$

with $s=T-t+1$, where $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ are defined in (41), $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$ are defined in (45), and $\mathbb{E}_{I_{t+1}}[\cdot]$ denotes taking expectation over I_{t+1} that decides the value of $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$.

we then reduce bounding the rounding gap (second term in (37)) to bounding the "dual convergence". Following previous notations, we denote by $\{\tilde{x}^*_\tau\}_{\tau=t}^T$ an optimal solution to $\bar{V}^{\mathrm{Off}}_{\boldsymbol{c}}(I_t)$ (10). Note that $\tilde{x}^*_t\in[0,1]$ can be fractional. We denote by $\tilde{x}^{\mathrm{round}}_t$ as the rounding of $\tilde{x}^*_t$ as in (34). We bound the rounding gap $\sum_{t=1}^T\mathbb{E}[G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)]$, where the formulation of $G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)$ is given in (36). Clearly, when $\tilde{x}^*_t=\tilde{x}^{\mathrm{round}}_t$, $G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)=0$. Thus, the rounding gap arises from the fact that $\tilde{x}^*_t\in[0,1]$ can be different from $\tilde{x}^{\mathrm{round}}_t\in\{0,1\}$.

A key observation can be summarized as follows: i) if $\boldsymbol{c}\geq\tilde{\boldsymbol{a}}_t$, then both $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ are well-defined in (41) and $\tilde{x}^*_t\neq\tilde{x}^{\mathrm{round}}_t$ happens if only $\tilde{\boldsymbol{a}}_t^\top\tilde{\boldsymbol{\mu}}_1\leq\tilde{r}_t\leq\tilde{\boldsymbol{a}}_t^\top\tilde{\boldsymbol{\mu}}_2$; ii) if $\boldsymbol{c}\geq\tilde{\boldsymbol{a}}_t$ does not hold and only $\tilde{\boldsymbol{\mu}}_1$ is well-defined in (41), then $\tilde{x}^*_t\neq\tilde{x}^{\mathrm{round}}_t$ happens if only $\tilde{\boldsymbol{a}}_t^\top\tilde{\boldsymbol{\mu}}_1\leq\tilde{r}_t$. Therefore, we can again involve the dual variables $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ (if well-defined) in bounding the rounding gap $G_{\boldsymbol{c}}(I_t)$ for any $\boldsymbol{c}\geq 0$, which reduces bounding the rounding gap into bounding the "dual convergence". We formalize the above arguments in the following lemma, where the proof is relegated to Appendix C.

LEMMA 8. *For any $\boldsymbol{c}$, it holds that*

$$\mathbb{E}_{I_t}\left[G_{\boldsymbol{c}}(I_t)\right]\leq\frac{\kappa_2}{s-1}+\kappa_2\cdot\mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top\tilde{\boldsymbol{\mu}}_1-\tilde{\boldsymbol{a}}_t^\top\hat{\boldsymbol{\mu}}_1)^2]\right]+\kappa_2\cdot\mathbb{E}_{\tilde{\boldsymbol{a}}}\left[\mathbb{1}_{\{\boldsymbol{c}\geq\tilde{\boldsymbol{a}}_t\}}\cdot\mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{a}}_t^\top\hat{\boldsymbol{\mu}}_2)^2]\right]$$

for a constant κ_2 that is independent of $\boldsymbol{c}$ and t, where $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ (if $\boldsymbol{c}\geq\tilde{\boldsymbol{a}}_t$ and well-defined) are defined in (41).

4.3. Bound on Dual Convergence and Policy

In this section, we first bound the "dual convergence" and then propose our $\hat{M}-$estimator. By further utilizing the second-order growth condition established in Lemma 6, we can show the following bound over the term $\mathbb{E}[(\boldsymbol{a}_t^\top \tilde{\boldsymbol{\mu}}_1 - \boldsymbol{a}_t^\top \hat{\boldsymbol{\mu}}_1)^2]$. We only state the "dual convergence" result for $\hat{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_1$. Since the bound is independent of $\boldsymbol{c}$, it is clear that the same bound also holds for $\hat{\boldsymbol{\mu}}_2$ and $\tilde{\boldsymbol{\mu}}_2$ whenever well-defined (i.e. $\boldsymbol{c} \geq \tilde{\boldsymbol{a}}_t$).

Lemma 9. *For any $\boldsymbol{c} \geq 0$, let $\hat{\boldsymbol{\mu}}_1$ be defined in (45) and $\tilde{\boldsymbol{\mu}}_1$ be defined in (41). Then, it holds that*

$$\mathbb{E}[(\boldsymbol{a}_t^\top \tilde{\boldsymbol{\mu}}_1 - \boldsymbol{a}_t^\top \hat{\boldsymbol{\mu}}_1)^2] \leq \frac{\delta_1}{s-1}$$

as long as $s-1 = T-t \geq t_0$, where δ_1 and t_0 are constants that are independent of s and $\boldsymbol{c}$.

The key of the proof of Lemma 9 is to regard $\tilde{\boldsymbol{\mu}}_1$ as the solution of the sample average approximation (definition in (41)) of the stochastic programming (43) with parameter $\boldsymbol{c}$, whose optimal solution is $\hat{\boldsymbol{\mu}}_1$ by definition (45). The second-order growth condition in Lemma 6 enables us to apply Theorem 2.1 in Shapiro (1993) showing that the gap between $\tilde{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_1$ is at the order of $\sqrt{\frac{1}{s-1}}$ with high probability, as $s-1 = T-t \to \infty$. This result is also known as asymptotic normality of sample average solution in the stochastic programming literature. Note that in order to apply Theorem 2.1 in Shapiro (1993), there are some additional conditions need to be satisfied. We verified that all these conditions are satisfied by our problem in the proof of Lemma 9, which is relegated to Appendix C. However, the high probability result cannot be directly translated into a bound over the L_2 norm. In order to obtain the bound over the L_2 norm, we further utilize the method developed in Li and Ye (2021). As a result, a combination of the methods from Shapiro (1993) and Li and Ye (2021) enables us to bound the L_2 norm at the order of $\frac{1}{s-1}$, which improves the $\frac{\log\log(s-1)}{s-1}$ bound established in Theorem 1 in Li and Ye (2021).

We now present our $\hat{M}-$estimator to complete our algorithm and provide the final regret bound. From Lemma 5, we know that $M_{c,\boldsymbol{a}_t}(I_{t+1})$ is close to $\boldsymbol{a}_t^\top \tilde{\boldsymbol{\mu}}_1$ and from the "dual convergence" established in Lemma 9, we know that $\hat{\boldsymbol{\mu}}_1$ is a good "approximate" of $\tilde{\boldsymbol{\mu}}_1$. Therefore, we use $\boldsymbol{a}_t^\top \hat{\boldsymbol{\mu}}_1$ as our $\hat{M}-$estimator, which is formalized in Algorithm 1. We provide the following regret bound for the $\hat{M}-$estimator in Algorithm 3 and the proof is relegated to Appendix C.

Theorem 2. *Suppose that the estimator $\hat{M}$ is given in Algorithm 3 and we denote by π the policy given in Algorithm 1. Then, it holds that*

$$\mathbb{E}_{\tilde{\boldsymbol{c}}_t^\pi}[Myopic_t(\pi, \tilde{\boldsymbol{c}}_t^\pi)] \leq \frac{\kappa_3}{T-t}$$

for a constant $\kappa_3 > 0$ that is independent of t.

Therefore, by applying Lemma 1 and Theorem 2, we can see that the regret of the policy given in Algorithm 3 is upper bounded by $O(\log T)$.

Algorithm 3 Algorithm achieving $O(\log T)$ Regret

```
1: Input: the remaining inventory $\tilde{\boldsymbol{c}}_t^\pi$ and size $\tilde{\boldsymbol{a}}_t$.
2: Obtain $\tilde{\boldsymbol{\mu}}_1$ by solving
       $$\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Fld}}_{\tilde{\boldsymbol{c}}_t^\pi, t+1}(\boldsymbol{\mu}) := \left(\frac{\tilde{\boldsymbol{c}}_t^\pi}{T-t}\right)^\top \boldsymbol{\mu} + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})\sim F}[\tilde{r} - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}]^+$$
3: Output: $\hat{M}_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t} = \boldsymbol{a}_t^\top \tilde{\boldsymbol{\mu}}_1$.
```

5. Detailed Comparison to Assumptions in Existing Literature

We conclude by making comparisons between our assumptions and the assumptions made in the existing literature, which can be summarized into the following two conditions:
(i). The non-degeneracy assumption over the ex-ante relaxation $\bar{V}^{\mathrm{Fld}}$, which requires the optimal solution to $\bar{V}^{\mathrm{Fld}}$ to be unique and strict complementary slackness condition being satisfied.
(ii). The second-order growth condition over the dual function L^{Fld}.
We first summarize the assumptions made in Li and Ye (2021) in the language of our paper as follows:

Assumption 4. *(in Li and Ye (2021)) The following conditions have to be satisfied:*
(i). $\tilde{r}_t$ and $\|\tilde{\boldsymbol{a}}_t\|_2$ are always bounded for each $t\in[T]$.
(ii). $\mathbf{C}$ scales linearly in T and $C_i/T \in (\underline{d}, \bar{d})$ with $\bar{d} \geq \underline{d} > 0$, for each $i\in[m]$.
(iii). The matrix $\mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}}\tilde{\boldsymbol{a}}^\top]$ is positive definite.
(iv). There exists a set $\Omega_{\boldsymbol{\mu}}$ containing all possible optimal dual variable to the offline optimum $\bar{V}^{\mathrm{Off}}$ (10) such that for any $\boldsymbol{\mu}\in\Omega_{\boldsymbol{\mu}}$ and any $\boldsymbol{c}\in\Omega_{\boldsymbol{c}} = [\underline{d}\cdot T, \bar{d}\cdot T]^m$, it holds that

$$\underline{\alpha}\cdot|\boldsymbol{a}^\top\boldsymbol{\mu} - \boldsymbol{a}^\top\boldsymbol{\mu}^*(\boldsymbol{c})| \leq |F(\boldsymbol{a}^\top\boldsymbol{\mu}|\boldsymbol{a}) - F(\boldsymbol{a}^\top\boldsymbol{\mu}^*(\boldsymbol{c})|\boldsymbol{a})| \leq \bar{\alpha}\cdot|\boldsymbol{a}^\top\boldsymbol{\mu} - \boldsymbol{a}^\top\boldsymbol{\mu}^*(\boldsymbol{c})|$$

for any $\boldsymbol{a}\in\mathcal{A}$, where $\boldsymbol{\mu}^*(\boldsymbol{c})$ is the optimal dual solution to the ex-ante relaxation $\bar{V}^{\mathrm{Fld}}_{1,\boldsymbol{c}}$ (11).
(v). For any $\boldsymbol{c}\in\Omega$, the optimal dual solution $\boldsymbol{\mu}^*(\boldsymbol{c})$ to the ex-ante relaxation $\bar{V}^{\mathrm{Fld}}_{1,\boldsymbol{c}}$ (11) satisfies the strict complementary condition.

It is shown in Proposition 2 of Li and Ye (2021) that condition (iv) in Assumption 4 implies the second-order growth condition of the dual function, while condition (iii), (iv) and (v) all together imply the non-degeneracy assumption. We summarize the assumptions made in Bray (2022) as follows:

Assumption 5. *(in Bray (2022)) The following conditions have to be satisfied:*
(i). We have $f(r|\boldsymbol{a}) \leq \bar{\alpha}$ for any $\boldsymbol{a}\in\mathcal{A}$ and any r in the support.
(ii). For any $\boldsymbol{a}\in\mathcal{A}$, $\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})\sim F}[\tilde{r}|\tilde{\boldsymbol{a}} = \boldsymbol{a}] \leq \beta$ for a constant $\beta > 0$.

(iii). $\|\boldsymbol{a}\|_2 \le \bar{d}$ *for any* $\boldsymbol{a} \in \mathcal{A}$.

(iv). The optimal dual solution to the Lagrangian problem $\min_{\boldsymbol{\mu}\ge 0} L^{\text{Fld}}_{\boldsymbol{c},1}$ *of the ex-ante relaxation* $\bar{V}^{\text{Fld}}_{1,\boldsymbol{c}}$ (11) *is unique and the strictly complementary slackness condition is satisfied, when* $\boldsymbol{c}$ *belongs to a neighborhood of* $\boldsymbol{C}$*, which scales linearly in* T.

(v). The Hessian matrix of $L^{\text{Fld}}_{\boldsymbol{C},1}(\boldsymbol{\mu}^*)$ *over* $\boldsymbol{\mu}^*$*, where* $\boldsymbol{\mu}^* = \text{argmin}_{\boldsymbol{\mu}\ge 0} L^{\text{Fld}}_{\boldsymbol{C},1}$ *is full rank (equivalently, positive definite).*

(vi). The Hessian matrix of $L^{\text{Fld}}_{\boldsymbol{C},1}(\boldsymbol{\mu})$ *over* $\boldsymbol{\mu}$ *is Lipschitz continuous when* $\boldsymbol{\mu}$ *belongs to a neighborhood of* $\boldsymbol{\mu}^* = \text{argmin}_{\boldsymbol{\mu}\ge 0} L^{\text{Fld}}_{\boldsymbol{C},1}$.

Note that condition (v) and (vi) in Assumption 5 together imply the second-order growth condition over the dual function, while the non-degeneracy assumption is stated in condition (iv) in Assumption 5. In particular, Balseiro et al. (2021) has summarized the assumptions in to the following two conditions:

Assumption 6 **(Assumption 2 in Balseiro et al. (2021))**. *The following conditions have to be satisfied:*

(i). The binding constraints for $\bar{V}^{\text{Fld}}_{1,\boldsymbol{c}}$ *remains the same as the binding constraints for* $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$*, as long as* $\boldsymbol{c}$ *belongs to a neighborhood of the initial capacity* $\boldsymbol{C}$*, where we denote by* $\mathcal{J}$ *the set of resource constraints that are binding.*

(ii). There exists a constant $\kappa > 0$ *such that*

$$\bar{V}^{\text{Fld}}_{1,\boldsymbol{c}} - \bar{V}^{\text{Fld}}_{1,\boldsymbol{C}} \ge \left(\nabla_{\boldsymbol{C}} \bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}\right)^\top \left(\frac{\boldsymbol{c}}{T} - \frac{\boldsymbol{C}}{T}\right) - \kappa \cdot \left(\frac{\boldsymbol{c}_{\mathcal{J}}}{T} - \frac{\boldsymbol{C}_{\mathcal{J}}}{T}\right)^2$$

for all $\boldsymbol{c}$ *belonging to a neighborhood of the initial capacity* $\boldsymbol{C}$*, where* $\boldsymbol{c}_{\mathcal{J}}$ *denotes* $\mathcal{J}$ *part of the vector* $\boldsymbol{c}$ *and* $\boldsymbol{C}_{\mathcal{J}}$ *denotes* $\mathcal{J}$ *part of the vector* $\boldsymbol{C}$.

It has been shown that a sufficient condition to guarantee condition (i) in Assumption 6 is that strict complementary slackness condition is satisfied by $\bar{V}^{\text{Fld}}_{1,\boldsymbol{C}}$ (SC 8 in Balseiro et al. (2021)), and a sufficient condition to guarantee condition (ii) in Assumption 6 is that the second-order growth condition is satisfied dual function L^{Fld}.

In Section 4, we need the second-order growth condition, but without the non-degeneracy assumption, and we derive a $O(\log T)$ regret bound following our myopic regret framework. In constrast, in Section 3, we get rid of both the second-order growth condition and the non-degeneracy assumption and consider our problem under the most general setting. Our main result is a $O(\log^2 T)$ regret bound.

References

S. Agrawal, Z. Wang, and Y. Ye. A dynamic near-optimal algorithm for online linear programming. *Operations Research*, 62(4):876–890, 2014.

A. Arlotto and I. Gurvich. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*, 9 (3):231–260, 2019.

A. Arlotto and X. Xie. Logarithmic regret in the dynamic and stochastic knapsack problem with equal rewards. *Stochastic Systems*, 10(2):170–191, 2020.

A. Asadpour, X. Wang, and J. Zhang. Online resource allocation with limited flexibility. *Management Science*, 66(2):642–666, 2020.

J. Baek and W. Ma. Bifurcating constraints to improve approximation ratios for network revenue management with reusable resources. *Operations Research*, 2022.

S. Balseiro, O. Besbes, and D. Pizarro. Survey of dynamic resource constrained reward collection problems: Unified model and analysis. *Available at SSRN 3963265*, 2021.

S. R. Balseiro and S. Xia. Uniformly bounded regret in dynamic fair allocation. *arXiv preprint arXiv:2205.12447*, 2022.

S. R. Balseiro, H. Lu, and V. Mirrokni. The best of many worlds: Dual mirror descent for online allocation problems. *Operations Research*, 2022.

O. Besbes, Y. Kanoria, and A. Kumar. The multisecretary problem with many types. *arXiv preprint arXiv:2205.09078*, 2022.

J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems.* Springer Science & Business Media, 2013.

R. Bray. Does the multisecretary problem always have bounded regret? *Available at SSRN 3497056*, 2019.

R. Bray. Logarithmic regret in multisecretary and online linear programming problems with continuous valuations. *arXiv preprint arXiv:1912.08917*, 2022.

N. Buchbinder and J. Naor. Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research*, 34(2):270–286, 2009.

P. Bumpensanti and H. Wang. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science*, 66(7):2993–3009, 2020.

F. H. Clarke. *Optimization and nonsmooth analysis.* SIAM, 1990.

H. Esfandiari, M. Hajiaghayi, V. Liaghat, and M. Monemizadeh. Prophet secretary. *SIAM Journal on Discrete Mathematics*, 31(3):1685–1701, 2017.

T. S. Ferguson. Who solved the secretary problem? *Statistical science*, 4(3):282–289, 1989.

D. Freund and S. Banerjee. Good prophets know when the end is near. *Available at SSRN 3479189*, 2019.

D. Freund and J. Zhao. Overbooking with bounded loss. In *Mathematics of Operations Research, forthcoming*, 2022.

A. Gupta and M. Molinaro. How experts can solve lps online. In *European Symposium on Algorithms*, pages 517–529. Springer, 2014.

P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather Modification; University of California Press: Berkeley, CA, USA*, page 221, 1967.

S. Jasin and S. Kumar. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research*, 37(2):313–345, 2012.

S. Jasin and S. Kumar. Analysis of deterministic lp-based booking limit and bid price controls for revenue management. *Operations Research*, 61(6):1312–1320, 2013.

J. Jiang and J. Zhang. Online resource allocation with stochastic resource consumption. *arXiv preprint arXiv:2012.07933*, 2020.

J. Jiang, X. Li, and J. Zhang. Online stochastic optimization with wasserstein based non-stationarity. *arXiv preprint arXiv:2012.06961*, 2020.

T. Kesselheim, A. Tönnis, K. Radke, and B. Vöcking. Primal beats dual on online packing lps in the random-order model. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 303–312, 2014.

X. Li and Y. Ye. Online linear programming: Dual convergence, new algorithms, and regret bounds. *Operations Research*, 2021.

X. Li, C. Sun, and Y. Ye. Simple and fast algorithm for binary integer and online linear programming. *Advances in Neural Information Processing Systems*, 33:9412–9421, 2020.

O. L. Mangasarian and T.-H. Shiau. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM Journal on Control and Optimization*, 25(3):583–595, 1987.

A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22–es, 2007.

M. Molinaro and R. Ravi. The geometry of online packing linear programs. *Mathematics of Operations Research*, 39(1):46–59, 2014.

M. I. Reiman and Q. Wang. An asymptotically optimal policy for a quantity-based network revenue management problem. *Mathematics of Operations Research*, 33(2):257–282, 2008.

A. Shapiro. Perturbation analysis of optimization problems in banach spaces. *Numerical Functional Analysis and Optimization*, 13(1-2):97–116, 1992.

A. Shapiro. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research*, 18(4):829–845, 1993.

K. Talluri and G. Van Ryzin. An analysis of bid-price controls for network revenue management. *Management science*, 44(11-part-1):1577–1593, 1998.

A. Vera and S. Banerjee. The bayesian prophet: A low-regret framework for online decision making. *Management Science*, 67(3):1368–1391, 2021.

A. Vera, S. Banerjee, and I. Gurvich. Online allocation and pricing: Constant regret via bellman inequalities. *Operations Research*, 69(3):821–840, 2021.

Appendix A: Useful Known Results

We proceed now to establish Lipschitz continuity of solutions of linear systems with respect to right-hand side perturbations.

Lemma 10 **(Theorem 2.2 of Mangasarian and Shiau (1987))**. *Consider two linear systems*

$$\hat{A}\boldsymbol{x} \le \boldsymbol{b}^1 \tag{46}$$

and

$$\hat{A}\boldsymbol{x} \le \boldsymbol{b}^2. \tag{47}$$

For any solution $\boldsymbol{x}^1$ *that satisfies linear system* (46)*, there exists a solution* $\boldsymbol{x}^2$ *that satisfies linear system* (47) *such that*

$$\|\boldsymbol{x}^1 - \boldsymbol{x}^2\|_\infty \le \mu \cdot \|\boldsymbol{b}^1 - \boldsymbol{b}^2\|_\infty$$

where

$$\mu = \sup_{\boldsymbol{v}} \left\{ \|\boldsymbol{v}\|_1 \left| \begin{array}{l} \|\boldsymbol{v}^\top \hat{A}\|_1 = 1,\ \boldsymbol{v} \ge 0 \\ \textit{Rows of } \hat{A} \textit{ corresponding non-zero elements} \\ \textit{of } \boldsymbol{v} \textit{ are linear independent.} \end{array} \right. \right\}$$

We state the well-known Bernstein's inequality in the following lemma.

Lemma 11 **(Bernstein's Inequality)**. *Let* $X_1, \dots, X_K$ *be independent zero-mean random variables. Suppose that* $|X_k| \le M$ *almost surely for all* $k \in [K]$*. Then, for all positive* $\epsilon > 0$*, it holds that*

$$P\left(\frac{1}{K} \cdot \sum_{i=1}^{K} X_k \ge \epsilon\right) \le \exp\left(-\frac{\frac{1}{2} \cdot K^2\epsilon^2}{\sum_{k=1}^{K} Var(X_k) + \frac{1}{3} \cdot MK\epsilon}\right)$$

We also state the well-known Hoeffding's inequality in the following lemma.

Lemma 12 **(Hoeffding's Inequality)**. *Let* $X_1, \dots, X_K$ *be independent random variables such that* $a_k \le X_k \le b_k$ *almost surely, for each* $k \in [K]$*. Denote by* $S_K = \frac{1}{K} \cdot \sum_{k=1}^{K} X_k$*. Then, for any* $\epsilon > 0$*, it holds that*

$$P\left(|S_k - \mathbb{E}[S_k] \ge \epsilon|\right) \le 2\exp\left(-\frac{2K^2\epsilon^2}{\sum_{k=1}^{K}(b_k - a_k)^2}\right)$$

We then state the results from Huber (1967) under our notations.

Assumption 7. *Suppose the following conditions hold:*

(N-1). For each fixed $\boldsymbol{\mu}$*, the function* $\psi((r, \boldsymbol{a}), \boldsymbol{\mu})$ *is separable.*

(N-2). Denote $\lambda(\boldsymbol{\mu}) = \mathbb{E}_{(r,\boldsymbol{a})}[\psi((r, \boldsymbol{a}), \boldsymbol{\mu})]$*, then we have* $\lambda(\boldsymbol{\mu}^*) = 0$.

Denote

$$u((r, \boldsymbol{a}), \boldsymbol{\mu}, d) = \sup_{\|\boldsymbol{\mu}' - \boldsymbol{\mu}\|_2 \le d} |\psi((r, \boldsymbol{a}), \boldsymbol{\mu}') - \psi((r, \boldsymbol{a}), \boldsymbol{\mu})|$$

(N-3). There are strictly positive numbers a, b, c_1, c_2, d_0 *such that*

(i). $|\lambda(\boldsymbol{\mu})| \ge a \cdot \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|$ *for* $\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\| \le d_0$.

(ii). $\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[u((\tilde{r}, \tilde{\boldsymbol{a}}), \boldsymbol{\mu}, d)] \le b \cdot d$ *for* $\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\| + d \le d_0$.

(iii). $\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[u((\tilde{r}, \tilde{\boldsymbol{a}}), \boldsymbol{\mu}, d)^2] \le c_1 \cdot d$ *for* $\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\| + d \le d_0$.

(iv). $u((r, \boldsymbol{a}), \boldsymbol{\mu}, d) \le c_2$ *for any* $(r, \boldsymbol{a})$.

Denote by

$$Z_n(\boldsymbol{\mu}',\boldsymbol{\mu}'')=\frac{\sum_{j=1}^n[\psi((r_j,\boldsymbol{a}_j),\boldsymbol{\mu}')-\psi((r_j,\boldsymbol{a}_j),\boldsymbol{\mu}'')+\lambda(\boldsymbol{\mu}')-\lambda(\boldsymbol{\mu}'')]}{\sqrt{n}+n|\lambda(\boldsymbol{\mu}')|}.$$

Then we have the following result from Huber (1967). Note that the original statement in Huber (1967) only concerns the convergence of $\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} Z_n(\boldsymbol{\mu},\boldsymbol{\mu}^*)$ to 0 as $n\to\infty$. We now specify the constant terms in their bound and characterize the convergence rate, which will be helpful in our other proofs. The proof simply follows the proof in Huber (1967), except that we make specific the constant terms, and we include here for completeness.

Lemma 13 **(Lemma 3 in Huber (1967))**. *The conditions in Assumption 7 imply that*

$$\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} Z_n(\boldsymbol{\mu},\boldsymbol{\mu}^*)\to 0$$

in probability as $n\to\infty$. Moreover, for any $\epsilon>0$, it holds that

$$P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} Z_n(\boldsymbol{\mu},\boldsymbol{\mu}^*)\ge 2\epsilon\right)\le c\cdot\epsilon^{-2}n^{-\gamma'}+n^{\gamma'-1}\cdot\left(\frac{c}{b^2q(1-q)}+\frac{c}{9b^2q^2(1-q)^2}\right)\cdot(\frac{\gamma'\cdot\log n}{|\log(1-q)|}+1)\cdot(2M)^m$$

as long as $n\ge n_0$, where n_0 satisfying $n_0^{\gamma'-\frac{1}{2}}=\frac{2b}{\epsilon}$ and $\gamma'\in(\frac{1}{2},1)$ is an arbitrary number. Moreover, we set $M\ge(3b)/(\epsilon a)$ and $q=1/M$.

Proof of Lemma 13. For the sake of simplicity, and without loss of generality, we choose the coordinate system such that $\boldsymbol{\mu}^*=0$. We also use $\boldsymbol{x}_j$ to denote $(r_j,\boldsymbol{a}_j)$ for each $j\in[n]$. The idea of the proof is to divide the cube $\|\boldsymbol{\mu}\|\le d_0$ into a slowly increasing number of smaller cubes and to bound $Z_n(\boldsymbol{\mu},0)$ in probability on each of those smaller cubes.

Put $q=1/M$, where $M\ge 2$ is an integer to be chosen later, and consider the concentric cubes

$$C_k=\{\boldsymbol{\mu}:\|\boldsymbol{\mu}\|\le(1-q)^k\cdot d_0\},\quad k=0,1,\dots,k_0.$$

Subdivide the difference $C_{k-1}\backslash C_k$ into smaller cubes with edges of length $2d=(1-q)^{k-1}q\cdot d_0$ such that the coordinates of their centers $\boldsymbol{\xi}$ are odd multiples of d, and

$$|\boldsymbol{\xi}|=(1-q)^{k-1}(1-\frac{q}{2})\cdot d_0.$$

For each value of k there are less than $(2M)^m$ such smalle cubes, so there are $N<k_0\cdot(2M)^m$ cubes contained in $C_0\backslash C_{k_0}$; number them $C_{(1)},\dots,C_{(N)}$.

Now let $\epsilon>0$ be given. We shall show that for a proper choice of M and of $k_0=k_0(n)$, the right-hand side of

$$P\left(\sup_{\boldsymbol{\mu}\in C_0} Z_n(\boldsymbol{\mu},0)\ge 2\epsilon\right)\le P\left(\sup_{\boldsymbol{\mu}\in C_{k_0}} Z_n(\boldsymbol{\mu},0)\ge 2\epsilon\right)+\sum_{l=1}^N P\left(\sup_{\mu\in C_{(l)}} Z_n(\mu,0)\ge 2\epsilon\right)\tag{48}$$

tends to 0 with increasing n, which establishes the final result.

Actually, we shall choose

$$M\ge(3b)/(\epsilon a),\ \text{which implies } q\le(\epsilon a)/(3b),\tag{49}$$

and $k_0=k_0(n)$ is defined by

$$(1-q)^{k_0}\cdot d_0\le n^{-\gamma'}<(1-q)^{k_0-1}\cdot d_0\tag{50}$$

where $\frac{1}{2} < \gamma < 1$ is an arbitrary fixed number. Thus, we have

$$k_0(n) - 1 < \frac{\gamma' \cdot (\log n + \log d_0)}{|\log(1-q)|} \le k_0(n), \tag{51}$$

hence

$$N \le (\frac{\gamma' \cdot (\log n + \log d_0)}{|\log(1-q)|} + 1) \cdot (2M)^m. \tag{52}$$

Now take any of the cubes $C_{(l)}$, with center ξ and edges of length $2d$ according to $2d = (1-q)^{k-1} q \cdot d_0$ and $|\xi| = (1-q)^{k-1}(1 - \frac{q}{2}) \cdot d_0$. For $\boldsymbol{\mu} \in C_{(l)}$, we have then by (N-3),

$$|\lambda(\boldsymbol{\mu})| \ge a \cdot \|\boldsymbol{\mu}\| \ge a \cdot (1-q)^k \cdot d_0 \tag{53}$$

and

$$|\lambda(\boldsymbol{\mu}) - \lambda(\boldsymbol{\xi})| \le \mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}, d)] \le bd \le b(1-q)^k q \cdot d_0. \tag{54}$$

We have

$$Z_n(\boldsymbol{\mu}, 0) \le Z_n(\boldsymbol{\mu}, \boldsymbol{\xi}) + \frac{\left|\sum_{j=1}^n [\psi(\boldsymbol{x}_j, \boldsymbol{\xi}) - \psi(\boldsymbol{x}_j, 0) - \lambda(\boldsymbol{\xi})]\right|}{\sqrt{n} + n|\lambda(\boldsymbol{\mu})|}, \tag{55}$$

hence

$$\sup_{\boldsymbol{\mu} \in C_{(l)}} Z_n(\boldsymbol{\mu}, 0) \le U_n + V_n \tag{56}$$

with

$$U_n = \frac{\sum_{j=1}^n [u(\boldsymbol{x}_j, \boldsymbol{\xi}, d) + \mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}, d)]]}{na(1-q)^k \cdot d_0}, \tag{57}$$

and

$$V_n = \frac{\sum_{j=1}^n [\psi(\boldsymbol{x}_j, \boldsymbol{\xi}) - \psi(\boldsymbol{x}_j, 0) - \lambda(\boldsymbol{\xi})]}{na(1-q)^k \cdot d_0}. \tag{58}$$

Thus,

$$P(U_n \ge \epsilon) = P\left(\sum_{j=1}^n [u(\boldsymbol{x}_j, \boldsymbol{\xi}, d) - \mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}, d)]] \ge \epsilon na(1-q)^k \cdot d_0 - 2n\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}, d)]\right). \tag{59}$$

In view of (54) and (49),

$$\epsilon a(1-q)^k \cdot d_0 - 2\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}}, \boldsymbol{\xi}, d)] \ge \epsilon a(1-q)^k \cdot d_0 - 2bq(1-q)^k \cdot d_0 \ge bq(1-q)^k \cdot d_0, \tag{60}$$

hence (N-3) (iv) and Hoeffding's inequality (Lemma 12) yields

$$\begin{aligned} P(U_n \ge \epsilon) &\le \exp(-\frac{2b^2q^2(1-q)^{2k} d_0 n}{c_2^2}) \le \exp(-\frac{2b^2q^2(1-q)^{2k_0} d_0 n}{c_2^2}) \\ &\le \exp(-\frac{2b^2q^2(1-q)^2 d_0 n^{1-2\gamma}}{c_2^2}). \end{aligned} \tag{61}$$

In a similar way,

$$P(V_n \ge \epsilon) \le \frac{c}{9b^2q^2(1-q)^2} \cdot \frac{1}{n(1-q)^{k-1} \cdot d_0}. \tag{62}$$

Hence, we obtain from (50), (56), (61) and (62) that

$$P\left(\sup_{\boldsymbol{\mu} \in C_{(j)}} Z_n(\boldsymbol{\mu}, 0) \ge 2\epsilon\right) \le n^{\gamma' - 1} \cdot \left(\frac{c}{b^2 q(1-q)} + \frac{c}{9b^2q^2(1-q)^2}\right). \tag{63}$$

Furthermore,

$$\sup_{\boldsymbol{\mu}\in C_{k_0}} Z_n(\boldsymbol{\mu},0) \leq \frac{\sum_{j=1}^{n}[u(\boldsymbol{x}_j,0,d)+\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}},0,d)]]}{\sqrt{n}} \tag{64}$$

with $d=(1-q)^{k_0}\cdot d_0 \leq n^{-\gamma'}$. Hence,

$$P\left(\sup_{\boldsymbol{\mu}\in C_{k_0}} Z_n(\boldsymbol{\mu},0)\geq 2\epsilon\right) \leq P\left(\sum_{j=1}^{n}[u(\boldsymbol{x}_j,0,d)-\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}},0,d)]]\geq 2\sqrt{n}\epsilon - 2n\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}},0,d)]\right). \tag{65}$$

Since $\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}},0,d)]\leq bd \leq bn^{-\gamma'}$, set n_0 such that $n_0^{\gamma'-\frac{1}{2}}=\frac{2b}{\epsilon}$. Then, for $n\geq n_0$, we have

$$\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}},0,d)]\leq \frac{\epsilon}{2\sqrt{n}} \Rightarrow 2\sqrt{n}\epsilon - 2n\mathbb{E}_{\tilde{\boldsymbol{x}}}[u(\tilde{\boldsymbol{x}},0,d)]\geq \sqrt{n}\epsilon;$$

thus, by Chebyshev's inequality,

$$P\left(\sup_{\boldsymbol{\mu}\in C_{k_0}} Z_n(\boldsymbol{\mu},0)\geq 2\epsilon\right)\leq c\cdot\epsilon^{-2}\cdot n^{-\gamma'}. \tag{66}$$

Now, putting (48), (52), (63), and (66) together, we obtain

$$P\left(\sup_{\boldsymbol{\mu}\in C_0} Z_n(\boldsymbol{\mu},0)\geq 2\epsilon\right)\leq c\cdot\epsilon^{-2}\cdot n^{-\gamma'}+n^{\gamma'-1}\cdot\left(\frac{c}{b^2q(1-q)}+\frac{c}{9b^2q^2(1-q)^2}\right)\cdot\left(\frac{\gamma'\cdot(\log n+\log d_0)}{|\log(1-q)|}+1\right)\cdot(2M)^m$$

as long as $n\geq n_0$ where n_0 satisfying $n_0^{\gamma'-\frac{1}{2}}=\frac{2b}{\epsilon}$, which completes our proof of the lemma. □

Appendix B: Missing Proofs for Section 3

Proof of Lemma 3. Our proof can be classifed into three steps. We fix $\{\hat{q}_j^*\}_{j=1}^n$ as an optimal solution to (13). In the first step, we discretize both the convex optimization problem (12) and (13) into two LPs with a granularity K such that $\{\hat{q}_j^*\}_{j=1}^n$ is an optimal solution to the discretized LP of (13). In the second step, we show that we can select one optimal solution to the discretized LP of (12), such that the gap between the selected optimal solution and $\{\hat{q}_j^*\}_{j=1}^n$ can be bounded by a constant independent of the granularity K. In the final step, we show that as the granularity K grows to infinity, there exists a subsequence of K such that the selected optimal solution to the discretized LP of (12) will converge to an optimal solution of (12), which completes our proof.

We now do the first step to discretize the convex optimization problem (12) and (13). For each $j\in[n]$, we denote by a function

$$G_j(q)=\int_{q'=1-q}^{1}F_j^{-1}(q')dq'.$$

For any integer $K\geq n+1$, we denote by a set $\{0,\frac{1}{K-n},\frac{2}{K-n},\dots,1\}\cup\{\hat{q}_j^*\}_{j=1}^n$ and let q_k^K to be the k-th smallest element in this set, for $k=1,2,\dots,K+1$. For each $j\in[n]$, we denote by $\hat{G}_j^K(\cdot)$ the piece-wise linear interpolation of G_j based on the values at points $\{q_k^K\}_{k=1}^{K+1}$. From the concavity of the function $G_j(\cdot)$, it is clear that we have

$$\hat{G}_j^K(q)\leq G_j(q)\leq \hat{G}_j^K(q)+\frac{\max_{j\in[n]}\{u_j\}}{K-n},\ \forall q\in[0,1],\ \text{and}\ \hat{G}_j^K(\hat{q}_j^*)=G_j(\hat{q}_j^*). \tag{67}$$

Then, we know that $\{\hat{q}_j^*\}_{j=1}^n$ is an optimal solution to the following optimization problem

$$\begin{aligned}\hat{V}_{t,\boldsymbol{c}}^K = \max\ & \sum_{j=1}^{n}p_j\cdot s\cdot\hat{G}_j^K(q_j) \\ \text{s.t.}\ & \sum_{j=1}^{n}p_j\cdot s\cdot a_{j,i}\cdot q_j\leq c_i,\quad \forall i\in[m] \\ & q_j\in[0,1],\quad \forall j\in[n]\end{aligned} \tag{68}$$

because optimization problem (68) differs from (13) only by having a pointwise-dominated objective function, and $\{\hat{q}_j^*\}_{j=1}^n$ attains the optimal objective value from (13) in the dominated problem (68). Without loss of generality, we assume that for each $j \in [n]$, the piece-wise linear functions $\hat{G}_j^K(\cdot)$ share the same set of end points of the piece-wise linear intervals, and we denote by $\{q_k^K\}_{k=1}^{K+1}$ the set of end points, with $q_1^K = 0$ and $q_{K+1}^K = 1$. Then, we have the following linear programming as a re-formulation of the discretization $\hat{V}_{t,\boldsymbol{c}}^K$ of the convex problem $\bar{V}_{t,\boldsymbol{c}}^{\mathrm{Fld}}$ (13):

$$\begin{aligned} \hat{V}_{t,\boldsymbol{c}}^K = \max\ & \sum_{j=1}^n \sum_{k=1}^K \beta_{j,k} \cdot x_{k,j}^K \\ \text{s.t.}\ & \sum_{j=1}^n \sum_{k=1}^K a_{j,i} \cdot x_{k,j}^K \le c_i, \quad \forall i \in [m] \\ & 0 \le x_{k,j}^K \le p_j \cdot s \cdot (q_{k+1}^K - q_k^K), \quad \forall j \in [n], \forall k \in [K] \end{aligned} \tag{69}$$

where $\beta_{j,k}$ is the coefficient that is inherited from the piece-wise linear function $\hat{G}_j^K$. Here, the variable $x_{k,j}^K$ can be interpreted as the number of queries,with type j and reward realization quantile lying in the interval $[q_k^K, q_{k+1}^K]$, being served in the relaxation $\bar{V}_{t,\boldsymbol{c}}^{\mathrm{Fld}}$ (11). We also denote by the following linear programming as a discretization of the convex problem $\bar{V}_{\boldsymbol{c}}^{\mathrm{Semi}}(\boldsymbol{d})$ (12), for each $\boldsymbol{d}$:

$$\begin{aligned} \bar{V}_{\boldsymbol{c}}^K(\boldsymbol{d}) = \max\ & \sum_{j=1}^n \sum_{k=1}^K \beta_{j,k} \cdot x_{k,j}^K \\ \text{s.t.}\ & \sum_{j=1}^n \sum_{k=1}^K a_{j,i} \cdot x_{k,j}^K \le c_i, \quad \forall i \in [m] \\ & 0 \le x_{k,j}^K \le d_j \cdot (q_{k+1}^K - q_k^K), \quad \forall j \in [n], \forall k \in [K] \end{aligned} \tag{70}$$

We now do the second step and compare one optimal solution of $\hat{V}_{t,\boldsymbol{c}}^K$ to another optimal solution of $\bar{V}_{\boldsymbol{c}}^K(\boldsymbol{d})$. Our result is formalized in the following claim. Our analysis follows the analysis in Mangasarian and Shiau (1987) over the Lipschitz continuity of solutions of linear programming, and we further show an equivalence between two linear systems to obtain a bound that is independent of the granularity K.

Claim 2. *There exists a constant μ such that for any K, for any optimal solution $\{\hat{x}_{k,j}^K\}$ of $\hat{V}_{t,\boldsymbol{c}}^K$ (69), we can select one optimal solution $\{\bar{x}_{k,j}^K\}$ of $\bar{V}_{\boldsymbol{c}}^K(\boldsymbol{d})$ (70) such that*

$$\left| \sum_{k=1}^K \hat{x}_{k,j}^K - \bar{x}_{k,j}^K \right| \le \mu \cdot \max_{j' \in [n]} \{(p_{j'} \cdot s - d_{j'})\}, \quad \forall j \in [n]. \tag{71}$$

where

$$\mu = \sup_{\boldsymbol{v}^1 \in \mathbb{R}^m, \boldsymbol{v}^2 \in \mathbb{R}^n} \left\{ \left\| \begin{matrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \end{matrix} \right\| \;\middle|\; \begin{array}{l} \|(\boldsymbol{v}^1)^\top A + (\boldsymbol{v}^2)^\top\|_1 = 1, \quad \boldsymbol{v}^1 \ge 0,\ \boldsymbol{v}^2 \ge 0 \\ \textit{Rows of } \begin{bmatrix} A \\ I_{n\times n} \end{bmatrix} \textit{ corresponding non-zero elements} \\ \textit{of } \begin{pmatrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \end{pmatrix} \textit{ are linear independent.} \end{array} \right\} \tag{72}$$

$A = (a_{i,j})_{\forall i \in [m], j \in [n]} \in \mathbb{R}^{m\times n}$ and I_n is an identity matrix with a size $n \times n$.

We do the third step to complete our proof. For any granularity K, we let $\{\hat{x}^K_{k,j}\}$ be a solution of $\hat{V}^K_{t,\boldsymbol{c}}$ (69) satisfying

$$\hat{x}^K_{k,j} = \begin{cases} p_j \cdot s \cdot (q^K_{k+1} - q^K_k), & \text{if } q^K_k \geq 1 - \hat{q}^*_j \\ \left(q^K_{k+1} - (1 - \hat{q}^*_j)\right) \cdot p_j \cdot s, & \text{if } q^K_k < 1 - \hat{q}^*_j \leq q^K_{k+1} \\ 0, & \text{if } q^K_{k+1} < 1 - \hat{q}^*_j. \end{cases}$$

Since $\{\hat{q}^*_j\}_{j=1}^n$ is an optimal solution to $\hat{V}^K_{t,\boldsymbol{c}}$ under the formulation (68), we must have $\{\hat{x}^K_{k,j}\}$ is an optimal solution of $\hat{V}^K_{t,\boldsymbol{c}}$ under the formulation (69). Then, we denote by $\{\bar{x}^K_{k,j}\}$ one optimal solution of $\bar{V}^K_{\boldsymbol{c}}(\boldsymbol{d})$ (70), as specified in Claim 2. We further construct

$$\bar{q}^K_j = \frac{\sum_{k=1}^K \bar{x}^K_{k,j}}{d_j}, \quad \forall j \in [n]. \tag{73}$$

We denote by $\hat{\boldsymbol{q}}^* = (\hat{q}^*_1, \dots, \hat{q}^*_n)$ and $\bar{\boldsymbol{q}}^K = (\bar{q}^K_1, \dots, \bar{q}^K_n)$. From definition, we know that $\hat{\boldsymbol{q}}^*, \bar{\boldsymbol{q}}^K \in [0,1]^K$. Therefore, there exists a point $\bar{\boldsymbol{q}}'$ and a sequence of integers $\{K_1, \dots, K_w, \dots\}_{w=1,2,\dots}$ such that

$$\bar{\boldsymbol{q}}' = \lim_{w \to \infty} \bar{\boldsymbol{q}}^{K_w}. \tag{74}$$

We show in the following claim that $\bar{\boldsymbol{q}}'$ is an optimal solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\boldsymbol{d})$ (12).

Claim 3. *Let $\bar{\boldsymbol{q}}'$ be constructed in (74). Then, $\bar{\boldsymbol{q}}'$ is an optimal solution to $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\boldsymbol{d})$ (12).*

For any integer w, from Claim 2, we know that

$$\|\hat{\boldsymbol{q}}^* - \bar{\boldsymbol{q}}^{K_w}\|_\infty = \left\| \frac{\sum_{k=1}^{K_w} \hat{x}^{K_w}_{k,j}}{p_j \cdot s} - \frac{\sum_{k=1}^{K_w} \bar{x}^{K_w}_{k,j}}{d_j} \right\|_\infty \leq \left\| \frac{\sum_{k=1}^{K_w} \hat{x}^{K_w}_{k,j}}{p_j \cdot s} - \frac{\sum_{k=1}^{K_w} \bar{x}^{K_w}_{k,j}}{p_j \cdot s} \right\|_\infty + \left\| \frac{\sum_{k=1}^{K_w} \bar{x}^{K_w}_{k,j}}{p_j \cdot s} - \frac{\sum_{k=1}^{K_w} \bar{x}^{K_w}_{k,j}}{d_j} \right\|_\infty$$
$$\leq \max_{j \in [n]} \left\{ \frac{\mu}{p_j} \right\} \cdot \max_{j' \in [n]} \{(p_{j'} - d_{j'}/s)\} + \max_{j \in [n]} \left\{ \left| \frac{d_j}{p_j \cdot s} - 1 \right| \right\}$$

Therefore, from (74) and Claim 3, there exists an optimal solution $\bar{\boldsymbol{q}}'$ of $\bar{V}^{\text{Semi}}_{\boldsymbol{c}}(\boldsymbol{d})$ (12) such that

$$\|\hat{\boldsymbol{q}}^* - \bar{\boldsymbol{q}}'\|_\infty \leq \max_{j \in [n]} \left\{ \frac{\mu}{p_j} \right\} \cdot \max_{j' \in [n]} \{(p_{j'} - d_{j'}/s)\} + \max_{j \in [n]} \left\{ \left| \frac{d_j}{p_j \cdot s} - 1 \right| \right\}.$$

Our proof is thus completed. □

Proof of Claim 2. We denote the linear programming $\hat{V}^K_{t,\boldsymbol{c}}$ (69) by

$$\hat{V}^K_{t,\boldsymbol{c}} = \max\ \boldsymbol{h}^\top \boldsymbol{x} \tag{75}$$
$$\text{s.t. } \boldsymbol{l}' \leq \mathcal{A}\boldsymbol{x} \leq \boldsymbol{u}'$$

and denote the linear programming $\bar{V}^K_{\boldsymbol{c}}(\boldsymbol{d})$ (70) by

$$\bar{V}^K_{\boldsymbol{c}}(\boldsymbol{d}) = \max\ \boldsymbol{h}^\top \boldsymbol{x} \tag{76}$$
$$\text{s.t. } \boldsymbol{l}'' \leq \mathcal{A}\boldsymbol{x} \leq \boldsymbol{u}''$$

Fix an optimal solution $\hat{\boldsymbol{x}}^K$ of $\hat{V}^K_{t,\boldsymbol{c}}$. We denote by J_1, J_2 and J_3 three row index sets of $\mathcal{A}$ such that

$$\mathcal{A}_{J_1} \hat{\boldsymbol{x}}^K = \boldsymbol{u}'_{J_1}, \mathcal{A}_{J_2} \hat{\boldsymbol{x}}^K = \boldsymbol{l}'_{J_2}, \text{ and } \boldsymbol{u}'_{J_3} \mathcal{A}_{J_3} \hat{\boldsymbol{x}}^K < \boldsymbol{b}'_{J_3}.$$

We further fix an optimal solution $\bar{\boldsymbol{x}}^{K'}$ of $\bar{V}^K_{\boldsymbol{c}}(\boldsymbol{d})$ and denote by $J_1 = J_{1,1} \cup J_{1,2}$, $J_2 = J_{2,1} \cup J_{2,2}$ such that

$$\mathcal{A}_{J_{1,1}} \bar{\boldsymbol{x}}^{K'} = \boldsymbol{u}''_{J_{1,1}}, \ \mathcal{A}_{J_{1,2}} \bar{\boldsymbol{x}}^{K'} < \boldsymbol{u}''_{J_{1,2}}, \ \mathcal{A}_{J_{2,1}} \bar{\boldsymbol{x}}^{K'} = \boldsymbol{l}''_{J_{2,1}} \text{ and } \mathcal{A}_{J_{2,2}} \bar{\boldsymbol{x}}^{K'} > \boldsymbol{l}''_{J_{2,2}}.$$

Then, we denote by a set of linear equalities and linear inequalities

$$\begin{aligned} &\mathcal{A}_{J_{1,1}}\boldsymbol{x} = \boldsymbol{u}''_{J_{1,1}} \\ &\mathcal{A}_{J_{1,2}}\bar{\boldsymbol{x}}^{K'} \le \mathcal{A}_{J_{1,2}}\boldsymbol{x},\ \mathcal{A}_{J_{1,2}}\boldsymbol{x} \le \boldsymbol{u}''_{J_{1,2}} \\ &\mathcal{A}_{J_{2,1}}\boldsymbol{x} = \boldsymbol{l}''_{J_{2,1}} \\ &\mathcal{A}_{J_{2,2}}\bar{\boldsymbol{x}}^{K'} \ge \mathcal{A}_{J_{2,2}}\boldsymbol{x},\ \mathcal{A}_{J_{2,2}}\boldsymbol{x} \ge \boldsymbol{l}''_{J_{1,2}} \\ &\boldsymbol{l}''_{J_3} \le \mathcal{A}_{J_3}\boldsymbol{x} \le \boldsymbol{u}''_{J_3} \end{aligned} \tag{77}$$

It is clear that $\bar{\boldsymbol{x}}^{K'}$ satisfies the linear system (77). On the other hand, $\hat{\boldsymbol{x}}^{K}$ satisfies the following linear system:

$$\begin{aligned} &\mathcal{A}_{J_{1,1}}\boldsymbol{x} = \boldsymbol{u}'_{J_{1,1}} \\ &\boldsymbol{u}'_{J_{1,2}} - \boldsymbol{u}''_{J_{1,2}} + \mathcal{A}_{J_{1,2}}\bar{\boldsymbol{x}}^{K'} \le \mathcal{A}_{J_{1,2}}\boldsymbol{x},\ \mathcal{A}_{J_{1,2}}\boldsymbol{x} \le \boldsymbol{u}'_{J_{1,2}} \\ &\mathcal{A}_{J_{2,1}}\boldsymbol{x} = \boldsymbol{l}'_{J_{2,1}} \\ &\boldsymbol{l}'_{J_{2,2}} - \boldsymbol{l}''_{J_{2,2}} + \mathcal{A}_{J_{2,2}}\bar{\boldsymbol{x}}^{K'} \ge \mathcal{A}_{J_{2,2}}\boldsymbol{x},\ \ \mathcal{A}_{J_{2,2}}\boldsymbol{x} \ge \boldsymbol{l}'_{J_{2,2}} \\ &\boldsymbol{l}'_{J_3} \le \mathcal{A}_{J_3}\boldsymbol{x} \le \boldsymbol{u}'_{J_3} \end{aligned} \tag{78}$$

Our remaining analysis can be classified into two steps. For the first step, we show that any variable $\boldsymbol{x}$ that satisfies the linear system (77) turns out to be an optimal solution of $\bar{V}^K_{\boldsymbol{c}}(\boldsymbol{d})$ (76). For the second step, we show that for $\hat{\boldsymbol{x}}^K$ that satisfies the linear system (78), we can find a variable $\bar{\boldsymbol{x}}^K$ satisfying the linear system (77) such that (71) holds.

We now prove the first step. Since $\hat{\boldsymbol{x}}^K$ is an optimal solution of $\hat{V}^K_{t,\boldsymbol{c}}$ (75), from the KKT optimality condition, we know that there exists dual variables $\boldsymbol{\mu}_{J_1} \ge 0$, $\boldsymbol{v}_{J_2} \le 0$ such that

$$\mathcal{A}_{J_1}^\top \boldsymbol{\mu}_{J_1} + \mathcal{A}_{J_2}^\top \boldsymbol{v}_{J_2} = \boldsymbol{h}.$$

Then, for any $\boldsymbol{x}$ that satisfies the linear system (77), we have

$$\begin{aligned} \boldsymbol{h}^\top \boldsymbol{x} &= \boldsymbol{\mu}_{J_1}^\top \mathcal{A}_{J_1}\boldsymbol{x} + \boldsymbol{v}_{J_2}^\top \mathcal{A}_{J_2}\boldsymbol{x} = \boldsymbol{\mu}_{J_{1,1}}^\top \mathcal{A}_{J_{1,1}}\boldsymbol{x} + \boldsymbol{\mu}_{J_{1,2}}^\top \mathcal{A}_{J_{1,2}}\boldsymbol{x} + \boldsymbol{v}_{J_{2,1}}^\top \mathcal{A}_{J_{2,1}}\boldsymbol{x} + \boldsymbol{v}_{J_{2,2}}^\top \mathcal{A}_{J_{2,2}}\boldsymbol{x} \\ &\ge \boldsymbol{\mu}_{J_{1,1}}^\top \mathcal{A}_{J_{1,1}}\bar{\boldsymbol{x}}^{K'} + \boldsymbol{\mu}_{J_{1,2}}^\top \mathcal{A}_{J_{1,2}}\bar{\boldsymbol{x}}^{K'} + \boldsymbol{v}_{J_{2,1}}^\top \mathcal{A}_{J_{2,1}}\bar{\boldsymbol{x}}^{K'} + \boldsymbol{v}_{J_{2,2}}^\top \mathcal{A}_{J_{2,2}}\bar{\boldsymbol{x}}^{K'} = \boldsymbol{\mu}_{J_1}^\top \mathcal{A}_{J_1}\bar{\boldsymbol{x}}^{K'} + \boldsymbol{v}_{J_2}^\top \mathcal{A}_{J_2}\bar{\boldsymbol{x}}^{K'} \\ &= \boldsymbol{h}^\top \bar{\boldsymbol{x}}^{K'} = \bar{V}^K_{t,\boldsymbol{c}}(\boldsymbol{d}) \end{aligned}$$

where the first inequality follows from both $\boldsymbol{x}$, $\bar{\boldsymbol{x}}^{K'}$ satisfies the linear system (77) and thus $\mathcal{A}_{J_{1,1}}\boldsymbol{x} = \mathcal{A}_{J_{1,1}}\bar{\boldsymbol{x}}^{K'} = \boldsymbol{u}''_{J_{1,1}}$, $\mathcal{A}_{J_{2,1}}\boldsymbol{x} = \mathcal{A}_{J_{2,1}}\bar{\boldsymbol{x}}^{K'} = \boldsymbol{l}''_{J_{2,1}}$, $\mathcal{A}_{J_{1,2}}\bar{\boldsymbol{x}}^{K'} \le \mathcal{A}_{J_{1,2}}\boldsymbol{x}$ and $\mathcal{A}_{J_{2,2}}\bar{\boldsymbol{x}}^{K'} \ge \mathcal{A}_{J_{2,2}}\boldsymbol{x}$.

It only remains to show the second step. We prove by exploiting the special structure of the linear systems (77) and (78). Note that under our setting, we have

$$\mathcal{A} = \begin{bmatrix} A & \ldots & A \\ & I_{nK} & \end{bmatrix}$$

where I_{nK} denotes an identity matrix with size $nK \times nK$ and $A = (a_{i,j})_{i\in[m],j\in[n]} \in \mathbb{R}^{m\times n}$. Also, the linear system (77) can be rewritten as

$$\boldsymbol{l}^1 \le \mathcal{A}\boldsymbol{x} \le \boldsymbol{u}^1 \tag{79}$$

and the linear system (78) can be rewritten as

$$\boldsymbol{l}^2 \le \mathcal{A}\boldsymbol{x} \le \boldsymbol{u}^2. \tag{80}$$

Then, for the solution $\hat{\boldsymbol{x}}^K$ of the linear system (80), we construct $\hat{\boldsymbol{y}} \in \mathbb{R}^n$ with

$$\hat{y}_j = \sum_{k=1}^{K} \hat{x}^K_{k,j}, \quad \forall j \in [n].$$

It is clear that $\hat{\boldsymbol{y}}$ is a solution to the following linear system

$$\hat{\boldsymbol{l}}^2 \le \begin{bmatrix} A \\ I_n \end{bmatrix} \boldsymbol{y} \le \hat{\boldsymbol{u}}^2 \tag{81}$$

where $\hat{\boldsymbol{l}}^2 \in \mathbb{R}^{n+m}$ satisfying

$$\hat{l}^2_i = l^2_i, \ \forall i \in [m] \text{ and } \hat{l}^2_{n+j} = \sum_{k=1}^{K} l^2_{j+kn}, \ \forall j \in [n]$$

and $\hat{\boldsymbol{u}}^2 \in \mathbb{R}^{n+m}$ satisfying

$$\hat{u}^2_i = u^2_i, \ \forall i \in [m] \text{ and } \hat{u}^2_{n+j} = \sum_{k=1}^{K} u^2_{j+kn}, \ \forall j \in [n].$$

In the same way, we denote by $\hat{\boldsymbol{l}}^1 \in \mathbb{R}^{n+m}$ satisfying

$$\hat{l}^1_i = l^1_i, \ \forall i \in [m] \text{ and } \hat{l}^1_{n+j} = \sum_{k=1}^{K} l^1_{j+kn}, \ \forall j \in [n]$$

and $\hat{\boldsymbol{u}}^1 \in \mathbb{R}^{n+m}$ satisfying

$$\hat{u}^1_i = u^1_i, \ \forall i \in [m] \text{ and } \hat{u}^1_{n+j} = \sum_{k=1}^{K} u^1_{j+kn}, \ \forall j \in [n].$$

We consider the linear system

$$\hat{\boldsymbol{l}}^1 \le \begin{bmatrix} A \\ I_n \end{bmatrix} \boldsymbol{y} \le \hat{\boldsymbol{u}}^1 \tag{82}$$

From Lemma 10, we know that for the solution $\hat{\boldsymbol{y}}$ that satisfies linear system (81), there exists a solution $\bar{\boldsymbol{y}}$ that satisfies linear system (82) and there also exists a constant μ such that

$$\|\hat{\boldsymbol{y}} - \bar{\boldsymbol{y}}\|_\infty \le \mu \cdot \|(\hat{\boldsymbol{l}}^1, \hat{\boldsymbol{u}}^1) - (\hat{\boldsymbol{l}}^2, \hat{\boldsymbol{u}}^2)\|_\infty \le \mu \cdot \max_{j' \in [n]} \{(p_{j'} \cdot s - d_{j'})\}. \tag{83}$$

Note that here the constant μ depends solely on A and I_n and is independent of the granularity K. We now construct a solution to the linear system (79) from $\bar{\boldsymbol{y}}$ to complete our proof. For each $j \in [n]$ and each $k \in [K]$, we define

$$\bar{x}^K_{k,j} = l^1_{j+kn} + (\bar{y}_j - \hat{l}^1_{n+j}) \cdot \frac{u^1_{j+kn} - l^1_{j+kn}}{\hat{u}^1_{n+j} - \hat{l}^1_{n+j}}.$$

It is clear to see that $\bar{\boldsymbol{x}}^K$ satisfies linear system (79) and satisfies

$$\sum_{k=1}^{K} \bar{x}^K_{k,j} = \bar{y}_j, \quad \forall j \in [n]. \tag{84}$$

Our proof of the second step is completed from (83), (74) and the fact that $\bar{\boldsymbol{x}}^K$ satisfies linear system (79). From the conclusion of the first step, we know that $\bar{\boldsymbol{x}}^K$ is an optimal solution to $\bar{V}^K_{\boldsymbol{c}}(\boldsymbol{d})$ and our proof of Claim 2 is thus completed. □

Proof of Claim 3. For any integer w, it is clear to see that $\bar{\boldsymbol{q}}^{K_w}$ is a feasible solution to $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d})$ (12), which implies that the limiting point $\bar{\boldsymbol{q}}'$ is a feasible solution to $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d})$ (12) since the feasible set is closed. We now prove optimality.

It is direct to see from the construction of the piece-wise linear function $\hat{G}_j^K(\cdot)$ that

$$\bar{V}_{\boldsymbol{c}}^{K_w}(\boldsymbol{d}) = \sum_{j=1}^{n}\sum_{k=1}^{K_w} \beta_{j,k}^{K_w} \cdot \bar{x}_{k,j}^{K_w} \le \sum_{j=1}^{n} d_j \cdot \int_{q=1-\bar{q}_j^{K_w}}^{1} F_j^{-1}(q)dq \le \bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}).$$

where the first inequality follows from the definition $\bar{q}_j^{K_w} = \frac{\sum_{k=1}^{K_w} \bar{x}_{k,j}^{K_w}}{p_j \cdot s}$.

We now show that $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}) \le \bar{V}_{\boldsymbol{c}}^{K_w}(\boldsymbol{d}) + \frac{(\sum_{j=1}^n d_j)\cdot \max_{j\in[n]}\{u_j\}}{K_w}$. For an optimal solution $\{\bar{q}_j^*\}_{j=1}^n$ of $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d})$. We construct for any $k \in [K_w]$ and $j \in [n]$

$$x_{k,j}^{K_w} = \begin{cases} p_j \cdot s \cdot (q_{k+1}^{K_w} - q_k^{K_w}), & \text{if } q_k^{K_w} \ge 1 - \bar{q}_j^* \\ \left(q_{k+1}^{K_w} - (1-\bar{q}_j^*)\right) \cdot p_j \cdot s, & \text{if } q_k^{K_w} < 1 - \bar{q}_j^* \le q_{k+1}^{K_w} \\ 0, & \text{if } q_{k+1}^{K_w} < 1 - \bar{q}_j^*. \end{cases}$$

It is clear to see that $\{x_{k,j}^{K_w}\}$ is a feasible solution to $\bar{V}_{\boldsymbol{c}}^{K_w}(\boldsymbol{d})$ (70) and it holds that

$$\begin{aligned} \bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}) &= \sum_{j=1}^{n} d_j \cdot G_j(\bar{q}_j^*) \le \sum_{j=1}^{n} d_j \cdot \hat{G}_j^{K_w}(\bar{q}_j^*) + \frac{(\sum_{j=1}^n d_j)\cdot \max_{j\in[n]}\{u_j\}}{K_w - n} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{K_w} d_j \cdot \beta_{j,k}^{K_w} \cdot x_{k,j}^{K_w} + \frac{(\sum_{j=1}^n d_j)\cdot \max_{j\in[n]}\{u_j\}}{K_w - n} \\ &\le \bar{V}_{\boldsymbol{c}}^{K_w}(\boldsymbol{d}) + \frac{(\sum_{j=1}^n d_j)\cdot \max_{j\in[n]}\{u_j\}}{K_w - n} \end{aligned}$$

where the first inequality follows from (67). Therefore, we conclude that

$$\bar{V}_{\boldsymbol{c}}^{K_w}(\boldsymbol{d}) \le \sum_{j=1}^{n} p_j \cdot s \cdot \int_{q=1-\bar{q}_j^{K_w}}^{1} F_j^{-1}(q)dq \le \bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}) \le \bar{V}_{\boldsymbol{c}}^{K_w}(\boldsymbol{d}) + \frac{(\sum_{j=1}^n d_j)\cdot \max_{j\in[n]}\{u_j\}}{K_w - n}.$$

Note that

$$\sum_{j=1}^{n} d_j \cdot \int_{q=1-\bar{q}'_j}^{1} F_j^{-1}(q)dq = \lim_{w\to\infty} \sum_{j=1}^{n} d_j \cdot \int_{q=1-\bar{q}_j^{K_w}}^{1} F_j^{-1}(q)dq.$$

We have

$$\begin{aligned} \left| \sum_{j=1}^{n} d_j \cdot \int_{q=1-\bar{q}'_j}^{1} F_j^{-1}(q)dq - \bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}) \right| &= \lim_{w\to\infty} \left| \sum_{j=1}^{n} d_j \cdot \int_{q=1-\bar{q}_j^{K_w}}^{1} F_j^{-1}(q)dq - \bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d}) \right| \\ &\le \lim_{w\to\infty} \frac{(\sum_{j=1}^n d_j)\cdot \max_{j\in[n]}\{u_j\}}{K_w - n} = 0 \end{aligned}$$

which implies that $\bar{\boldsymbol{q}}'$ is an optimal solution to $\bar{V}_{\boldsymbol{c}}^{\text{Semi}}(\boldsymbol{d})$ (12). Our proof is therefore completed. □

Proof of Claim 1. Denote by functions

$$G_j(q) := \int_{q'=q}^{1} F_j^{-1}(q')dq', \quad \forall j \in [n].$$

It is clear that

$$G'_j(q) = -F_j^{-1}(q) \text{ and } G''_j(q) = -\frac{1}{f(q|\boldsymbol{a}_j)} \in [-\frac{1}{\alpha}, 0].$$

Therefore, from the concavity of $G_j(\cdot)$, we have

$$\int_{q=q_1}^{q_2} F_j^{-1}(q)dq = G_j(q_1) - G_j(q_2) \le F_j^{-1}(q_1)\cdot(q_2 - q_1) + \frac{(q_2-q_1)^2}{\alpha}$$

which completes our proof. □

Appendix C: Missing Proofs for Section 4

Proof of Lemma 4. By plugging (35) into (8), we have that

$$\begin{aligned}\mathrm{Myopic}_t(\pi, \tilde{\boldsymbol{c}}_t^{\pi}) &= \mathbb{E}_{(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)}\left[\mathbb{E}_{I_{t+1}}[\bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^{\pi}}(I_t)] - \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^{\pi} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^{\pi}}(I_{t+1})]] - \tilde{r}_t \cdot \tilde{x}_t^{\pi}\right] \\ &= \mathbb{E}_{(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)}\left[\mathbb{E}_{I_{t+1}}[(\tilde{x}_t^{\mathrm{round}} - \tilde{x}_t^{\pi}) \cdot (\tilde{r}_t + \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^{\pi} - \tilde{\boldsymbol{a}}_t}(I_{t+1})) + (\tilde{x}_t^{\pi} - \tilde{x}_t^{\mathrm{round}}) \cdot \bar{V}^{\mathrm{Off}}_{\tilde{\boldsymbol{c}}_t^{\pi}}(I_{t+1}) + G_{\tilde{\boldsymbol{c}}_t^{\pi}}(I_t)]\right] \\ &= \mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)}[(\tilde{x}_t^{\mathrm{round}} - \tilde{x}_t^{\pi}) \cdot (\tilde{r}_t - M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}))]\right] + \mathbb{E}[G_{\tilde{\boldsymbol{c}}_t^{\pi}}(I_t)]\end{aligned} \tag{85}$$

From the definition of $\tilde{x}_t^{\mathrm{round}}$ and $\tilde{x}_t^{\pi}$, we know that $\tilde{x}_t^{\mathrm{round}} - \tilde{x}_t^{\pi} \in \{-1, 1\}$ if and only if $\hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t} \leq \tilde{r}_t \leq M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1})$ and $\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t$, or $M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}) \leq \tilde{r}_t \leq \hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}$ and $\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t$. Thus, we have that

$$\begin{aligned}&\mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{(\tilde{r}_t, \tilde{\boldsymbol{a}}_t)}[(\tilde{x}_t^{\mathrm{round}} - \tilde{x}_t^{\pi}) \cdot (\tilde{r}_t - M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}))]\right] \\ \leq &\mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{\tilde{r}_t \sim F(\cdot|\tilde{\boldsymbol{a}}_t)}[\mathbb{1}_{\{\hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t} \leq \tilde{r}_t \leq M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1})\}} \cdot (M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}) - \hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t})]\right]\right] \cdot \mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t\}} \\ &+ \mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{\tilde{r}_t \sim F(\cdot|\tilde{\boldsymbol{a}}_t)}[\mathbb{1}_{\{\hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t} \geq \tilde{r}_t \geq M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1})\}} \cdot (\hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t} - M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}))]\right]\right] \cdot \mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t\}} \\ \leq &\bar{\alpha} \cdot \mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[(\hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t} - M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}))^2\right]\right] \cdot \mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t\}} \\ \leq &2\bar{\alpha} \cdot \mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t\}} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[(\hat{M}_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t} - \mathbb{E}_{I_{t+1}}[M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1})])^2\right] \\ &+ 2\bar{\alpha} \cdot \mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^{\pi} \geq \tilde{\boldsymbol{a}}_t\}} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}\left[(\mathbb{E}_{I_{t+1}}[M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1})] - M_{\tilde{\boldsymbol{c}}_t^{\pi}, \tilde{\boldsymbol{a}}_t}(I_{t+1}))^2\right]\right]\end{aligned} \tag{86}$$

where the second inequality follows from Assumption 3 which implies that the density of $\tilde{r}_t$ is upper bounded by $\bar{\alpha}$ and the third inequality follows from $(a+b)^2 \leq 2a^2 + 2b^2$ for any a, b. Plugging (86) into (85), we complete our proof of the lemma. □

Proof of Lemma 6. Note that we have

$$L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}) = \frac{\boldsymbol{c}^{\top}\boldsymbol{\mu}}{s-1} + \mathbb{E}_{\tilde{\boldsymbol{a}}}\left[\mathbb{E}_{\tilde{r} \sim F_{\tilde{\boldsymbol{a}}}}[\tilde{r} - \tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}]^{+}|\tilde{\boldsymbol{a}}\right] = \frac{\boldsymbol{c}^{\top}\boldsymbol{\mu}}{s-1} + \mathbb{E}_{\tilde{\boldsymbol{a}}}\left[\int_{\tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}}^{\infty} (\tilde{r} - \tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}) f(r|\tilde{\boldsymbol{a}}) dr\right]$$

which implies that

$$\frac{\partial L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu})}{\partial \mu_i} = \frac{c_i}{s-1} - \mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{a}_i \cdot (1 - F(\tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}|\tilde{\boldsymbol{a}}))], \quad \forall i \in [m]$$

and

$$\frac{\partial^2 L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu})}{\partial \mu_i \partial \mu_{i'}} = \mathbb{E}_{\tilde{\boldsymbol{a}}}[f(\tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}|\tilde{\boldsymbol{a}}) \cdot \tilde{a}_i \cdot \tilde{a}_{i'}], \quad \forall i, i' \in [m]$$

where we use $\tilde{a}_i$ to denote the i-th element of vector $\tilde{\boldsymbol{a}}$. Moreover, note that for any $\boldsymbol{\mu} \geq 0$, from Taylor's theorem, there exists a $\boldsymbol{\mu}'$ that lies on the intersection between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^*$ such that

$$L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}) - L^{\mathrm{Fld}}_{\boldsymbol{c}}(\mu^*) = (\boldsymbol{\mu} - \boldsymbol{\mu}^*)^{\top} \frac{\partial L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}^*)}{\partial \boldsymbol{\mu}} + \frac{1}{2} \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}^*)^{\top} \cdot \frac{\partial^2 L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}')}{\partial \boldsymbol{\mu}^2} \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}^*)$$

where $\frac{\partial L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}^*)}{\partial \boldsymbol{\mu}} = (\frac{\partial L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}^*)}{\partial \mu_1}, \ldots, \frac{\partial L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}^*)}{\partial \mu_m})$, and $\frac{\partial^2 L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}')}{\partial \boldsymbol{\mu}^2}$ is the Hessian matrix that equals $\mathbb{E}_{\tilde{\boldsymbol{a}}}[f(\tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}'|\tilde{\boldsymbol{a}}) \cdot \tilde{\boldsymbol{a}} \cdot \tilde{\boldsymbol{a}}^{\top}]$. Note that Ω is assumed to be a convex set in Assumption 3 and both $\boldsymbol{\mu}, \boldsymbol{\mu}^* \in \Omega$. Then, we have $\boldsymbol{\mu}' \in \Omega$ and Assumption 3 implies that $\mathbb{E}_{\tilde{\boldsymbol{a}}}[f(\tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}'|\tilde{\boldsymbol{a}})] \geq \underline{\alpha}$. The smallest positive eigenvalue of $\frac{\partial^2 L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}')}{\partial \boldsymbol{\mu}^2}$ is therefore lower bounded by $\underline{\alpha} \cdot \underline{\beta}$. Also, from the optimality of $\boldsymbol{\mu}^*$, we must have

$$(\boldsymbol{\mu} - \boldsymbol{\mu}^*)^{\top} \frac{\partial L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}^*)}{\partial \boldsymbol{\mu}} \geq 0$$

for any $\boldsymbol{\mu} \geq 0$. Thus, it holds that

$$\frac{\underline{\alpha}\underline{\beta}}{2} \cdot \|\mathcal{P}_S(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\|_2^2 \leq \frac{1}{2} \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}^*)^{\top} \cdot \frac{\partial^2 L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}')}{\partial \boldsymbol{\mu}^2} \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}^*) \leq L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}) - L^{\mathrm{Fld}}_{\boldsymbol{c}, t+1}(\boldsymbol{\mu}^*)$$

which completes our proof. □

Proof of Lemma 7. From (42), we know that

$$
\begin{aligned}
M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})-\mathbb{E}[M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})] &\leq \boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\mathbb{E}[\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1], \quad \text{if } M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})\geq\mathbb{E}[M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})]\\
\mathbb{E}[M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})]-M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}) &\leq \mathbb{E}[\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2]-\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1, \quad \text{if } M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})\leq\mathbb{E}[M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})]
\end{aligned}
$$

which implies that

$$
\mathrm{Var}(M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1}))\leq 2\mathrm{Var}(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1)+2\mathrm{Var}(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2)+2(\mathbb{E}[\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1]-\mathbb{E}[\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2])^2
$$

Note that we have

$$
\mathrm{Var}(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1)=\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1+\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_1])^2]\leq 2\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2]+2(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_1])^2
$$

Similarly, we have

$$
\mathrm{Var}(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2)=\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2+\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_2])^2]\leq 2\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2]+2(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_2])^2
$$

Also, we have

$$
\begin{aligned}
(\mathbb{E}[\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2]-\mathbb{E}[\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1])^2 &= (\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_2]-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2+\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1+\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_1])^2\\
&\leq 3(\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_2]-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2+3(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2+3(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_1])^2
\end{aligned}
$$

Thus, we have that

$$
\begin{aligned}
\mathrm{Var}(M_{\boldsymbol{c},\boldsymbol{a}_t}(I_{t+1})) &\leq 10(\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_1]-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2+10(\boldsymbol{a}_t^\top\mathbb{E}[\tilde{\boldsymbol{\mu}}_2]-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2+6(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2\\
&\quad+4\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2]+4\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2] \qquad (87)\\
&\leq 14\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_1-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2]+14\mathbb{E}[(\boldsymbol{a}_t^\top\tilde{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)^2]+6(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2.
\end{aligned}
$$

It only remains to bound the term $(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2-\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_1)^2$. We have the following claim, where the proof is relegated to the end of this proof.

Claim 4. *It holds that*

$$
(\boldsymbol{a}_t^\top(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2))^2\leq\frac{2}{(s-1)\underline{\alpha\beta}}\cdot\bar{d}^3\cdot m^{1/2}\cdot\gamma
$$

where $\bar{d}=\max_{j\in[n]}\{\|\boldsymbol{a}_j\|_2\}$ *and* $\gamma=\max_{i\in[m],j\in[n]:a_{j,i}>0}\frac{u_j}{a_{j,i}}$.

Therefore, our proof is completed by combining (87) and Claim 4. □

Proof of Claim 4. From Lemma 6, by substituting $\hat{\boldsymbol{\mu}}_1$ into $\boldsymbol{\mu}^*$, and substituting $\hat{\boldsymbol{\mu}}_2$ into $\boldsymbol{\mu}$, we have that

$$
\begin{aligned}
\frac{\underline{\alpha\beta}}{2}\cdot\|\mathcal{P}_S(\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_1)\|_2^2 &\leq L^{\text{Fld}}_{\boldsymbol{c},t+1}(\hat{\boldsymbol{\mu}}_2)-L^{\text{Fld}}_{\boldsymbol{c},t+1}(\hat{\boldsymbol{\mu}}_1)=L^{\text{Fld}}_{\boldsymbol{c},t+1}(\hat{\boldsymbol{\mu}}_2)-L^{\text{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\hat{\boldsymbol{\mu}}_2)+L^{\text{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\hat{\boldsymbol{\mu}}_2)-L^{\text{Fld}}_{\boldsymbol{c},t+1}(\hat{\boldsymbol{\mu}}_1)\\
&\leq L^{\text{Fld}}_{\boldsymbol{c},t+1}(\hat{\boldsymbol{\mu}}_2)-L^{\text{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\hat{\boldsymbol{\mu}}_2)=\frac{1}{s-1}\cdot\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2
\end{aligned}
$$

where the second inequality follows from $L^{\text{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\hat{\boldsymbol{\mu}}_2)\leq L^{\text{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\hat{\boldsymbol{\mu}}_1)\leq L^{\text{Fld}}_{\boldsymbol{c},t+1}(\hat{\boldsymbol{\mu}}_1)$ by noting that $\hat{\boldsymbol{\mu}}_2\in\arg\min_{\boldsymbol{\mu}\geq 0}L^{\text{Fld}}_{\boldsymbol{c}-\boldsymbol{a}_t,t+1}(\boldsymbol{\mu})$. Thus, we have

$$
(\boldsymbol{a}_t^\top(\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_1))^2\leq\bar{d}^2\cdot\|\mathcal{P}_S(\hat{\boldsymbol{\mu}}_2-\hat{\boldsymbol{\mu}}_1)\|_2^2\leq\frac{2}{(s-1)\underline{\alpha\beta}}\cdot(\boldsymbol{a}_t^\top\hat{\boldsymbol{\mu}}_2)\cdot\bar{d}^2\leq\frac{2}{(s-1)\underline{\alpha\beta}}\cdot\|\boldsymbol{a}_t\|_2\cdot\|\hat{\boldsymbol{\mu}}_2\|_2\cdot\bar{d}^2
$$

Clearly, we must have $\hat{\mu}_{2,i}\leq\gamma$ for each $i\in[m]$ with $\gamma=\max_{i\in[m],j\in[n]:a_{j,i}>0}\frac{u_j}{a_{j,i}}$, which implies that $\|\hat{\boldsymbol{\mu}}_2\|_2\leq\sqrt{m}\gamma$. Thus, it holds that

$$
(\boldsymbol{a}_t^\top(\hat{\boldsymbol{\mu}}_1-\hat{\boldsymbol{\mu}}_2))^2\leq\frac{2}{(s-1)\underline{\alpha\beta}}\cdot\bar{d}^3\cdot m^{1/2}\cdot\gamma
$$

which completes our proof. □

Proof of Lemma 8. We first consider the setting where $\boldsymbol{c} \geq \tilde{\boldsymbol{a}}_t$ and thus both $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ are well-defined in (41). We show that $\tilde{x}_t^* \neq \tilde{x}_t^{\text{round}}$ happens only if $\boldsymbol{a}_t^\top \tilde{\boldsymbol{\mu}}_1 \leq \tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2$. Note that the value of $\tilde{x}_t^*$ can be determined in the following way:

$$\tilde{x}_t^* = \underset{\phi \in [0,1]}{\arg\max}\, \phi \cdot \tilde{r}_t + \bar{V}^{\text{Off}}_{\boldsymbol{c} - \phi \cdot \tilde{\boldsymbol{a}}_t}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c}}(I_{t+1}) \tag{88}$$

Note that for any $\phi \in [0,1]$, we must have

$$-\phi \cdot \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 \leq \bar{V}^{\text{Off}}_{\boldsymbol{c} - \phi \cdot \tilde{\boldsymbol{a}}_t}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c}}(I_{t+1}) \leq -\phi \cdot \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$$

which can be proved following the same intuition of Lemma 5. Therefore, when $\tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$, we must have $\tilde{x}_t^* = \tilde{x}_t^{\text{round}} = 0$, and when $\tilde{r}_t \geq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2$, we must have $\tilde{x}_t^* = \tilde{x}_t^{\text{round}} = 1$. We conclude that $G_{\boldsymbol{c}}(I_t) = 0$ when $\tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$ or $\tilde{r}_t \geq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2$. We now assume that $\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 \leq \tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2$ and we further consider two cases as follows:
Case 1: If $\tilde{x}_t^{\text{round}} = 0$, then we have

$$G_{\boldsymbol{c}}(I_t) = \tilde{r}_t \cdot \tilde{x}_t^* + \bar{V}^{\text{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^*}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c}}(I_{t+1})$$

Note that we have

$$\bar{V}^{\text{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^*}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c}}(I_{t+1}) \leq -\tilde{x}_t^* \cdot \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$$

By noting $\tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2$ and $\tilde{x}_t^* \leq 1$, we have that

$$G_{\boldsymbol{c}}(I_t) \leq \tilde{x}_t^* \cdot (\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1) \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$$

which gives us an upper bound for the first case.
Case 2: If $\tilde{x}_t^{\text{round}} = 1$, then we have

$$G_{\boldsymbol{c}}(I_t) = \tilde{r}_t \cdot (\tilde{x}_t^* - 1) + \bar{V}^{\text{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^*}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t}(I_{t+1}).$$

Note that we have

$$\bar{V}^{\text{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t \cdot \tilde{x}_t^*}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c} - \tilde{\boldsymbol{a}}_t}(I_{t+1}) \leq (1 - \tilde{x}_t^*) \cdot \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2.$$

By noting $\tilde{r}_t \geq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$ and $0 \leq 1 - \tilde{x}_t^* \leq 1$, we have that

$$G_{\boldsymbol{c}}(I_t) \leq (1 - \tilde{x}_t^*) \cdot (\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1) \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$$

which gives us an upper bound for the second case. Therefore, on both cases, we show that if $\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 \leq \tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2$, it holds that

$$G_{\boldsymbol{c}}(I_t) \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1. \tag{89}$$

which implies that

$$\begin{aligned}
\mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{\tilde{r}_t \sim F(\cdot|\tilde{\boldsymbol{a}}_t)}[G_{\boldsymbol{c}}(I_t)]\right]\right] &\leq \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{\tilde{r}_t \sim F(\cdot|\tilde{\boldsymbol{a}}_t)}[\mathbb{1}_{\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 \leq \tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2} \cdot (\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1)]\right]\right] \\
&\leq \bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}\left[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1)^2\right]\right].
\end{aligned}$$

We further note that

$$\begin{aligned}
(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1)^2 &= (\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2 + \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1 + \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1)^2 \\
&\leq 3(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2)^2 + 3(\tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1)^2 + 3(\tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1)^2.
\end{aligned}$$

We use Claim 4 to bound the term $(\tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1)^2$. Then, we have

$$\mathbb{E}_{I_t}\left[G_{\boldsymbol{c}}(I_t)\right] \leq \frac{\kappa_2}{s-1} + \kappa_2 \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1)^2]\right] + \kappa_2 \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2)^2]\right] \tag{90}$$

for a constant κ_2.

We then consider the setting where $\boldsymbol{c} \geq \tilde{\boldsymbol{a}}_t$ does not hold and only $\tilde{\boldsymbol{\mu}}_1$ is well-defined in (41). Still, when $\tilde{r}_t \leq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$, we must have $\tilde{x}_t^* = \tilde{x}_t^{\text{round}} = 0$, which implies $G_{\boldsymbol{c}}(I) = 0$. We now assume that $\tilde{r}_t \geq \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$. Then, since

$$-\phi \cdot \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 \leq \bar{V}^{\text{Off}}_{\boldsymbol{c}-\phi\cdot\tilde{\boldsymbol{a}}_t}(I_{t+1}) - \bar{V}^{\text{Off}}_{\boldsymbol{c}}(I_{t+1}) \leq -\phi \cdot \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$$

for any $\phi[0,1]$, we have

$$G_{\boldsymbol{c}}(I_t) \leq \tilde{x}_t^* \cdot (\tilde{r}_t - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1) \leq \tilde{r}_t - \tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1$$

by noting $\tilde{x}_t^* \in [0,1]$. Denote by j_t as the type of query t. It holds that

$$\mathbb{E}_{\tilde{r}_t\sim F(\cdot|\tilde{\boldsymbol{a}}_t)}[G_{\boldsymbol{c}}(I_t)] \leq \frac{1}{2}\cdot(u_{j_t} - \boldsymbol{a}_{j_t}^\top \tilde{\boldsymbol{\mu}}_1)^2 \leq (u_{j_t} - \boldsymbol{a}_{j_t}^\top \hat{\boldsymbol{\mu}}_1)^2 + (\boldsymbol{a}_{j_t}^\top \hat{\boldsymbol{\mu}}_1 - \boldsymbol{a}_{j_t}^\top \tilde{\boldsymbol{\mu}}_1)^2 \leq u_{j_t}\cdot(u_{j_t} - \boldsymbol{a}_{j_t}^\top \hat{\boldsymbol{\mu}}_1) + (\boldsymbol{a}_{j_t}^\top \hat{\boldsymbol{\mu}}_1 - \boldsymbol{a}_{j_t}^\top \tilde{\boldsymbol{\mu}}_1)^2.$$

It is clear to see that $\min_{\mu\in\Omega} L^{\text{Fld}}_{\boldsymbol{c},t+1}(\mu)$ is the dual problem of $\max_{\boldsymbol{x}} \bar{V}^{\text{Fld}}_{t+1,\boldsymbol{c}}$. Denote by $\bar{\boldsymbol{x}} = \{\bar{x}_j(r)\} \in \text{argmax}_{\boldsymbol{x}} \bar{V}^{\text{Fld}}_{t+1,\boldsymbol{c}}$ such that $\bar{\boldsymbol{x}}$ and $\hat{\boldsymbol{\mu}}_1$ is an optimal primal-dual pair. Then, when $c_i < a_{j_t,i}$ for some $i \in [m]$, we must have

$$\mathbb{E}_{r\sim F(\cdot|\boldsymbol{a}_{j_t})}[\bar{x}_{j_t}(r)] \leq \frac{1}{p_{j_t}\cdot(s-1)}$$

and as a result we have

$$\alpha\cdot(u_{j_t} - \boldsymbol{a}_{j_t}^\top \tilde{\boldsymbol{\mu}}_1) \leq P(\boldsymbol{a}_{j_t}^\top \tilde{\boldsymbol{\mu}}_1 \leq r \leq u_{j_t} | r \sim F_{j_t}) = \mathbb{E}_{r\sim F(\cdot|\boldsymbol{a}_{j_t})}[\bar{x}_{j_t}(r)] \leq \frac{1}{p_{j_t}\cdot(s-1)}$$

where $\alpha > 0$ is a lower bound on the density function specified in Assumption 1. Then we have

$$\mathbb{E}_{I_{t+1}}\left[\mathbb{E}_{\tilde{r}_t\sim F(\cdot|\tilde{\boldsymbol{a}}_t)}[G_{\boldsymbol{c}}(I_t)]\right] \leq \frac{u_{j_t}}{\alpha\cdot p_{j_t}\cdot(s-1)} + \mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1)^2].$$

Therefore, we have

$$\mathbb{E}_{I_t}[G_{\boldsymbol{c}}(I_t)] \leq \frac{\kappa_2}{s-1} + \kappa_2 \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1)^2]\right] \tag{91}$$

for some constant $\kappa_2 > 0$. Our proof is completed by combining (90) and (91). □

Proof of Lemma 9. From (44), $\mathcal{S}$ refers to the space spanned by $\mathcal{A}$, which is a subspace of $\mathbb{R}^m$. We further denote by $\mathcal{S}^\perp$ the orthogonal complement subspace of S in the Euclidean space $\mathbb{R}^m$, i.e., $\mathbb{R}^m = \mathcal{S} + \mathcal{S}^\perp$ and $\boldsymbol{u}^\top \boldsymbol{v} = 0$ for any $\boldsymbol{u} \in \mathcal{S}$ and any $\boldsymbol{v} \in \mathcal{S}^\perp$.

Following basics in linear algebra, any vector $\boldsymbol{\mu} \in \mathbb{R}^m$ can be decomposed uniquely as $\boldsymbol{\mu} = \boldsymbol{\mu}_S + \boldsymbol{\mu}_{S^\perp}$, where $\boldsymbol{\mu}_S \in \mathcal{S}$ and $\boldsymbol{\mu}_{S^\perp} \in \mathcal{S}^\perp$. Then the projection of $\boldsymbol{\mu}$ to the subspace $\mathcal{S}$ can be given as $\mathcal{P}_S(\boldsymbol{\mu}) = \boldsymbol{\mu}_S$. Following this decomposition, for any $\boldsymbol{c} \geq 0$, we have that

$$\begin{aligned}\min_{\boldsymbol{\mu}\geq\Omega} L^{\text{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu}) &= \min_{\boldsymbol{\mu}_S+\boldsymbol{\mu}_{S^\perp}\in\Omega} \left(\frac{\boldsymbol{c}}{T-t}\right)^\top \boldsymbol{\mu}_S + \left(\frac{\boldsymbol{c}}{T-t}\right)^\top \boldsymbol{\mu}_{S^\perp} + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}_S]^+ \\ &= \min_{\boldsymbol{\mu}_S\in\mathcal{S}} \left(\frac{\boldsymbol{c}}{T-t}\right)^\top \boldsymbol{\mu}_S + h(\boldsymbol{\mu}_S) + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^\top \boldsymbol{\mu}_S]^+\end{aligned}$$

where we note that $\boldsymbol{a}^\top \boldsymbol{\mu}_{S^\perp} = 0$ for any $\boldsymbol{a} \in \mathcal{A}$ and

$$h(\boldsymbol{\mu}_S) = \min_{\boldsymbol{\mu}_{S^\perp}\in S^\perp} \left(\frac{\boldsymbol{c}}{T-t}\right)^\top \boldsymbol{\mu}_{S^\perp} \text{ s.t. } \boldsymbol{\mu}_{S^\perp} + \boldsymbol{\mu}_S \in \Omega.$$

Note that if the optimization problem that defines $h(\boldsymbol{\mu}_S)$ is infeasible for some $\boldsymbol{\mu}_S$, then we simply set $h(\boldsymbol{\mu}_S) = +\infty$. Denote by $\hat{\mathcal{S}} \subset \mathcal{S} \cap \Omega$ the convex set such that $h(\boldsymbol{\mu})$ is finite for all $\boldsymbol{\mu} \in \hat{\mathcal{S}}$. Therefore, the dual problem $\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu})$ can be transferred into a minimization problem over the set $\hat{\mathcal{S}}$, i.e.,

$$\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Fld}}_{\boldsymbol{c},t+1}(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}\in\hat{\mathcal{S}}} \hat{G}_{\boldsymbol{c}}(\boldsymbol{\mu}) := \left(\frac{\boldsymbol{c}}{T-t}\right)^{\top} \boldsymbol{\mu} + h(\boldsymbol{\mu}) + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}]^{+}$$

Following the same way, the minimization of the sample average dual function $L^{\mathrm{Off}}_{\boldsymbol{c},I_{t+1}}(\boldsymbol{\mu})$ in (40) can be formulated as

$$\min_{\boldsymbol{\mu}\in\Omega} L^{\mathrm{Off}}_{\boldsymbol{c},I_{t+1}}(\boldsymbol{\mu}) = \min_{\boldsymbol{\mu}\in\hat{\mathcal{S}}} \hat{G}_{\boldsymbol{c},I_{t+1}}(\boldsymbol{\mu}) := \left(\frac{\boldsymbol{c}}{T-t}\right)^{\top} \boldsymbol{\mu} + h(\boldsymbol{\mu}) + \frac{1}{T-t} \cdot \sum_{\tau=t+1}^{T} [\tilde{r}_{\tau} - \tilde{\boldsymbol{a}}_{\tau}^{\top}\boldsymbol{\mu}]^{+}.$$

We denote by

$$\boldsymbol{\mu}^*(\boldsymbol{c}') \in \arg\min_{\mu\in\hat{\mathcal{S}}} \hat{G}_{\boldsymbol{c}'}(\boldsymbol{\mu}) \text{ and } \tilde{\boldsymbol{\mu}}(\boldsymbol{c}') \in \arg\min_{\boldsymbol{\mu}\in\hat{\mathcal{S}}} \hat{G}_{\boldsymbol{c}',I_{t+1}}(\boldsymbol{\mu}) \tag{92}$$

for any $\boldsymbol{c}' \geq 0$, where $\tilde{\boldsymbol{\mu}}(\boldsymbol{c}')$ is a random variable whose value depends on I_{t+1}. Clearly, we have that

$$\mu^*(\boldsymbol{c}) = \mathcal{P}_S(\hat{\boldsymbol{\mu}}_1), \text{ and } \tilde{\boldsymbol{\mu}}(\boldsymbol{c}) = \mathcal{P}_S(\tilde{\boldsymbol{\mu}}_1)., \text{ and } \tilde{\boldsymbol{\mu}}(\boldsymbol{c} - \boldsymbol{a}_t) = \mathcal{P}_S(\tilde{\boldsymbol{\mu}}_2) \tag{93}$$

Thus, in order to bound $\mathbb{E}[(\boldsymbol{a}_t^{\top}\tilde{\boldsymbol{\mu}}_1 - \boldsymbol{a}_t^{\top}\hat{\boldsymbol{\mu}}_1)^2]$, it is sufficient to bound

$$\mathbb{E}[(\boldsymbol{\mu}^*(\boldsymbol{c}') - \tilde{\boldsymbol{\mu}}(\boldsymbol{c}'))^2] \tag{94}$$

for any $\boldsymbol{c}' \geq 0$, where the $\mathbb{E}[\cdot]$ is taken over I_{t+1}. In what follows, we will consider bounding (94). Since we will assume that $\boldsymbol{c}'$ is now fixed, for notation simplicity, we will drop $\boldsymbol{c}'$ in the expression of $\boldsymbol{\mu}^*(\boldsymbol{c}')$, $\tilde{\boldsymbol{\mu}}(\boldsymbol{c}')$, $\hat{G}_{\boldsymbol{c}'}(\boldsymbol{\mu})$ and $\hat{G}_{\boldsymbol{c}',I_{t+1}}(\boldsymbol{\mu})$. We simply denote $\boldsymbol{\mu}^*$, $\tilde{\boldsymbol{\mu}}$, $\hat{G}(\boldsymbol{\mu})$ and $\hat{G}_{I_{t+1}}(\boldsymbol{\mu})$.

We first note that the function $h(\boldsymbol{\mu})$ is a convex function over $\boldsymbol{\mu} \in \hat{\mathcal{S}}$. We also note that the function $\hat{G}$ is simply a re-formulation of the function $\hat{L}$ after projecting the decision variable onto the subspace $\mathcal{S}$. Thus, Lemma 6 implies the following second-order growth condition for the function $\hat{G}$, i.e.,

$$\hat{G}(\boldsymbol{\mu}) - \hat{G}(\boldsymbol{\mu}^*) \geq \frac{\underline{\alpha}\underline{\beta}}{2} \cdot \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2^2 \text{ for any } \boldsymbol{\mu} \in \hat{\mathcal{S}} \subset \mathcal{S}. \tag{95}$$

which implies that $\boldsymbol{\mu}^*$ is unique. Thus, from Lemma 2.1 in Shapiro (1992), we know that

$$\|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2 \leq \frac{2}{\underline{\alpha}\underline{\beta}} \cdot \sup_{\boldsymbol{\mu}\in\hat{\mathcal{S}}\cap B(\boldsymbol{\mu}^*,r_0)} \left\{ \frac{\hat{G}_{I_{t+1}}(\boldsymbol{\mu}) - \hat{G}(\boldsymbol{\mu}) - \hat{G}_{I_{t+1}}(\boldsymbol{\mu}^*) + \hat{G}(\boldsymbol{\mu}^*)}{\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2} \right\}$$

where $r_0 = \|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2$ and $B(\boldsymbol{\mu}^*, r_0) = \{\boldsymbol{\mu} : \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2 \leq r_0\}$. Therefore, in order to further bound $\|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2$, it is sufficient to bound the term

$$\sup_{\boldsymbol{\mu}\in\hat{\mathcal{S}}\cap B(\boldsymbol{\mu}^*,r_0)} \left\{ \frac{\frac{1}{T-t}\sum_{\tau=t+1}^{T}[\tilde{r}_{\tau} - \tilde{\boldsymbol{a}}_{\tau}^{\top}\boldsymbol{\mu}]^{+} - \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}]^{+} - \frac{1}{T-t}\sum_{\tau=t+1}^{T}[\tilde{r}_{\tau} - \tilde{\boldsymbol{a}}_{\tau}^{\top}\boldsymbol{\mu}^*]^{+} + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^{\top}\boldsymbol{\mu}^*]^{+}}{\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2} \right\}.$$

We denote by the function

$$g(\boldsymbol{\mu}, (r, \boldsymbol{a})) := [r - \boldsymbol{a}^{\top}\boldsymbol{\mu}]^{+} + \bar{\alpha}\bar{d}^2 \cdot \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2^2.$$

Moreover, we denote by

$$\tilde{\delta}(\boldsymbol{\mu}) = \frac{1}{T-t} \cdot \sum_{\tau=t+1}^{T} g(\boldsymbol{\mu}, (\tilde{r}_{\tau}, \tilde{\boldsymbol{a}}_{\tau})) - \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g(\boldsymbol{\mu}, (\tilde{r}, \tilde{\boldsymbol{a}}))] \tag{96}$$

which is a random variable that depends on I_{t+1}. Then, the above upper bound over $\|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2$ can be re-formulated as

$$\|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2 \leq \frac{2}{\underline{\alpha}\underline{\beta}} \cdot \sup_{\boldsymbol{\mu} \in \hat{\mathcal{S}} \cap B(\boldsymbol{\mu}^*, r_0)} \left\{ \frac{\tilde{\delta}(\boldsymbol{\mu}) - \tilde{\delta}(\boldsymbol{\mu}^*)}{\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2} \right\}.$$

We denote by $\tilde{E}$ the set of $\boldsymbol{\mu}$ such that $\tilde{\delta}(\boldsymbol{\mu})$ is non-differentiable. For simplicity, we denote by $\hat{B}(\boldsymbol{\mu}^*, r_0) = \hat{\mathcal{S}} \cap B(\boldsymbol{\mu}^*, r_0)$. Note that $\tilde{\delta}(\boldsymbol{\mu})$ is Liptschitz continuous. Then, from the mean value theorem for Liptschitz function (Clarke (1990) p41), we have that

$$\|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2^2 \leq \frac{4}{\underline{\alpha}^2 \underline{\beta}^2} \cdot \sup_{\boldsymbol{\mu} \in \hat{B}(\boldsymbol{\mu}^*, r_0) \setminus E} \{\|\nabla \tilde{\delta}(\boldsymbol{\mu})\|_2^2\} = \frac{4}{\underline{\alpha}^2 \underline{\beta}^2} \cdot (\|\nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2 + \sup_{\boldsymbol{\mu} \in \hat{B}(\boldsymbol{\mu}^*, r_0) \setminus \tilde{E}} \{\|\nabla \tilde{\delta}(\boldsymbol{\mu}) - \nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2\}) \quad (97)$$

We have the following claim over the term $\|\nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2$, which is proved at the end of this proof.

Claim 5. *It holds that*

$$\mathbb{E}[\|\nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2] \leq \frac{\bar{d}^2}{T - t}$$

where the expectation $\mathbb{E}[\cdot]$ *takes over* $I_{t+1} = \{(\tilde{r}_{t+1}, \tilde{\boldsymbol{a}}_{t+1}), \ldots, (\tilde{r}_T, \tilde{\boldsymbol{a}}_T)\}$ *that defines* $\tilde{\delta}(\cdot)$ *in* (96).

Regarding the term $\sup_{\boldsymbol{\mu} \in \hat{B}(\boldsymbol{\mu}^*, r_0) \setminus \tilde{E}} \{\|\nabla \tilde{\delta}(\boldsymbol{\mu}) - \nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2\}$, following the steps in Shapiro (1993), we can show that

$$\sup_{\boldsymbol{\mu} \in B(\mu^*, r_0) \setminus \tilde{E}} \{ \frac{\|\nabla \tilde{\delta}(\boldsymbol{\mu}) - \nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2}{\frac{1}{\sqrt{T-t}} + r_0} \} \to 0 \text{ in probability as } T - t \to \infty.$$

As a result, there exists a constant t_0 such that as long as $T - t \geq t_0$, we have that

$$\sup_{\boldsymbol{\mu} \in \hat{B}(\boldsymbol{\mu}^*, r_0) \setminus \tilde{E}} \{\|\nabla \tilde{\delta}(\boldsymbol{\mu}) - \nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2\} \leq \frac{1}{2} \cdot (\frac{1}{T - t} + r_0^2)$$

with a high probability. We formalize the above step in the following claim for completeness, by making explicit the constant term t_0, as well as the "high probability".

Claim 6. *It holds that*

$$P\left(\sup_{\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2 \leq r_0} \frac{\|\nabla \tilde{\delta}(\boldsymbol{\mu}) - \nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2}{\frac{1}{\sqrt{T-t}} + 3\bar{\alpha}\bar{d}^2 \cdot r_0} \geq \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2} \right) \leq \delta(\bar{\alpha}, \bar{d}, m, \gamma) \cdot \frac{\log(T - t)}{(T - t)^{\frac{12}{25}}} \quad (98)$$

as long as $T - t \geq t_1(\bar{\alpha}, \bar{d})$, *where* $\delta_1(\bar{\alpha}, \bar{d}, m, \gamma)$ *is a constant that depends solely on the parameters* $\bar{\alpha}, \bar{d}, m, \gamma$ *and* $t_1(\bar{\alpha}, \bar{d})$ *is a constant that depends solely on the parameters* $\bar{\alpha}, \bar{d}$.

The proof of Claim 6 is relegated to the end of this proof. Denote by

$$\tilde{z} = \sup_{\|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|_2 \leq r_0} \frac{\|\nabla \tilde{\delta}(\boldsymbol{\mu}) - \nabla \tilde{\delta}(\boldsymbol{\mu}^*)\|_2}{\frac{1}{\sqrt{T-t}} + 3\bar{\alpha}\bar{d}^2 \cdot r_0}.$$

We also denote by three events

$$\mathcal{E}_1 = \{\tilde{z} \leq \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2}\} \text{ and } \mathcal{E}_2 = \{\frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2} < \tilde{z} \text{ and } r_0 \leq \frac{1}{(T - t)^{\frac{7}{25}}}\} \text{ and } \mathcal{E}_3 = \{\frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2} < \tilde{z} \text{ and } r_0 > \frac{1}{(T - t)^{\frac{7}{25}}}\}$$

Clearly, we have

$$\mathbb{E}[r_0^2] = P(\mathcal{E}_1) \cdot \mathbb{E}[r_0^2 | \mathcal{E}_1] + P(\mathcal{E}_2) \cdot \mathbb{E}[r_0^2 | \mathcal{E}_2] + P(\mathcal{E}_2) \cdot \mathbb{E}[r_0^2 | \mathcal{E}_2] \quad (99)$$

Now, from (97), Claim 5, and the definition of the event $\mathcal{E}_1$, we have

$$\begin{aligned} P(\mathcal{E}_1)\cdot\mathbb{E}[r_0^2|\mathcal{E}_1] &\leq \frac{4}{\underline{\alpha}^2\underline{\beta}^2}\cdot P(\mathcal{E}_1)\cdot\mathbb{E}[\|\nabla\tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2|\mathcal{E}_1] + P(\mathcal{E}_1)\cdot\frac{1}{36\bar{\alpha}^2\bar{d}^4}\cdot\mathbb{E}\left[\left(\frac{1}{\sqrt{T-t}}+3\bar{\alpha}\bar{d}^2\cdot r_0\right)^2|\,\mathcal{E}_1\right] \\ &\leq \frac{4}{\underline{\alpha}^2\underline{\beta}^2}\cdot\mathbb{E}[\|\nabla\tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2] + \frac{1}{2}\cdot P(\mathcal{E}_1)\cdot\mathbb{E}[r_0^2|\mathcal{E}_1] + \frac{1}{18\bar{\alpha}^2\bar{d}^4\cdot(T-t)} \end{aligned} \tag{100}$$

which implies that

$$P(\mathcal{E}_1)\cdot\mathbb{E}[r_0^2|\mathcal{E}_1] \leq \frac{8\bar{d}^2}{\underline{\alpha}^2\underline{\beta}^2\cdot(T-t)} + \frac{1}{9\bar{\alpha}\bar{d}^2\cdot(T-t)} \tag{101}$$

For the term $P(\mathcal{E}_2)\cdot\mathbb{E}[r_0^2|\mathcal{E}_2]$, from the definition of the event $\mathcal{E}_2$ and Claim 6, we have

$$P(\mathcal{E}_2)\cdot\mathbb{E}[r_0^2|\mathcal{E}_2] \leq P(\mathcal{E}_2)\cdot\frac{1}{(T-t)^{\frac{14}{25}}} \leq \delta(\bar{\alpha},\bar{d},m,\gamma)\cdot\frac{\log(T-t)}{(T-t)^{\frac{12}{25}}}\cdot\frac{1}{(T-t)^{\frac{14}{25}}} \leq \frac{1}{T-t} \tag{102}$$

as long as $T-t > t_2(\bar{\alpha},\bar{d},m,\gamma)$, where $t_2(\bar{\alpha},\bar{d},m,\gamma)$ is a constant that depends solely on parameters $\bar{\alpha},\bar{d},m,\gamma$.

We now want to bound the probability that event $\mathcal{E}_3$ happens. Note that we have

$$\frac{\underline{\alpha}\underline{\beta}}{2}\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2 \leq \frac{1}{T-t}\sum_{\tau=t+1}^{T}[\tilde{r}_\tau-\tilde{\boldsymbol{a}}_\tau^\top\tilde{\boldsymbol{\mu}}]^+ - \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r}-\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\mu}}]^+ - \frac{1}{T-t}\sum_{\tau=t+1}^{T}[\tilde{r}_\tau-\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*]^+ + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r}-\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*]^+. \tag{103}$$

By modifying the approach in Li and Ye (2021), we have the following lemma, which follows Proposition 2 and Proposition 3 of Li and Ye (2021) and is proved at the end of this proof.

Lemma 14. *It holds that*

$$P\left(\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2 \geq \epsilon^2\right) \leq K\cdot\log(\frac{1}{\epsilon})\exp(-\epsilon^2(T-t)).$$

for a constant K that only depends on parameters $\underline{\alpha},\underline{\beta},\bar{\alpha},\bar{d}$

By setting $\epsilon = \frac{1}{(T-t)^{\frac{7}{25}}}$ in Lemma 14, we have that

$$P(\mathcal{E}_3) \leq P\left(\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2 \geq \left(\frac{1}{(T-t)^{\frac{7}{25}}}\right)^2\right) \leq K'\cdot\log(T-t)\cdot\exp(-(T-t)^{\frac{11}{25}}) \leq \frac{K''}{T-t} \tag{104}$$

as long as $T-t > t_3$, where K', K'', t_3 are constants. Therefore, combining (99), (101), (102), and (104), we have

$$\mathbb{E}[\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2] = \mathbb{E}[r_0^2] \leq \frac{\delta_2}{T-t}$$

as long as $T-t > t_4$, where δ_2 and t_4 are constants. Therefore, our proof is completed. □

Proof of Claim 5. Note that

$$\nabla\tilde{\delta}(\boldsymbol{\mu}^*) = -\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\tilde{\boldsymbol{a}}_\tau\cdot\mathbb{1}_{\{\tilde{r}_\tau\geq\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*\}} + \mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}}\cdot(1-F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*|\tilde{\boldsymbol{a}}))]$$

Clearly, for each $\tau = t+1,\ldots,T$, we have

$$\mathbb{E}_{(\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau)}[\tilde{\boldsymbol{a}}_\tau\cdot\mathbb{1}_{\{\tilde{r}_\tau\geq\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*\}}] = \mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}}\cdot(1-F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*|\tilde{\boldsymbol{a}}))].$$

Thus, we have

$$\mathbb{E}[\|\nabla\tilde{\delta}(\boldsymbol{\mu}^*)\|_2^2] = \frac{1}{T-t}\cdot\mathrm{Var}(\tilde{\boldsymbol{a}}_\tau\cdot\mathbb{1}_{\{\tilde{r}_\tau\geq\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*\}}) \leq \frac{\bar{d}^2}{T-t}$$

which completes our proof. □

Proof of Claim 6. We denote by

$$\psi((r,\boldsymbol{a}),\boldsymbol{\mu}) = \nabla g(\boldsymbol{\mu},(r,\boldsymbol{a})) - \nabla g(\boldsymbol{\mu}^*,(r,\boldsymbol{a}))$$

We also denote by

$$\lambda(\boldsymbol{\mu}) = \mathbb{E}_{(r,\boldsymbol{a})}[\psi((r,\boldsymbol{a}),\boldsymbol{\mu})]$$

and

$$u((r,\boldsymbol{a}),\boldsymbol{\mu},d) = \sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2 \le d} \|\psi((r,\boldsymbol{a}),\boldsymbol{\mu}) - \psi((r,\boldsymbol{a}),\boldsymbol{\mu}^*)\|_2$$

Clearly, we have that

$$\lambda(\boldsymbol{\mu}^*) = 0 \tag{105}$$

Thus, condition (N-2) in Huber (1967) is satisfied. Also, for any $\boldsymbol{\mu}$, from definition of the functions ψ and λ, we have that

$$\|\lambda(\boldsymbol{\mu})\|_2 = \|2\bar{\alpha}\bar{d}^2(\boldsymbol{\mu}-\boldsymbol{\mu}^*) + \mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}} \cdot (F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}|\tilde{\boldsymbol{a}}) - F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*|\tilde{\boldsymbol{a}}))]\|_2$$

Note that

$$F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}|\tilde{\boldsymbol{a}}) - F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*|\tilde{\boldsymbol{a}}) \le \bar{\alpha} \cdot |\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu} - \tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*| \le \bar{\alpha}\bar{d} \cdot \|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2$$

Thus, we have

$$\begin{aligned} \|\lambda(\boldsymbol{\mu})\|_2 &\ge 2\bar{\alpha}\bar{d}^2 \cdot \|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2 - \|\mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}} \cdot (F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}|\tilde{\boldsymbol{a}}) - F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}|\tilde{\boldsymbol{a}}))]\|_2 \\ &\ge 2\bar{\alpha}\bar{d}^2 \cdot \|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2 - \bar{\alpha}\bar{d}^2 \cdot \|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2 = \bar{\alpha}\bar{d}^2 \cdot \|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2 \end{aligned} \tag{106}$$

for any $d_0 \ge 0$ and $\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2 \le d_0$. Thus, we verify that condition (N-3) (i) in Huber (1967) holds. Moreover, we have

$$\|\psi((r,\boldsymbol{a}),\boldsymbol{\mu}) - \psi((r,\boldsymbol{a}),\boldsymbol{\mu}^*)\|_2 = \|\boldsymbol{a} \cdot \mathbb{1}_{\{r \ge \boldsymbol{a}^\top\boldsymbol{\mu}^*\}} - \boldsymbol{a} \cdot \mathbb{1}_{\{r \ge \boldsymbol{a}^\top\boldsymbol{\mu}\}}\| \le \|\boldsymbol{a} \cdot (\mathbb{1}_{\{\boldsymbol{a}^\top\boldsymbol{\mu} \le r \le \boldsymbol{a}^\top\boldsymbol{\mu}^*\}} + \mathbb{1}_{\{\boldsymbol{a}^\top\boldsymbol{\mu}^* \le r \le \boldsymbol{a}^\top\boldsymbol{\mu}\}})\| \le \bar{d} \tag{107}$$

which implies that

$$\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[u((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\mu},d)] \le \mathbb{E}_{\tilde{\boldsymbol{a}}}[\|\tilde{\boldsymbol{a}}\|_2 \cdot 2\bar{\alpha}\bar{d}d] \le 2\bar{\alpha}\bar{d}^2 \cdot d \tag{108}$$

Similarly, we have

$$\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[u((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\mu},d)^2] \le \mathbb{E}_{\tilde{\boldsymbol{a}}}[\|\tilde{\boldsymbol{a}}\|_2^2 \cdot 2\bar{\alpha}\bar{d}d] \le 2\bar{\alpha}\bar{d}^3 \cdot d$$

Thus, we verify conditions (N-3) (ii) and (iii) in Huber (1967) for any $d_0 \ge 0$. Clearly, $\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\psi((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\mu}^*)]$ is finite, which verifies condition (N-4) in Huber (1967). Thus, denote by

$$Z_{T-t}(\boldsymbol{\mu},\boldsymbol{\mu}^*) = \frac{\|\frac{1}{T-t} \cdot \sum_{\tau=t+1}^{T}(\psi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}) - \psi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}^*)) - \lambda(\boldsymbol{\mu}) + \lambda(\boldsymbol{\mu}^*)\|}{\frac{1}{\sqrt{T-t}} + \|\lambda(\boldsymbol{\mu})\|_2}$$

By Lemma 13, for any $\epsilon > 0$ and any $\gamma' \in (\frac{1}{2},1)$, we have

$$\begin{aligned} &P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\| \le d_0} Z_{T-t}(\boldsymbol{\mu},\boldsymbol{\mu}^*) \ge 2\epsilon\right) \le 2\bar{\alpha}\bar{d}^3\epsilon^{-2}(T-t)^{-\gamma'} \\ &+ (T-t)^{\gamma'-1}\left(\frac{2\bar{\alpha}\bar{d}^3}{4\bar{\alpha}^2\bar{d}^4 q(1-q)} + \frac{2\bar{\alpha}\bar{d}^3}{36\bar{\alpha}^2\bar{d}^4 q^2(1-q)^2}\right)\left(\frac{\gamma'(\log(T-t)+\log d_0)}{|\log(1-q)|} + 1\right)(2M)^m \end{aligned} \tag{109}$$

as long as $T-t \ge n_0$, where n_0 satisfying $n_0^{\gamma'-\frac{1}{2}} = \frac{4\bar{\alpha}\bar{d}^2}{\epsilon}$ and $\gamma' \in (\frac{1}{2},1)$ is an arbitrary number. Moreover, we set $M \ge (6\bar{\alpha}^2\bar{d}^2)/(\epsilon\bar{\alpha}\bar{d}^2) = 6\bar{\alpha}/\epsilon$ and $q = 1/M$.

We first specify $\epsilon = \frac{\underline{\alpha}\underline{\beta}}{24\bar{\alpha}\bar{d}^2}$ and $\gamma' = 13/25$. Then, (109) will become

$$P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} Z_{T-t}(\boldsymbol{\mu},\boldsymbol{\mu}^*) \ge \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2}\right) \le 2\bar{\alpha}\bar{d}^3\left(\frac{12\bar{\alpha}\bar{d}^2}{\underline{\alpha}\underline{\beta}}\right)^2 (T-t)^{-\frac{13}{25}} \tag{110}$$
$$+ (T-t)^{-\frac{12}{25}}\left(\frac{2\bar{\alpha}\bar{d}^3}{4\bar{\alpha}^2\bar{d}^4 q(1-q)} + \frac{2\bar{\alpha}\bar{d}^3}{36\bar{\alpha}^2\bar{d}^4 q^2(1-q)^2}\right)\left(\frac{13(\log(T-t)+\log d_0)}{25|\log(1-q)|}+1\right)(2M)^m$$

as long as $T-t \ge (96\bar{\alpha}^2\bar{d}^4/(\underline{\alpha}\underline{\beta}))^{50}$, $M = 144\bar{\alpha}^2\bar{d}^2/(\underline{\alpha}\underline{\beta})$ and $q = 1/M = \underline{\alpha}\underline{\beta}/(144\bar{\alpha}^2\bar{d}^2)$. Denote by $\delta(\bar{\alpha},\bar{d},m)$ a constant that depends solely on $\bar{\alpha}$, $\bar{d}$ and m, i.e.,

$$\delta(\bar{\alpha},\bar{d},m) = 2\bar{\alpha}\bar{d}^3\left(\frac{12\bar{\alpha}\bar{d}^2}{\underline{\alpha}\underline{\beta}}\right)^2 + \left(\frac{2\bar{\alpha}\bar{d}^3}{4\bar{\alpha}^2\bar{d}^4 q(1-q)} + \frac{2\bar{\alpha}\bar{d}^3}{36\bar{\alpha}^2\bar{d}^4 q^2(1-q)^2}\right)\left(\frac{13}{25|\log(1-q)|}+1\right)(2M)^m \tag{111}$$

with $M = 144\bar{\alpha}^2\bar{d}^2/(\underline{\alpha}\underline{\beta})$ and $q = 1/M = \underline{\alpha}\underline{\beta}/(144\bar{\alpha}^2\bar{d}^2)$. Then, from (110), we have

$$P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} Z_{T-t}(\boldsymbol{\mu},\boldsymbol{\mu}^*) \ge \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2}\right) \le \delta(\bar{\alpha},\bar{d},m)\cdot\frac{\log(T-t)+\log d_0}{(T-t)^{\frac{12}{25}}} \tag{112}$$

as long as $T-t \ge (96\bar{\alpha}^2\bar{d}^4/(\underline{\alpha}\underline{\beta}))^{50}$. Note that

$$\nabla\delta(\boldsymbol{\mu}) - \nabla\delta(\boldsymbol{\mu}^*) = \frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}(\psi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}) - \psi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}^*)) - \lambda(\boldsymbol{\mu}) + \lambda(\boldsymbol{\mu}^*).$$

We know that (112) implies that

$$P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} \frac{\|\nabla\delta(\boldsymbol{\mu}) - \nabla\delta(\boldsymbol{\mu}^*)\|_2}{\frac{1}{\sqrt{T-t}} + \|\lambda(\boldsymbol{\mu})\|_2} \ge \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2}\right) \le \delta(\bar{\alpha},\bar{d},m)\cdot\frac{\log(T-t)+\log d_0}{(T-t)^{\frac{12}{25}}}. \tag{113}$$

From Assumption 1, we have $\|\tilde{\boldsymbol{\mu}}\|_2 \le \gamma\sqrt{m}$ and $\|\boldsymbol{\mu}^*\|_2 \le \gamma\sqrt{m}$, which implies that $r_0 = \|\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*\|_2 \le 2\gamma\sqrt{m}$. Therefore, we can simply set $d_0 = 2\gamma\sqrt{m}$ and we know that r_0 is always upper bounded by d_0, which implies that

$$P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le r_0} \frac{\|\nabla\delta(\boldsymbol{\mu}) - \nabla\delta(\boldsymbol{\mu}^*)\|_2}{\frac{1}{\sqrt{T-t}} + \|\lambda(\boldsymbol{\mu})\|_2} \ge \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2}\right) \le P\left(\sup_{\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|\le d_0} \frac{\|\nabla\delta(\boldsymbol{\mu}) - \nabla\delta(\boldsymbol{\mu}^*)\|_2}{\frac{1}{\sqrt{T-t}} + \|\lambda(\boldsymbol{\mu})\|_2} \ge \frac{\underline{\alpha}\underline{\beta}}{12\bar{\alpha}\bar{d}^2}\right)$$
$$\le \delta(\bar{\alpha},\bar{d},m)\cdot\frac{\log(T-t)+\log(2\gamma\sqrt{m})}{(T-t)^{\frac{12}{25}}} \tag{114}$$

where the first inequality follows from $r_0 \le d_0 = 2\gamma\sqrt{m}$ almost surely and the second inequality follows from (113) by setting $d_0 = 2\gamma\sqrt{m}$. Therefore, we denote by $\delta(\bar{\alpha},\bar{d},m,\gamma) = \delta(\bar{\alpha},\bar{d},m)\cdot\log(2\gamma\sqrt{m})$ and our proof of (98) is completed by noting that

$$\|\lambda(\boldsymbol{\mu})\|_2 = \|2\bar{\alpha}\bar{d}^2(\boldsymbol{\mu}-\boldsymbol{\mu}^*) + \mathbb{E}_{\tilde{\boldsymbol{a}}}[\tilde{\boldsymbol{a}}\cdot(F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}|\tilde{\boldsymbol{a}}) - F(\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*|\tilde{\boldsymbol{a}}))]\|_2 \le 3\bar{\alpha}\bar{d}^2\cdot\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2.$$

□

Proof of Lemma 14. Following Li and Ye (2021), we denote by

$$\phi((r,\boldsymbol{a}),\boldsymbol{\mu}) = -\boldsymbol{a}^\top\cdot\mathbb{1}(r > \boldsymbol{a}^\top\boldsymbol{\mu}).$$

Then, from Lemma 1 of Li and Ye (2021), we have

$$\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\mu}}]^+ - \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\tilde{r} - \tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*]^+ = \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\phi((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\mu}^*)]\cdot(\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}^*) + \mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}\left[\int_{\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\mu}}}^{\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*}(\mathbb{1}(\tilde{r} > v) - \mathbb{1}(\tilde{r} > \tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*))dv\right]. \tag{115}$$

Also, from Lemma 2 of Li and Ye (2021), we have

$$\begin{aligned}\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}([\tilde{r}_\tau-\tilde{\boldsymbol{a}}_\tau^\top\tilde{\boldsymbol{\mu}}]^+-[\tilde{r}_\tau-\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*]^+)=&\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\phi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}^*)\cdot(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*)\\&+\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\int_{\tilde{\boldsymbol{a}}_\tau^\top\tilde{\boldsymbol{\mu}}}^{\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*}(\mathbb{1}(\tilde{r}_\tau>v)-\mathbb{1}(\tilde{r}_\tau>\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*))dv\end{aligned}\tag{116}$$

Plugging (115) and (116) into (103), we have that

$$\begin{aligned}\frac{\alpha\beta}{2}\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\le&\underbrace{\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\phi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}^*)\cdot(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\phi((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\mu}^*)]\cdot(\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*)}_{I}\\&\underbrace{+\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\int_{\tilde{\boldsymbol{a}}_\tau^\top\tilde{\boldsymbol{\mu}}}^{\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*}(\mathbb{1}(\tilde{r}_\tau>v)-\mathbb{1}(\tilde{r}_\tau>\tilde{\boldsymbol{a}}_\tau^\top\boldsymbol{\mu}^*))dv-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}\left[\int_{\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\mu}}}^{\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*}(\mathbb{1}(\tilde{r}>v)-\mathbb{1}(\tilde{r}>\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*))dv\right]}_{II}\end{aligned}\tag{117}$$

We proceed by bounding the term I and term II separately.

Bound I: by Matrix Hoeffding's inequality, we have that

$$P\left(\|\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\phi((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\phi((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\mu}^*)]\|_2\ge\epsilon\right)\le m\cdot\exp\left(-\frac{\epsilon^2\cdot(T-t)}{\bar{d}^2}\right)$$

which implies that

$$P\left(|I|\ge\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2\right)\le m\cdot\exp\left(-\frac{\epsilon^2\cdot(T-t)}{\bar{d}^2}\right)\tag{118}$$

Bound II: We define a function

$$\eta((r,\boldsymbol{a}),\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)=\int_{\boldsymbol{a}^\top\boldsymbol{\mu}_1}^{\boldsymbol{a}^\top\boldsymbol{\mu}_2}(\mathbb{1}(r>v)-\mathbb{1}(r>\boldsymbol{a}^\top\boldsymbol{\mu}^*))dv$$

for any $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2\in\hat{\mathcal{S}}$. We have

$$II=\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)]\tag{119}$$

We utilize a splitting scheme to split the set $\{\boldsymbol{\mu}:\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2\le\sqrt{m}\cdot\gamma\}$ into disjoint cubes, similar to Huber (1967) and Li and Ye (2021). It is clear to see that $\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2\le\sqrt{m}\cdot\gamma$. The idea is to divide the region $\{\boldsymbol{\mu}:\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2\le\sqrt{m}\cdot\gamma\}$ into a slowly increasing number of smaller cubes. We consider the concentric cubes:

$$C_k=\{\boldsymbol{\mu}:\|\boldsymbol{\mu}-\boldsymbol{\mu}^*\|_2\le(1-q)^k\cdot\sqrt{m}\gamma\},\quad k=1,\dots,k_0$$

where q is a constant such that $\bar{\alpha}\cdot\bar{d}^2\cdot\frac{q(1+q)}{(1-q)^2}\le\frac{\alpha\beta}{32}$, and k_0 is a constants such that $(1-q)^{k_0}\cdot\sqrt{m}\gamma=\epsilon$. We then further divide the region $C_{k-1}\setminus C_k$ into cubes $\{\Omega_{kl}\}_{l=1}^{l_k}$ with edges of length $2(1-q)^k\cdot q\cdot\sqrt{m}\gamma$ such that the centers of these cubes $\boldsymbol{\xi}_{kl}$ satisfies

$$\|\boldsymbol{\xi}_{kl}\|_2=(1-q)^{k-1}(1-\frac{q}{2})\cdot\sqrt{m}\gamma,\quad l=1,\dots,k_l.$$

In total, there are no more than $k_0\cdot(\frac{2}{q})^m$ number of cubes. We now denote by $\tilde{\Omega}$ the cube that contains $\tilde{\boldsymbol{\mu}}$ and denote by $\tilde{\boldsymbol{\xi}}$ the center of the cube $\tilde{\Omega}$. Note that both $\tilde{\Omega}$ and $\tilde{\boldsymbol{\xi}}$ are random, where the randomness arises from $\tilde{\boldsymbol{\mu}}$.

$$\begin{aligned}\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)]=&\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})]\\&+\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)]\end{aligned}$$

which implies that

$$\begin{aligned}
&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)]\geq\epsilon^2+\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{8}\right)\\
\leq&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})]\geq\frac{\epsilon^2}{2}+\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{16}\right)\\
&+P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)]\geq\frac{\epsilon^2}{2}+\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{16}\right)
\end{aligned}\tag{120}$$

We have the following result.

Claim 7. *It holds that*

$$\begin{aligned}
&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})]\geq\frac{\epsilon^2}{2}+\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}\|_2^2}{16}\right)\\
&\leq k_0\cdot\left(\frac{2}{q}\right)^m\cdot\exp\left(-\frac{\epsilon^2(T-t)}{2q^2\bar{d}^2}\right)
\end{aligned}$$

where q is a constant such that $\bar{\alpha}\cdot\bar{d}^2\cdot\frac{q(1+q)}{(1-q)^2}\leq\frac{\underline{\alpha}\beta}{32}$, and k_0 is a constants such that $(1-q)^{k_0}\cdot\sqrt{m}\gamma=\epsilon$.

We also have the following result.

Claim 8. *It holds that*

$$\begin{aligned}
&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)]\geq\frac{\epsilon^2}{2}+\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{16}\right)\\
&\leq k_0\cdot\left(\frac{2}{q}\right)^m\cdot\exp\left(-\frac{\epsilon^2\cdot(T-t)q^2}{2\bar{d}^2}\right)
\end{aligned}$$

where q is a constant such that $\bar{\alpha}\cdot\bar{d}^2\cdot\frac{q(1+q)}{(1-q)^2}\leq\frac{\underline{\alpha}\beta}{32}$, and k_0 is a constants such that $(1-q)^{k_0}\cdot\sqrt{m}\gamma=\epsilon$.

Combining Claim 7, Claim 8 and (120), we have that

$$\begin{aligned}
&P\left(\mathrm{II}\geq\epsilon^2+\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{8}\right)\\
=&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\boldsymbol{\mu}^*)]\geq\epsilon^2+\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{8}\right)\\
\leq&k_0\cdot\left(\frac{2}{q}\right)^m\cdot\exp\left(-\frac{\epsilon^2(T-t)}{2q^2\bar{d}^2}\right)+k_0\cdot\left(\frac{2}{q}\right)^m\cdot\exp\left(-\frac{\epsilon^2\cdot(T-t)q^2}{2\bar{d}^2}\right)
\end{aligned}\tag{121}$$

where q is a constant such that $\bar{\alpha}\cdot\bar{d}^2\cdot\frac{q(1+q)}{(1-q)^2}\leq\frac{\underline{\alpha}\beta}{32}$, and k_0 is a constants such that $(1-q)^{k_0}\cdot\sqrt{m}\gamma=\epsilon$. Our bound of the term II is now completed.

We now use the bound on I and II to prove the probability bound on the event $\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\geq\epsilon^2$. Combining (118) and (121), we have

$$\begin{aligned}
&P\left(\mathrm{I}+\mathrm{II}\geq2\epsilon^2+2\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{4}\right)\\
\leq&P\left(\mathrm{I}\geq\epsilon^2+\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{8}\right)+P\left(\mathrm{II}\geq\epsilon^2+\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\beta\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{8}\right)\\
\leq&m\cdot\exp\left(-\frac{\epsilon^2\cdot(T-t)}{\bar{d}^2}\right)+k_0\cdot\left(\frac{2}{q}\right)^m\cdot\left(\exp\left(-\frac{\epsilon^2(T-t)}{2q^2\bar{d}^2}\right)+\exp\left(-\frac{\epsilon^2\cdot(T-t)q^2}{2\bar{d}^2}\right)\right)
\end{aligned}$$

where c_2 is a constant. Further note that

$$\frac{\underline{\alpha}\underline{\beta}}{2}\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\le \mathrm{I}+\mathrm{II},$$

and

$$\frac{\underline{\alpha}\underline{\beta}}{2}\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\ge 2\epsilon^2+2\epsilon\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2+\frac{\underline{\alpha}\underline{\beta}\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{4}$$

implies that

$$\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\ge\epsilon^2\cdot\frac{4(2+\sqrt{2\underline{\alpha}\underline{\beta}+4})^2}{\underline{\alpha}^2\underline{\beta}^2}.$$

Therefore, we have that

$$\begin{aligned}P\left(\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\ge\epsilon^2\cdot\frac{4(2+\sqrt{2\underline{\alpha}\underline{\beta}+4})^2}{\underline{\alpha}^2\underline{\beta}^2}\right)\le& m\cdot\exp\left(-\frac{\epsilon^2\cdot(T-t)}{\bar{d}^2}\right)\\&+k_0\cdot\left(\frac{2}{q}\right)^m\cdot\left(\exp\left(-\frac{\epsilon^2(T-t)}{2q^2\bar{d}^2}\right)+\exp\left(-\frac{\epsilon^2\cdot(T-t)q^2}{2\bar{d}^2}\right)\right)\end{aligned}$$

By substituting $\epsilon'=\epsilon\cdot\frac{2(2+\sqrt{2\underline{\alpha}\underline{\beta}+4})}{\underline{\alpha}\underline{\beta}}$ and noting that $k_0=O(\log(\frac{1}{\epsilon}))$, we have that

$$P\left(\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2\ge(\epsilon')^2\right)\le K\cdot\log\left(\frac{1}{\epsilon'}\right)\exp\left(-(\epsilon')^2(T-t)\right)$$

for a constant K that only depends on parameters $\underline{\alpha},\underline{\beta},\bar{\alpha},\bar{d}$, which completes our proof. □

Proof of Claim 7. We denote by

$$g((r,\boldsymbol{a}),\boldsymbol{\mu}_1,d)=\sup_{\boldsymbol{\mu}_2:\|\boldsymbol{\mu}_2-\boldsymbol{\mu}_1\|_2\le d}\eta((r,\boldsymbol{a}),\boldsymbol{\mu}_1,\boldsymbol{\mu}_2).$$

Then we have

$$\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})]\le\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\tilde{d})+\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]$$

where $\tilde{d}$ denotes the edge length of the cube $\tilde{\Omega}$. We now consider the cubes $\{\Omega_{kl}\}$ and the center concentric cubes C_{k_0} separately.

(i). If the cube $\tilde{\Omega}\in\{\Omega_{kl}\}$, then, we note that

$$\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\le\mathbb{E}_{\tilde{\boldsymbol{a}}}\left[\sup_{\tilde{\boldsymbol{\mu}}_2:\|\tilde{\boldsymbol{\mu}}_2-\tilde{\boldsymbol{\xi}}\|_2\le\tilde{d}}\int_{\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\mu}}_2}^{\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*}\int_{\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\mu}}_2}^{\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\xi}}}(\mathbb{1}(r>v)-\mathbb{1}(r>\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*))dvdF(r|\tilde{\boldsymbol{a}})\right]\le\bar{\alpha}\bar{d}^2\cdot\tilde{d}\cdot(\|\boldsymbol{\mu}^*-\tilde{\boldsymbol{\xi}}\|_2+\tilde{d})$$

By definition, we have $\tilde{d}=q\cdot\|\boldsymbol{\mu}^*-\tilde{\boldsymbol{\xi}}\|_2$ and $\|\boldsymbol{\mu}^*-\tilde{\boldsymbol{\mu}}\|_2\ge\|\boldsymbol{\mu}^*-\tilde{\boldsymbol{\xi}}\|_2-\tilde{d}$. Then, we have

$$\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\le\bar{\alpha}\cdot\bar{d}^2\cdot\frac{q(1+q)}{(1-q)^2}\cdot\|\boldsymbol{\mu}^*-\tilde{\boldsymbol{\mu}}\|_2^2\le\frac{\underline{\alpha}\underline{\beta}\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{32}$$

by specifying q such that $\bar{\alpha}\cdot\bar{d}^2\cdot\frac{q(1+q)}{(1-q)^2}\le\frac{\underline{\alpha}\underline{\beta}}{32}$.

(ii). If the cube $\tilde{\Omega}=C_{k_0}$, then by noting $\tilde{\boldsymbol{\xi}}=\boldsymbol{\mu}^*$ and $(1-q)^{k_0}\cdot\sqrt{m}\gamma=\epsilon$, we have that

$$\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\le\mathbb{E}_{\tilde{\boldsymbol{a}}}\left[\sup_{\boldsymbol{\mu}_2:\|\boldsymbol{\mu}_2-\tilde{\boldsymbol{\xi}}\|_2\le\tilde{d}}\int_{\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}_2}^{\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*}\int_{\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}_2}^{\tilde{\boldsymbol{a}}^\top\tilde{\boldsymbol{\xi}}}(\mathbb{1}(r>v)-\mathbb{1}(r>\tilde{\boldsymbol{a}}^\top\boldsymbol{\mu}^*))dvdF(r|\tilde{\boldsymbol{a}})\right]\le\bar{\alpha}\cdot\bar{d}^2\cdot\epsilon^2$$

Therefore, for both (i) and (ii), it holds that

$$
\begin{aligned}
&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\xi}})]\geq\bar{\alpha}\bar{d}^2\epsilon^2+\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}+\frac{\underline{\alpha}\underline{\beta}\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{16}\right)\\
\leq&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\tilde{d})+\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\geq\bar{\alpha}\bar{d}^2\epsilon^2+\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}+\frac{\underline{\alpha}\underline{\beta}\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2^2}{16}\right)\\
\leq&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\tilde{d})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\geq\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}\right)\\
\leq&P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\tilde{d})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\geq\frac{\epsilon\cdot\tilde{d}}{2q}\right)
\end{aligned}
$$

We apply the Hoeffding inequality to bound the above probability for each cube Ω_{kl}, and C_{k_0}. It holds that

$$g((r,\boldsymbol{a}),\boldsymbol{\xi}_{kl},d_{kl})\leq\bar{d}\cdot d_{kl},\ \text{and}\ g((r,\boldsymbol{a}),\boldsymbol{\xi}_{k_0},d_{k_0})\leq\bar{d}\cdot d_{k_0}\quad\forall(r,\boldsymbol{a})$$

where d_{kl} denotes the length of the edge of the cube Ω_{kl} and d_{k_0} denotes the length of the edge of the cube C_{k_0}. By Hoeffding's inequality, we have

$$P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\xi}_{kl},d_{kl})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\xi}_{kl},d_{kl})]\geq\frac{\epsilon\cdot d_{kl}}{2q}\right)\leq\exp\left(-\frac{\epsilon^2(T-t)}{2q^2\bar{d}^2}\right)$$

for each cube Ω_{kl}, and

$$P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\xi}_{k_0},d_{k_0})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\xi}_{k_0},d_{k_0})]\geq\frac{\epsilon\cdot d_{k_0}}{2q}\right)\leq\exp\left(-\frac{\epsilon^2(T-t)}{2q^2\bar{d}^2}\right)$$

for the cube C_{k_0}. Note that

$$\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\tilde{\boldsymbol{\xi}},\tilde{d})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\tilde{\boldsymbol{\xi}},\tilde{d})]\geq\frac{\epsilon\cdot\tilde{d}}{2q}$$

implies that

$$\frac{1}{T-t}\cdot\sum_{\tau=t}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\xi}_{k_0},d_{k_0})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\xi}_{k_0},d_{k_0})]\geq\frac{\epsilon\cdot d_{k_0}}{2q},$$

or there exists at least one Ω_{kl} such that

$$\frac{1}{T-t}\cdot\sum_{\tau=t}^{T}g((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\xi}_{kl},d_{kl})-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[g((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\xi}_{kl},d_{kl})]\geq\frac{\epsilon\cdot d_{kl}}{2q}.$$

Applying the union bound, our proof is completed. □

Proof of Claim 8. Note that

$$\eta((r,\boldsymbol{a}),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)\leq\bar{d}\cdot\|\tilde{\boldsymbol{\xi}}-\boldsymbol{\mu}^*\|_2$$

Then, if $\tilde{\Omega}\in\{\Omega_{kl}\}$, we have that

$$\eta((r,\boldsymbol{a}),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)\leq\bar{d}\cdot\|\tilde{\boldsymbol{\xi}}-\boldsymbol{\mu}^*\|_2\leq\frac{\bar{d}\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{q}$$

and if $\tilde{\Omega}=C_{k_0}$, we have that

$$\eta((r,\boldsymbol{a}),\tilde{\boldsymbol{\xi}},\boldsymbol{\mu}^*)\leq\bar{d}\cdot\|\tilde{\boldsymbol{\xi}}-\boldsymbol{\mu}^*\|_2=0$$

by noting $\tilde{\boldsymbol{\xi}}=\boldsymbol{\xi}_{k_0}=\boldsymbol{\mu}^*$. Now, for each Ω_{kl}, by Hoeffding's inequality, we have

$$P\left(\frac{1}{T-t}\cdot\sum_{\tau=t+1}^{T}\eta((\tilde{r}_\tau,\tilde{\boldsymbol{a}}_\tau),\boldsymbol{\xi}_{kl},\boldsymbol{\mu}^*)-\mathbb{E}_{(\tilde{r},\tilde{\boldsymbol{a}})}[\eta((\tilde{r},\tilde{\boldsymbol{a}}),\boldsymbol{\xi}_{kl},\boldsymbol{\mu}^*)]\geq\frac{\epsilon\cdot\|\tilde{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\|_2}{2}\right)\leq\exp\left(-\frac{\epsilon^2\cdot(T-t)q^2}{2\bar{d}^2}\right).$$

Our proof is completed from the union bound over all Ω_{kl}. □

Proof of Theorem 2. From (4), we have

$$
\begin{aligned}
\mathrm{Myopic}_t(\pi, \tilde{\boldsymbol{c}}_t^\pi) \leq & 2\bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \mathrm{Var}(M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1}))] + \mathbb{E}_{I_t}[G_{\tilde{\boldsymbol{c}}_t^\pi}(I_t)] \\
& + 2\bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \left(\hat{M}_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t} - \mathbb{E}_{I_{t+1}}[M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1})]\right)^2\right]
\end{aligned} \tag{122}
$$

Then, from Lemma 7 and Lemma 8, we know that

$$
\begin{aligned}
\mathrm{Myopic}_t(\pi, \tilde{\boldsymbol{c}}_t^\pi) \leq & \frac{2\bar{\alpha}\kappa_1 + \kappa_2}{T-t} + (2\bar{\alpha}\kappa_1 + \kappa_2) \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_1)^2]\right] \\
& + (2\bar{\alpha}\kappa_1 + \kappa_2) \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{1}_{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t} \cdot \mathbb{E}_{I_{t+1}}[(\tilde{\boldsymbol{a}}_t^\top \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{a}}_t^\top \hat{\boldsymbol{\mu}}_2)^2]\right] \\
& + 2\bar{\alpha} \cdot \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \left(\hat{M}_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t} - \mathbb{E}_{I_{t+1}}[M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1})]\right)^2\right]
\end{aligned} \tag{123}
$$

where $\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\mu}}_2$ are defined in (41) and $\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2$ are defined in (45), with $\boldsymbol{c} = \tilde{\boldsymbol{c}}_t^\pi$. Moreover, from Lemma 5 and the definition of $\hat{M}_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}$ in Algorithm 3, we know

$$
\begin{aligned}
& \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \left(\hat{M}_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t} - \mathbb{E}_{I_{t+1}}[M_{\tilde{\boldsymbol{c}}_t^\pi, \tilde{\boldsymbol{a}}_t}(I_{t+1})]\right)^2\right] \leq \mathbb{E}_{\tilde{\boldsymbol{a}}_t}\left[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot \left((\hat{\boldsymbol{\mu}}_1 - \mathbb{E}[\tilde{\boldsymbol{\mu}}_1])^2 + (\hat{\boldsymbol{\mu}}_1 - \mathbb{E}[\tilde{\boldsymbol{\mu}}_2])^2\right)\right] \\
& \leq \mathbb{E}[(\hat{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_1)^2] + 2 \cdot \mathbb{E}[(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^2] + 2 \cdot \mathbb{E}[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot (\hat{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_2)^2]
\end{aligned}
$$

We apply Claim 4 to bound the term $\mathbb{E}[(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^2]$ and we apply the "dual convergence" established in Lemma 9 to bound $\mathbb{E}[(\hat{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_1)^2]$ and $\mathbb{E}[\mathbb{1}_{\{\tilde{\boldsymbol{c}}_t^\pi \geq \tilde{\boldsymbol{a}}_t\}} \cdot (\hat{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_2)^2]$ at the order of $O(\frac{1}{T-t})$. Therefore, we prove our result. □