Multi-channel Autobidding with Budget and ROI Constraints

Yuan Deng [1] **Negin Golrezaei** [2] **Patrick Jaillet** [2] **Jason Cheuk Nam Liang** [2] **Vahab Mirrokni** [1]

Abstract

In digital online advertising, advertisers procure ad impressions simultaneously on multiple platforms, or so-called *channels*, such as Google Ads, Meta Ads Manager, etc., each of which consists of numerous ad auctions. We study how an advertiser maximizes total conversion (e.g. ad clicks) while satisfying aggregate *return-on-investment (ROI)* and budget constraints across all channels. In practice, an advertiser does not have control over, and thus cannot globally optimize, which individual ad auctions she participates in for each channel, and instead authorizes a channel to procure impressions on her behalf: the advertiser can only utilize two levers on each channel, namely setting a per-channel budget and per-channel target ROI. In this work, we first analyze the effectiveness of each of these levers for solving the advertiser's global multi-channel problem. We show that when an advertiser only optimizes over per-channel ROIs, her total conversion can be arbitrarily worse than what she could have obtained in the global problem. Further, we show that the advertiser can achieve the global optimal conversion when she only optimizes over per-channel budgets. In light of this finding, under a bandit feedback setting that mimics real-world scenarios where advertisers have limited information on ad auctions in each channels and how channels procure ads, we present an efficient learning algorithm that produces per-channel budgets whose resulting conversion approximates that of the global optimal problem.

1. Introduction

In today's online advertising world, advertisers (including but not limited to small businesses, marketing practitioners, non-profits, etc.) have been embracing an expanding array of advertising platforms such as search engines, social media platforms, web publisher display, etc., which present a plenitude of channels for advertisers to procure ad impressions and obtain traffic. In this growing multi-channel environment, the booming online advertising activities have fueled extensive research and technological advancements in *attribution analytics* to answer questions like which channels are more effective in targeting certain users? Or, which channels produce more user conversion (e.g. ad clicks) or *return-on-investment* (ROI) with the same amount of investments? (see (Kannan et al., 2016) for a comprehensive survey on attribution analytics). Yet, this area of research has largely left out a crucial phase in the workflow of advertisers' creation of a digital ad campaign, namely how advertisers interact with advertising channels, which is the physical starting point of a campaign.

To illustrate the significance of advertiser-channel interactions, consider for example a small business who is relatively well-informed through attribution research that Google Ads and Meta ads are the two most effective channels for its products. The business instantiates its ad campaigns through interacting with the platforms' ad management interfaces (see Figure 1), on which the business utilizes levers such as specifying budget and a target ROI[1] to control campaigns. Channels then inputs these specified parameters into their *autobidding* procedures, where they procure impressions on the advertiser's behalf to procure ads through automated blackbox algorithms. Eventually, channels report performance metrics such as expenditure and conversion back to the advertiser once the campaign ends. Therefore, one of the most important decisions advertisers need to make involves how to optimize over these levers provided by channels. Unfortunately, this has rarely been addressed in attribution analytics and relevant literature. Hence, this works contributes to filling this vacancy by addressing two themes:

> *How effective are these channel levers for advertisers to achieve their conversion goals? And how should advertisers optimize over such levers?*

To answer these questions, we study a setting where an advertiser simultaneously procures ads on multiple channels,

[1]Google Research [2]Massachusetts Institute of Technology. Correspondence to: Jason Cheuk Nam Liang <jcnliang@mit.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

[1]Target ROI is the numerical inverse of CPA or cost per action on Google Ads, and cost per result goal in Meta Ads.

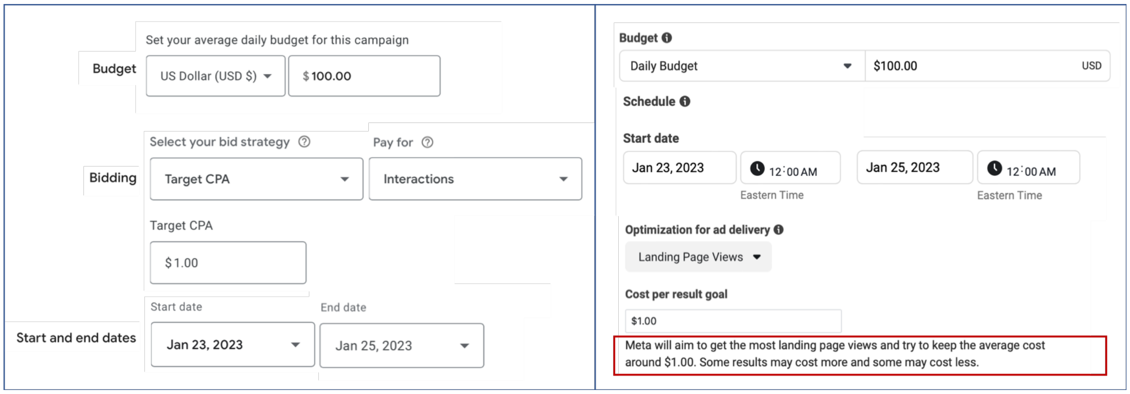

Figure 1. Interfaces on Google Ads (left) and Meta Ads Manager (right) for creating ad campaigns that allow advertisers to set per-channel budgets and ROIs. CPA, or cost per action on Google Ads, as well as cost per result goal on Meta Ads Manager, are effectively the inverse value for an advertiser's per-channel target ROI. Meta Ads Manager highlights its autobidding procedure maximizes total conversion while respecting advertisers' per-channel target ROI (see red box highlighted), supporting the GL-OPT and CH-OPT models in Eq. (1), (3).

each of which consists of multiple ad auctions that sell ad impressions. The advertiser's *global optimization problem* is to maximize total conversion over all channels, while respecting a global budget constraint that limits total spend, and a global ROI constraint that ensures total conversion is at least the target ROI times total spend. However, as channels operate as independent entities and conduct autobidding procurement on behalf of advertisers, there are no realistic means for an advertiser to implement the global optimization problem via optimizing over individual auctions. Instead, advertisers can only use two levers, namely a per-channel ROI and per-channel budget, to influence how channels should autobid for impressions. Our goal is to understand how effective are these levers by comparing the total conversion via optimizing levers versus the globally optimal conversion, and also present methodologies to help advertisers optimize over the usage of these levers. We summarize our contributions as followed:

Main contributions.

1. Modelling ad procurement through per-channel ROI and budget levers. In Section 2 we develop a novel model for online advertisers to optimize over the per-channel ROI and budget levers to maximize total conversion over channels while respecting a global ROI and budget constraint. This multi-channel optimization model closely imitates real-world practices (see Figure 1 for evidence), and to the best of our knowledge is the first of its kind to characterize advertisers' interactions with channels to run ad campaigns.

2. Solely optimizing per-channel budgets are sufficient to maximize conversion. In Theorem 3.2 of Section 3, we show that solely optimizing for per-channel ROIs is inadequate to optimize conversion across all channels, possibly resulting in arbitrarily worse total conversions compared to the global optimal where advertisers can optimize over individual auctions. In contrast, in Theorem 3.3 and Corollary 3.4 we show solely optimizing for per-channel budgets allows advertisers to achieve the global optimal.

3. Algorithm to optimize per-channel budget levers. Under a realistic bandit feedback structure where advertisers can only observe the total conversion and spend in each channel after making a per-channel budget decision, in Section 4 we develop an algorithm that augments stochastic gradient descent (SGD) with the upper-confidence bound (UCB) algorithm, and eventually outputs a per-channel budget profile with which advertisers can achieve $\mathcal{O}(T^{-1/3})$ approximation accuracy in total conversion compared to that of the optimal per-channel budget profile within T iterations. Our algorithm relates to constrained convex optimization with uncertain constraints and bandit feedback under a "one point estimation" regime, and to the best of our knowledge, our proposed algorithm is the first to handle such a setting; see discussions in Remark 4.9 of Section 4.

Related works. We review literature that relates to key themes of this work: autobidding, budget and ROI management, and constrained optimization with bandit feedback.

1. Autobidding. The autobidding model has been formally developed in (Aggarwal et al., 2019), and has been analyzed through the lens of welfare efficiency or price of anarchy in (Deng et al., 2021; Balseiro et al., 2021a; Deng et al., 2022b; Mehta, 2022), as well as individual advertiser fairness in (Deng et al., 2022a). The autobidding model has also been compared to classic quasi-linear utility models in (Balseiro et al., 2021b). The autobidding model considered in these papers assume advertisers can directly optimize over individual auctions, whereas in this work we address a more realistic setting that mimics practice where advertisers can only use levers provided by channels, and let channels procure ads on their behalf.

2. Budget and ROI management. Budget and ROI management strategies have been widely studied in the context of mechanism design and online learning. (Balseiro et al., 2017) studies the "system equilibria" of a range of budget management strategies in terms of the platforms' profits and advertisers' utility; (Balseiro & Gur, 2019; Balseiro et al., 2022) study online bidding algorithms (called pacing) that help advertisers achieve high utility in repeated second-price auctions while maintaining a budget constraint, whereas (Feng et al., 2022) studies similar algorithms but considers respecting a long term ROI constraint in addition to a fixed budget. All of these works on budget and ROI management focus on bidding strategies in a single repeated auction where advertisers' decisions are bid values submitted directly to the auctions. In contrast, this work studies advertisers making decisions on how to adjust per-channel ROI and budget levers while leaving the bidding to channels' blackbox algorithms.

3. Constrained optimization with bandit feedback. Lemma 4.5 of Section 4 shows the advertiser's optimization problem for per-channel budgets is a constrained convex problem with bandit feedback. Due to space limitations, here we only review two very recent works (Usmanova et al., 2019; Nguyen & Balasubramanian, 2022) that are most relevant to this paper, and refer readers to other related works in our extended literature review in Appendix A.1. Both works consider a similar convex setting to ours where the objective and constraints are only available through noisy function value evaluations. Although the two works achieve $\mathcal{O}(T^{-1})$ and $\mathcal{O}(T^{-1/2})$ respective approximation accuracy to the optimal solution, which contrasts our $\mathcal{O}(T^{-1/3})$ accuracy, these works impose several assumptions that are stronger than the ones that we consider. First, the objective and constraint functions are smooth (i.e. the gradients are Lipschitz continuous) and (Nguyen & Balasubramanian, 2022) further assume strong convexity. But in our work, our objectives and constraints are piece-wise linear and do not satisfy such salient properties. Second, and most importantly, both works consider a setting with "two point estimations" that allows the optimizer to access the objective and constraint function values twice in each iteration, enabling more efficient estimations. This work, however, lies in the one-point setting where we can only access function values once per iteration. Finally, we remark that the optimal accuracy/oracle complexity for the one-point setting for constrained (non-smooth) convex optimization with bandit feedback and unknown constraints remains an open question; see Remark 4.9 for more details.

See Appendix A.1 for an extended literature review.

2. Preliminaries

Advertisers' global optimization problem. Consider an advertiser running a digital ad campaign to procure ad impressions on $M \in \mathbb{N}$ platforms such as Google Ads, Meta Ads Manager etc., each of which we call a *channel*. Each channel j consists of $m_j \in \mathbb{N}$ parallel ad auctions, each of which corresponds to the sale of an ad impression.[2] An ad auction $n \in [m_j]$ is associated with a value $v_{j,n} \geq 0$ that represents the expected conversion (e.g. number of clicks) of the impression on sale, and a cost $d_{j,n} \geq 0$ that is required for the purchase of the impression. For example, the cost in a single slot second-price auction is the highest competing bid from competitors in the market, and in a posted price auction the cost is simply the posted price by the seller of the impression. Writing $\boldsymbol{v}_j = (v_{j,n})_{n\in[m_j]}$ and $\boldsymbol{d}_j = (d_{j,n})_{n\in[m_j]}$, we assume that $\boldsymbol{z}_j := (\boldsymbol{v}_j, \boldsymbol{d}_j)$ is sampled from some unknown discrete distribution $\boldsymbol{p}_j$ supported on unknown finite set $F_j \subseteq \mathbb{R}_+^{m_j} \times \mathbb{R}_+^{m_j}$.

The advertiser's goal is to maximize total conversion of procured ad impressions, while subject to a *return-on-investment (ROI)* constraint that states total conversion across all channels is no less than γ times total spend for some pre-specified target ROI $0 < \gamma < \infty$, as well as a budget constraint that states total spend over all channels is no greater than the total budget $\rho \geq 0$. Mathematically, the advertiser's *global optimization problem* is:

$$\begin{aligned} \text{GL-OPT} = &\max_{\boldsymbol{x}_j\in[0,1]^{m_j},\forall j\in[M]} \sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{v}_j^\top \boldsymbol{x}_j\right] \\ &\text{s.t. } \textstyle\sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{v}_j^\top \boldsymbol{x}_j - \gamma \boldsymbol{d}_j^\top \boldsymbol{x}_j\right] \geq 0 \\ &\quad\;\; \textstyle\sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{d}_j^\top \boldsymbol{x}_j\right] \;\leq\; \rho\,. \end{aligned} \tag{1}$$

Here, the decision variable $\boldsymbol{x}_j \in [0,1]^{m_j}$ is a vector where $x_{j,n}$ denotes whether the impression in auction n for channel j is procured. We remark that $\boldsymbol{x}$ depends on the realization of $\boldsymbol{z} = (\boldsymbol{v}_j, \boldsymbol{d}_j)_{j\in[M]}$ and is also random. We note that the ROI and budget constraints are taken in expectation because an advertiser procures impressions from a very large number of auctions in total, and thus the advertiser typically only requires to satisfy constraints in an average sense. GL-OPT is a widely adopted formulation for autobidding practices in modern online advertising; see e.g. (Aggarwal et al., 2019; Balseiro et al., 2021a; Deng et al., 2021; 2022b). In Section 5 we discuss more general advertiser objectives.

Our overarching goal is to develop methodologies that enable an advertiser to achieve total campaign conversion that match GL-OPT. However, directly optimizing GL-OPT may not be plausible as we discuss in the following.

Advertisers' levers to solve their global problems. To

[2] Ad auctions for each channel may be run by the channel itself or other external ad inventory suppliers such as web publishers.

solve the global optimization problem GL-OPT, ideally advertisers would like to optimize over individual auctions across all channels. However, in reality channels operate as independent entities, and typically do not provide means for general advertisers to participate in specific individual auctions at their discretion. Instead, channels provide advertisers with specific *levers* to express their ad campaign goals on spend and conversion. In this work, we focus on two of the most widely used levers, namely the per-channel ROI target and per-channel budget (see illustration in Fig. 1). After an advertiser inputs these parameters to a channel, the channel then procures ads on behalf of the advertiser through autonomous programs (we call this programmatic process *autobidding*) to help advertiser achieve procurement results that match with the inputs. We elaborate on this later.

Formally, we consider the setting where for each channel j, an advertiser is allowed to input a per-channel target ROI $0 \leq \gamma_j < \infty$, and a per-channel budget $\rho_j \in [0, \rho]$ where we recall $\rho > 0$ is the total advertiser budget for a certain campaign. Then, the channel uses these inputs in its autobidding algorithm to procure ads, and returns the total conversion $V_j(\gamma_j, \rho_j; \boldsymbol{z}_j) \geq 0$, as well as total spend $D_j(\gamma_j, \rho_j; \boldsymbol{z}_j) \geq 0$ to the advertiser, where we recall $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j) \in F_j$ is the realized vector of value-cost pairs in channel j; V_j and D_j will be further specified later.

As the advertiser has the freedom of choice to input either per-channel target ROI's, budgets, or both, we consider three options: 1. input only a per-channel target ROI; 2. input only a per-channel budget; 3. input both per-channel target ROI and budgets. Such options correspond to the following decision sets for $(\gamma_j, \rho_j)_{j\in[M]}$:

$$
\begin{aligned}
&\textbf{Per-channel budget only option: } \mathcal{I}_B = \\
&\{(\gamma_j, \rho_j)_{j\in[M]} \in \mathbb{R}_+^{2\times M} : \gamma_j = 0, \rho_j \in [0, \rho] \text{ for } \forall j\}. \\
&\textbf{Per-channel target ROI only option: } \mathcal{I}_R = \\
&\{(\gamma_j, \rho_j)_{j\in[M]} \in \mathbb{R}_+^{2\times M} : \gamma_j \geq 0, \rho_j = \infty \text{ for } \forall j\}. \\
&\textbf{General option: } \mathcal{I}_G = \\
&\{(\gamma_j, \rho_j)_{j\in[M]} : \gamma_j \geq 0, \rho_j \in [0, \rho] \text{ for } \forall j\}.
\end{aligned}
\tag{2}
$$

The advertiser's goal in practice is to maximize their total conversion of procured ad impressions through optimizing over per-channel budgets and target ROIs, while subject to the global ROI and budget constraint similar to those in GL-OPT. Mathematically, for any option $\mathcal{I} \in \{\mathcal{I}_B, \mathcal{I}_R, \mathcal{I}_G\}$, the advertiser's optimization problem can be written as

$$
\begin{aligned}
\text{CH-OPT}(\mathcal{I}) = &\max_{(\gamma_j,\rho_j)_j\in\mathcal{I}} \sum_{j\in M} \mathbb{E}\left[V_j(\gamma_j, \rho_j; \boldsymbol{z}_j)\right] \\
\text{s.t. } &\textstyle\sum_{j\in M} \mathbb{E}\left[V_j(\gamma_j, \rho_j; \boldsymbol{z}_j) - \gamma D_j(\gamma_j, \rho_j; \boldsymbol{z}_j)\right] \geq 0 \\
&\textstyle\sum_{j\in[M]} \mathbb{E}\left[D_j(\gamma_j, \rho_j; \boldsymbol{z}_j)\right] \leq \rho \, ,
\end{aligned}
\tag{3}
$$

where the expectation is taken w.r.t. randomness in $\boldsymbol{z}_j$. We remark that for any channel $j \in [M]$, the number of auctions m_j as well as the distribution $\boldsymbol{p}_j$ are fixed and not a function of the input parameters γ_j, ρ_j.

The functions (V_j, D_j) that map per-channel target ROI and budgets γ_j, ρ_j to the total conversion and expenditure are specified by various factors including but not limited to channel j's autobidding algorithms deployed to procure ads on advertisers' behalf, or the auction mechanisms that sell impressions. In this work, we study a general setup that closely mimics industry practices: we assume that on the behalf of the advertiser, each channel aims to optimize their conversion over all m_j auctions while respecting the advertiser's input (i.e., per-channel target ROI and budgets). (See e.g. Meta Ads Manager in Figure 1 specifically highlights the channel's autobidding procurement methodology which supports this setup). Hence, each channel j's optimization problem can be written as

$$
\begin{aligned}
\boldsymbol{x}_j^*(\gamma_j, \rho_j; \boldsymbol{z}_j) \;=\; &\arg\max_{\boldsymbol{x}\in[0,1]^{m_j}} \; \boldsymbol{v}_j^\top \boldsymbol{x} \\
&\text{s.t.} \quad \boldsymbol{v}_j^\top \boldsymbol{x} \geq \gamma_j \boldsymbol{d}_j^\top \boldsymbol{x}, \quad \boldsymbol{d}_j^\top \boldsymbol{x} \leq \rho_j \, ,
\end{aligned}
\tag{4}
$$

where $\boldsymbol{x} = (x_n)_{n\in[M_j]} \in [0,1]^{m_j}$ denotes the vector of probabilities to win each of the parallel auctions, i.e. $x_n \in [0,1]$ is the probability to win auction $n \in [m_j]$ in channel j. In light of this representation, the corresponding conversion and spend functions are given by

$$
\begin{aligned}
&V_j(\gamma_j, \rho_j; \boldsymbol{z}_j) = \boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\gamma_j, \rho_j; \boldsymbol{z}_j) \\
&V_j(\gamma_j, \rho_j) = \mathbb{E}[V_j(\gamma_j, \rho_j; \boldsymbol{z}_j)] \\
&D_j(\gamma_j, \rho_j; \boldsymbol{z}_j) = \boldsymbol{d}_j^\top \boldsymbol{x}_j^*(\gamma_j, \rho_j; \boldsymbol{z}_j) \\
&D_j(\gamma_j, \rho_j) = \mathbb{E}[D_j(\gamma_j, \rho_j; \boldsymbol{z}_j)] \, .
\end{aligned}
\tag{5}
$$

Here, the expectation is taken w.r.t. randomness in $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j)$. We assume that for any (γ_j, ρ_j) and realization $\boldsymbol{z}_j$, $V_j(\gamma_j, \rho_j; \boldsymbol{z}_j)$ is bounded above by some absolute constant $\overline{V} \in (0, \infty)$ almost surely. We remark that Eq.(5) assumes channels are able to achieve optimal procurement performance. Later in Section 5, we briefly discuss setups where channels does not optimally solve for Eq.(4).

Key questions and organization of this paper.

1. (Section 3) How effective are the per-channel ROI and budget levers to help advertisers achieve the globally optimal conversion GL-OPT while respecting the global ROI and budget constraints? In particular, for each option $\mathcal{I} \in \{\mathcal{I}_B, \mathcal{I}_R, \mathcal{I}_G\}$ defined in Eq. (2), what is the discrepancy between CH-OPT($\mathcal{I}$) versus the optimal GL-OPT?
2. (Section 4) How can advertisers optimize per-channel target ROIs and budgets to solve for CH-OPT($\mathcal{I}$)?

3. On the Efficacy of Per-channel Levers

In this section, we examine the effectiveness of the per-channel target ROI and per-channel budget levers in achieving the global optimal GL-OPT. In particular, we study if the optimal solution to the channel problem CH-OPT($\mathcal{I}$) defined in Eq. (3) for $\mathcal{I} \in \{\mathcal{I}_B, \mathcal{I}_R, \mathcal{I}_G\}$ is equal to the global optimal GL-OPT.

Our first result in this section is the following Lemma 3.1 which shows that GL-OPT serves as a theoretical upper bound for an advertiser's conversion through optimizing CH-OPT($\mathcal{I}$) with any option $\mathcal{I}$.

Lemma 3.1 (GL-OPT is the theoretical upper bound for conversion)**.** *For any option $\mathcal{I} \in \{\mathcal{I}_B, \mathcal{I}_R, \mathcal{I}_G\}$, we have* GL-OPT $\geq$ CH-OPT($\mathcal{I}$).

The proof of Lemma 3.1 is deferred to Appendix B.1. In light of the theoretical upper bound GL-OPT, we are now interested in the gap between GL-OPT and CH-OPT($\mathcal{I}$) for option $\mathcal{I} \in \{\mathcal{I}_B, \mathcal{I}_R, \mathcal{I}_G\}$. In the following Theorem 3.2, we show that there exists a problem instance under which the ratio $\frac{\text{CH-OPT}(\mathcal{I}_R)}{\text{GL-OPT}}$ nears 0, implying the per-channel ROIs alone fail to help advertisers optimize conversion.

Theorem 3.2 (Per-channel ROI only option fails to optimize conversion)**.** *Consider an advertiser with a (global) target ROI of $\gamma = 1$ procuring impressions from $M = 2$ channels with 1 and 2 auctions, respectively. The advertiser has unlimited budget $\rho = \infty$, and chooses the per-channel target ROI only option $\mathcal{I}_R$ defined in Eq. (2). Assume there is only one realization of value-cost pairs $\boldsymbol{z} = (\boldsymbol{v}_j, \boldsymbol{d}_j)_{j\in[M]}$ (i.e. the support $F = F_1 \times F_2$ is a singleton), and the realization is presented in the following table, where $X > 0$ is some arbitrary parameter. Then, $\lim_{X\to\infty} \frac{\text{CH-OPT}(\mathcal{I}_R)}{\text{GL-OPT}} = 0$.*

	Channel 1	*Channel 2*	
	Auction 1	*Auction 2*	*Auction 3*
Value $v_{j,n}$	*1*	X	$2X$
Spend $d_{j,n}$	*0*	$1 + X$	$2(1 + X)$

See proof in Appendix B.2. In contrast, the budget only option $\mathcal{I}_B$ in fact allows an advertiser's conversion to reach the theoretical upper bound GL-OPT through solely optimizing for per-channel budgets. This is formalized in the following theorem (see proof in Appendix B.3).

Theorem 3.3 (Per-channel budget suffices to achieve optimal conversion)**.** *For the budget only option $\mathcal{I}_B$ defined in Eq.(2), we have* GL-OPT = CH-OPT($\mathcal{I}_B$) *for any global target ROI $\gamma > 0$ and total budget $\rho > 0$, even for $\rho = \infty$.*

As an immediate extension of Theorem 3.3, the following Corollary 3.4 states per-channel ROI's in fact become redundant once advertisers optimize for per-channel budgets.

Corollary 3.4 (Redundancy of per-channel ROIs)**.** *For the general option $\mathcal{I}_G$ defined in Eq.(2) where an advertiser sets both per-channel ROI and budgets, we have* GL-OPT = CH-OPT($\mathcal{I}_G$) *for any aggregate ROI $\gamma > 0$ and total budget $\rho > 0$, even for $\rho = \infty$. Further, there exists an optimal solution $(\gamma_j, \rho_j)_{j\in[M]}$ to* CH-OPT($\mathcal{I}_G$)*, s.t. $\gamma_j = 0$ for all $j \in [M]$.*

In light of the redundancy of per-channel ROIs as illustrated in Corollary 3.4, in the rest of the paper we will fix $\gamma_j = 0$ for any channel $j \in [M]$, and omit γ_j in all relevant notations; e.g. we will write $D_j(\rho_j; \boldsymbol{z}_j)$ and $D_j(\rho_j)$, instead of $D_j(\gamma_j, \rho_j; \boldsymbol{z}_j)$ and $D_j(\gamma_j, \rho_j)$. Equivalently, we will only consider the per-channel budget only option $\mathcal{I}_B$.

4. Optimization Algorithm for Per-channel Budgets under Bandit Feedback

In this section, we develop an efficient algorithm to solve for per-channel budgets that optimize CH-OPT($\mathcal{I}_B$) defined in Eq. (3), which achieves the theoretical optimal conversion, namely GL-OPT, as illustrated in Theorem 3.3. In particular, we consider algorithms that run over $T > 0$ periods, where each period for example corresponds to the duration of 1 hour or 1 day. At the end of T periods, the algorithm produces some per-channel budget profile $(\rho_j)_{j\in[M]} \in [0, \rho]^M$ that approximates CH-OPT($\mathcal{I}_B$), and satisfies aggregate ROI and budget constraints, namely

$$\textstyle\sum_{j\in M} V_j(\rho_j) \geq \gamma \sum_{j\in M} D_j(\rho_j),\ \ \sum_{j\in[M]} D_j(\rho_j) \leq \rho\,,$$

where we recall $(V_j(\rho_j), D_j(\rho_j))$ are defined in Eq. (5).

The algorithm proceeds as follows: at the beginning of period $t \in [T]$, the advertiser sets per-channel budgets $(\rho_{j,t})_{j\in[M]}$, while simultaneously values and costs $\boldsymbol{z}_t = (\boldsymbol{z}_{j,t})_{j\in[M]} = (\boldsymbol{v}_{j,t}, \boldsymbol{d}_{j,t})_{j\in[M]}$, where $(\boldsymbol{v}_{j,t}, \boldsymbol{d}_{j,t}) \in \mathbb{R}_+^{m_j} \times \mathbb{R}_+^{m_j}$ are sampled (independently in each period) from finite support $F = F_1 \times \ldots F_M$ according to discrete distributions $(\boldsymbol{p}_j)_{j\in[M]}$. Each channel j then takes as input $\rho_{j,t} \in [0, \rho]$ and procures ads on behalf of the advertiser, and reports the total realized conversion $V_j(\rho_{j,t}; \boldsymbol{z}_{j,t})$ as well as total spend $D_j(\rho_{j,t}; \boldsymbol{z}_{j,t})$ to the advertiser (see definitions in Eq. (5)). For simplicity we assume any realization $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j) \in F_j$ admits an ordering $\frac{v_{j,1}}{d_{j,1}} > \cdots > \frac{v_{j,m_j}}{d_{j,m_j}}$ for all channels $j \in [M]$.

Bandit feedback: We highlight that the advertiser receives *bandit feedback* from channels, i.e. the advertiser only observes the numerical values $V_j(\rho_{j,t}; \boldsymbol{z}_{j,t})$ and $D_j(\rho_{j,t}; \boldsymbol{z}_{j,t})$, but does not get to observe $V_j(\rho'_j; \boldsymbol{z}'_j)$ and $D_j(\rho'_j; \boldsymbol{z}'_j)$ evaluated at any other per-channel budget $\rho'_j \neq \rho_{j,t}$ and realized value-cost pairs $\boldsymbol{z}'_j \neq \boldsymbol{z}_{j,t}$.

We also make two mild assumptions: In Assumption 4.1, we assume that any channel will deplete input per-channel budgets. This is a natural assumption that mimics practical scenarios, e.g. marketing for small businesses that have moderate-sized budgets. In Assumption 4.2, we assume for any realization of value-cost pairs $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j)$ in a channel $j \in [M]$, there always exists an auction $n \in [m_j]$ in this channel whose value-to-cost ratio is at least γ, i.e. $v_{j,n} > \gamma d_{j,n}$.

Assumption 4.1 (Moderate budgets)**.** Assume $\rho < \infty$, and for any channel $j \in [M]$, value-cost realization $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j) \in F_j$, and per-channel budget $\rho_j \in [0, \rho]$, the optimal solution $\boldsymbol{x}_j^*(\rho_j; \boldsymbol{z}_j)$ defined in Eq. (4) is budget binding, i.e. $D_j(\rho_j; \boldsymbol{z}_j) = \boldsymbol{d}_j^\top \boldsymbol{x}_j^*(\rho_j; \boldsymbol{z}_j) = \rho_j$.

Assumption 4.2 (Strictly feasible global ROI constraints)**.** Fix any channel $j \in [M]$ and any realization of value-cost pairs $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j) \in F_j$. Then, the channel's optimization problem in Eq. (4) is strictly feasible, i.e. the set $\left\{\boldsymbol{x}_j \in [0,1]^{m_j} : \boldsymbol{v}_j^\top \boldsymbol{x}_j > \gamma \boldsymbol{d}_j^\top \boldsymbol{x}_j\right\}$ is nonempty.

4.1. Optimize per-channel budgets with SGD-UCB

Here, we describe our algorithm to solve for optimal per-channel budgets w.r.t. CH-OPT$(\mathcal{I}_B)$. Similar to most algorithms for constrained optimization, we take a dual stochastic gradient descent (SGD) approach; see a comprehensive survey on dual descent methods in (Bertsekas, 2014). First, we consider the Lagrangian functions w.r.t. CH-OPT$(\mathcal{I}_B)$ where we let $\boldsymbol{c} = (\lambda, \mu) \in \mathbb{R}_+^2$ be the dual variables corresponding to the ROI and budget constraints, respectively:

$$\begin{aligned} &\mathcal{L}_j(\rho_j, \boldsymbol{c}; \boldsymbol{z}_j) = (1+\lambda) V_j(\rho_j; \boldsymbol{z}_j) - (\lambda\gamma + \mu)\rho_j \\ &\mathcal{L}_j(\rho_j, \boldsymbol{c}) = \mathbb{E}\left[\mathcal{L}_j(\rho_j, \boldsymbol{c}; \boldsymbol{z}_j)\right] . \end{aligned} \tag{6}$$

Then, in each period $t \in [T]$ given dual variables $\boldsymbol{c}_t = (\lambda_t, \gamma_t)$, SGD decides on a primal decision, i.e. per-channel budget $(\rho_{j,t})_{j\in[M]}$ by optimizing the following

$$\rho_{j,t} \;=\; \arg\max_{\rho_j \in [0,\rho]} \mathcal{L}_j(\rho_j, \boldsymbol{c}_t; \boldsymbol{z}_{j,t}) . \tag{7}$$

Having observed the realized values $(V_j(\rho_{j,t}; \boldsymbol{z}_{j,t}))_{j\in[M]}$ (note that spend is $(\rho_{j,t})_{j\in[M]}$ under Assumption 4.1), we calculate the current period violation in budget and ROI constraints, namely $g_{1,t} := \sum_{j\in M} \left(V_j(\rho_{j,t}; \boldsymbol{z}_{j,t}) - \gamma\rho_{j,t}\right)$ and $g_{2,t} = \rho - \sum_{j\in[M]} \rho_{j,t}$. Next, we update dual variables via $\Pi_{[0,C_F]}\left(\lambda_t - \eta g_{1,t}\right)$ and $\mu_{t+1} = \Pi_{[0,C_F]}\left(\mu_t - \eta g_{2,t}\right)$, where Π is the projection operator, η is some pre-specified step size, and C_F is some dual variable upper bound specified in Eq. (9).[3]

However, we cannot realistically find the primal decisions by solving Eq. (7) since the function $\mathcal{L}_j(\cdot, \boldsymbol{c}_t; \boldsymbol{z}_{j,t})$ is unknown due to the bandit feedback structure. Therefore, we provide a modification to SGD to handle this issue. We briefly note that although bandit feedback prevents naively applying SGD to our problem, this may not be the case in other online advertising scenarios that involve relevant learning tasks, underlining the challenges of our problem; see Remark A.1 in Appendix A.2 for comparisons with related works.

[3]One can also employ more general mirror descent dual variable updates; see e.g. (Balseiro et al., 2022).

To handle bandit feedback, we take a natural approach to augment SGD with the celebrated *upper-confidence bound (UCB)* algorithm; see intro to UCB and multi-arm bandits in (Slivkins et al., 2019). In particular, we first discretize our per-channel budget decision set $[0, \rho]$ into granular "arms" separated by distance $\delta > 0$:

$$\mathcal{A}(\delta) = \{a_k\}_{k\in[K]} \text{ where } a_k = (k-1)\delta . \tag{8}$$

for $K := \lceil \rho/\delta \rceil + 1$. In the following we will use the terms "per-channel budget" and "arm" interchangeably. In the spirit of UCB, in each period t we maintain some estimate $\left(\overline{V}_j(a_k)\right)_{j\in[M]}$ of the conversions $\left(V_j(a_k)\right)_{j\in[M]}$ as well as an upper confidence bound $\texttt{UCB}_{j,t}(a_k)$ for each arm a_k using historical payoffs from periods in which arm a_k is pulled. Finally, we update primal decisions for each channel $j \in [M]$ using the "best arm" $\rho_{j,t} = \arg\max_{a_k \in \mathcal{A}(\delta)} \;(1 + \lambda_t)\left(\overline{V}_{j,t}(a_k) + \texttt{UCB}_{j,t}(a_k)\right) - (\lambda_t\gamma + \mu_t)\, a_k$.

Finally, to ensure aggregate ROI and budget constraint satisfaction, we maintain variables that check ROI and budget balances, namely $S_{1,t}$ and $S_{2,t}$, to record the cumulative ROI and spend across all channels up until period t. When the ROI balance check $S_{1,t}$ is too negative, or the budget balance check is too large, we "stop" the algorithm, and naively set some pre-defined small per-channel budget $\underline{\rho} \in (0, \rho)$ (later chosen in Theorem 4.8) during all periods after the "stopping time" denoted as τ_A. We remark that similar approaches to ensure constraint satisfaction has been introduced in e.g. (Balseiro et al., 2022; Feng et al., 2022).

We summarize our algorithm, called SGD-UCB, in Algorithm 1.[4]

4.2. Analyzing the SGD-UCB algorithm

In this subsection, we analyze the performance of SGD-UCB in Algorithm 1, and present accuracy guarantees on the final output $\overline{\boldsymbol{\rho}}_T = \left(\frac{1}{T}\sum_{t\in[T]} \rho_{j,t}\right)_{j\in[M]}$. The backbone of our analysis strategy is to show the cumulative loss over T periods, namely $T \cdot$ GL-OPT $- \mathbb{E}\left[\sum_{t\in[T]} \sum_{j\in[M]} V_j(\rho_{j,t})\right]$ consists of three main parts, namely the "stopping error" due to some condition for the while loop being violated and naively setting a small per-channel budget $\underline{\rho}$ after the "stopping time" τ_A (see step 10); the error induced by UCB in our algorithm; and the error due to SGD (or what is typically viewed as the deviations from complementary slackness); see following Proposition 4.3. Then we further bound each part.

[4]There has been very recent works that combine SGD with adversarial bandit type algorithms such as EXP3 (Castiglioni et al., 2022), or with Thompson sampling which is another well-known algorithm for stochastic bandit problems (e.g. (Ding et al., 2021)), and works that employ SGD in bandit problems (e.g. (Han et al., 2021)). Yet to the best of our knowledge, our approach to integrate SGD with UCB is novel.

Algorithm 1 SGD-UCB

1: **Input:** Budget discretization set of arms $\mathcal{A}(\delta)$ defined in Eq.(8). Step size $\eta > 0$. Initialize $N_{j,1}(a_k) = \overline{V}_{j,1}(a_k) = 0$ for all $j \in [M]$ and $k \in [K]$, and dual variables $\lambda_1 = \mu_1 = 0$. Set $\underline{\rho} \in (0, \rho/M)$, $\beta > 0$ and dual variable upper bound

$$C_F = M\overline{V} \max\left\{\frac{1}{\beta\underline{\rho}}, \frac{1}{\rho - M\underline{\rho}}\right\} \tag{9}$$

where $\overline{V} \geq \max_{j\in[M]} \max_{\rho_j\in[0,\rho]} \max_{\boldsymbol{z}_j\in F_j} V_j(\rho_j, \boldsymbol{z}_j)$ is the conversion upper bound.

2: Set initial constraint balance checks: $S_{1,t} = S_{2,t} = 0$ for $t = 1$, and start period counter $t = 1$.

3: **while** $t \leq T$ and $S_{1,t} - \gamma M\rho + \beta\underline{\rho}(T - t) \geq 0$ and $S_{2,t} + M\rho + M\underline{\rho}(T - t) \leq \rho T$ **do**

4: **Set per-channel budget.** For each channel $j \in [M]$: If $t \leq K$, set $\rho_{j,t} = a_t$. Else if $t > K$, set $\rho_{j,t} =$

$$\arg\max_{a_k\in\mathcal{A}(\delta)} \overline{V}_{j,t}(a_k) + \text{UCB}_{j,t}(a_k) - \frac{(\lambda_t\gamma + \mu_t)a_k}{1+\lambda_t},$$

where $\text{UCB}_{j,t}(a_k) = \sqrt{\frac{2\log(T)}{N_{j,t}(a_k)}}$.

5: Observe realized conversion $\{V_j(\rho_{j,t}; \boldsymbol{z}_{j,t})\}_{j\in[M]}$, and update for each arm $k \in [K]$ and channel $j \in [M]$

$$N_{j,t+1}(a_k) = N_{j,t}(a_k) + \mathbb{I}\{\rho_{j,t} = a_k\}$$
$$\overline{V}_{j,t+1}(a_k) = \frac{N_{j,t}(a_k)\overline{V}_{j,t}(a_k) + V_j(\rho_{j,t}; \boldsymbol{z}_{j,t})\mathbb{I}\{\rho_{j,t} = a_k\}}{N_{j,t+1}(a_k)}$$

6: **Update dual variables.** Calculate $g_{1,t} = \sum_{j\in M}(V_j(\rho_{j,t}; \boldsymbol{z}_{j,t}) - \gamma\rho_{j,t})$ and $g_{2,t} = \rho - \sum_{j\in[M]}\rho_{j,t}$. Then, set

$$\begin{aligned} \lambda_{t+1} &= \Pi_{[0,C_F]}\left(\lambda_t - \eta g_{1,t}\right), \\ \mu_{t+1} &= \Pi_{[0,C_F]}\left(\mu_t - \eta g_{2,t}\right). \end{aligned} \tag{10}$$

7: Update balance check: $S_{1,t+1} = S_{1,t} + g_{1,t}$ and $S_{2,t+1} = S_{2,t} + \sum_{j\in[M]}\rho_{j,t}$.

8: Increment period counter $t \leftarrow t + 1$.

9: **end while**

10: Record $\tau_A = t - 1$ and for all $t = \tau_A + 1 \ldots T$ set $\rho_{j,t} = \underline{\rho}$ for all $j \in [M]$.

11: Output $\overline{\boldsymbol{\rho}}_T = \left(\frac{1}{T}\sum_{t\in[T]}\rho_{j,t}\right)_{j\in[M]}$.

Proposition 4.3 (Regret decomposition)**.** *For any channel $j \in [M]$ define $\rho_j^*(t) = \arg\max_{\rho_{j\in[M]}} \mathcal{L}_j(\rho_j; \boldsymbol{c}_t)$ to be the optimal per-channel budget w.r.t. dual variables $\boldsymbol{c}_t = (\lambda_t, \mu_t)_{t\in[T]}$. Then $T \cdot$ GL-OPT $- \sum_{t\in[T]}\sum_{j\in[M]} V_j(\rho_{j,t})$ is bounded by*

$$\underbrace{M\overline{V}(T - \tau_A)}_{\text{Stopping error}} + \underbrace{\sum_{t\in[\tau_A]}(\lambda_t g_{1,t} + \mu_t g_{2,t})}_{\text{SGD complementary slackness deviations}}$$
$$+ \sum_{j\in[M]}\underbrace{\sum_{t\in[\tau_A]}\mathcal{L}_j(\rho_j^*(t), \lambda_t, \mu_t) - \mathcal{L}_j(\rho_{j,t}, \lambda_t, \mu_t)}_{\text{UCB error}}.$$

where $\tau_A \in [T]$ is defined in step 10 of Algorithm 1.

Recall the definitions of $g_{1,t}$ and $g_{2,t}$ in step 5 of Algorithm 1, and the fact that the conversion $V_j(\rho_j; \boldsymbol{z}_j)$ is bounded above by absolute constant $\overline{V} \in (0, \infty)$ almost surely.

We bound the stopping error together with SGD complementary slackness violations in the following Lemma 4.4, which follows standard analyses for SGD; see proof in Appendix C.2.

Lemma 4.4 (Bounding stopping error and complementary slackness deviations)**.** *Assume Assumptions 4.1 and 4.2 hold. Recall $\eta > 0$ is the step size. Then we have $M\overline{V}(T - \tau_A) + \sum_{t\in[\tau_A]}(\lambda_t g_{1,t} + \mu_t g_{2,t}) \leq \mathcal{O}\left(\eta T + \frac{1}{\eta}\right)$.*

Challenges in bounding UCB error due to adversarial contexts and continuum-arm dicretization. Bounding our UCB error is much more challenging than doing so in classic stochastic multi-arm bandit settings: first, our setup involves discretizing a continuum of arms i.e. our discretization in Eq.(8) for $[0, \rho]$; second, and more importantly, the dual variables $\{\boldsymbol{c}_t\}_{t\in[T]}$ are effectively *adversarial contexts* since they are updated via SGD instead of being stochastically sampled from some nice distribution, and correspondingly the Lagrangian function $\mathcal{L}_j(a_k, \boldsymbol{c}_t; \boldsymbol{z}_t)$ can be viewed as a reward function that maps any arm-context pair $(a_k, \boldsymbol{c}_t)$ to (stochastic) payoffs. Both continuum-arms and adversarial contexts have been notorious in making reward function estimations highly inefficient; see e.g. discussions in (Agrawal, 1995; Agarwal et al., 2014). We further elaborate on specific challenges that adversarial contexts bring about:

1. Boundedness of rewards. In classic stochastic multi-arm bandtis and UCB, losses in total rewards grow linearly with the magnitude of rewards. In our setting, the reward function, i.e. the Lagrangian function $\mathcal{L}_j(a_k, \boldsymbol{c}_t; \boldsymbol{z}_t)$, scales linearly with the magnitude of contexts (see Eq. (6), so large contexts (i.e. large dual variables) may lead to large losses.

2. Context-dependent exploration-exploitation tradeoffs. The typical trade-off for arm exploration and exploitation in our setting depends on the particular values of the contexts (i.e. the dual variables), which means there may exist "bad" contexts that lead to poor tradeoffs that require significantly more explorations to achieve accurate estimates of arm rewards than other "good" contexts. We elaborate more in Lemma 4.7 and discussions thereof.

We first handle continuum arm discretization via showing the specific form of conversion functions $V(\rho_j; \boldsymbol{z})$ in Eq. (4) induces salient structures for the Lagrangian function, namely it is continuous, piecewise linear, concave, and unimodal[5]; we present the proof in Appendix C.3

Lemma 4.5 (Structural properties)**.** • *For any channel $j \in [M]$ and per-channel budget ρ_j the conversion function $V_j(\rho_j)$ is continuous, piece-wise linear, strictly increasing,*

[5] A function $f : \mathbb{R} \to \mathbb{R}$ is unimodal if $\exists y^*$ such that $f(y)$ strictly increases when $y \leq y^*$ and strictly decreases when $y \geq y^*$.

and concave. In particular, $V_j(\rho_j)$ *takes the form*

$$V_j(\rho_j) = \sum_{n\in[S_j]} (s_{j,n}\rho_j + b_{j,n})\,\mathbb{I}\{\rho_j \in [r_{j,n-1}, r_{j,n}]\}\,,$$

where $S_j \in \mathbb{N}$ *and* $\{(s_{j,n}, b_{j,n}, r_{j,n})\}_{n\in[S_j]}$ *only depend on the support* F_j *and distribution* $\boldsymbol{p}_j$ *from which value and costs are sampled. These parameters satisfy* $s_{j,1} > \cdots > s_{j,S_j} > 0$ *and* $0 = r_{j,0} < r_{j,1} < \cdots < r_{j,S_j} = \rho$*, as well as* $b_{j,n} \geq 0$ *s.t.* $s_{j,n}r_{j,n} + b_{j,n} = s_{j,n+1}r_{j,n} + b_{j,n+1}$ *for all* $n \in [S_j - 1]$*, with* $b_{j,1} = 0$*. This implies* $V_j(\rho_j)$ *is continous in* ρ_j*.*

- *For any dual variables* $\boldsymbol{c} = (\lambda, \mu) \in \mathbb{R}_+^2$, $\mathcal{L}_j(\rho_j, \boldsymbol{c})$ *defined in Eq.* (6) *is continuous, piece-wise linear, concave, and unimodal in* ρ_j*. In particular,*

$$\mathcal{L}_j(\rho_j, \boldsymbol{c}) = \sum_{n=1}^{S_j} \left(\sigma_{j,n}(\boldsymbol{c})\rho_j + b'_{j,n}\right)\mathbb{I}\{\rho_j \in [r_{j,n-1}, r_{j,n}]\}$$

where the slope $\sigma_{j,n}(\boldsymbol{c}) = (1+\lambda)s_{j,n} - (\mu + \gamma\lambda)$ *and* $b'_{j,n} = (1+\lambda)b_{j,n}$*. This implies* $\arg\max_{\rho_j\geq 0} \mathcal{L}_j(\rho_j, \boldsymbol{c}) = \max\{r_{j,n} : n = 0, 1 \ldots, S_j\,, \sigma_{j,n}(\boldsymbol{c}) \geq 0\}$*.*

In fact, for any realized value-cost pairs $\boldsymbol{z}$, the "realization versions" of the conversion and Lagrangians functions, namely $V_j(\rho_j; \boldsymbol{z})$ and $\mathcal{L}_j(\rho_j, \boldsymbol{c}; \boldsymbol{z})$, also satisfy the same properties as those of $V_j(\rho_j)$ and $\mathcal{L}_j(\rho_j, \boldsymbol{c})$. We provide a visual illustration for these properties in Figure 2.

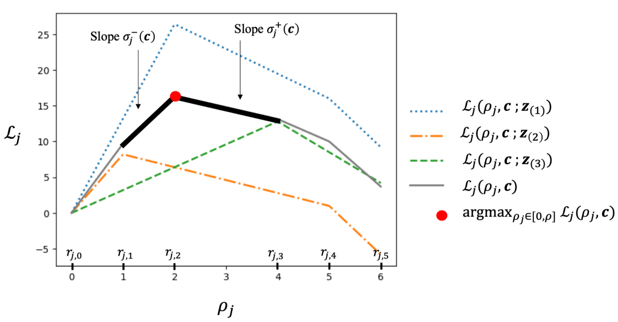

Figure 2. Illustration of Lagrangian functions defined in Eq. (6) with $M_j = 2$ auctions in channel j, and support F_j that contains 3 elements, $\boldsymbol{z}_{(1)} = ((8,2),(2,3))$, $\boldsymbol{z}_{(2)} = ((3,4),(1,4))$, $\boldsymbol{z}_{(3)} = ((8,1),(4,2))$, and context $\boldsymbol{c} = (\lambda, \mu) = (4,2)$. Under Lemma 4.5, $S_j = 5$, where the "turning points" $r_{j,0} \ldots r_{j,S_j}$ are indicated on the x-axis, and the optimal budget w.r.t. $\boldsymbol{c}$ is $\arg\max_{\rho_j\in[0,\rho]} \mathcal{L}_j(\rho_j; \boldsymbol{c}) = r_{j,2}$. The adjacent slopes in Eq. (11) are $\sigma_j^-(\boldsymbol{c}) = \sigma_{j,2}(\boldsymbol{c})$, and $\sigma_j^+(\boldsymbol{c}) = \sigma_{j,3}(\boldsymbol{c})$.

We now handle the reward boundedness issue for the Lagrangian functions that arise from adversarial contexts: in Lemma 4.6 (proof in Appendix C.4), we show the Lagrangian functions are bounded by some absolute constants:

Lemma 4.6 (Bounding Lagrangian functions)**.** *For any* $t \in [T]$, $j \in [M]$ *and* $\rho_j \in [0, \rho]$ *we have* $-(1+\gamma)\,\rho C_F \leq \mathcal{L}_j(\rho_j, \lambda_t, \mu_t) \leq (1 + C_F)\overline{V}$*, where the dual variables* $(\lambda_t, \mu_t)_{t\in[T]}$ *are generated from Algorithm 1.*

We now address the context-dependent exploration-exploitation tradeoff. To illustrate (e.g. Figure 2), define the slopes that are adjacent to the optimal per-channel budget w.r.t. $\boldsymbol{c} = (\lambda, \mu)$ as followed: assume the nth "turning point" $r_{j,n} = \arg\max_{\rho_j\in[0,\rho]} \mathcal{L}_j(\rho_j, \boldsymbol{c})$, then

$$\sigma_j^-(\boldsymbol{c}) = \sigma_{j,n}(\boldsymbol{c}) \;\text{ and }\; \sigma_j^+(\boldsymbol{c}) = \sigma_{j,n+1}(\boldsymbol{c}) \qquad (11)$$

Similar to standard exploration-exploitation tradeoffs in bandits, the flatter the slope (e.g. $\sigma_j^-(\boldsymbol{c})$ is close to 0), the more pulls required to accurately estimate rewards for sub-optimal arms on the slope, but the lower the loss in conversion for pulling sub-optimal arms. Our setting is challenging because the magnitude of this tradeoff, or equivalently adjacent slopes $\sigma_j^-(\boldsymbol{c})$ and $\sigma_j^+(\boldsymbol{c})$, depend on the adversarial contexts. In Lemma 4.7 we bound the UCB error by handling this context-dependent tradeoff through separately analyzing periods when the adjacent slopes $\sigma_j^-(\boldsymbol{c})$ and $\sigma_j^+(\boldsymbol{c})$ are less or greater than some parameter $\underline{\sigma} > 0$ chosen later, and characterize the context-dependent tradeoff using $\underline{\sigma}$.

Lemma 4.7 (Bounding UCB error)**.** *Assume the discretization width* δ *satisfies* $\delta < \underline{r}_j := \min_{n\in[S_j]} r_{j,n} - r_{j,n-1}$*, where* S_j *and* $\{r_{j,n}\}_{n=0}^{S_j}$ *are defined in Lemma 4.5. Then the UCB error in Proposition 4.3 is upper bounded by* $\mathcal{O}\left(\delta T + \underline{\sigma} T + \frac{1}{\underline{\sigma}\delta}\right)$*, where* $\underline{\sigma} > 0$ *is any positive number.*

See Appendix C.5 for the proof. Finally, we put together Proposition 4.3, Lemma 4.4 and 4.7, and obtain the main result Theorem 4.8 whose proof we detail in Appendix C.6

Theorem 4.8 (Putting everything together)**.** *Assume assumptions 4.1 and 4.2 hold. Take step size* $\eta = \Theta(1/\sqrt{T})$*, discretization width* $\delta = \Theta(T^{-1/3})$ *and* $\beta = \underline{\rho} = \frac{1}{\log(T)}$ *in Algorithm 1, as well as* $\underline{\sigma} = \Theta(T^{-1/3})$ *in Lemma 4.7. Then, for large enough* T *we have* $T \cdot \text{GL-OPT} - \mathbb{E}\left[\sum_{t\in[T]}\sum_{j\in[M]} V_j(\rho_{j,t})\right] \leq \mathcal{O}(T^{2/3})$*. Further, recalling* $\overline{\boldsymbol{\rho}}_T$ *is the final output of Algorithm 1, we have*

$$\text{GL-OPT} - \sum_{j\in[M]} \mathbb{E}\left[V_j(\overline{\rho}_{j,T})\right] \leq \mathcal{O}(T^{-1/3}) \text{ and}$$

constraint satisfaction: $\rho - \sum_{j\in[M]} \mathbb{E}[\overline{\rho}_{j,T}] \geq 0$*, as well as* $\sum_{j\in[M]} \mathbb{E}[V_j(\overline{\rho}_{j,T}) - \gamma\overline{\rho}_{j,T}] \geq 0$*.*

We make an important remark that distinguishes our result in Theorem 4.8 with related literature on convex optimization:

Remark 4.9. In light of Lemma 4.5, the advertiser's optimization problem CH-OPT($\mathcal{I}_B$) in Eq. (3) effectively becomes a convex problem (see Proposition C.5 in Appendix C.7). Hence it may be tempting for one to directly employ off-the-shelf convex optimization algorithms. However, our problem involves stochastic bandit feedback, and more importantly, *uncertain constraints*, meaning that we cannot analytically determine whether a primal decision satisfies the constraints of the problem. For example, in CH-OPT($\mathcal{I}_B$), for some primal decision $(\rho_j)_{j\in[M]}$, we cannot determine whether the ROI constraint $\sum_{j\in M}\mathbb{E}\left[V_j(\gamma_j,\rho_j;\boldsymbol{z}_j)-\gamma D_j(\gamma_j,\rho_j;\boldsymbol{z}_j)\right]\geq 0$ holds because the distribution $(\boldsymbol{p}_j)_{j\in[M]}$ from which $\boldsymbol{z}$ is sampled is unknown. To the best of our knowledge, there are only two recent works that handle a similar stochastic bandit feedback, and uncertain constraint setting, namely (Usmanova et al., 2019) and (Nguyen & Balasubramanian, 2022). Nevertheless, our setting is more challenging because these works consider a "two-point estimation" regime where one can make function evaluations to the objective and constraints twice each period, whereas our setting involves "one-point estimation" such that we can only make function calls once per period. We note the optimal oracle complexities for unknown constraint convex optimization with one-point bandit feedback, remains an open problem.[6]

5. Additional Discussions

See more details on following discussions in Appendix A.3.

General advertiser objectives. In GL-OPT and CH-OPT($\mathcal{I}$) we can consider more general objectives, namely $\sum_{j\in[M]}\mathbb{E}\left[\boldsymbol{v}_j^\top\boldsymbol{x}_j-\alpha\boldsymbol{d}_j^\top\boldsymbol{x}_j\right]$ and $\sum_{j\in M}\mathbb{E}\left[V_j(\gamma_j,\rho_j;\boldsymbol{z}_j)-\alpha V_j(\gamma_j,\rho_j;\boldsymbol{z}_j)\right]$ for some private cost $\alpha\in[0,\gamma]$. Our results in Section 3 still hold, and Algorithm 1 can still produce per-channel budgets that are approximately optimal via introducing α into the Lagrangian. Note $\alpha=0$ recovers models in the previous section, whereas $\alpha=1$ yields classic quasi-linear utilities.

Ad auctions selling multiple impressions. In GL-OPT and channels' autobidding problem Eq.(4), a single impression is sold per auction. Yet, our insights in Section 3 also hold for auctions that sell multiple impressions from which advertisers can at most procure 1, e.g. position auctions such as VCG or generalized second price (GSP). To see this, use our original notation $(\boldsymbol{v}_j,\boldsymbol{d}_j)$ to represent the concatenation of all value-cost pairs of individual impressions across all auctions in channel j, while $\boldsymbol{x}_j$ is the vector of indicators for individual impressions. Further, our SGD-UCB algorithm produces accurate lever estimates for auctions whose induced conversion function $V_j(\rho_j)$ still possesses properties in Lemma 4.5. This holds for position auctions whose marginal cost for winning a higher position increases, e.g. VCG (but not GSP).

Non-optimal autobidding in channels. Eq. (4) assumes channels adopt "optimal autobidding", and natural questions regarding non-optimal autobidding lead to interesting future research. In such a scenario, an advertiser's (bandit) feedback in channel j is $V_j(\gamma_j,\rho_j;\boldsymbol{z}_j)-\epsilon_j$ for some loss $\epsilon_j>0$. One potential resolution is to treat such loss as adversarial corruptions to bandit rewards, and augment SGD with bandit algorithms robust to adversarial corruptions; e.g. (Lykouris et al., 2018; Gupta et al., 2019).

[6] See Table 4.1 in (Larson et al., 2019) for best known complexity bounds for one-point bandit feedback setups.

References

Agarwal, A., Hsu, D., Kale, S., Langford, J., Li, L., and Schapire, R. Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning*, pp. 1638–1646. PMLR, 2014.

Aggarwal, G., Badanidiyuru, A., and Mehta, A. Autobidding with constraints. In *International Conference on Web and Internet Economics*, pp. 17–30. Springer, 2019.

Agrawal, R. The continuum-armed bandit problem. *SIAM journal on control and optimization*, 33(6):1926–1951, 1995.

Audet, C. and Dennis Jr, J. E. A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010, 2004.

Audet, C. and Tribes, C. Mesh-based nelder–mead algorithm for inequality constrained optimization. *Computational Optimization and Applications*, 71(2):331–352, 2018.

Balseiro, S., Kim, A., Mahdian, M., and Mirrokni, V. Budget management strategies in repeated auctions. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 15–23, 2017.

Balseiro, S., Golrezaei, N., Mahdian, M., Mirrokni, V., and Schneider, J. Contextual bandits with cross-learning. *Advances in Neural Information Processing Systems*, 32, 2019a.

Balseiro, S., Golrezaei, N., Mirrokni, V., and Yazdanbod, S. A black-box reduction in mechanism design with private cost of capital. *Available at SSRN 3341782*, 2019b.

Balseiro, S., Deng, Y., Mao, J., Mirrokni, V., and Zuo, S. Robust auction design in the auto-bidding world. *Advances in Neural Information Processing Systems*, 34: 17777–17788, 2021a.

Balseiro, S. R. and Gur, Y. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

Balseiro, S. R., Deng, Y., Mao, J., Mirrokni, V. S., and Zuo, S. The landscape of auto-bidding auctions: Value versus utility maximization. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 132–133, 2021b.

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

Bertsekas, D. P. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

Besbes, O., Gur, Y., and Zeevi, A. Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems*, 27, 2014.

Braverman, M., Mao, J., Schneider, J., and Weinberg, M. Selling to a no-regret buyer. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 523–538, 2018.

Bűrmen, Á., Puhan, J., and Tuma, T. Grid restrained nelder-mead algorithm. *Computational optimization and applications*, 34(3):359–375, 2006.

Castiglioni, M., Celli, A., Marchesi, A., Romano, G., and Gatti, N. A unifying framework for online optimization with long-term constraints. *arXiv preprint arXiv:2209.07454*, 2022.

Cheung, W. C., Simchi-Levi, D., and Zhu, R. Learning to optimize under non-stationarity. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1079–1087. PMLR, 2019.

Deng, Y., Schneider, J., and Sivan, B. Prior-free dynamic auctions with low regret buyers. *Advances in Neural Information Processing Systems*, 32, 2019.

Deng, Y., Mao, J., Mirrokni, V., and Zuo, S. Towards efficient auctions in an auto-bidding world. In *Proceedings of the Web Conference 2021*, pp. 3965–3973, 2021.

Deng, Y., Golrezaei, N., Jaillet, P., Liang, J. C. N., and Mirrokni, V. Fairness in the autobidding world with machine-learned advice. *arXiv preprint arXiv:2209.04748*, 2022a.

Deng, Y., Mao, J., Mirrokni, V., Zhang, H., and Zuo, S. Efficiency of the first-price auction in the autobidding world. *arXiv preprint arXiv:2208.10650*, 2022b.

Ding, Q., Hsieh, C.-J., and Sharpnack, J. An efficient algorithm for generalized linear bandit: Online stochastic gradient descent and Thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pp. 1585–1593. PMLR, 2021.

Drutsa, A. Optimal non-parametric learning in repeated contextual auctions with strategic buyer. In *International Conference on Machine Learning*, pp. 2668–2677. PMLR, 2020.

Dzahini, K. J., Kokkolaras, M., and Le Digabel, S. Constrained stochastic blackbox optimization using a progressive barrier and probabilistic estimates. *Mathematical Programming*, pp. 1–58, 2022.

Fasano, G., Liuzzi, G., Lucidi, S., and Rinaldi, F. A linesearch-based derivative-free approach for nonsmooth constrained optimization. *SIAM journal on optimization*, 24(3):959–992, 2014.

Feng, Z., Padmanabhan, S., and Wang, D. Online bidding algorithms for return-on-spend constrained advertisers. *arXiv preprint arXiv:2208.13713*, 2022.

Flaxman, A. D., Kalai, A. T., and McMahan, H. B. Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*, 2004.

Golrezaei, N., Jaillet, P., and Liang, J. C. N. Incentive-aware contextual pricing with non-parametric market noise. *arXiv preprint arXiv:1911.03508*, 2019a.

Golrezaei, N., Javanmard, A., and Mirrokni, V. Dynamic incentive-aware learning: Robust pricing in contextual auctions. *Advances in Neural Information Processing Systems*, 32, 2019b.

Golrezaei, N., Jaillet, P., Liang, J. C. N., and Mirrokni, V. Bidding and pricing in budget and ROI constrained markets. *arXiv preprint arXiv:2107.07725*, 2021a.

Golrezaei, N., Manshadi, V., Schneider, J., and Sekar, S. Learning product rankings robust to fake users. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 560–561, 2021b.

Gupta, A., Koren, T., and Talwar, K. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pp. 1562–1578. PMLR, 2019.

Han, Y., Liang, Z., Wang, Y., and Zhang, J. Generalized linear bandits with local differential privacy. *Advances in Neural Information Processing Systems*, 34:26511–26522, 2021.

Kannan, P., Reinartz, W., and Verhoef, P. C. The path to purchase and attribution modeling: Introduction to special section, 2016.

Larson, J., Menickelly, M., and Wild, S. M. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019.

Luo, H., Wei, C.-Y., Agarwal, A., and Langford, J. Efficient contextual bandits in non-stationary worlds. In *Conference On Learning Theory*, pp. 1739–1776. PMLR, 2018.

Lykouris, T., Mirrokni, V., and Paes Leme, R. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 114–122, 2018.

Mehta, A. Auction design in an auto-bidding setting: Randomization improves efficiency beyond vcg. In *Proceedings of the ACM Web Conference 2022*, pp. 173–181, 2022.

Nguyen, A. and Balasubramanian, K. Stochastic zeroth-order functional constrained optimization: Oracle complexity and applications. *INFORMS Journal on Optimization*, 2022.

Pourmohamad, T. and Lee, H. K. The statistical filter approach to constrained optimization. *Technometrics*, 62 (3):303–312, 2020.

Slivkins, A. et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

Usmanova, I., Krause, A., and Kamgarpour, M. Safe convex learning under uncertain constraints. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2106–2114. PMLR, 2019.

Varian, H. R. Position auctions. *international Journal of industrial Organization*, 25(6):1163–1178, 2007.

Appendices for
Multi-channel Autobidding with Budget and ROI constraints

A. Additional material

A.1. Extended Literature Review

Generally speaking, our work focuses on advertisers' impression procurement process or the interactions between advertisers and impression sellers, which has been addressed in a vast amount of literature in mechanism design and online learning; see e.g. (Braverman et al., 2018; Deng et al., 2019; Golrezaei et al., 2019b;a; Balseiro et al., 2019b; Golrezaei et al., 2021a) to name a few. In addition to the literature review on constrained optimization under bandit feedback in the introduction section, we also discuss additional related works in this area here.

Constrained optimization under bandit feedback. Section 4 where we develop an algorithm to optimize over per-channel target budgets relates to the area of convex constrained optimization with bandit feedback (also referred to as zero-order or gradient-less feedback) since in light of Lemma 4.5 in Section 4 our problem of interest is also constrained and convex. First, there has been a plenitude of algorithms developed for deterministic constrained convex optimization under a bandit feedback structures where function evaluations for the objective and constraints are non-stochastic. Such algorithms include filter methods (Audet & Dennis Jr, 2004; Pourmohamad & Lee, 2020), barrier-type methods (Fasano et al., 2014; Dzahini et al., 2022), as well as Nelder-Mead type algorithms (Bűrmen et al., 2006; Audet & Tribes, 2018); see (Nguyen & Balasubramanian, 2022) and references therein for a comprehensive survey. In contrast to these works, our optimization algorithm developed in Section 4 handles noisy bandit feedback. Regarding works that also address stochastic settings, (Flaxman et al., 2004) presents online optimization algorithms under the *known constraint* regime, which assumes the optimizer can evaluate whether all constraints are satisfied, i.e. constraints are analytically available. Further, the algorithm achieves a $\mathcal{O}(T^{-1/4})$ accuracy. In this work, our setting is more complex as the optimizer (i.e. the advertiser) cannot tell whether the ROI constrained is satisfied (due to unknown value and cost distributions in each channels' auctions). Yet our proposed algorithm can still achieve a more superior $\mathcal{O}(T^{-1/3})$ accuracy due to our specific problem structure.

A.2. Additional materials for Section 4

Remark A.1. Our problem of interest to apply SGD under bandit feedback is more difficult than similar problems in related works that study online bidding strategies under budget and ROI constraints; see e.g. (Balseiro et al., 2017; 2022; Feng et al., 2022). To illustrate, consider for instance (Balseiro et al., 2017) in which a budget constrained advertiser's primal decision at period t is to submit a bid value b_t after observing her value v_t. The advertiser competes with some unknown highest competing bid d_t in the market, and after submitting bid b_t, does not observe d_t if she does not win the competition, which involves a semi-bandit feedback structure. Nevertheless, the corresponding Lagrangian under SGD takes the special form $\mathcal{L}_j(b, \mu_t; \boldsymbol{z}_t) = (v_t - (1+\mu_t)d_t)\,\mathbb{I}\{b \geq d_t\}$ where μ_t is the dual variable w.r.t. the budget constraint. This simply allows an advertiser to optimize for her primal decision by bidding $\arg\max_{b\geq 0} \mathcal{L}_j(b, \boldsymbol{c}_t; \boldsymbol{z}_t) = \frac{v_t}{1+\mu_t}$. So even though (Balseiro et al., 2017; 2022; Feng et al., 2022) study dual SGD under bandit feedback, the special structures of their problem instances permits SGD to effectively optimize for primal decisions in each period, as opposed to Eq. (7) in our setting in which we cannot directly solve for the primal decision due to unknown conversion functions.

A.3. Additional materials for Section 5

A.3.1. More general advertiser objectives

In GL-OPT as well as CH-OPT$(\mathcal{I})$ we can also consider more general objectives, namely $\max_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_M} \sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{v}_j^\top \boldsymbol{x}_j - \alpha \boldsymbol{d}_j^\top \boldsymbol{x}_j\right]$ and $\max_{(\gamma_j,\rho_j)_{j\in[M]}\in\mathcal{I}} \sum_{j\in M} \mathbb{E}\left[V_j(\gamma_j,\rho_j;\boldsymbol{z}_j) - \alpha V_j(\gamma_j,\rho_j;\boldsymbol{z}_j)\right]$ for some private cost $\alpha \in [0,\gamma]$[7] in GL-OPT and CH-OPT$(\mathcal{I})$, respectively. When $\alpha = 0$, we recover our considered models in the previous section, whereas in when $\alpha = 1$, we obtain the classic quasi-linear utility. We remark that this private cost model has been introduced and studied in related literature; see (Balseiro et al., 2019b) and references therein. Nevertheless, when each channel's autobidding problem remains as is in Eq.(4), i.e. channels still aim to maximize conversion which

[7] If $\alpha > \gamma$ the ROI constraints in GL-OPT as well as CH-OPT$(\mathcal{I})$ become redundant.

causes a misalignment between advertiser objectives and channel behavior, it is not difficult to see in our proofs that all our results still hold in Section 3, and our UCB-SGD algorithm still produces estimates of the same order of accuracy via introducing α into the Lagrangian. In other words, even if channels aim to maximize total conversion for advertisers, advertisers can optimize for GL-OPT with a private cost α through optimizing CH-OPT($\mathcal{I}$) that also incorporates the same private cost.

A.3.2. Ad auctions selling multiple impressions.

Recall in GL-OPT and each channel's autobidding problem in Eq. (4) we implicitly assumed each auction sells of a single impression. Interestingly, our results in Section 3 that states solely optimizing per-channel budgets would be sufficient for the advertiser to achieve GL-OPT also holds in the scenario when each single auction corresponds to the sale of multiple impressions in which an advertiser can at most procure 1, or in other words, the auctions are position auctions such as VCG or generalized second price (GSP) auctions. To see this, we can use our original notation $(\boldsymbol{v}_j, \boldsymbol{d}_j)$ to represent the concatenation of all value-cost pairs of all individual impressions across all auctions in channel j. Correspondingly the decision $\boldsymbol{x}_j$ is the vector of indicators that determine whether an individual impression is procured, but with the additional linear constraint that says the sum of indicators within any auction sum up to less than 1 to represent an advertiser can win at most 1 impression. A similar representation can also be used for each channel's autobidding problem in Eq. (4). Then, it is not difficult to see the results and proofs of Theorem 3.3 and Corollary 3.4 still hold valid.

Regarding the applicability of our proposed UCB-SGD algorithm in Algorithm 1, it suffices to check if the position auctions induce a conversion function $V_j(\rho_j)$ that satisfies properties illustrated in Lemma 4.5, in particular whether $V_j(\rho_j; \boldsymbol{z}_j)$ is continuous, concave, piecewise linear, and strictly increasing. We claim that this is true for VCG auctions. Fix some VCG auction n in channel j with L positions, each corresponding to a click through rate (CTR) θ_ℓ that represents the probability of a user viewing that position (so it is natural to assume $\theta_1 > \theta_2 > \ldots \theta_L$; see (Varian, 2007) for more details. If an advertiser procures position ℓ, she acquires value $v_{(\ell)} = \theta_\ell \cdot v$ where v represents her expected conversion due to a user viewing her ad (i.e. the value of winning any position), and incurs expenditure $p_{(\ell)} = \sum_{\ell'=\ell}^{L}(\theta_{\ell'} - \theta_{\ell'+1})d^{(\ell')}$, where $d^{(\ell')}$ is the ℓ'th highest competing bid in the market, and we denote $\theta_{L+1} = 0$. Now, this single VCG auction can be viewed as L separated impressions (each sold in a separate single-impression auction) with value-cost pairs $\left(v_{(1)} - v_{(2)}, p_{(1)} - p_{(2)}\right), \ldots \left(v_{(L)} - v_{(L+1)}, p_{(L)} - p_{(L+1)}\right)$, where we denote $v_{(L+1)} = p_{(L+1)} = 0$. Correspondingly, the indicator decision variables $x_{n,j}$ in GL-OPT and the channel's autobidding problem Eq. (4) for the original multi-impression auction n of channel j can be viewed as a vector of indicator decisions $\boldsymbol{x} \in \{0,1\}^L$ for each of these separated impressions. Now, the proof of Lemma 4.5 still holds if we can show that procuring any separated impression ℓ also implies procuring impression $\ell+1, \ell+2 \ldots L$, or equivalently, $\frac{v_{(\ell)} - v_{(\ell+1)}}{p_{(\ell)} - p_{(\ell+1)}} > \frac{v_{(\ell')} - v_{(\ell'+1)}}{p_{(\ell')} - p_{(\ell'+1)}}$ for any $\ell' > \ell$ (since the advertiser's problem is the LP relaxation of the 0-1 knapsack problem as discussed in the proof of Lemma 4.5). It is thus easy to see this holds because

$$\frac{v_{(\ell)} - v_{(\ell+1)}}{p_{(\ell)} - p_{(\ell+1)}} = \frac{(\theta_\ell - \theta_{\ell+1})\, v}{(\theta_\ell - \theta_{\ell+1}) d^{(\ell)}} = \frac{v}{d^{(\ell)}} \tag{12}$$

which decreases in ℓ for any ℓ because $d^{(\ell)}$ is the ℓth highest competing bid. In other words, marginal cost increases as one procures higher positions. Hence, the proof of Lemma 4.5 holds w.r.t. these separated impressions, and thus for VCG auctions, the induced conversion satisfies properties in Lemma 4.5. The insight here is that for any position auctions whose marginal cost increases as the advertiser procures higher positions, properties in Lemma 4.5 hold. However, this is not true for auctions like generalized second price (GSP).

A.3.3. Non-optimal autobidding in channels.

We recall in previous sections we assumed that each channel adopt "optimal autobidding" that solves Eq. (4) to optimality. This raises the natural question that whether our findings will still hold when channels do not procure ads optimally, perhaps because of non-stationary environments (Besbes et al., 2014; Luo et al., 2018; Cheung et al., 2019), or the presence of strategic market participants who aim to manipulate the market (Golrezaei et al., 2019a; Drutsa, 2020; Golrezaei et al., 2021b;a). In such a scenario, an advertiser's (bandit) conversion feedback in a channel j would be $V(\gamma_j, \rho_j; \boldsymbol{z}_j) - \epsilon_j$ for some channel-specific and possibly adversarial loss $\epsilon_j > 0$. One potential resolution is to treat such ϵ_j as adversarial corruptions to bandit rewards, and instead of integrating vanilla UCB with SGD as in Algorithm 1, augment SGD with bandit algorithms that are robust to corruptions; see e.g. (Lykouris et al., 2018; Gupta et al., 2019). Nevertheless, it remains an open question to prove how such augmentation would perform in our specific bandit-feedback constrained optimization setup. This leads to potential research directions of both practical and theoretical significance.

B. Proofs for Section 3

B.1. Proof of Lemma 3.1

Fix any option $\mathcal{I} \in \{\mathcal{I}_B, \mathcal{I}_R, \mathcal{I}_G\}$ defined in Eq. (2), and let $(\widetilde{\boldsymbol{\gamma}}, \widetilde{\boldsymbol{\rho}}) \in \mathcal{I}$ be the optimal solution to CH-OPT$(\mathcal{I})$. Note that for the per-channel ROI only option $\mathcal{I}_R$, we have $\widetilde{\rho}_j = \infty$ and for the per-channel budget only we have $\widetilde{\gamma}_j = 0$ for all $j \in [M]$. Further, for any realization of value-cost pairs over all auctions $\boldsymbol{z} = (\boldsymbol{v}_j, \boldsymbol{d}_j)_{j\in[M]}$, recall the optimal solution $\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ to $V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ for each channel $j \in [M]$ as defined in Eq. (4).

Due to feasibility of $(\widetilde{\boldsymbol{\gamma}}, \widetilde{\boldsymbol{\rho}}) \in \mathcal{I}$ for CH-OPT$(\mathcal{I})$, we have

$$\sum_{j\in M} \mathbb{E}\left[V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \geq \gamma \sum_{j\in M} \mathbb{E}\left[D_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \Longrightarrow \sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \geq \gamma \sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{d}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right]$$

where we used the definitions $V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j) = \boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ and $D_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j) = \boldsymbol{d}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ in Eq. (5). This implies $\left(\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right)_{j\in[M]}$ satisfies the ROI constraint in GL-OPT. A similar analysis implies $\left(\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right)_{j\in[M]}$ also satisfies the budget constraint in GL-OPT. Therefore, $\left(\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right)_{j\in[M]}$ is feasible to GL-OPT. So

$$\text{GL-OPT} \geq \sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] = \sum_{j\in M} \left[V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] = \text{CH-OPT}(\mathcal{I})\,. \tag{13}$$

where the final equality follows from the assumption that $(\widetilde{\boldsymbol{\gamma}}, \widetilde{\boldsymbol{\rho}}) \in \mathcal{I}$ is the optimal solution to CH-OPT$(\mathcal{I})$. □

B.2. Proof of Theorem 3.2

Recall the value and cost instance:

	Channel 1	Channel 2	
	Auction 1	Auction 2	Auction 3
Value $v_{j,n}$	1	X	$2X$
Spend $d_{j,n}$	0	$1+X$	$2(1+X)$

Let $\widetilde{\boldsymbol{\gamma}} = (\widetilde{\gamma}_1, \widetilde{\gamma}_2)$ be the optimal solution to CH-OPT$(\mathcal{I}_R)$ and recall under the option $\mathcal{I}_R$, we let per-channel budgets to be infinity. It is easy to see that $\widetilde{\gamma}_1$ can be any arbitrary nonnegative number because the advertiser always wins auction 1, and $\widetilde{\gamma}_2 > \frac{X}{1+X}$: if otherwise $\widetilde{\gamma}_2 \leq \frac{X}{1+X}$, then the optimal outcome of channel 2 is to win both auctions 2 and 3. However, in this case, the advertiser wins all auctions and acquires total value $1 + X + 2X = 1 + 3X$, and incurs total spend $0 + (1 + X) + 2(1 + X) = 3 + 3X$, which violates the ROI constraint in CH-OPT$(\mathcal{I}_R)$ because $\frac{1+3X}{3+3X} < 1$. Therefore the advertiser can only win auction 1, or in other words $\widetilde{\gamma}_2 > \frac{X}{1+X}$. This implies that the optimal objective to CH-OPT$(\mathcal{I}_R)$ is 1. On the other hand, it is easy to see that the optimal solution to GL-OPT is to only win auctions 1 and 2, yielding an optimal value of $1 + X$. Therefore $\frac{\text{CH-OPT}(\mathcal{I}_R)}{\text{GL-OPT}} = \frac{1}{1+X}$. Taking $X \to \infty$ yeilds the desired result. □

B.3. Proof of Theorem 3.3

In light of Lemma 3.1, we only need to show CH-OPT$(\mathcal{I}_B) \geq$ GL-OPT. Let $\widetilde{\boldsymbol{x}}(\boldsymbol{z}) = \{\widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j))\}_{j\in[N]}$ be the optimal solution to GL-OPT, and define $\widetilde{\gamma}_j = 0$ and $\widetilde{\rho}_j = \mathbb{E}\left[\boldsymbol{d}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j))\right]$ to be the corresponding expected spend for each channel j under the optimal solution $\widetilde{\boldsymbol{x}}(\boldsymbol{z})$ to GL-OPT, respectively.

We first argue that $(\widetilde{\gamma}_j, \widetilde{\rho}_j)_{j\in[M]}$ is feasible to CH-OPT$(\mathcal{I}_B)$. Recall the optimal solution $\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ to $V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ for each channel $j \in [M]$ as defined in Eq. (4), as well as the definitions $V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j) = \boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ and $D_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j) = \boldsymbol{d}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ in Eq. (5). Then, we have

$$\mathbb{E}\left[D_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \;=\; \mathbb{E}\left[\boldsymbol{d}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j); \boldsymbol{z}_j\right] \;\overset{(i)}{\leq}\; \widetilde{\rho}_j \;=\; \mathbb{E}\left[\boldsymbol{d}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j)\right]\,, \tag{14}$$

where (i) follows from feasibility of $\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ to $V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$. Summing over $j \in [M]$ we conclude that $(\widetilde{\gamma}_j, \widetilde{\rho}_j)_{j\in[M]}$ satisfies the budget constraint in CH-OPT$(\mathcal{I}_B)$:

$$\sum_{j\in[M]} \mathbb{E}\left[D_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \leq \sum_{j\in[M]} \mathbb{E}\left[\boldsymbol{d}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j)\right] \overset{(i)}{\leq} \rho\,. \tag{15}$$

Here (i) follows from feasibility of $\widetilde{\boldsymbol{x}}(\boldsymbol{z}) = \{\widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j))\}_{j\in[N]}$ to GL-OPT since it is the optimal solution.

On the other hand, we have

$$V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j) \;=\; \boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j) \;\overset{(i)}{\geq}\; \boldsymbol{v}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j) \tag{16}$$

where (i) follows from optimality of $\boldsymbol{x}_j^*(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$ to $V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)$. Hence, we have

$$\sum_{j\in M} \mathbb{E}\left[V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \;\geq\; \sum_{j\in M} \mathbb{E}\left[\boldsymbol{v}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j)\right] \;\overset{(i)}{\geq}\; \gamma \sum_{j\in M} \mathbb{E}\left[\boldsymbol{d}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j)\right] \;\overset{(ii)}{\geq}\; \gamma \sum_{j\in[M]} \mathbb{E}\left[D_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \tag{17}$$

where (i) follows from feasibility of $\widetilde{\boldsymbol{x}}(\boldsymbol{z}) = \{\widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j))\}_{j\in[N]}$ to GL-OPT since it is the optimal solution; (ii) follows from Eq. (14). Hence combining Eq. (15) (17) we can conclude that $(\widetilde{\gamma}_j, \widetilde{\rho}_j)_{j\in[M]}$ is feasible to CH-OPT$(\mathcal{I}_B)$.

Finally, we have CH-OPT$(\mathcal{I}_B) \geq \sum_{j\in M} \mathbb{E}\left[V_j(\widetilde{\gamma}_j, \widetilde{\rho}_j; \boldsymbol{z}_j)\right] \geq \sum_{j\in M} \mathbb{E}\left[\boldsymbol{v}_j^\top \widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j)\right] =$ GL-OPT, where the last inequality follows from Eq. (16), and the final equality is because we assumed $\widetilde{\boldsymbol{x}}(\boldsymbol{z}) = \{\widetilde{\boldsymbol{x}}_j(\boldsymbol{z}_j))\}_{j\in[N]}$ is the optimal solution to GL-OPT. □

B.4. Proof of Corollary 3.4

In light of Lemma 3.1, we only need to show CH-OPT$(\mathcal{I}_G) \geq$ GL-OPT. Let $(\widetilde{\boldsymbol{\gamma}}, \widetilde{\boldsymbol{\rho}}) \in \mathcal{I}_B$ be the optimal solution to CH-OPT$(\mathcal{I}_B)$, and by definition of $\mathcal{I}_B$ in Eq. (2) we have $\widetilde{\gamma}_j = 0$ for all $j \in [M]$. Since $(\widetilde{\boldsymbol{\gamma}}, \widetilde{\boldsymbol{\rho}})$ is feasible to CH-OPT$(\mathcal{I}_B)$, it is also feasible to CH-OPT$(\mathcal{I}_G)$ since these two problems share the same ROI and budget constraints. Because they also share the same objectives, we have

$$\text{CH-OPT}(\mathcal{I}_G) \geq \text{CH-OPT}(\mathcal{I}_B) = \text{GL-OPT} \tag{18}$$

where the final equality follows from Theorem 3.3. □

C. Proofs for Section 4

C.1. Proof of Proposition 4.3

Let $(\rho_j^*)_{j\in[M]}$ be the optimal per-channel budgets to CH-OPT$(\mathcal{I}_B)$, and define $\bar{\mu}_T = \frac{1}{\tau_A}\sum_{t\in[\tau_A]} \mu_t$ as well as $\bar{\lambda}_T = \frac{1}{\tau_A}\sum_{t\in[\tau_A]} \lambda_t$. Then

$$\begin{aligned}
& T \cdot \text{GL-OPT} - \sum_{t\in[T]}\sum_{j\in[M]} V_j(\rho_{j,t}) \\
\overset{(i)}{\leq}\;& M\overline{V}(T-\tau_A) + \tau_A \text{CH-OPT}(\mathcal{I}_B) - \sum_{t\in[\tau_A]}\sum_{j\in[M]} V_j(\rho_{j,t}) \\
\leq\;& M\overline{V}(T-\tau_A) + \tau_A \cdot \Big(\mathcal{L}_j(\rho_j^*, \bar{\lambda}_T, \bar{\mu}_T) + \rho\bar{\mu}_T\Big) - \sum_{t\in[\tau_A]}\sum_{j\in[M]} V_j(\rho_{j,t}) \\
\overset{(ii)}{\leq}\;& M\overline{V}(T-\tau_A) + \rho\sum_{t\in[\tau_A]}\mu_t + \sum_{t\in[\tau_A]}\sum_{j\in[M]} \mathcal{L}_j(\rho_j^*, \lambda_t, \mu_t) - \sum_{t\in[\tau_A]}\sum_{j\in[M]} \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t) - \lambda_t\left(V_j(\rho_{j,t}) - \gamma\rho_{j,t}\right) + \mu_t\rho_{j,t} \\
\overset{(iii)}{\leq}\;& M\overline{V}(T-\tau_A) + \sum_{j\in[M]}\sum_{t\in[\tau_A]} \mathcal{L}_j(\rho_j^*(t), \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t) + \sum_{t\in[\tau_A]} \left(\lambda_t g_{1,t} + \mu_t g_{2,t}\right)\,.
\end{aligned} \tag{19}$$

Here, (i) follows from Theorem 3.3 that states GL-OPT $=$ CH-OPT$(\mathcal{I}_B)$ and CH-OPT$(\mathcal{I}_B)$ is apparently upper bounded by $M\overline{V}$; (ii) follows from the CH-OPT$(\mathcal{I}_B) = \sum_{j\in[M]} V_j(\rho_j^*)$ and the definition of the Lagrangian in Eq. (6); in (iii) we define $\rho_j^*(t) = \arg\max_{\rho_j\geq 0} \mathcal{L}_j(\rho_j, \boldsymbol{c}_t)$ to be the optimal budget that maximizes the Lagrangian w.r.t. the dual variables $\boldsymbol{c}_t = (\lambda_t, \mu_t)$. □

C.2. Proof for Lemma 4.4

Recall $g_{1,t} = \sum_{j\in[M]} (V_{j,t}(\rho_{j,t}; \boldsymbol{z}_{j,t}) - \gamma\rho_{j,t})$ and $g_{2,t} = \rho - \sum_{j\in[M]} \rho_{j,t}$ defined in Algorithm 1. Also recall $\tau_A \in [T]$ defined in step 10 of Algorithm 1. In the following, we will show

$$
\begin{aligned}
& M\overline{V}(T-\tau_A) + \sum_{t\in[\tau_A]} (\lambda_t g_{1,t} + \mu_t g_{2,t}) \\
\leq\ & C_F \max\{M\overline{V}, \rho\} + M^2\overline{V}\rho \cdot \max\left\{\frac{1}{\beta\underline{\rho}}, \frac{1}{\rho - M\underline{\rho}}\right\} + \frac{(\gamma M^2\bar{V}^2 + \rho^2)}{2} \cdot \eta T + \frac{1}{2\eta}C_F^2 \ = \ \mathcal{O}\Big(\eta T + \frac{1}{\eta}\Big)\,,
\end{aligned}
\tag{20}
$$

where we recall $C_F = M\overline{V}\max\left\{\frac{1}{\beta\underline{\rho}}, \frac{1}{\rho - M\underline{\rho}}\right\}$ defined in Eq. (9).

From Lemma C.4, we have for any $t \in [T]$, and $\lambda, \mu \in [0, C_F]$,

$$
\begin{aligned}
\sum_{\tau\in[t]} (\lambda_\tau - \lambda)\, g_{1,\tau} &\leq \frac{\eta M^2\bar{V}^2}{2} \cdot t + \frac{1}{2\eta}\lambda^2 \\
\sum_{\tau\in[t]} (\mu_\tau - \mu)\, g_{2,\tau} &\leq \frac{\eta\rho^2}{2} \cdot t + \frac{1}{2\eta}\mu^2\,,
\end{aligned}
\tag{21}
$$

where we used the fact that $\lambda_1 = \mu_1 = 0$ in Algorithm 1.

Suppose that $\tau_A = T$ and thus $M\overline{V}(T-\tau_A) = 0$. Then, considering $\lambda = \mu = 0$ in Eq. (21), we have

$$
\sum_{t\in[\tau_A]} \lambda_t g_{1,t} \leq \frac{\eta M^2\bar{V}^2}{2} \cdot T \text{ and } \sum_{t\in[\tau_A]} \mu_t g_{2,t} \leq \frac{\eta\rho^2}{2} \cdot T\,.
\tag{22}
$$

Thus, Eq. (20) holds.

If $\tau_A < T$, then according to Algorithm 1, we either have $S_{1,\tau_A} - \gamma M\rho + \beta\underline{\rho}(T-\tau_A) < 0$ or $S_{2,\tau_A} + M\rho + M\underline{\rho}(T-\tau_A) > \rho T$, where we recall $S_{1,\tau_A} = \sum_{t\in[\tau_A-1]} g_{1,t}$ and $S_{2,\tau_A} = \sum_{t\in[\tau_A-1]}\sum_{j\in[M]} \rho_{j,t} = \sum_{t\in[\tau_A-1]}(\rho - g_{2,t})$:

- If $S_{1,\tau_A} - \gamma M\rho + \beta\underline{\rho}(T-\tau_A) < 0$, then we have $\sum_{t\in[\tau_A-1]} g_{1,t} < \gamma M\rho - \beta\underline{\rho}(T-\tau_A)$. Hence, considering $\lambda = \frac{M\overline{V}}{\beta\underline{\rho}} \in [0, C_F]$ in Eq. (21), we have

$$
\begin{aligned}
& M\overline{V}(T-\tau_A) + \sum_{t\in[\tau_A]} \lambda_t g_{1,t} \\
\leq\ & M\overline{V}(T-\tau_A) + \lambda_{\tau_A} g_{1,\tau_A} + \sum_{t\in[\tau_A-1]} \lambda g_{1,t} + \frac{\eta M^2\bar{V}^2}{2} \cdot (\tau_A - 1) + \frac{1}{2\eta}\lambda^2 \\
<\ & \lambda_{\tau_A} g_{1,\tau_A} + M\overline{V}(T-\tau_A) - M\overline{V}(T-\tau_A) + \frac{\gamma M^2\overline{V}\rho}{\beta\underline{\rho}} + \frac{\eta M^2\bar{V}^2}{2} \cdot (\tau_A - 1) + \frac{1}{2\eta}\lambda^2 \\
\leq\ & C_F M\overline{V} + \frac{\gamma M^2\overline{V}\rho}{\beta\underline{\rho}} + \frac{\eta M^2\bar{V}^2}{2} \cdot T + \frac{1}{2\eta}C_F^2,
\end{aligned}
\tag{23}
$$

  where the final inequality uses the fact that $\tau_A \leq T$, $\lambda \leq C_F$, and $g_{1,t} \leq M\overline{V}$ for any $t \in [T]$. Hence, similar to Eq. (22) by further taking $\mu = 0$ in Eq.(21) we show that Eq. (20) holds.

- If $S_{2,\tau_A} + M\rho + M\underline{\rho}(T-\tau_A) > \rho T$, then we have $\sum_{t\in[\tau_A-1]}(\rho - g_{2,t}) > \rho T - M\rho - M\underline{\rho}(T-\tau_A)$, or equivalently $\sum_{t\in[\tau_A-1]} g_{2,t} < M\underline{\rho}(T-\tau_A) + M\rho - \rho(T-\tau_A) \leq -(\rho - M\underline{\rho})(T-\tau_A) + M\rho$. Hence, considering $\mu = \frac{M\overline{V}}{\rho - M\underline{\rho}} \in [0, C_F]$ in Eq.(21) we have

$$\begin{aligned} M\overline{V}(T-\tau_A) + \sum_{t\in[\tau_A]} \mu_t g_{1,t} &\leq M\overline{V}(T-\tau_A) + \mu_{\tau_A} g_{2,\tau_A} + \sum_{t\in[\tau_A-1]} \mu g_{2,t} + \frac{\eta\rho^2}{2}\cdot\tau_A + \frac{1}{2\eta}\mu^2 \\ &< \mu_{\tau_A} g_{2,\tau_A} + M\overline{V}(T-\tau_A) - M\overline{V}(T-\tau_A) + \frac{M^2\overline{V}\rho}{\rho-\underline{\rho}} + \frac{\eta\rho^2}{2}\cdot\tau_A + \frac{1}{2\eta}\mu^2 \\ &\leq C_F\rho + \frac{M^2\overline{V}\rho}{\rho-\underline{\rho}} + \frac{\eta\rho^2}{2}\cdot T + \frac{1}{2\eta}C_F^2, \end{aligned} \tag{24}$$

where the final inequality uses the fact that $\tau_A \leq T$, $\lambda \leq C_F$, and $g_{2,t} \leq \rho$ for any $t \in [T]$. Hence, similar to Eq. (22) by further taking $\lambda = 0$ in Eq.(21) we show that Eq. (20) holds.

□

C.3. Proof of Lemma 4.5

We first show for any realization $\boldsymbol{z} = (\boldsymbol{z}_j)_{j\in[M]} = (\boldsymbol{v}_j, \boldsymbol{d}_j)_{j\in[M]}$, the conversion function $V_j(\rho_j;\boldsymbol{z}_j)$ is piecewise linear, strictly inreasing, and concave for any $j \in [M]$.

Fix any channel j which consists of m_j parallel auctions, and recall that we assumed the ordering $\frac{v_{j,1}}{d_{j,1}} > \frac{v_{j,2}}{d_{j,2}} > \cdots > \frac{v_{j,m_j}}{d_{j,m_j}}$ for any realization $\boldsymbol{z}_j$. Then, with the option where the per-channel ROI is set to 0 (i.e. omitted) $V_j(\rho_j;\boldsymbol{z}_j)$ is exactly the LP relaxation of a 0-1 knapsack, whose optimal solution $\boldsymbol{x}_j^*(\rho_j;\boldsymbol{z}_j)$ is well known to be unique, and takes the form for any auction index $n \in [m_j]$:

$$x_{j,n}^*(\rho_j;\boldsymbol{z}_j) = \begin{cases} 1 & \text{if } \sum_{n'\in[n]} d_{j,n'} \leq \rho_j \\ \frac{\rho_j - \sum_{n'\in[n-1]} d_{j,n'}}{d_{j,n}} & \text{if } \sum_{n'\in[n]} d_{j,n'} > \rho_j \\ 0 & \text{otherwise} \end{cases}$$

where we denote $d_{j,0} = 0$. With this form, it is easy to see

$$V_j(\rho_j;\boldsymbol{z}_j) = \boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\rho_j;\boldsymbol{z}_j) = \sum_{n\in[m_j]} \left(\frac{v_{j,n}}{d_{j,n}}\rho_j + b_{j,n}\right) \mathbb{I}\left\{d_{j,0} + \cdots + d_{j,n-1} \leq \rho_j \leq d_{j,0} + \cdots + d_{j,n}\right\} \tag{25}$$

where we denote $d_{j,0} = 0$ and also $b_{j,n} = \sum_{n'\in[n-1]} v_{j,n'} - \frac{v_{j,n}}{d_{j,n}} \cdot \left(\sum_{n'\in[n-1]} d_{j,n'}\right)$ and $v_{j,0} = 0$. It is easy to check that any two line segments, say $[X_{n-1}, X_n]$ and $[X_n, X_{n+1}]$ where we write $X_n = d_{j,0} + \cdots + d_{j,n}$, intersect at $\rho_j = X_n$, because $\frac{v_{j,n}}{d_{j,n}}\rho_j + b_{j,n} = \frac{v_{j,n+1}}{d_{j,n+1}}\rho_j + b_{j,n+1}$ at $\rho_j = X_n$. Hence, from Eq. (25) we can conclude $V_j(\rho_j;\boldsymbol{z}_j)$ is continuous, which further implies it is piecewise linear and strictly increasing. Further, the ordering $\frac{v_{j,1}}{d_{j,1}} > \frac{v_{j,2}}{d_{j,2}} > \cdots > \frac{v_{j,m_j}}{d_{j,m_j}}$ implies that the slopes on each segment $[X_n, X_{n+1}]$ decreases as n increases, which implies $V_j(\rho_j;\boldsymbol{z}_j)$ is concave.

Since $V_j(\rho_j) = \mathbb{E}\left[V_j(\rho_j;\boldsymbol{z}_j)\right]$, where the expectation is taken w.r.t. randomness in $\boldsymbol{z}_j$, and since the $\boldsymbol{z}_j$ is sampled from some discrete distribution $\boldsymbol{p}_j$ on finite support F_j, $V_j(\rho_j)$ is simply a weighted average over all $(V_j(\rho_j;\boldsymbol{z}_j))_{\boldsymbol{z}_j\in F_j}$ with weights in $\boldsymbol{p}_j$, so $V_j(\rho_j)$ is also continuous, piecewise linear, strictly increasing, and concave, and thus can be written as in Lemma 4.5 with parameters $\{(s_{j,n}, b_{j,n}, r_{j,n})\}_{n\in[S_j]}$ that only depend on the support F_j and distribution $\boldsymbol{p}_j$.

Finally, according to the definition of $\mathcal{L}_j(\rho_j, \boldsymbol{c}) = \mathbb{E}\left[\mathcal{L}_j(\rho_j, \boldsymbol{c};\boldsymbol{z}_j)\right]$ and $\mathcal{L}_j(\rho_j, \boldsymbol{c};\boldsymbol{z}_j) = (1+\lambda)V_j(\rho_j;\boldsymbol{z}_j) - (\lambda\gamma + \mu)\rho_j$ as defined in Eq. (6), we have

$$\mathcal{L}_j(\rho_j, \boldsymbol{c}) = (1+\lambda)V_j(\rho_j) - (\lambda\gamma + \mu)\rho_j \tag{26}$$

which implies $\mathcal{L}_j(\rho_j, \boldsymbol{c})$ is continuous, piecewise linear, and concave because $V_j(\rho_j)$ is continuous, piecewise linear, and concave as shown above. Combining Eq. (26) and the representation of $V_j(\rho_j)$ in Lemma (4.5), we have

$$\mathcal{L}_j(\rho_j, \boldsymbol{c}) = \sum_{n\in[S_j]} (\sigma_{j,n}(\boldsymbol{c})\rho_j + (1+\lambda)b_{j,n})\,\mathbb{I}\{r_{j,n-1} \leq \rho_j \leq r_{j,n}\}\,. \tag{27}$$

where the slope $\sigma_{j,n}(\boldsymbol{c}) = (1+\lambda)sj,n - (\mu+\gamma\lambda)$ decreases in n. Thus at the point $r_{j,n^*} = \max\{r_{j,n} : n = 0, 1\ldots, S_j\ , \sigma_{j,n}(\boldsymbol{c}) \geq 0\}$ in which the slope to the right turns negative for the first time, $\mathcal{L}_j(\rho_j, \boldsymbol{c})$ takes its maximum value $\max_{\rho_j\geq 0}\mathcal{L}_j(\rho_j, \boldsymbol{c})$, because to the left of r_{j,n^*}, namely the region $[0, r_{j,n^*}]$, $\mathcal{L}_j(\rho_j, \boldsymbol{c})$ strictly increases because slopes are positive; and to the right of r_{j,n^*}, namely the region $[r_{j,n^*}, \rho]$, $\mathcal{L}_j(\rho_j, \boldsymbol{c})$ strictly decreases because slopes are negative. □

C.4. Proof for Lemma 4.6

Recall the definition of the Lagrangian function $\mathcal{L}_j(\rho_j, \boldsymbol{c}; \boldsymbol{z}_j) = (1+\lambda)V_j(\rho_j; \boldsymbol{z}_j) - (\lambda\gamma+\mu)\rho_j$ in Eq.(6). Then, since $V_j(\rho_j; \boldsymbol{z}_j) \leq \overline{V}$, and $\lambda_t, \mu_t \in [0, C_F]$ for any period $t \in [T]$ and per-channel budget $\rho_j \in [0, \rho]$, we can conclude $-(1+\gamma)\,\rho C_F \ \leq\ \mathcal{L}_j(\rho_j, \lambda_t, \mu_t) \ \leq\ (1+C_F)\overline{V}$. □

C.5. Proof for Lemma 4.7

In the following, instead of bounding $\sum_{t\in[\tau_A]}\mathcal{L}_j(\rho^*_{j,t}, \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t)$, we bound $\sum_{t\in[T]}\mathcal{L}_j(\rho^*_{j,t}, \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t)$ where we consider the hypothetical scenario in which we ignore the termination criteria for the while loop in Algorithm 1, and continue to set per-channel budgets based on steps 4-6 in the algorithm until the end of period T. This is due to the fact that $\sum_{t\in[T]}\mathcal{L}_j(\rho^*_{j,t}, \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t) \geq \sum_{t\in[\tau_A]}\mathcal{L}_j(\rho^*_{j,t}, \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t)$.

We fix some channel $j \in [M]$ and omit the subscript j when the context is clear. Also, we first introduce some definitions that will be used throughout our proof. Fix some positive constant $\underline{\sigma} > 0$ whose value we choose later, and recall a_k denotes the kth arm in the discretized budget set $\mathcal{A}(\delta)$ as we defined in Eq. (8). Then we define the following

$$\begin{aligned}
&\Delta_k(\boldsymbol{c}) = \max_{\rho_j\in[0,\rho]}\mathcal{L}_j(\rho_j, \boldsymbol{c}) - \mathcal{L}_j(a_k, \boldsymbol{c})\\
&\mathcal{C}_n = \left\{\boldsymbol{c} \in \{\boldsymbol{c}_t\}_{t\in[T]} : r_{j,n} = \arg\max_{\rho_j\geq 0}\mathcal{L}_j(\rho_j, \boldsymbol{c})\right\} \text{ for } n = 0\ldots S_j\\
&\mathcal{C}(\underline{\sigma}) = \left\{\boldsymbol{c} \in \{\boldsymbol{c}_t\}_{t\in[T]} : \sigma_j^-(\boldsymbol{c}) > \underline{\sigma},\ |\sigma_j^+(\boldsymbol{c})| > \underline{\sigma}\right\} \text{ for } n = 0, \ldots, S_j\\
&m_k(\boldsymbol{c}) = \frac{8\log(T)}{\Delta_k^2(\boldsymbol{c})} \text{ for } \forall(k, \boldsymbol{c}) \text{ s.t. } \Delta_k(\boldsymbol{c}) > 0\,.
\end{aligned} \tag{28}$$

Here, the "adjacent slopes" $\sigma_j^-(\boldsymbol{c})$ and $\sigma_j^+(\boldsymbol{c})$, which are defined in Eq.(11), represent the slopes that are adjacent to the optimal budget $\arg\max_{\rho_j\in[0,\rho]}\mathcal{L}_j(\rho_j, \boldsymbol{c})$ for any context $\boldsymbol{c} = (\lambda, \mu)$. Further, S_j and $\{r_{j,n}\}_{j\in[S_j]}$ are defined in Lemma 4.5. Here we state in words the meanings of $\Delta_k(\boldsymbol{c})$, $\mathcal{C}(\underline{\sigma})$ and $\mathcal{C}_n$, respectively.

- $\Delta_k(\boldsymbol{c})$ denotes the loss in contextual bandit rewards when pulling arm a_k under context $\boldsymbol{c}$.
- $\mathcal{C}_n$ is the set including all context $\boldsymbol{c}_t$ under which the optimal per-channel budget $\arg\max_{\rho_j\geq 0}\mathcal{L}_j(\rho_j, \boldsymbol{c}_t)$ is taken at the nth "turning point" $r_{j,n}$ (see Lemma 4.5).
- $\mathcal{C}(\underline{\sigma})$ is the set of all contexts, in which the adjacent slopes to the optimal point w.r.t. the context $\boldsymbol{c}$, namely $\arg\max_{\rho_j\geq 0}\mathcal{L}_j(\rho_j, \boldsymbol{c})$, have magnitude greater than $\underline{\sigma}$, or in other words, the adjacent slopes are steep.

On a related note, for any context $\boldsymbol{c}$, we define the following "adjacent regions" that sandwich the optimal budget w.r.t.$\boldsymbol{c}$

$$\mathcal{U}_j^-(\boldsymbol{c}) = [r_{j,n-1}, r_{j,n}] \text{ and } \mathcal{U}_j^+(\boldsymbol{c}) = [r_{j,n}, r_{j,n+1}] \quad \text{if} \quad \boldsymbol{c} \in \mathcal{C}_n\,. \tag{29}$$

In other words, if $\boldsymbol{c} \in \mathcal{C}_n$, per the definition of $\mathcal{C}_n$ above, $\arg\max_{\rho_j\in[0,\rho]}\mathcal{L}_j(\rho_j, \boldsymbol{c})$ is located at the nth "turning point" $r_{j,n}$, then $\mathcal{U}_j^-(\boldsymbol{c})$ and $\mathcal{U}_j^-(\boldsymbol{c})$ are respectively the left and right regions surrounding $r_{j,n}$.

With the above definitions, we demonstrate how to bound the UCB-error. Define $N_{k,t} = \sum_{\tau \leq t-1} \mathbb{I}\{\rho_{j,\tau} = a_k\}$ to be the number of times arm k is pulled up to time t, then we can decompose the UCB error as followed

$$
\begin{aligned}
&\sum_{t>K} \mathcal{L}_j(\rho_j^*(t), \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t) = X_1 + X_2 + X_3 \quad \text{where} \\
&X_1 = \sum_{t>K: \boldsymbol{c}_t \notin \mathcal{C}(\underline{\sigma})} \sum_{k\in[K]} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} \leq m_k(\boldsymbol{c}_t)\} \\
&X_2 = \sum_{t>K: \boldsymbol{c}_t \in \mathcal{C}(\underline{\sigma})} \sum_{k\in[K]} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} \leq m_k(\boldsymbol{c}_t)\} \\
&X_3 = \sum_{k\in[K]} \sum_{t>K} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} > m_k(\boldsymbol{c}_t)\}\,.
\end{aligned}
\tag{30}
$$

In Section C.5.1, we show that $X_1 \leq \widetilde{\mathcal{O}}(\delta T + \underline{\sigma} T + \frac{1}{\delta})$; in Section C.5.2 we show that $X_2 \leq \widetilde{\mathcal{O}}(\delta T + \frac{1}{\delta\underline{\sigma}})$; in Section C.5.3 we show that $X_3 \leq \widetilde{\mathcal{O}}(\frac{1}{\delta T})$.

Remark C.1. In the following sections C.5.1, C.5.2 and C.5.3 where we bound X_1, X_2, and X_3, respectively, we assume the optimal per-channel $\rho_j^*(t) = \arg\max_{\rho_j \in [0,\rho]} \mathcal{L}_j(\rho_j, \boldsymbol{c}_t)$ lies in the arm set $\mathcal{A}(\delta)$ for all t. This is because otherwise, we can consider the following decomposition of the UCB error in period t as followed:

$$
\mathcal{L}_j(\rho_j^*(t), \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t) = \mathcal{L}_j(\rho_j^*(t), \boldsymbol{c}_t) - \mathcal{L}_j(a_t^*, \boldsymbol{c}_t) + \mathcal{L}_j(a_t^*, \boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t}, \boldsymbol{c}_t) \quad \text{where } a_t^* = \arg\max_{a_k \in \mathcal{A}(\delta)} \mathcal{L}_j(a_k, \boldsymbol{c}_t)
$$

The first term will yield an error in the order of $\mathcal{O}(\delta)$ due to the Lagrangian function being unimodal, piecewise linear liner, which implies $|a_t^* - \rho_j^*(t)| \leq \delta$ so that $\mathcal{L}_j(\rho_j^*(t), \boldsymbol{c}_t) - \mathcal{L}_j(a_t^*, \boldsymbol{c}_t) = \mathcal{O}(\delta)$. Hence, this "discretization error" will accumulate to a magnitude of $\mathcal{O}(\delta T)$ over T periods, which leads to an additional error that is already accounted for in the statement of the lemma.

C.5.1. Bounding X_1.

Our strategy to bound X_1 consists of 4 steps, namely bounding the loss of arm a_k at each context $\boldsymbol{c} \notin \mathcal{C}(\underline{\sigma})$ when $a_k \in \mathcal{U}_j^-(\boldsymbol{c})$ lies on the left adjacent region of the optimal budget; $a_k < \min \mathcal{U}_j^-(\boldsymbol{c})$ lies to the left of the left adjacent region; $a_k \in \mathcal{U}_j^+(\boldsymbol{c})$ lies on the right adjacent region of the optimal budget; and $a_k > \max \mathcal{U}_j^+(\boldsymbol{c})$ lies to the right of the right adjacent region. Here we recall the adjacent regions are defined in Eq.(29).

Step 1: $a_k \in \mathcal{U}_j^-(\boldsymbol{c}_t)$. For arm k such that $a_k \in \mathcal{U}_j^-(\boldsymbol{c}_t)$, recall Lemma 4.5 that $\mathcal{L}_j(a, \boldsymbol{c}_t)$ is linear in a for $a \in \mathcal{U}_j^-(\boldsymbol{c}_t)$, so $\Delta_k(\boldsymbol{c}_t) = \sigma_j^-(\boldsymbol{c}_t) \cdot (\rho_j^*(t) - a_k) \leq \underline{\sigma}\rho$ where we used the condition that $\boldsymbol{c}_t \notin \mathcal{C}(\underline{\sigma})$ so the adjacent slopes have magnitude at most $\underline{\sigma}$, and $\rho_j^*(t) \leq \rho$. Thus, summing over all such k we get

$$
\begin{aligned}
&\sum_{t>K: \boldsymbol{c}_t \notin \mathcal{C}(\underline{\sigma})} \sum_{k\in[K]: a_k \in \mathcal{U}_j^-(\boldsymbol{c}_t)} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} \leq m_k(\boldsymbol{c}_t)\} \\
\leq &\sum_{t>K: \boldsymbol{c}_t \notin \mathcal{C}(\underline{\sigma})} \sum_{k\in[K]: a_k \in \mathcal{U}_j^-(\boldsymbol{c}_t)} \underline{\sigma}\rho \cdot \mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} \leq m_k(\boldsymbol{c}_t)\} \;\leq\; \underline{\sigma}\rho T = \mathcal{O}(\underline{\sigma} T)\,.
\end{aligned}
\tag{31}
$$

Step 2: $a_k < \min \mathcal{U}_j^-(\boldsymbol{c}_t)$. For arm k such that $a_k < \min \mathcal{U}_j^-(\boldsymbol{c}_t)$, we further split contexts into groups $\mathcal{C}_n$ for $n = 0 \ldots S_j$ (defined in Eq. (28)) based on whether the corresponding optimal budget w.r.t. the Lagrangian at the context is taken at the nth "turning point" (see Figure 2 of illustration). Then, for each context group n by defining $k' := \max\{k : a_k < r_{j,n-1}\}$

to be the arm closest to and less than $r_{j,n-1}$, we have

$$\begin{aligned}
&\sum_{t>K:\boldsymbol{c}_t\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\ \sum_{k\in[K]:a_k<\min\mathcal{U}_j^-(\boldsymbol{c}_t)} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t}=a_k, N_{k,t}\le m_k(\boldsymbol{c}_t)\}\\
\overset{(i)}{=}\ &\sum_{t>K:\boldsymbol{c}_t\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\ \sum_{k\in[K]:a_k<r_{j,n-1}} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t}=a_k, N_{k,t}\le m_k(\boldsymbol{c}_t)\}\\
=\ &\sum_{t>K}\sum_{\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\ \sum_{k\in[K]:a_k<r_{j,n-1}} \Delta_k(\boldsymbol{c})\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c},\rho_{j,t}=a_k, N_{k,t}\le m_k(\boldsymbol{c})\}\\
\overset{(ii)}{\le}\ &\sum_{t>K}\sum_{\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\left(\Delta_{k'}(\boldsymbol{c})\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c}\}+\sum_{k\in[K]:a_k<r_{j,n-1}-\delta}\Delta_k(\boldsymbol{c})\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c},\rho_{j,t}=a_k, N_{k,t}\le m_k(\boldsymbol{c})\}\right)\\
\overset{(iii)}{\le}\ &\left((1+C_F)s_{j,n-1}\delta+\rho\underline{\sigma}\right)T+\sum_{k\in[K]:a_k<r_{j,n-1}-\delta}\ \sum_{\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c})
\end{aligned}\tag{32}$$

where in the final equality we defined $Y_k(\boldsymbol{c})=\sum_{t>K}\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c},\rho_{j,t}=a_k,N_{k,t}\le m_k(\boldsymbol{c})\}$. In (i) we used the fact that the left end of the left adjacent region, i.e. $\min\mathcal{U}_j^-(\boldsymbol{c}_t)$ is exactly $r_{j,n-1}$ because for context $\boldsymbol{c}_t\in\mathcal{C}_n$ the optimal budget $\arg\max_{\rho_j\in[0,\rho]}\mathcal{L}_j(\rho_j,\boldsymbol{c}_t)$ is at the nth turning point; in (ii) we used the definition $k':=\max\{k:a_k<r_{j,n-1}\}$ where we recall arms are indexed such that $a_1<a_2<\cdots<a_K$. Note that in (ii) we separate out the arm $a_{k'}$ because its distance to the optimal per-channel may be less than δ since it is the closest arm, and thus we ensure all other arms indexed by $k\in[K]:a_k<r_{j,n-1}-\delta$, are at least δ away from the optimal per-channel budget; (iii) follows from the fact that under a context $\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})$, we have $\arg\max_{\rho_j\in[0,\rho]}\mathcal{L}_j(\rho_j,\boldsymbol{c})=r_{j,n}$ so

$$\begin{aligned}
\Delta_{k'}(\boldsymbol{c})\ &=\ \mathcal{L}_j(r_{j,n},\boldsymbol{c})-\mathcal{L}_j(r_{j,n-1},\boldsymbol{c})+\mathcal{L}_j(r_{j,n-1},\boldsymbol{c})-\mathcal{L}_j(a_{k'},\boldsymbol{c})\\
&=\ \sigma_j^-(\boldsymbol{c})(r_{j,n}-r_{j,n-1})+\sigma_{j,n-1}(\boldsymbol{c})(r_{j,n-1}-a_{k'})\\
&\overset{(iv)}{\le}\ \underline{\sigma}\rho+\sigma_{j,n-1}(\boldsymbol{c})\delta\\
&\overset{(v)}{\le}\ \underline{\sigma}\rho+(1+C_F)s_{j,n-1}\delta\,,
\end{aligned}$$

where in (iv) we used $\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})$ implies $\sigma_j^-(\boldsymbol{c})\le\underline{\sigma}$, as well as all $r_{j,n}\le\rho$ for any n and the fact that k' lies on the line segment between points $r_{j,n-2}$ and $r_{j,n-1}$ since $\delta<\min_{n'\in[S_j]}r_{j,n'}-r_{j,n'-1}$; in (v) we recall $\sigma_{j,n-1}(\boldsymbol{c})=(1+\lambda)s_{j,n-1}-(\mu+\gamma\lambda)\le(1+C_F)s_{j,n-1}$ where C_F is defined in Lemma 4.6.

We now bound $\sum_{\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c})$ in Eq. (32). It is easy to see the following inequality for any sequence of context $\boldsymbol{c}_{(1)},\dots,\boldsymbol{c}_{(\ell)}\in\{\boldsymbol{c}_t\}_{t\in[T]}$ (This is a slight generalization of an inequality result shown in (Balseiro et al., 2019a)):

$$Y_k(\boldsymbol{c}_{(1)})+\cdots+Y_k(\boldsymbol{c}_{(\ell)})\le\max_{\ell'=1\dots\ell}m_k(\boldsymbol{c}_{(\ell')})\,.\tag{33}$$

This is because

$$\begin{aligned}
\sum_{\ell'\in[\ell]}Y_k(\boldsymbol{c}_{(\ell')})\ &=\ \sum_{t>K}\sum_{\ell'\in[\ell]}\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c}_{(\ell')},\rho_{j,t}=a_k,N_{k,t}\le m_k(\boldsymbol{c}_{(\ell')})\}\\
&\le\ \sum_{t>K}\sum_{\ell'\in[\ell]}\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c}_{(\ell')},\rho_{j,t}=a_k,N_{k,t}\le\max_{\ell'=1\dots\ell}m_k(\boldsymbol{c}_{(\ell')})\}\\
&=\ \sum_{t>K}\mathbb{I}\{\boldsymbol{c}_t\in\{\boldsymbol{c}_{(\ell')}\}_{\ell'\in[\ell]},\rho_{j,t}=a_k,N_{k,t}\le\max_{\ell'=1\dots\ell}m_k(\boldsymbol{c}_{(\ell')})\}\\
&\le\ \max_{\ell'=1\dots\ell}m_k(\boldsymbol{c}_{(\ell')})\,.
\end{aligned}$$

For simplicity denote $L=|\mathcal{C}_n/\mathcal{C}(\underline{\sigma})|$, and order contexts in $\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})$ as $\{\boldsymbol{c}_{(\ell)}\}_{\ell\in[L]}$ s.t. $\Delta_k(\boldsymbol{c}_{(1)})>\Delta_k(\boldsymbol{c}_{(2)})>\cdots>\Delta_k(\boldsymbol{c}_{(L)})$, or equivalently $m_k(\boldsymbol{c}_{(1)})<m_k(\boldsymbol{c}_{(2)})<\cdots<m_k(\boldsymbol{c}_{(L)})$ according to Eq.(28). Then multiplying Eq. (33) by by

$\Delta_k(\boldsymbol{c}_{(\ell)}) - \Delta_k(\boldsymbol{c}_{(\ell+1)})$ (which is strictly positive due to the ordering of contexts), and summing $\ell = 1 \dots L$ we get

$$\begin{aligned}
\sum_{\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})} \Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c}) = \sum_{\ell\in[L]} \Delta_k(\boldsymbol{c}_{(\ell)})Y_k(\boldsymbol{c}_{(\ell)}) \ &\leq \ \sum_{\ell\in[L]} m_k(\boldsymbol{c}_{(\ell)}) \left(\Delta_k(\boldsymbol{c}_{(\ell)}) - \Delta_k(\boldsymbol{c}_{(\ell+1)})\right) \\
&\overset{(i)}{=} \ 8\log(T) \sum_{\ell\in[L-1]} \frac{\Delta_k(\boldsymbol{c}_{(\ell)}) - \Delta_k(\boldsymbol{c}_{(\ell+1)})}{\Delta_k^2(\boldsymbol{c}_{(\ell)})} \ \overset{(ii)}{\leq} \ 8\log(T)\int_{\Delta_k(\boldsymbol{c}_{(L)})}^{\infty} \frac{dz}{z^2} \\
&= \ \frac{8\log(T)}{\Delta_k(\boldsymbol{c}_{(L)})} \ \overset{(iii)}{=} \ \frac{8\log(T)}{\min_{\boldsymbol{c}\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})}\,.
\end{aligned} \tag{34}$$

Here (i) follows from the definition of $m_k(\boldsymbol{c})$ in Eq. (28) where $m_k(\boldsymbol{c}) = \frac{8\log(T)}{\Delta_k^2(\boldsymbol{c})}$; both (ii) and (iii) follow from the ordering of contexts so that $\Delta_k(\boldsymbol{c}_{(1)}) > \Delta_k(\boldsymbol{c}_{(2)}) > \dots > \Delta_k(\boldsymbol{c}_{(L)})$. Note that for any $\boldsymbol{c} \in \mathcal{C}_n/\mathcal{C}(\underline{\sigma})$ and arm k such that $a_k < r_{j,n-1}$, we have

$$\begin{aligned}
\Delta_k(\boldsymbol{c}) \ &= \ \mathcal{L}_j(r_{j,n},\boldsymbol{c}) - \mathcal{L}_j(r_{j,n-1},\boldsymbol{c}) + \mathcal{L}_j(r_{j,n-1},\boldsymbol{c}) - \mathcal{L}_j(a_k,\boldsymbol{c}) \\
&> \ \mathcal{L}_j(r_{j,n-1},\boldsymbol{c}) - \mathcal{L}_j(a_k,\boldsymbol{c}) \\
&\overset{(i)}{\geq} \ \sigma_{j,n-1}(\boldsymbol{c})(r_{j,n-1} - a_k) \\
&\overset{(ii)}{\geq} \ \left(\sigma_{j,n-1}(\boldsymbol{c}) - \sigma_{j,n}(\boldsymbol{c})\right)\left(r_{j,n-1} - a_k\right) \\
&\overset{(iii)}{=} \ (1+\lambda)\left(s_{j,n-1} - s_{j,n}\right)\left(r_{j,n-1} - a_k\right) \\
&> \ \left(s_{j,n-1} - s_{j,n}\right)\left(r_{j,n-1} - a_k\right),
\end{aligned} \tag{35}$$

where in (i) we recall the slope $\sigma_{j,n-1}(\boldsymbol{c})$ is defined in Lemma 4.5 and further (i) follows from concavity of $\mathcal{L}_j(\rho j, \boldsymbol{c})$ in the first argument ρ_j; in (ii) we used the fact that $\sigma_{j,n}(\boldsymbol{c}) \geq 0$ since the optimal budget $\arg\max_{\rho_j\in[0,\rho]} \mathcal{L}_j(\rho_j,\boldsymbol{c})$ is taken at the nth turning point, and is the largest turning point whose left slope is non-negative from Lemma 4.5; (iii) follows from the definition $\sigma_{j,n'}(\boldsymbol{c}) = (1+\lambda)s_{j,n'} - (\mu + \gamma\lambda)$ for any n'.

Finally combining Eqs. (32), (34) and (35), and summing over $n = 1 \dots S_j$ we get

$$\begin{aligned}
&\sum_{t>K:\boldsymbol{c}_t\notin\mathcal{C}(\underline{\sigma})} \ \sum_{k\in[K]:a_k<\min\mathcal{U}_j^-(\boldsymbol{c}_t)} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} \leq m_k(\boldsymbol{c}_t)\} \\
&= \sum_{n\in[S_j]} \ \sum_{t>K:\boldsymbol{c}_t\in\mathcal{C}_n/\mathcal{C}(\underline{\sigma})} \ \sum_{k\in[K]:a_k<\min\mathcal{U}_j^-(\boldsymbol{c}_t)} \Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t} = a_k, N_{k,t} \leq m_k(\boldsymbol{c}_t)\} \\
&\leq \sum_{n\in[S_j]} \left((1+C_F)s_{j,n-1}\delta + \rho\underline{\sigma}\right)T + \sum_{n\in[S_j]} \ \sum_{k\in[K]:a_k<r_{j,n-1}-\delta} \frac{8\log(T)}{\left(s_{j,n-1} - s_{j,n}\right)\left(r_{j,n-1} - a_k\right)} \\
&\overset{(i)}{\leq} \sum_{n\in[S_j]} \left((1+C_F)s_{j,n-1}\delta + \rho\underline{\sigma}\right)T + \sum_{n\in[S_j]} \ \sum_{\ell=1}^{K} \frac{8\log(T)}{\left(s_{j,n-1} - s_{j,n}\right)\ell\delta} \\
&\leq \sum_{n\in[S_j]} \left((1+C_F)s_{j,n-1}\delta + \rho\underline{\sigma}\right)T + \frac{8\log(T)\log(K)}{\delta} \sum_{n\in[S_j]} \frac{1}{\left(s_{j,n-1} - s_{j,n}\right)} \\
&= \widetilde{\mathcal{O}}(\delta T + \underline{\sigma}T + \frac{1}{\delta})\,.
\end{aligned} \tag{36}$$

Note that (i) follows because for all $a_k < r_{j,n-1} - \delta$, the a_k's distances from $r_{j,n-1}$ are at least $\delta, 3\delta, 2\delta \dots$. In the last equation, we hide all logarithmic factors using the notation $\widetilde{\mathcal{O}}$, and note that the constants C_F, $(s_{j,n})_{n\in S_j}$, S_j are all absolute constants that depend only on the support F_j and corresponding sampling distribution $\boldsymbol{p}_j$ for value-cost pairs; see definitions of these absolute constants in Lemmas 4.5 and 4.6.

Step 3 and 4: $a_k \in \mathcal{U}_j^+(\boldsymbol{c}_t)$ or $a_k > \max\ \mathcal{U}_j^+(\boldsymbol{c}_t)$. The cases where arm $a_k \in \mathcal{U}_j^+(\boldsymbol{c}_t)$ and $a_k > \max\mathcal{U}_j^+(\boldsymbol{c}_t)$ are symmetric to $a_k \in \mathcal{U}_j^-(\boldsymbol{c}_t)$ and $a_k < \min\mathcal{U}_j^+(\boldsymbol{c}_t)$, respectively, and we omit from this paper.

Therefore, combining Eqs. (31) and (36) we can conclude

$$X_1 \leq \widetilde{\mathcal{O}}(\delta T + \underline{\sigma} T + \frac{1}{\delta})\,. \tag{37}$$

C.5.2. BOUNDING X_2.

We first rewrite X_2 as followed

$$\begin{aligned}
X_2 &= \sum_{t>K:\boldsymbol{c}_t\in\mathcal{C}(\underline{\sigma})}\sum_{k\in[K]}\Delta_k(\boldsymbol{c}_t)\mathbb{I}\{\rho_{j,t}=a_k, N_{k,t}\leq m_k(\boldsymbol{c}_t)\}\\
&= \sum_{t>K}\sum_{n\in[S_j]}\sum_{k\in[K]}\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c},\rho_{j,t}=a_k, N_{k,t}\leq m_k(\boldsymbol{c})\}\\
&\overset{(i)}{=} \sum_{n\in[S_j]}\sum_{k\in[K]}\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c})\\
&\overset{(ii)}{=} \sum_{n\in[S_j]}\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\sum_{k\in\{k_n^-,k_n^+\}}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c}) + \sum_{n\in[S_j]}\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\sum_{k\in[K]/\{k_n^-,k_n^+\}}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c})\\
&\overset{(iii)}{\leq} T\delta(1+C_F)\sum_{n\in[S_j]}(s_{j,n}+s_{j,n+1}) + \sum_{n\in[S_j]}\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\sum_{k\in[K]/\{k_n^-,k_n^+\}}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c})\,.
\end{aligned}\tag{38}$$

where in (i) we define $Y_k(\boldsymbol{c}) = \sum_{t>K}\mathbb{I}\{\boldsymbol{c}_t=\boldsymbol{c},\rho_{j,t}=a_k,N_{k,t}\leq m_k(\boldsymbol{c})\}$; in (ii) we separate out two arms k_n^- and k_n^+ defined as followed: recall for context $\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})$, the optimal budget $\arg\max_{\rho_j\in[0,\rho]}\mathcal{L}_j(\rho_j,\boldsymbol{c}) = r_{j,n}$ is taken at the nth turning point per the definition of $\mathcal{C}_n$ in Eq. (28), and thereby we defined $k_n^- := \max\{k\in[K]: a_k < r_{j,n}\}$ to be the arm closest to and no greater than $r_{j,n}$, whereas $k_n^+ := \min\{k\in[K]: a_k > r_{j,n}\}$ to be the arm closest to and no less than $r_{j,n}$; in (iii), for small enough $\delta < \min_{n'\in[S_j]} r_{j,n'} - r_{j,n'-1}$, we know that k_n^- lies on the line segment between $r_{j,n-1}$ and $r_{j,n}$, so $\Delta_{k_n^-}(\boldsymbol{c}) = \sigma_j^-(\boldsymbol{c})(r_{j,n}-a_{k_n^-}) \leq \sigma_j^-(\boldsymbol{c})\delta \leq (1+C_F)s_{j,n-1}\delta$, where in the final inequality follows from the definition of $\sigma_j^-(\boldsymbol{c}) = \sigma_{j,n}(\boldsymbol{c}) = (1+\lambda)s_{j,n} - (\mu+\gamma\lambda) \leq (1+\lambda)s_{j,n} \leq (1+C_F)s_{j,n}$ where C_F is defined in Eq. (4.6). A similar bound holds for $\Delta_{k_n^+}(\boldsymbol{c})$.

Then, following the same logic as Eqs. (33), (34), (35) in Section C.5.1 where we bound X_1, we can bound $\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c})$ as followed for any arm $k\in[K]/\{k_n^-,k_n^+\}$, i.e. arms who are at least δ away from the optimal per-channel budget w.r.t. $\boldsymbol{c}$:

$$\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c}) \leq \frac{8\log(T)}{\min_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})}\,. \tag{39}$$

Now, the set $k\in[K]/\{k_n^-,k_n^+\}$ in Eq. (38) can be further split into two subsets, namely $\{k\in[K]: a_k < r_{j,n}-\delta\}$ and $\{k\in[K]: a_k > r_{j,n}+\delta\}$ due to the definitions $k_n^- := \max\{k\in[K]: a_k < r_{j,n}\}$ and $k_n^+ := \min\{k\in[K]: a_k > r_{j,n}\}$. Therefore, for any k s.t. $a_k < r_{j,n}-\delta$ and any $\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})$,

$$\Delta_k(\boldsymbol{c}) = \mathcal{L}_j(r_{j,n},\boldsymbol{c}) - \mathcal{L}_j(a_k,\boldsymbol{c}) \geq \sigma_j^-(\boldsymbol{c})(r_{j,n}-a_k) \geq \underline{\sigma}(r_{j,n}-a_k)\,,$$

where the final inequality follows from the definition of $\mathcal{C}(\underline{\sigma})$ in Eq. (28) such that $\sigma_j^-(\boldsymbol{c}) \geq \underline{(}\sigma)$ for $\boldsymbol{c}\in\mathcal{C}(\underline{\sigma})$. Hence combining this with Eq. (39) we have

$$\sum_{k\in[K]:a_k<r_{j,n}-\delta}\sum_{\boldsymbol{c}\in\mathcal{C}_n\cap\mathcal{C}(\underline{\sigma})}\Delta_k(\boldsymbol{c})Y_k(\boldsymbol{c}) \leq \sum_{k\in[K]:a_k<r_{j,n}-\delta}\frac{8\log(T)}{\underline{\sigma}(r_{j,n}-a_k)} \overset{(i)}{\leq} \sum_{\ell=1}^{K}\frac{8\log(T)}{\underline{\sigma}\ell\delta} \leq \frac{8\log(T)\log(K)}{\underline{\sigma}\delta}\,, \tag{40}$$

where (i) follows because for all $a_k < r_{j,n}-\delta$, the a_k's distances from $r_{j,n-1}$ are at least $\delta, 3\delta, 2\delta \ldots$. Symmetrically, we can show an identical bound for the set $\{k\in[K]: a_k > r_{j,n}+\delta\}$. Hence, combining Eqs. (38) and (40) we can conclude

$$X_2 \leq \widetilde{\mathcal{O}}\left(\delta T + \frac{1}{\delta\underline{\sigma}}\right)\,. \tag{41}$$

Here, similar to our bound in Eq. (36) for bounding X_1, we hide all logarithmic factors using the notation $\widetilde{\mathcal{O}}$, and note that the constants C_F, $(s_{j,n})_{n\in S_j}$, S_j are all absolute constants that depend only on the support F_j and corresponding sampling distribution $\boldsymbol{p}_j$ for value-cost pairs; see definitions of these absolute constants in Lemma 4.5 and 4.6.

C.5.3. Bounding X_3.

We first define

$$\bar{\mathcal{L}} = (1+\gamma)\,\rho C_F + (1+C_F)\bar{V} \tag{42}$$

where C_F is specified in Lemma 4.6. Recalling the definition $\Delta_k(\boldsymbol{c}) = \max_{\rho_j\in[0,\rho]} \mathcal{L}_j(\rho_j,\boldsymbol{c}) - \mathcal{L}_j(a_k,\boldsymbol{c})$ in Eq. (28), and $-(1+\gamma)\rho C_F \le \mathcal{L}_j(\rho_j,\boldsymbol{c}) \le (1+C_F)\bar{V}$ for any $\rho_j\in[0,\rho]$ and context $\boldsymbol{c}$ (see Lemma 4.6), it is easy to see

$$\Delta_k(\boldsymbol{c}) \le \bar{\mathcal{L}} \quad \forall k\in[K], \forall \boldsymbol{c}. \tag{43}$$

Then we bound X_3 as followed

$$\begin{aligned} X_3 \;&=\; \sum_{k\in[K]}\sum_{t>K} \mathbb{E}\left[\Delta_k(\boldsymbol{c})\mathbb{I}\{\rho_{j,t}=a_k, N_{k,t}>m_k(\boldsymbol{c})\}\right] \\ &\overset{(i)}{\le}\; \bar{\mathcal{L}}\cdot \sum_{k\in[K]}\sum_{t>K} \mathbb{P}\left(\rho_{j,t}=a_k, N_{k,t}>m_k(\boldsymbol{c}_t)\right) \\ &\overset{(ii)}{\le}\; \bar{\mathcal{L}}\cdot \sum_{k\in[K]}\sum_{t>K} \mathbb{P}\Big(\hat{V}_{j,t}(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k + \texttt{UCB}_{j,t}(a_k) \ge \hat{V}_{j,t}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t) + \texttt{UCB}_{j,t}(\rho_j^*(t)), \\ &\qquad\qquad N_{k,t}>m_k(\boldsymbol{c}_t)\Big)\,, \end{aligned} \tag{44}$$

where (i) follows from Eq. (43); in (ii), recall that we choose arm $\rho_{j,t}=a_k$ because the estimated UCB rewards of arm a_k is greater than that of any other arm including $\rho_j^*(t)$ according to the UCB-SGD (Algorithm 1), or mathematically, $\hat{V}_{j,t}(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k + \texttt{UCB}_{j,t}(a_k) \ge \hat{V}_{j,t}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t) + \texttt{UCB}_{j,t}(\rho_j^*(t))$. Here we also used the fact that $\rho_j^*(t)$ lies in the arm set $\mathcal{A}(\delta)$ for all t (see Remark C.1).

Now let $\hat{R}_n(a_k)$ denote the average conversion of arm k over its first n pulls, i.e.

$$\hat{R}_n(a_k) = \hat{V}_{j,\tau}(a_k) \text{ for } \tau = \min\{t\in[T] : N_{k,t}=n\} \tag{45}$$

where we recall $\hat{V}_{j,\tau}(a_k)$ is the estimated conversion for arm a_k in channel j during period τ as defined in Algorithm 1. In other words, τ is the period during which arm a_k is pulled for the nth time so $\hat{R}_n(a_k) = \hat{V}_{j,\tau}(a_k)$.

Hence, we continue with Eq. (44) as followed:

$$\begin{aligned} &\mathbb{P}\Big(\hat{V}_{j,t}(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k + \texttt{UCB}_{j,t}(a_k) \ge \hat{V}_{j,t}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t) + \texttt{UCB}_{j,t}(\rho_j^*(t)),\; N_{k,t}>m_k(\boldsymbol{c}_t)\Big) \\ \le\;& \mathbb{P}\Big(\max_{n:m_k(\boldsymbol{c}_t)<n\le t}\Big\{\hat{R}_n(a_k) + \texttt{UCB}_n(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k\Big\} \\ &\qquad \ge \min_{n':1\le n'\le t}\Big\{\hat{R}_{n'}(\rho_j^*(t)) + \texttt{UCB}_{n'}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t)\Big\}\Big) \\ \le\;& \sum_{n=\lceil m_k(\boldsymbol{c}_t)\rceil+1}^{t}\sum_{n'=1}^{t} \mathbb{P}\left(\hat{R}_n(a_k) + \texttt{UCB}_n(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k > \hat{R}_{n'}(\rho_j^*(t)) + \texttt{UCB}_{n'}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t)\right) \end{aligned} \tag{46}$$

Now, when the event $\Big\{\hat{R}_n(a_k) + \texttt{UCB}_n(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k > \hat{R}_{n'}(\rho_j^*(t)) + \texttt{UCB}_{n'}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t)\Big\}$ occurs, it is easy to see that one of the following events must also occur:

$$\begin{aligned} &\mathcal{G}_{1,n} = \left\{\bar{R}_n(a_k) \ge V(a_k) + \texttt{UCB}_n(a_k)\right\} \quad \text{for } n \text{ s.t. } m_k(\boldsymbol{c}_t) < n \le t \\ &\mathcal{G}_{2,n'} = \left\{\bar{R}_{n'}(\rho_j^*(t)) \le V(\rho_j^*(t)) - \texttt{UCB}_n(\rho_j^*(t))\right\} \quad \text{for } n' \text{ s.t. } 1\le n'\le t \\ &\mathcal{G}_3 = \left\{V_j(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t) \;<\; V_j(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k + 2\cdot\texttt{UCB}_n(a_k)\right\} \end{aligned} \tag{47}$$

Note that for $n > m_k(\boldsymbol{c}_t)$, we have $\texttt{UCB}_n(a_k) = \sqrt{\frac{2\log(T)}{n}} < \sqrt{\frac{2\log(T)}{m_k(\boldsymbol{c}_t)}} = \frac{\Delta_k(\boldsymbol{c}_t)}{2}$ since we defined $m_k(\boldsymbol{c}) = \frac{8\log(T)}{\Delta_k^2(\boldsymbol{c})}$ in Eq. (28). Therefore

$$V_j(a_k) - \frac{\lambda_t\gamma + \mu_t}{1+\lambda_t}a_k + 2\cdot\texttt{UCB}_n(a_k) < \underbrace{V_j(a_k) - \frac{\lambda_t\gamma + \mu_t}{1+\lambda_t}a_k}_{=\mathcal{L}(a_k,\boldsymbol{c}_t)} + \Delta_k(\boldsymbol{c}_t) \overset{(i)}{=} \underbrace{V_j(\rho_j^*(t)) - \frac{\lambda_t\gamma + \mu_t}{1+\lambda_t}\rho_j^*(t)}_{=\mathcal{L}(\rho_j^*(t),\boldsymbol{c}_t)=\max_{a\in\mathcal{A}(\delta)}\mathcal{L}(a,\boldsymbol{c}_t)}$$

where (i) follows from the definition of $\Delta_k(\boldsymbol{c}) = \max_{a\in\mathcal{A}(\delta)}\mathcal{L}(a,\boldsymbol{c}) - \mathcal{L}(a_k,\boldsymbol{c})$ in Eq. (28) for any context $\boldsymbol{c}$. This implies that event $\mathcal{G}_3$ in Eq. (47) cannot hold for $n > m_k(\boldsymbol{c}_t)$. Therefore

$$\mathbb{P}\left(\hat{R}_n(a_k) + \texttt{UCB}_n(a_k) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}a_k > \hat{R}_{n'}(\rho_j^*(t)) + \texttt{UCB}_{n'}(\rho_j^*(t)) - \frac{\lambda_t\gamma+\mu_t}{1+\lambda_t}\rho_j^*(t)\right) \leq \mathbb{P}\left(\mathcal{G}_{1,n}\cup\mathcal{G}_{2,n'}\right). \quad (48)$$

From the standard UCB analysis and the Azuma Hoeffding's inequality, we have $\mathbb{P}(\mathcal{G}_{1,n}) \leq \frac{\bar{V}}{T^4}$ and $\mathbb{P}(\mathcal{G}_{2,n'}) \leq \frac{\bar{V}}{T^4}$. Hence combining Eqs. (44) (46), (48) we can conclude

$$\begin{aligned} X_3 &\leq \sum_{k\in[K]}\sum_{t>K}\sum_{n=\lceil m_k(\boldsymbol{c}_t)\rceil+1}^{t}\sum_{n'=1}^{t}\left(\mathbb{P}\left(\mathcal{G}_{1,n}\right) + \mathbb{P}\left(\mathcal{G}_{2,n'}\right)\right)\\ &\leq \sum_{k\in[K]}\sum_{t>K}\sum_{n=\lceil m_k(\boldsymbol{c}_t)\rceil+1}^{t}\sum_{n'=1}^{t}\frac{2\bar{V}}{T^4}\\ &\leq \frac{2K\bar{V}}{T} = \mathcal{O}\left(\frac{1}{\delta T}\right). \end{aligned} \quad (49)$$

□

C.6. Proof for Theorem 4.8

Starting from Proposition 4.3, we get

$$\begin{aligned} &T\cdot\text{GL-OPT} - \mathbb{E}\left[\sum_{t\in[T]}\sum_{j\in[M]}V_j(\rho_{j,t})\right]\\ &\leq M\overline{V}(T-\tau_A) + \sum_{j\in[M]}\mathbb{E}\left[\sum_{t\in[\tau_A]}\mathcal{L}_j(\rho_j^*(t),\boldsymbol{c}_t) - \mathcal{L}_j(\rho_{j,t},\boldsymbol{c}_t)\right] + \mathbb{E}\Big[\sum_{t\in[\tau_A]}(\lambda_t g_{1,t} + \mu_t g_{2,t})\Big]\\ &\overset{(i)}{\leq} M\overline{V}(T-\tau_A) + \mathcal{O}\left(\underline{\sigma}T + \delta T + \frac{1}{\underline{\sigma}\delta}\right) + \mathcal{O}\left(\eta T + \frac{1}{\eta}\right) \end{aligned} \quad (50)$$

where in (i) we applied Lemma 4.7 and 4.4. Taking $\eta = 1/\sqrt{T}$, $\delta = \underline{\sigma} = T^{-1/3}$ (i.e. $K = \mathcal{O}(T^{1/3})$ yields $T\cdot\text{GL-OPT} - \mathbb{E}\left[\sum_{t\in[T]}\sum_{j\in[M]}V_j(\rho_{j,t})\right] \leq \mathcal{O}(T^{2/3})$. According to Lemma 4.5, $V_j(\rho_j)$ is concave for all $j\in[M]$, so

$$\begin{aligned} \mathcal{O}(T^{-1/3}) &\geq \text{GL-OPT} - \frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left[\sum_{j\in[M]}V_j(\rho_{j,t})\right]\\ &\geq \text{GL-OPT} - \mathbb{E}\left[\sum_{j\in[M]}V_j\left(\frac{1}{T}\sum_{t\in[T]}\rho_{j,t}\right)\right]\\ &\geq \text{GL-OPT} - \mathbb{E}\left[\sum_{j\in[M]}V_j(\overline{\rho}_{j,T})\right] \end{aligned} \quad (51)$$

where in the final equality we used the definition $\bar{\boldsymbol{\rho}}_T$ as defined in Algorithm 1.

Regarding ROI constraint satisfaction, consider

$$\begin{aligned}
0 &\overset{(i)}{\leq} \frac{1}{T}\sum_{t\in[T]}\mathbb{E}\left[g_{1,t}\right] \\
&= \frac{1}{T}\sum_{t\in[T]}\sum_{j\in[M]}\mathbb{E}\left[V_j(\rho_{j,t};\boldsymbol{z}_{j,t}) - \gamma\rho_{j,t}\right] \\
&= \frac{1}{T}\sum_{t\in[T]}\sum_{j\in[M]}\mathbb{E}\left[V_j(\rho_{j,t}) - \gamma\rho_{j,t}\right] \\
&\overset{(ii)}{\leq} \sum_{j\in[M]}\mathbb{E}\left[V_j\left(\frac{1}{T}\sum_{t\in[T]}\rho_{j,t}\right) - \gamma\cdot\frac{1}{T}\sum_{t\in[T]}\rho_{j,t}\right] \\
&= \sum_{j\in[M]}\mathbb{E}[V_j\left(\overline{\rho}_{j,T}\right) - \gamma\overline{\rho}_{j,T}]\,.
\end{aligned} \tag{52}$$

where (i) follows from Lemma C.3; in (ii) we again applied concavity of $V_j(\rho_j)$. We omit the analysis for the budget constraint as it is similar to the above. ☐

C.7. Additional Results for Section 4

Proposition C.2. *Assume Assumption 4.2 holds, and recall $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j) \in F_j$ is any realization of values and costs for channel $j \in [M]$. Then, for any channel $j \in [M]$, we have $\min_{\boldsymbol{z}_j\in F_j}\frac{v_{j,1}}{d_{j,1}} > \gamma$, where we recall the ordering $\frac{v_{j,1}}{d_{j,1}} > \frac{v_{j,2}}{d_{j,2}} > \cdots > \frac{v_{j,m_j}}{d_{j,m_j}}$ for any element $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j) \in F_j$ (see Section 4). Further, there exists some $\widetilde{\rho} \in (0, \rho)$ s.t. for any per-channel budget $\rho_j \leq \widetilde{\rho}$, we have $V_j(\rho_j; \boldsymbol{z}_j) = \frac{v_{j,1}}{d_{j,1}}\rho_j > \gamma\rho_j$ for any $j \in [M]$.*

Proof. Under Assumption 4.2, it is easy to see for any realization of value-cost pairs $\boldsymbol{z}_j = (\boldsymbol{v}_j, \boldsymbol{d}_j)$ there always exists an auction $n \in [m_j]$ whose value-to-cost ratio is at least γ, i.e. $v_{j,n} > \gamma d_{j,n}$. Hence we know that $\frac{v_{j,1}}{d_{j,1}} \geq \frac{v_{j,n}}{d_{j,n}} > \gamma$. Now, in Eq. (25) within the proof of Lemma 4.5, we showed

$$V_j(\rho_j;\boldsymbol{z}_j) = \boldsymbol{v}_j^\top \boldsymbol{x}_j^*(\rho_j;\boldsymbol{z}_j) = \sum_{n\in[m_j]}\left(\frac{v_{j,n}}{d_{j,n}}\rho_j + b_{j,n}\right)\mathbb{I}\left\{d_{j,0} + \cdots + d_{j,n-1} \leq \rho_j \leq d_{j,0} + \cdots + d_{j,n}\right\}\,,$$

where $d_{j,0} = v_{j,0} = b_{j,1} = 0$. This implies that for any $\rho_j < d_{j,1}$, we have $V_j(\rho_j;\boldsymbol{z}_j) = \frac{v_{j,1}}{d_{j,1}}\rho_j > \gamma\rho_j$. Therefore, we can take $\widetilde{\rho} = \min_{j\in[M]}\min_{\boldsymbol{z}_j\in F_j} d_{j,1}$, which ensures that for any $\rho_j \leq \widetilde{\rho}$ and realization $\boldsymbol{z}_j \in F_j$ we have $V_j(\rho_j;\boldsymbol{z}_j) = \frac{v_{j,1}}{d_{j,1}}\rho_j > \gamma\rho_j$ for any channel $j \in [M]$. ☐

Lemma C.3 (Constraint satisfaction)**.** *Assume Assumption 4.2 holds, and consider $\beta = \underline{\rho} = \frac{1}{\log(T)}$ in Algorithm 1. Then, for large enough T we have*

$$\frac{1}{T}\sum_{t\in[T]} g_{1,t} \geq 0 \quad \textit{and} \quad \frac{1}{T}\sum_{t\in[T]}\sum_{j\in[M]}\rho_{j,t} \leq \rho\,,$$

where we recall $g_{1,t} = \sum_{j\in[M]}\left(V_j(\rho_{j,t};\boldsymbol{z}_{j,t}) - \gamma\rho_{j,t}\right)$.

Proof. Recall $\tau_A \in [T]$ defined in step 10 of Algorithm 1.

If $\tau_A = T$, then we know that Algorithm 1 does not exit the while loop, and therefore $S_{1,t} - \gamma M\rho + \beta\underline{\rho}(T-t) \geq 0$ for $t = T$, or equivalently $S_{1,T} \geq \gamma M\rho > 0$. Since we recall $S_{1,T} = \sum_{t\in[T-1]} g_{1,t}$, we can conclude that $\sum_{t\in[T]} g_{1,t} = S_{1,T} + g_{1,T} \geq M\rho + g_{1,T} \geq 0$ since $g_{1,T} \geq -\gamma M\rho$. Similarly, we also have $S_{2,t} + M\rho + \underline{\rho}(T-t) \leq \rho T$ for $t = T$, or equivalently $S_{2,T} \leq \rho T - M\rho$ where we used the fact that $\underline{\rho} = 1/\log(T) < \rho$ for large enough T and $M \geq 2$. Hence, recalling $S_{2,T} = \sum_{t\in[T-1]}\sum_{j\in[M]}\rho_{j,t}$, we can conclude that $\sum_{t\in[T]}\sum_{j\in[M]}\rho_{j,t} = S_{2,T} + \sum_{j\in[M]}\rho_{j,T} \leq \rho T - M\rho + \sum_{j\in[M]}\rho_{j,T} \leq \rho T$ since $\sum_{j\in[M]}\rho_{j,T} \leq M\rho$.

If $\tau_A < T$, then we know that at the "stopping time" τ_A, the while loop in Algorithm 1 has not yet exited, so we have

$$S_{1,\tau_A} - \gamma M\rho + \beta\underline{\rho}(T-\tau_A) \geq 0 \text{ and } S_{2,\tau_A} + M\rho + M\underline{\rho}(T-\tau_A) \leq \rho T \tag{53}$$

Hence,

$$\begin{aligned}
\sum_{t\in[T]} g_{1,t} &= \sum_{t\in[\tau_A-1]} g_{1,t} + g_{1,\tau_A} + \sum_{t=\tau_A+1}^{T} g_{1,t} \\
&\overset{(i)}{\geq} \gamma M\rho - \beta\underline{\rho}(T-\tau_A) + g_{1,\tau_A} + \sum_{t=\tau_A+1}^{T} g_{1,t} \\
&\geq \gamma M\rho - \beta\underline{\rho}(T-\tau_A) - \gamma M\rho + \sum_{t=\tau_A+1}^{T} g_{1,t} \\
&\overset{(ii)}{=} -\beta\underline{\rho}(T-\tau_A) + \sum_{t=\tau_A+1}^{T}\sum_{j\in[M]} \left(V_j(\underline{\rho};\boldsymbol{z}_{j,t}) - \gamma\underline{\rho}\right) \\
&\overset{(iii)}{\geq} -\beta\underline{\rho}(T-\tau_A) + \sum_{t=\tau_A+1}^{T}\sum_{j\in[M]} \left(\underline{\rho}\cdot \min_{\boldsymbol{z}_j\in F_j} \frac{v_{j,1}}{d_{j,1}} - \gamma\underline{\rho}\right) \\
&= -\beta\underline{\rho}(T-\tau_A) + (T-\tau_A)M\left(\underline{\rho}\cdot \min_{\boldsymbol{z}_j\in F_j} \frac{v_{j,1}}{d_{j,1}} - \gamma\underline{\rho}\right) \\
&\overset{(iv)}{\geq} 0
\end{aligned} \tag{54}$$

where (i) follows from $S_{1,\tau_A-1} = \sum_{t\in[\tau_A-2]} g_{1,t}$ and Eq. (53); (ii) follows from Algorithm 1 where we set $\rho_{j,t} = \underline{\rho}$ for all $j\in[M]$ and $t = \tau_A + 1 \dots T$; for (iii), assuming the jth channel's realized value cost pairs $\boldsymbol{z}_{j,t}$ is the element $\boldsymbol{z}_j \in F_j$, then Proposition C.2 says $V_j(\underline{\rho};\boldsymbol{z}_{j,t}) \geq \frac{v_{j,1}}{d_{j,1}}\underline{\rho}$ since $\underline{\rho} = \frac{1}{\log(T)} < \widetilde{\rho}$ for large enough T. Hence $V_j(\underline{\rho};\boldsymbol{z}_{j,t}) \geq \min_{\boldsymbol{z}_j\in F_j} \frac{v_{j,1}}{d_{j,1}}\underline{\rho}$; (iv) follows from the fact that $\min_{\boldsymbol{z}_j\in F_j} \frac{v_{j,1}}{d_{j,1}} > \gamma$ according to Proposition C.2, so $M\min_{\boldsymbol{z}_j\in F_j} \frac{v_{j,1}}{d_{j,1}} \geq M\gamma + \beta$ since $\beta = \frac{1}{\log(T)} \leq M\min_{\boldsymbol{z}_j\in F_j} \frac{v_{j,1}}{d_{j,1}} - M\gamma$ for large enough T.

Similarly, we have

$$\begin{aligned}
\sum_{t\in[T]}\sum_{j\in[M]} \rho_{j,t} &= \sum_{t\in[\tau_A-1]}\sum_{j\in[M]} \rho_{j,t} + \sum_{j\in[M]} \rho_{j,\tau_A} + \sum_{t=\tau_A+1}^{T}\sum_{j\in[M]} \rho_{j,t} \\
&\overset{(i)}{\leq} \rho T - M\rho - M\underline{\rho}(T-\tau_A) + \sum_{j\in[M]} \rho_{j,\tau_A} + M(T-\tau_A)\underline{\rho} \\
&\leq \rho T - M\rho - M\underline{\rho}(T-\tau_A) + M\rho + M(T-\tau_A)\underline{\rho} \\
&= \rho T
\end{aligned} \tag{55}$$

where (i) follows from $S_{2,\tau_A} = \sum_{t\in[\tau_A-1]}\sum_{j\in[M]} \rho_{j,t}$ and Eq. (53), as well as in Algorithm 1 we set $\rho_{j,t} = \underline{\rho}$ for all $j\in[M]$ and $t = \tau_A, \tau_A+1 \dots T$. □

Lemma C.4. *Let $(\lambda_t,\mu_t)_{t\in[T]}$ be the dual variables generated by Algorithm 1. Then for any $\lambda,\mu\in[0,C_F]$ and $t\in[T]$ we have*

$$\begin{aligned}
\sum_{\tau\in[t]} (\lambda_\tau - \lambda)\, g_{1,\tau} &\leq \frac{\eta M^2\bar{V}^2}{2}\cdot t + \frac{1}{2\eta}(\lambda-\lambda_1)^2 \\
\sum_{\tau\in[t]} (\mu_\tau - \mu)\, g_{2,\tau} &\leq \frac{\eta\rho^2}{2}\cdot t + \frac{1}{2\eta}(\mu-\mu_1)^2\,.
\end{aligned} \tag{56}$$

where we recall $g_{1,\tau} = \sum_{j\in[M]}(V_{j,\tau}(\rho_{j,\tau}) - \gamma\rho_{j,\tau})$ and $g_{2,\tau} = \rho - \sum_{j\in[M]}\rho_{j,\tau}$.

Proof. We will show Eq. (56). Starting with the first inequality w.r.t. λ_τ's, we have

$$(\lambda_\tau - \lambda)\, g_{1,\tau} = (\lambda_{\tau+1} - \lambda)\, g_{1,\tau} + (\lambda_\tau - \lambda_{\tau+1})\, g_{1,\tau} \tag{57}$$

Since $\lambda_{\tau+1} = \Pi_{[0,C_F]}\left(\lambda_\tau - \eta g_{1,\tau}\right)_+ = \arg\min_{\lambda\in[0,C_F]}\left(\lambda - (\lambda_\tau - \eta g_{1,\tau})\right)^2$, we have

$$(\lambda_{\tau+1} - (\lambda_\tau - \eta g_{1,\tau}))\cdot(\lambda - \lambda_{\tau+1}) \geq 0 \quad \forall\lambda\in[0,C_F]\,. \tag{58}$$

So we have

$$\begin{aligned}(\lambda_{\tau+1} - \lambda)\, g_{1,\tau} \;&\leq\; \frac{1}{\eta}(\lambda_{\tau+1} - \lambda_\tau)\cdot(\lambda - \lambda_{\tau+1})\\ &=\; \frac{1}{2\eta}\left((\lambda - \lambda_\tau)^2 - (\lambda - \lambda_{\tau+1})^2 - (\lambda_{\tau+1} - \lambda_\tau)^2\right)\,.\end{aligned} \tag{59}$$

Plugging the above back into Eq. (57) we get

$$\begin{aligned}(\lambda_\tau - \lambda)\, g_{1,\tau} \;&\leq\; (\lambda_\tau - \lambda_{\tau+1})\, g_{1,\tau} + \frac{1}{2\eta}\left((\lambda - \lambda_\tau)^2 - (\lambda - \lambda_{\tau+1})^2 - (\lambda_{\tau+1} - \lambda_\tau)^2\right)\\ &\leq\; \frac{\eta}{2}g_{1,\tau}^2 + \frac{1}{2\eta}\left((\lambda - \lambda_\tau)^2 - (\lambda - \lambda_{\tau+1})^2\right)\\ &\leq\; \frac{\eta M^2\bar{V}^2}{2} + \frac{1}{2\eta}\left((\lambda - \lambda_\tau)^2 - (\lambda - \lambda_{\tau+1})^2\right)\,,\end{aligned} \tag{60}$$

where the final inequality follows from the fact that $V_{j,\tau}(\rho_{j,\tau}) \leq \bar{V}$ for any $j\in[M]$ and $\tau\in[t]$ so $g_{1,\tau}\leq M\bar{V}$. Summing the above over $\tau = 1\ldots t$ and telescoping we get

$$\sum_{\tau\in[t]}(\lambda_\tau - \lambda)\, g_{1,\tau} \leq \frac{\eta M^2\bar{V}^2}{2}\cdot t + \frac{1}{2\eta}(\lambda - \lambda_1)^2 \quad \text{for } \forall\lambda\in[0,C_F]\,.$$

Following the same arguments above we can show

$$\sum_{\tau\in[t]}(\mu_\tau - \mu)\, g_{2,\tau} \leq \frac{\eta\rho^2}{2}\cdot T + \frac{1}{2\eta}(\mu - \mu_1)^2 \quad \text{for } \forall\mu\in[0,C_F]\,.$$

□

Proposition C.5. *Under Assumption 4.2, the advertiser's per-channel only budget optimization problem, namely* $\text{CH-OPT}(\mathcal{I}_B)$ *is a convex problem.*

Proof. Recalling the $\text{CH-OPT}(\mathcal{I}_B)$ in Eq. (3) and the definition of $\mathcal{I}_B$ in Eq. (2), we can write $\text{CH-OPT}(\mathcal{I}_B)$ as

$$\begin{aligned}\text{CH-OPT}(\mathcal{I}_B) = \max_{(\gamma_j)_{j\in[M]}\in\mathcal{I}} \quad & \sum_{j\in M} V_j(\rho_j)\\ \text{s.t.} \quad & \sum_{j\in M} V_j(\rho_j) \geq \gamma\sum_{j\in M}\rho_j\\ & \sum_{j\in[M]}\rho_j \leq \rho\,,\end{aligned} \tag{61}$$

Here we used the definition $V_j(\rho_j) = \mathbb{E}\left[V_j(\rho_j;\boldsymbol{z}_j)\right]$ in Eq. (5), and $D_j(\rho_j;\boldsymbol{z}_j) = \rho_j$ for any $\boldsymbol{z}_j$ under Assumption 4.2. According to Lemma 4.5, $V_j(\rho_j)$ is concave in ρ_j for any j, so the objective of $\text{CH-OPT}(\mathcal{I}_B)$ maximizes a concave function. For the feasibility region, assume ρ_j and ρ_j' are feasible, then defining $\rho_j'' = \theta\rho_j + (1-\theta)\rho_j'$ for any $\theta\in[0,1]$, we know that

$$\begin{aligned}\sum_{j\in M}\left(V_j(\rho_j'') - \gamma\rho_j''\right) \;&\overset{(i)}{\geq}\; \sum_{j\in M}\left(\theta V_j(\rho_j) + (1-\theta)V_j(\rho_j') - \gamma\rho_j''\right)\\ &=\; \theta\sum_{j\in M}\left(V_j(\rho_j) - \gamma\rho_j\right) + (1-\theta)\sum_{j\in M}\left(V_j(\rho_j') - \gamma\rho_j'\right)\\ &\overset{(ii)}{\geq}\; 0\end{aligned}$$

where (i) follows from concavity of $V_j(\rho_j)$ and (ii) follows from feasiblity of ρ_j and ρ_j'. On the other hand it is apparent that $\sum_{j\in[M]} \rho_j'' \leq \rho$. Hence we conclude that for any ρ_j and ρ_j' feasible, $\rho_j'' = \theta\rho_j + (1-\theta)\rho_j'$ is also feasible, so the feasible region of $\text{CH-OPT}(\mathcal{I}_B)$ is convex. This concludes the statement of the proposition. $\square$