Learning in Proportional Allocation Auctions Games

Younes Ben Mazziane[1], Cleque-Marlain Mboulou Moutoubi[1],
Eitan Altman[1,2] and Francesco De Pellegrini[1]

***Abstract*—The *Kelly* or *proportional allocation* mechanism is a simple and efficient auction-based scheme that distributes an infinitely divisible resource proportionally to the agents' bids. When agents are aware of the allocation rule, their interactions form a game, that has been extensively studied. This paper examines the less explored repeated Kelly game, focusing mainly on utilities that are logarithmic in the allocated resource fraction. We first derive this logarithmic form from fairness–throughput trade-offs in wireless network slicing, and then prove that the induced stage game admits a unique Nash equilibrium (NE). For the repeated play, we prove convergence to this NE under three behavioral models: (i) all agents use Online Gradient Descent (`OGD`), (ii) all agents use Dual Averaging with a quadratic regularizer (`DAQ`) (a variant of the Follow-the-Regularized leader algorithm), and (iii) all agents play myopic best responses (`BR`). Our convergence results hold even when agents use personalized learning rates in `OGD` and `DAQ` (e.g., tuned to optimize individual regret bounds), and they extend to a broader class of utilities that meet a certain sufficient condition. Finally, we complement our theoretical results with extensive simulations of the repeated Kelly game under several behavioral models, comparing them in terms of convergence speed to the NE, and per-agent time-average utility. The results suggest that `BR` achieves the fastest convergence and the highest time-average utility, and that convergence to the stage-game NE may fail under heterogeneous update rules.**

***Index Terms*—Kelly mechanism, auctions, game theory, learning in games, no-regret learning.**

I. Introduction

Decentralized resource allocation in large-scale systems is a fundamental problem extensively studied in network economics [1]. In this context, a resource owner seeks to distribute resources among multiple agents to optimize an objective, such as maximizing *social welfare*, namely, the aggregate net benefit of the agents, or their revenue. It is standard to assume that the resource owner may have partial or lack information about the agents' utilities or preferences. Instead, they depend on signals [2] provided by the agents, such as declared valuations, willingness to pay, or other indirect indicators of agents' preferences. Moreover, agents often act selfishly and strategically in order to maximize their benefits. This problem is prevalent in various technological domains, including bandwidth allocation in communication networks [3], task scheduling in cloud computing [4], energy distribution in smart grids [5], and pricing mechanisms in shared transportation systems [6].

[1]LIA, Avignon university, Avignon, France; [2]INRIA, Sophia Antipolis, France.

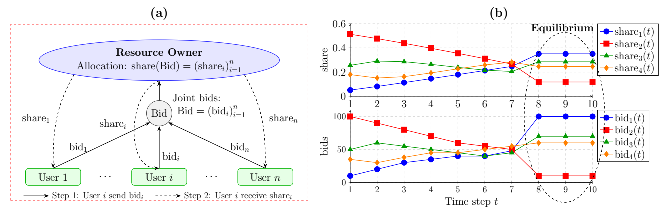


Fig. 1: Repeated resource allocation game.

The *Kelly* or *proportional allocation* mechanism stands out among decentralized resource allocation mechanisms for its simplicity and efficiency [1], [7]. In its basic form, agents submit bids to secure shares of a finite, infinitely divisible resource, with allocations distributed proportionally to their bids. In a generalized formulation [2], each user's allocation is determined by a *weighting function* of its bid, enabling diverse allocation strategies. In particular, the classic Kelly mechanism arises when this weighting function is simply the identity for all agents.

Many works have shown that the Kelly mechanism enjoys strong social-welfare optimality guarantees across several settings: (i) agents with unlimited budgets who are either *price takers* (unaware of the allocation rule) [7] or *price anticipator* (aware of it) [2], [8], [9], and (ii) environments in which price anticipator agents have budget constraints [10]–[12]. More specifically, price anticipator agents induce a competitive game with continuous action sets, and the guarantees in terms of social welfare hold exclusively at a Nash Equilibrium (NE) of this game, that we refer to as the *Kelly game* in the sequel.

In practice, however, agents are not necessarily aware of the utilities of other agents. A more realistic setting is when agents know only their own utilities but adapt their bids over repeated synchronous rounds based on feedback from previous rounds, e.g., the aggregate bid. This motivates the study of the repeated Kelly game. Figure 1 illustrates this setting, where at each round, agents compete over a new resource by submitting bids based on outcomes from previous rounds. They then receive a fraction of the resource according to the Kelly mechanism. In this setting, rational agents aim to maximize their time-average utility.

To our knowledge, only a few works have examined the repeated Kelly game [13]–[15], and they focus on the case where agents' utilities are linear in the fraction of the allocated resource. This case coincides with *Tullock* (rent-seeking) contests [16], also known as lottery contests [15]. In particular, [13] shows that if every player uses any *no-regret* bidding algorithm, then each player's average utility converges to their stage-game NE utility. [17] proves that under a specific bidding rule used by all agents, the sequence of actions converges to a NE of the stage game. [15] establishes convergence to this equilibrium when all agents employ fictitious-play updates. On the other hand, we study the repeated Kelly game with a general class of utilities that include logarithmic ones.

A. Contributions

Our contributions are summarized as follows:

1) We show that a practical scenario of interest induces a repeated Kelly game with logarithmic utilities.
2) We derive a tractable sufficient condition ensuring that the stage game satisfies Rosen's Diagonal Strict Concavity (**DSC**) with some vector $\boldsymbol{r} > \mathbf{0}$ ($\boldsymbol{r}$-DSC), equivalently, $\boldsymbol{r}$-monotonicity, and thus admits a unique Nash equilibrium. The condition reduces to verifying negativity of a scalar function, and we show it holds for logarithmic utilities.
3) For repeated Kelly games satisfying our sufficient $\boldsymbol{r}$-DSC condition, and with utilities that differ only by multiplicative factors, we prove convergence when all agents use either Online Gradient Descent (`OGD`) or Dual Averaging with a quadratic regularizer (`DAQ`).
4) We establish convergence of best-response dynamics in the repeated Kelly game under logarithmic utilities.
5) We conduct extensive numerical simulations to validate our theoretical results, and complement them with additional scenarios in which agents run heterogeneous learning dynamics.

We provide more details about our contributions.

Rosen's $\boldsymbol{r}$-DSC and uniqueness of the Nash equilibrium. A standard way to establish **DSC** is to show that a certain $n \times n$ matrix (with n the number of agents) is negative definite over the action set. For a dense matrix, checking negative definiteness typically requires $\mathcal{O}(n^2)$ memory and $\mathcal{O}(n^3)$ time. In contrast, Theorem 2 exploits the structure of the repeated Kelly game to reduce this verification to $\mathcal{O}(n)$ time and $\mathcal{O}(1)$ memory. This tractable condition also enables proving that there exists a vector $\boldsymbol{r}$ for which $\boldsymbol{r}$-**DSC** holds under logarithmic utilities. Moreover, proving **DSC** extends prior uniqueness guarantees of the NE to arbitrary convex action sets, which accommodates budget constraints and Kelly mechanisms with general weighting functions.

Convergence of no-regret learning to the NE. Previous results show that convergence of `OGD` is guaranteed when the stage game satisfies $\boldsymbol{r}$-**DSC** for some $\boldsymbol{r} > \mathbf{0}$, whereas convergence of `DAQ` requires the stronger condition **1**-**DSC**. Under our $\boldsymbol{r}$-**DSC** sufficient condition—which holds for logarithmic utilities—convergence of `OGD` follows immediately. However, these previous results impose a common learning rate across agents, and our **1**-**DSC** sufficient condition holds only for homogeneous logarithmic utilities. We show in Theorem 3 that, under affine heterogeneity in utilities (e.g., utilities share the same logarithmic form but differ by agent-specific multiplicative factors), `OGD` still converges to the stage-game NE when agents use regret-optimal learning rates. Under the same heterogeneity model, Theorem 4 further establishes that, assuming our $\boldsymbol{r}$-**DSC** sufficient condition (for some $\boldsymbol{r} > \mathbf{0}$), `DAQ` also converges under personalized regret-optimal learning rates.

Convergence of best response dynamics. When agents use best response dynamics, we model the iterates of agents as fixed point iterations. We then derive closed form expressions of the Jacobian of the fixed point operator, and we prove that it is a contraction. This proves convergence of the system to the NE of the stage game but also shows that the convergence speed is linear.

Numerical simulations. We simulate the bidding algorithms under both *homogeneous dynamics*, where agents use the same update rule, and *heterogeneous dynamics*, where two update rules coexist in the population. Under *homogeneous dynamics*, the simulations confirm our theoretical convergence results to the stage game NE, and indicate that, in terms of both convergence speed and time-average utility, `BR` performs best, followed by `OGD`, and then `DAQ`. Under *heterogeneous dynamics*, the results suggest that convergence to the stage game NE may fail. However, the resulting time-average utilities remain close across algorithms and near the NE utility, with `BR` consistently better in the considered settings.

B. Paper outline

The rest of the paper is organized as follows. Section II formally introduces the Kelly mechanism and the induced game. Section III-A derives the repeated Kelly game with α-fair utilities for bandwidth allocation in wireless networks, presents the proposed bidding algorithms, and establishes their convergence guarantees. Section IV complements the theoretical results with numerical simulations. Section V concludes the paper.

II. Problem Formulation

We consider a repeated allocation of a unit-sized divisible resource among n agents over T rounds according to the general allocation mechanism proposed in [2], which extends the proportionally fair Kelly mechanism introduced in [7].

Bidding. At each step t, each agent i submits a bid $b_{i,t}$ that must be at least a fixed positive constant $\tilde{\epsilon}_i$, i.e., $b_{i,t} \geq \tilde{\epsilon}_i > 0$. It must also respect the budget constraint, i.e., $b_{i,t} \leq \tilde{c}_i$, where c_i is the budget of agent i at each round t. This bid is based on previously submitted bids $\boldsymbol{b}_1, \dots, \boldsymbol{b}_{t-1}$, where $\boldsymbol{b}_s = (b_{i,s})_{i \in \mathcal{I}}$ and $\mathcal{I}$ is the set of agents.
Allocation: Based on the bids of each round t, the resource owner allocates fractions $x_{i,t}(\boldsymbol{b}_t)$ of the resource according to

$$x_{i,t}(\boldsymbol{b}_t) = \begin{cases} \frac{w_i(b_{i,t})}{\sum_{j=1}^n w_j(b_{j,t}) + \delta}, & \text{if } w_i(b_{i,t}) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $w_i : \mathbb{R}^+ \to \mathbb{R}^+$ are continuous, increasing functions governing how resources are distributed, and $\delta \geq 0$ is a reservation parameter [2]. If w_i is the identity function, this mechanism reduces to the classic Kelly mechanism.

Each agent i has a valuation function $V_i : [0,1] \to \mathbb{R}^n_{\geq 0}$, where $V_i(x_i)$ quantifies monetary benefit of acquiring a fraction x_i of the resource. The utility of agent i at each step t is determined by the function φ_i, defined as the value the agent derives from the allocated fraction minus the payment, i.e., $\varphi_i(\boldsymbol{b}_t) = V_i(x_{i,t}(\boldsymbol{b}_t)) - b_{i,t}$. The objective of each agent is devise an online bidding strategy $b_{i,1}, \dots, b_{i,T}$ to maximize its aggregate utility, i.e., $\sum_{t=1}^T \varphi_i(\boldsymbol{b}_t)$.

Following [2], define the change of variable $z_{i,t} = w_i(b_{i,t})$, and the function $p_i : \mathbb{R}^+ \to \mathbb{R}^+$ as the inverse of w_i, i.e., $p_i(z_i) \triangleq w_i^{-1}(z_i)$. We refer to p_i as the payment function for agent i. Under this change of variable, the allocation rule and the utility function become,

$$x_{i,t}(\boldsymbol{z}_t) = \begin{cases} \frac{z_{i,t}}{\sum_{j=1}^n z_{j,t} + \delta} & \text{if } z_{i,t} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

$$\varphi_i(\boldsymbol{z}) \triangleq V_i(x_i(\boldsymbol{z})) - p_i(z_i). \tag{3}$$

Note that both formulations are equivalent in the sense that the allocated fraction corresponding to a given payment is the same in each setting. In this paper, we will focus on the second formulation using $\boldsymbol{z}$.

When agents are aware that the Kelly mechanism governs resource allocation, the interaction between them forms a competitive repeated game. We define $\mathcal{G}$ as the stage game arising from this competition, where the set of players is $\mathcal{I}$ with utility functions $\boldsymbol{\varphi} = (\varphi_i)_{i \in \mathcal{I}}$ and action space constrained by budgets, denoted $\mathcal{R}$, and given by the cartesian product of $\mathcal{R}_i$ for $i \in \mathcal{I}$, where $\mathcal{R}_i \triangleq [\epsilon_i, c_i]$, where $\epsilon_i = p_i^{-1}(\tilde{\epsilon}_i)$ and $c_i = p_i^{-1}(\tilde{c}_i)$.

We make the following assumptions about the functions V_i and p_i.

Assumption 1. *Over the domain $[0,1]$, V_i is strictly increasing, concave, and twice continuously differentiable ($V_i \in \mathcal{C}^2([0,1])$). Over the domain $\mathbb{R}^n_{\geq 0}$, p_i is convex, increasing with respect to z_i for any i, and twice continuously differentiable, i.e., $\boldsymbol{p} \in \mathcal{C}^2(\mathbb{R}^n_{\geq 0})$.*

Note that the above assumption is standard [1], [2]. Moreover, it is natural for V_i and p_i to be increasing. Agents gain larger utility from receiving a larger share of the resource.

Under Assumption 1, the utility function φ_i is concave with respect to the i-th component and thus the *best response operator* is a function, that we denote as $\mathbf{BR} : \mathbb{R}^n \mapsto \mathbb{R}^n$. Note that the utility function of each agent in (3) depends only on their own bid, (i.e., $\mathcal{G}$ is an aggregative game), and thus by abuse of notation, we can write $\varphi_i(z_i, z_{-i}) = \varphi_i(z_i, s_i(\boldsymbol{z}))$ such that $s_i(\boldsymbol{z}) \triangleq \sum_{j \neq i} z_j + \delta$. The best response of a player i, denoted $\mathrm{BR}_i : \mathbb{R}_+ \mapsto \mathbb{R}_+$, is defined as,

$$\mathrm{BR}_i(s) = \min\left(\arg\max_{z_i \in \mathcal{R}_i} \varphi_i(z_i, s) \right), \tag{4}$$

and it holds that $\mathbf{BR}(\boldsymbol{z}) = (\mathrm{BR}_i(s_i(\boldsymbol{z})))_{i \in [n]}$. Let $(z_i, \boldsymbol{z}_{-i})$ denote the vector where agent i submits a bid z_i, while the other agents submit bids $\boldsymbol{z}_{-i}$.

Definition 1 (Nash Equilibrium)**.** *A strategy profile $\boldsymbol{z}^* = (z_1^*, z_2^*, \dots, z_n^*) \in \mathcal{R}$ is a* Nash Equilibrium (NE) *of $\mathcal{G}$ if, for every player $i \in \mathcal{I}$,*

$$\varphi_i(z_i^*, \boldsymbol{z}_{-i}^*) \geq \varphi_i(z_i, \boldsymbol{z}_{-i}^*) \quad \forall z_i \in \mathcal{R}_i \tag{5}$$

As a consequence of Assumption 1, the function φ_i is concave in its i-th component and twice continuously differentiable on $\mathbb{R}^n_{\geq 0}$, and the actions set $\mathcal{R}$ is non empty, closed, bounded, and convex. Existence of a Nash Equilibrium (NE) of the game $\mathcal{G}$ follows by [18, Thm. 1].

Theorem 1. *The set of Nash equillibria of $\mathcal{G}$, denoted $NE(\mathcal{G})$ is non-empty, i.e., $NE(\mathcal{G}) \neq \emptyset$.*

Notation. We use $\dot{V}_i(\cdot)$ and $\ddot{V}_i(\cdot)$ to denote the first and second derivatives of V_i, respectively. We use a similar notation for p_i. We use $\partial_j \varphi_i$ to designate the partial derivative of φ_i with respect to the bid of agent j.

III. Repeated Kelly Game

A. Motivation: Bandwidth allocation in wireless networks

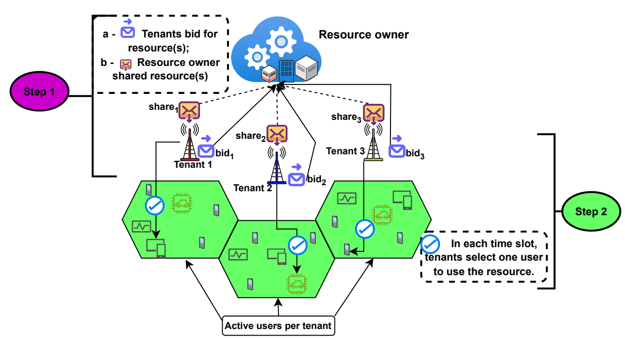


Fig. 2: Bandwidth Allocation between Tenants and Users

In this section, we show how a repeated Kelly game with logarithmic valuations V_i arises in bandwidth allocation among multiple *tenants* (e.g., virtual operators or service providers), each serving its own set of users. Over rounds $t \in \{1, \dots, T\}$, an infrastructure provider allocates a total bandwidth B according to the Kelly mechanism; given bids $z_j(t)$, tenant j receives,

$$B_j(t) = \frac{z_j(t)}{\sum_{k\in\mathcal{I}} z_k(t) + \delta}. \tag{6}$$

where $\mathcal{I}$ denotes the set of tenants and $\delta \geq 0$.

Within round t, time is divided into slots. At each slot, tenant j schedules exactly one user from its set $\mathcal{I}_j$; unscheduled users are silent. Let $S_j^\tau(t) \in \mathcal{I}_j$ denote the scheduled user at slot τ. When a user i is scheduled, its transmission rate, denoted $r_{j,i}^\tau(t)$, is proportional to the allocated bandwidth,

$$r_{j,i}^\tau(t) = \gamma_{j,i}^\tau(t) B_j(t) \mathbb{1}\left(S_j^\tau(t) = i\right), \tag{7}$$

where $\gamma_{j,i}^\tau(t) > 0$. For instance in [19], $\gamma_{j,i}^\tau(t) = \ln\left(1 + \frac{p_{j,i} h_{j,i}^\tau(t)}{N_0}\right)$, where $p_{j,i}$ is the transmission power, $h_{j,i}^\tau(t)$ is the channel state, and N_0 is the noise power.

A standard objective in this setting is the *Proportional-fair* metric [20]. Optimizing this objective enables balancing the overall throughput and fairness across users. We use $\mathrm{PropFair}_j(t)$ to denote the proportional fair metric of tenant j at round t, and it is expressed as,

$$\mathrm{PropFair}_j(t) \triangleq \sum_{i\in\mathcal{I}_j} \ln\left(\sum_\tau r_{j,i}^\tau(t)\right) \tag{8}$$

$$= N_j \ln(B_j(t)) + \sum_{i\in\mathcal{I}_j} \ln\left(\sum_\tau \gamma_{j,i}^\tau(t) \mathbb{1}\left(S_j^\tau(t) = i\right)\right), \tag{9}$$

where N_j is the number of users served by tenant j, i.e., $N_j = |\mathcal{I}_j|$. This decomposition makes the roles of bidding and scheduling transparent: the term $N_j \ln(B_j(t))$ depends only on the bidding process, while the second term is controlled by the scheduling policy and channel states, and is independent of the bids. Using the quasi-linear utility model [1], the stage utility of tenant j at round t is,

$$\varphi_j(z_j(t), z_{-j}(t)) = N_j \ln\left(\frac{z_j(t) B}{\sum_{k\in\mathcal{I}_j} z_k(t)}\right) + \sum_{i\in\mathcal{I}_j} \ln\left(\sum_\tau \gamma_{j,i}^\tau(t) \mathbb{1}\left(S_j^\tau(t) = i\right)\right) - z_j. \tag{10}$$

Therefore, the induced bidding interaction is a repeated Kelly game with logarithmic V_i's.

B. Single-Agent Formulation of the Online Bidding Problem

At round t of the repeated Kelly game, agent i faces uncertainty about others' aggregate bid $s_i(\boldsymbol{z}(t)) = \delta + \sum_{j\neq i} z_j(t)$. Before bidding, agent i only knows the history $(s_i(\boldsymbol{z}(1)), \ldots, s_{-i}(\boldsymbol{z}(t-1)))$. A bidding algorithm $\mathcal{A}_i$ maps this history to a bid $z_i^{\mathcal{A}_i}(t) \in \mathcal{R}_i$, and earns the payoff $\varphi_i(z_i^{\mathcal{A}_i}(t), s_i(\boldsymbol{z}(t)))$. Under Assumption1, this yields an Online Convex Optimization (OCO) [21] problem: in each of the T rounds, the agent chooses $z_i^{\mathcal{A}_i}(t) \in \mathcal{R}_i$, then a concave reward function $u_i^t : \mathcal{R}_i \mapsto \mathbb{R}$, defined as $u_i^t(z_i) \triangleq \phi_i(z_i, s_i(\boldsymbol{z}(t)))$ is revealed, and the agent receives $u_i^t\left(z_i^{\mathcal{A}_i}(t)\right)$. The objective is then to maximize the aggregate reward over rounds. In this framework, the main performance metric of an algorithm $\mathcal{A}_i$ is the *regret*, denoted as $\mathrm{Reg}_T^{(i)}(\mathcal{A}_i)$, and defined as the gap between the cumulative reward of the best fixed bid in hindsight and the agent's cumulative reward, i.e.,

$$\mathrm{Reg}_T^{(i)}(\mathcal{A}_i) \triangleq \max_{z_i\in\mathcal{R}_i} \sum_{t=1}^T u_i^t(z_i) - \sum_{t=1}^T u_i^t\left(z_i^{\mathcal{A}_i}(t)\right). \tag{11}$$

Define the constants D_i and G_i as upper bounds on the diameter of the decision set $\mathcal{R}_i$, and the derivatives of u_i^t's for any t, i.e.,

$$D_i \geq c_i - \epsilon_i, \quad G_i \geq \sup_{z\in\mathcal{R}} |\partial_i \varphi_i(z_i, s_i(\boldsymbol{z}))|, \tag{12}$$

Since the Kelly game yields finite diameter $D_i < \infty$ and gradient bound $G_i < \infty$ (because φ_i is continuous over the closed set $\mathcal{R}$), standard OCO methods achieve sublinear regret, $\mathrm{Reg}_T^{(i)} = o(T)$. Consequently, for any sequence of opponent aggregates $s_i(\boldsymbol{z}(t))$, the agent's time-average reward approaches that of the best fixed bid in hindsight.

C. Bidding algorithms

We consider four bidding algorithms. Two of them are adaptations of classical no-regret methods to the repeated Kelly game, namely Online Gradient Descent (`OGD`) [22] and Dual Averaging (`DA`) [23], an instance of the Follow-The-Regularized-Leader (FTRL) family of algorithms. The third algorithm is an instance of Regularized-Robbins–Monro (`RRM`) family of algorithms, recently studied in the context of repeated games [24], [25], and encompassing `DA` as a special case. The fourth algorithm is a myopic best-response scheme.

`OGD`, `DA`, and `RRM` are first-order methods: they only require the derivative of the stage utility with respect to the agent's bid. Specifically, at each step t, algorithm $\mathcal{A}_i \in \{\mathtt{OGD}, \mathtt{DA}, \mathtt{RRM}\}$ for agent i uses the gradient of the utility function evaluated at its current bid $z_i^{\mathcal{A}_i}(t)$, namely,

$$g_t^{(i),\mathcal{A}_i} \triangleq \partial_i u_i^t(z_i^{\mathcal{A}_i}(t)), \tag{13}$$

It also employs a learning rate (or step-size) $\eta_t^{(i)} > 0$ that is tuned at each step t.

Online Gradient Descent `OGD`. When $\mathcal{A}_i = \mathtt{OGD}$, agent i updates their bid by moving along the gradient/derivative of the reward function u_i^t at the current bid, then projects back to the feasible set $\mathcal{R}_i$ (minimum bid and budget constraints). Formally, the update at step $t+1$ is given by,

$$z_i^{\mathtt{OGD}}(t+1) = \Pi_{\mathcal{R}_i}\left(z_i^{\mathtt{OGD}}(t) + \eta_{t+1}^{(i)} g_t^{(i),\mathtt{OGD}}\right), \tag{14}$$

where $\Pi_{\mathcal{R}_i}$ is the euclidean projection over $\mathcal{R}_i$, which reduces to clipping,

$\Pi_{\mathcal{R}_i}(z) = \max(\min(z, c_i), \epsilon_i)$. Taking $\eta_t^{(i)} = D_i/(G_i\sqrt{t})$, yields, $\text{Reg}_T^{(i)}(\texttt{OGD}) \leq \frac{3}{2}G_i D_i\sqrt{T}$ [21][Thm. 3.1]. Similar regret guarantees hold when $\eta_t^{(i)}$ is constant over time; taking $\eta_t^{(i)} \equiv \eta^{(i)} = D_i/(G_i\sqrt{T})$ leads to, $\text{Reg}_T^{(i)}(\texttt{OGD}) \leq G_i D_i\sqrt{T}$ [26]. If u_i^t's are γ_i–strongly convex on $\mathcal{R}_i$, then choosing a more aggressive learning rate $\eta_t^{(i)} = 1/(\gamma_i t)$ yields, $\text{Reg}_T^{(i)}(\texttt{OGD}) \leq \frac{G_i^2}{\gamma_i}\left(1 + \log T\right)$.

Dual Averaging (DA). This algorithm employs a *regularizer*, i.e., a continuous strongly convex function, $h_i : \mathcal{R}_i \mapsto \mathbb{R}$. Let $g_{1:t}^{(i),\texttt{DA}}$ designates the sum of the gradients up to time t, i.e., $g_{1:t}^{(i),\texttt{DA}} = \sum_{s=1}^{t} g_s^{(i),\texttt{DA}}$. DA's update selects the bid z that maximizes $\left(z * g_{1:t}^{(i),\texttt{DA}} - \frac{1}{\eta_t^{(i)}}h_i(z)\right)$. In particular, if $h_i(z) = z^2/2$, then the update, denoted DAQ, is given by,

$$z_i^{\texttt{DAQ}}(t+1) = \Pi_{\mathcal{R}_i}\left(\eta_{t+1}^{(i)} g_{1:t}^{(i),\texttt{DAQ}}\right). \tag{15}$$

This update is also known as the lazy version of OGD, while (14) is known as the agile one. Taking an adaptive learning rate $\eta_t^{(i)} = D_i/(2G_i\sqrt{t})$ yields, $\text{Reg}_T^{(i)}(\texttt{DAQ}) \leq \sqrt{2}G_i D_i\sqrt{T}$ [23][Sec. 3.1]. If the number of rounds T is apriori known, then taking $\eta_t^{(i)} = D_i/(G_i\sqrt{T})$ yields, $\text{Reg}_T^{(i)}(\texttt{DAQ}) \leq G_i D_i\sqrt{T}$ [23][Sec. 3.2].

Remark 1. *In general, OGD and DAQ require an upper bound G_i on the gradient of the agent's utility (see (12)), for tuning $\eta_t^{(i)}$. The constant G_i depends on the budgets of the other players and may therefore be unknown in practice. In the case of logarithmic utilities, i.e., $V_i(\cdot) = a_i\ln(\cdot) + d_i$, and p_i is the identity function, we can write,*

$$|\partial_i\varphi_i(\boldsymbol{z})| = \left|a_i\left(\frac{1}{z_i} - \frac{1}{\sum_j z_j + \delta}\right) - 1\right| \leq \frac{a_i}{z_i} + 1 \leq \frac{a_i}{\epsilon_i} + 1. \tag{16}$$

for every $\boldsymbol{z} \in \mathcal{R}$. Thus taking $G_i = \frac{a_i}{\epsilon_i} + 1$ provides a bound that is independent of the other agents' budgets, which simplifies the use of these bidding algorithms in practice. Indeed, the learning rate for each agent i, using either OGD or DAQ, can be tuned as $\eta_t^{(i)} = \mathcal{O}\left(\frac{c_i\epsilon_i}{a_i\sqrt{T}}\right)$, leading to a regret $\mathcal{O}\left(\frac{c_i a_i}{\epsilon_i}\sqrt{T}\right)$.

Regularized-Robbins Monro (RRM). In the context of the repeated Kelly, an agent i using RRM maintains a cumulative weighted sum of gradients at each step t, denoted $y_i^{\texttt{RRM}}(t)$, which is then converted to the bid for that iteration, denoted as $z_i^{\texttt{RRM}}(t)$. Initially, $y_i^{\texttt{RRM}}(0) = 0$. At any step $t \geq 1$,

$$\begin{cases} y_i^{\texttt{RRM}}(t) = y_i^{\texttt{RRM}}(t-1) + \eta_t^{(i)} g_t^{(i),\texttt{RRM}}, \\ z_i^{\texttt{RRM}}(t) = Q_i(y_i^{\texttt{RRM}}(t-1)) : \\ Q_i(y) \triangleq \arg\max_{z_i \in \mathcal{R}_i}(z_i y - h_i(z_i)), \end{cases} \tag{17}$$

where $h_i : \mathcal{R}_i \mapsto \mathbb{R}$ is a continuous and K_i-strongly convex function for some $K_i > 0$. In particular, if $h_i(z) = \frac{z^2}{2\lambda_i}$, then the update, denoted RMQ, is given by,

$$z_i^{\texttt{RMQ}}(t) = \Pi_{\mathcal{R}_i}\left(\lambda_i y_i^{\texttt{RMQ}}(t)\right), \tag{18}$$

In particular, if $\eta_t^{(i)}$ is constant over time and $\lambda_i = 1$ for all agents, then RMQ and DAQ yield the same update. Thus RMQ in this case models selfish behavior as well.

Best-response (BR). When $\mathcal{A}_i = \texttt{BR}$, at each round t, agent i selects the bid that maximizes her payoff function φ_i, when the aggregate bid of the other agents is equal to its value in the previous round, $s_i(\boldsymbol{z}(t-1)) \triangleq \sum_{j \neq i} z_j$. Formally,

$$z_i^{\texttt{BR}}(t) = \text{BR}_i\left(s_i\left(\boldsymbol{z}^{\texttt{BR}}(t-1)\right)\right). \tag{19}$$

While BR lacks in general the no-regret guarantees of DAQ and OGD, it is simpler to implement: it uses the observed aggregate $s_i\left(\boldsymbol{z}^{\texttt{BR}}(t-1)\right)$ and the agent's own constraints, and requires no knowledge or estimation of other agents' budgets.

Remark 2. *When the V_i's are logarithmic, i.e., $V_i(\cdot) = a_i\ln(\cdot) + d_i$ with $a_i > 0$, and $p_i(z) = z$, straightforward calculations yield a closed-form expression for the best response operator,*

$$BR_i(s) = \Pi_{\mathcal{R}_i}\left(\frac{-s + \sqrt{s^2 + 4a_i s}}{2}\right). \tag{20}$$

More generally, [27] derives closed form expressions for the best-response operator for V_i's of the α-fair type with $\alpha \in \{0, 1, 2\}$.

D. Convergence guarantees

In this section, we focus utilities of the form, namely $V_i(\cdot) = a_i V(\cdot) + d_i$ with $a_i > 0$. Our results hold in particular when $V(\cdot) = \ln(\cdot)$. This model is motivated by the network-slicing setting described in Section III-A. We first provide a sufficient condition for the stage game $\mathcal{G}$ to satisfy *Strong Diagonal Strict Concavity* (**SDSC**), which implies Rosen's *Diagonal Strict Concavity* [18], or equivalently *monotonicity* [28]. As a consequence, the Nash equilibrium is unique. This property also serves as a key ingredient to establish convergence to the equilibrium under OGD and DAQ dynamics. Finally, we prove convergence of BR via a contraction argument.

For a vector $\boldsymbol{r} \in \mathbb{R}_{>0}^n$, **SDSC** is defined in terms of the $n \times n$ matrix $\boldsymbol{H_r}(\boldsymbol{z})$, whose (i,j)-entry is given by,

$$(\boldsymbol{H_r}(\boldsymbol{z}))_{i,j} \triangleq r_i\,\partial_{i,j}^2\varphi_i(\boldsymbol{z}) + r_j\,\partial_{j,i}^2\varphi_j(\boldsymbol{z}), \tag{21}$$

where the partial derivative is taken with respect to the actions of agent i and j, i.e., z_i and z_j, respectively.

Definition 2 (Strong Diagonal Strict Concavity)**.** *The game $\mathcal{G}$ satisfies Strong Diagonal Strict Concavity in $\boldsymbol{r}$ if and only if the matrix $\boldsymbol{H_r}(\boldsymbol{z})$ is negative definite for all $\boldsymbol{z} \in \mathcal{R}$,*

$$\max_{v:\|v\|=1}\{\boldsymbol{v}^\intercal\boldsymbol{H_r}(\boldsymbol{z})\boldsymbol{v}\} < 0. \tag{22}$$

In this case, we write that $\mathcal{G}$ satisfies **SDSC**$(\boldsymbol{r})$.

SDSC appears in Rosen's paper [18] as a sufficient condition for *diagonal strict concavity*, or equivalently for the game to be *monotone* [28]. This assumption is particularly useful for analyzing more general Kelly-type games with coupled action sets and for proving convergence of continuous-time dynamics. **SDSC** is also one of the main conditions for the convergence of RRM updates in a multi-agent setting to a NE [24], [25].

We prove in Theorem 2 that the Kelly game satisfies **SDSC** when utilities scale logarithmically with the allocated resource. First, we introduce the necessary notation. Define the functions f_i, g_i, and $\psi_{\boldsymbol{r},\boldsymbol{V}}$ as

$$f_i(x) = (1-x)^2 \ddot{V}_i(x) - 2(1-x)\dot{V}_i(x), \quad (23)$$

$$g_i(x) = -x(1-x)\ddot{V}_i(x) + (2x-1)\dot{V}_i(x), \quad (24)$$

$$\psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) \triangleq \left(\sum_{i\in\mathcal{I}} \frac{r_i g_i(x_i)^2}{k_i(x_i)}\right)\left(\sum_{i\in\mathcal{I}} \frac{1}{r_i k_i(x_i)}\right), \quad (25)$$

where $k_i(x) = g_i(x) - f_i(x) + \delta^2 L_i$ and $L_i = \min_{z_i\in\mathcal{R}_i} \ddot{p}_i(z_i)$. Further define the set $\Delta = \{\boldsymbol{x} > \boldsymbol{0} : \sum_{i\in\mathcal{I}} x_i \leq \frac{\sum_{k\in\mathcal{I}} c_k}{\sum_{k\in\mathcal{I}} c_k+\delta}\}$.

Theorem 2. *The following holds,*

1) *If there exists a vector* $\boldsymbol{r} > \boldsymbol{0}$ *such that* $\psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x})$ *is strictly smaller than* 1*, then* $\mathcal{G}$ *is* **SDSC**($\boldsymbol{r}$)*. Formally,*

$$\exists \boldsymbol{r} > \boldsymbol{0} : \sup_{\boldsymbol{x}\in\Delta} \psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) < 1 \implies \mathcal{G} \text{ is } \textbf{SDSC}(\boldsymbol{r}). \quad (26)$$

2) *If* $V_i(\cdot) = a_i \ln(\cdot) + d_i$*, then the condition* (26) *is satisfied for* $r_i = 1/a_i$*.*

The proof of Theorem 2 is presented in the supplementary material. In general, a negative definiteness numerical test for an $n\times n$ matrix requires $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ memory. Theorem 2 exploits the structure of the matrix $\boldsymbol{H}_{\boldsymbol{r}}(\boldsymbol{z})$ in the Kelly game to significantly reduce the verification, for a fixed $\boldsymbol{z}$, to $\mathcal{O}(n)$ time and $\mathcal{O}(1)$ memory via the condition (26). Moreover, this reduction turns **SDSC** test into bounding the maximum of a closed-form function, which is easier to deal with analytically; in particular, it enables the proof of **SDSC** when the V_i's are logarithmic.

Corollary 1. *The condition* (26) *is sufficient for the uniqueness of the Nash equilibrium of* $\mathcal{G}$*. We denote this unique equilibrium as* $\boldsymbol{z}^*$*.*

Corollary 1 follows directly from Theorem 2 using [18]. Uniqueness of the Nash equilibrium has been established under various conditions: when the V_i's satisfy Assumption 1 with identity payment function and no budget constraints [1, Thm. 2.2], or more generally when the p_i's satisfy Assumption 1 but are identical across agents [2, Prop. 2]. These results do not account for constraints. Under budget constraints, uniqueness was shown in [29, Thm. 1] for a common linear weighting function, while [30] allows heterogeneous linear coefficients but no budget constraints. By contrast, Theorem 1 and Corollary 1 establish uniqueness for logarithmic V_i, admits any p_i satisfying Assumption 1 (possibly heterogeneous), and incorporates budget constraints, thereby unifying and extending the above results for logarithmic V_i.

Leveraging the **SDSC** property of the game $\mathcal{G}$, Theorem 3 establishes the convergence of OGD to the unique NE when the utilities scale logarithmically in the allocated resource.

Theorem 3 (Convergence of OGD)**.** *Assume that the condition* (26) *holds and that the* V_i*'s are of the form* $V_i(\cdot) = a_i V(\cdot) + d_i$*, with* $a_i > 0$ *and* V *satisfying Assumption 1. If each agent updates their bid using OGD, i.e.,* $\mathcal{A}_i =$ OGD*,* $\forall i \in \mathcal{I}$*, with* $\eta_t^{(i)} = \alpha_i \eta_t^{(0)}$*,* $\alpha_i > 0$*,* $\sum_{t=1}^{\infty} \eta_t^{(0)} = \infty$*, and* $\sum_{t=1}^{\infty} \left(\eta_t^{(0)}\right)^2 < \infty$*, then the sequence of play* $\boldsymbol{z}^{OGD}(t) = (z_i^{OGD}(t))_{i\in\mathcal{I}}$ *converges to the unique NE of the stage game, i.e.,* $\lim_{t\to\infty} \boldsymbol{z}^{OGD}(t) = \boldsymbol{z}^*$*.*

Proof: Let $\boldsymbol{r}^*$ be the vector for which the condition (26) is satisfied.

If $\eta_t^{(i)} = \eta_t^{(0)}$, then by Theorem 2 the game $\mathcal{G}$ is $\boldsymbol{r}^*$-*monotone*, so the convergence of OGD follows directly from [28][Thm. 4].

To extend this result to heterogeneous learning rates $\eta_t^{(i)} = \alpha_i \eta_t^{(0)}$, consider the auxiliary game $\tilde{\mathcal{G}}$ with modified utilities $\tilde{\phi}_i = \alpha_i \phi_i$. The following holds,

- The OGD updates in the game $\tilde{\mathcal{G}}$ with $\tilde{\eta}_t^{(i)} = \eta_t^{(0)}$ coincide with the OGD updates in the game $\mathcal{G}$ with $\eta_t^{(i)} = \alpha_i \eta_t^{(0)}$.
- The games $\tilde{\mathcal{G}}$ and $\mathcal{G}$ share the same set of Nash equilibria.
- If $\mathcal{G}$ is $\boldsymbol{r}^*$-monotone, then the game $\tilde{\mathcal{G}}$ is $\boldsymbol{r}$-monotone with $r_i = r_i^*/\alpha_i$.

Combining the above statements with Theorem 2, we deduce that $\tilde{\mathcal{G}}$ is $\boldsymbol{r}$-monotone with $r_i = r_i^*/\alpha_i$. Thus, applying [28][Thm. 4] to $\tilde{\mathcal{G}}$ yields the desired convergence result, which completes the proof.

■

The step-size condition in Theorem 3 is satisfied whenever $\eta_t \propto t^{-\beta}$ with $\beta \in (1/2, 1]$. For logarithmic utilities, the induced optimization problem is convex, and the standard choice is $\eta_t \propto t^{-1/2}$, which yields the usual $O(\sqrt{T})$ regret but does not satisfy this condition. A simple workaround in the merely convex case is to take β close to 1/2, which preserves sublinear regret while complying with $\beta > 1/2$. Moreover, since each agent's action set is a compact interval $[\epsilon_i, c_i]$ with $\epsilon_i > 0$, the logarithmic utilities are in fact strongly concave on this domain, with parameter $\gamma_i > 0$ that goes to 0 when $\epsilon_i \to 0$. In this strongly concave setting, one may take $\eta_t \propto 1/t$ (corresponding to $\beta = 1$), which both satisfies the corollary's condition and yields the standard logarithmic regret guarantees.

Now we turn our attention to the analysis of DAQ and RMQ. Theorem 4 shows that RMQ converges to the NE of the stage game. Moreover, when the utilities scale logarithmically in the allocated resource, the theorem quantifies the convergence gap of DAQ to the NE, via the metric $\overline{\mathrm{Gap}}_T(\mathcal{A})$, used in [24], and defined as,

$$\overline{\mathrm{Gap}}_T(\mathcal{A}) = \frac{1}{T}\sum_{t=1}^{T}\mathrm{Gap}\left(\boldsymbol{z}^{\mathcal{A}}(t)\right): \tag{27}$$

$$\mathrm{Gap}(\boldsymbol{z}) = \sum_{i\in\mathcal{I}}\frac{\partial_i\varphi_i(\boldsymbol{z})}{a_i}\left(z_i^* - z_i\right). \tag{28}$$

Indeed, when the V_i's are logarithmic, the game $\mathcal{G}$ is **SDSC**($\boldsymbol{r}^*$) with $r_i^* = 1/a_i$, as shown in Theorem 2. Thus, $\mathrm{Gap}(\boldsymbol{z}) > 0$ for all $\boldsymbol{z} \neq \boldsymbol{z}^*$, with equality if and only if $\boldsymbol{z} = \boldsymbol{z}^*$ [24].

Theorem 4 (Convergence of RMQ and DAQ)**.** *Assume that the condition (26) holds and that the V_i's are of the form $V_i(\cdot) = a_i V(\cdot) + d_i$, with $a_i > 0$ and V is function that satisfies Assumption 1. Further assume that $\epsilon_i = \epsilon$ for all agents. The following holds,*

1) *If all agents update their bids according to RMQ, i.e., $\forall i$, $\mathcal{A}_i$ = RMQ, $\eta_t^{(i)} = \alpha_i\eta_t^{(0)}$, $\alpha_i > 0$, and $\sum_{t=1}^{\infty}(\eta_t^{(0)})^2/\sum_{t=1}^{\infty}\eta_t^{(0)} \to 0$, then the vector of bids $\boldsymbol{z}^{RMQ}(t)$ converges to the unique NE of the stage game, i.e., $\lim_{t\to\infty}\boldsymbol{z}^{RMQ}(t) = \boldsymbol{z}^*$.*
2) *If all agents update their bids according to DAQ and $V(\cdot) = \ln(\cdot)$, with $\eta_t^{(i)} = \frac{\epsilon c_i}{a_i\sqrt{T}}$, p_i is the identity function, and $a_i \geq \epsilon$, then*

$$\overline{\mathrm{Gap}}_T(DAQ) \leq \frac{1}{2\epsilon\sqrt{T}}\left(\sum_{i\in\mathcal{I}} c_i + 4|\mathcal{I}|\max_{i\in\mathcal{I}} c_i\right). \tag{29}$$

Proof: Let $\boldsymbol{r}^*$ be the vector for which the condition (26).

We first prove the convergence of RMQ in (18).

When for all agents, $\eta_t^{(i)} = \eta_t^{(0)}$, and $\sum_{t=1}^{\infty}(\eta_t^{(0)})^2/\sum_{t=1}^{\infty}\eta_t^{(0)} \to 0$, convergence of the bids vector, when all agents use RRM—with RMQ as particular case—is guaranteed by [24][Thm. 4.6] under two conditions: 1) The game $\mathcal{G}$ is **SDSC**(**1**), and 2) For any player i and for any sequence $\{y_n\} \subset \mathbb{R}$ such that $Q_i(y_n) \to \ell$, $F_i(\ell, y_n) \to 0$, where F_i is what is called the *Fenchel conjugate* in [24], and defined as, $F_i(\ell, y) = h_i^*(y) - \tilde{h}_i(y, \ell)$, where $h_i^*(y) = \max_{z\in\mathcal{R}_i}\left\{\tilde{h}_i(y,z)\right\}$, and h_i^* is the convex conjugate of h_i. It is easy to prove that the second condition holds for quadratic regularizers, i.e., any h such that $h(z) = \frac{z^2}{2\lambda}$, and $\lambda > 0$ [31]. Thus, when $\boldsymbol{r}^* = \mathbf{1}$ convergence follows directly from [24][Thm. 4.6].

To extend this result to arbitrary $\boldsymbol{r}^* > \mathbf{0}$ and when $\eta_t^{(i)} = \alpha_i\eta_t^{(0)}$, we define the payoff functions $\tilde{\phi}_i = r_i^*\phi_i$, and the corresponding stage game $\tilde{\mathcal{G}}$. The following holds,

- The games $\tilde{\mathcal{G}}$ and $\mathcal{G}$ share the same set of Nash equilibria.
- If the game $\mathcal{G}$ is **SDSC**($\boldsymbol{r}^*$), then the game $\tilde{\mathcal{G}}$ is **SDSC**(**1**).
- RMQ updates with $\tilde{\eta}_t^{(i)} = \eta_t^{(0)}$ and regularizer $\tilde{h}_i(z) = \frac{z^2}{2\tilde{\lambda}_i}$, with $\tilde{\lambda}_i = \frac{\lambda_i\alpha_i}{r_i^*}$ in the repeated $\tilde{\mathcal{G}}$, coincides with RMQ updates in the repeated $\mathcal{G}$ with $\eta_t^{(i)} = \alpha_i\eta_t^{(0)}$, and $h_i(z) = \frac{z^2}{2\lambda_i}$.

The statements above combined with Theorem 2 yields the convergence of the RMQ updates with homogeneous $\tilde{\eta}_t^{(i)}$ regularizers $\tilde{h}_i$ in the repeated $\tilde{\mathcal{G}}$. This implies the convergence of RMQ updates in $\mathcal{G}$ with heterogeneous $\eta_t^{(i)}$'s.

We now prove the gap bound for DAQ when $V(\cdot) = \ln(\cdot)$, when $\eta_t^{(i)} = \frac{\epsilon c_i}{a_i\sqrt{T}}$. By Theorem 2, r_i^* is equal to $1/a_i$ for logarithmic V. Moreover, because $\eta_t^{(i)}$ is constant over time, RMQ, with $\lambda_i = 1$, and DAQ updates coincide. Similarly to the convergence proof of RMQ, we consider the proxy RMQ updates in $\tilde{\mathcal{G}}$ with $\tilde{\eta}_t^{(i)} = \eta_t^{(0)} = \frac{1}{\sqrt{T}}$, for any agent i, $\tilde{h}_i(z) = \frac{z^2}{2\tilde{\lambda}_i}$, $\tilde{\lambda}_i = \alpha_i a_i$, and $\alpha_i = \frac{\epsilon c_i}{a_i}$. Applying [24, Thm. 6.2] and [24, Cor. 6.3] to these RMQ updates in $\tilde{\mathcal{G}}$ yields,

$$\overline{\mathrm{Gap}}_T(\mathrm{DAQ}) \leq \frac{1}{\sqrt{T}}\left(\tilde{\Omega} + \frac{\tilde{G}^2}{2\tilde{K}}\right), \tag{30}$$

where

$$\tilde{\Omega} \triangleq \max_{\boldsymbol{z}\in\mathcal{R}}\sum_i \tilde{h}_i(z_i) - \min_{\boldsymbol{z}\in\mathcal{R}}\sum_i \tilde{h}_i(z_i), \tag{31}$$

$$\tilde{K} \triangleq \min_{i\in\mathcal{I}}\frac{1}{\tilde{\lambda}_i}, \quad \tilde{G} \triangleq \sup_{\boldsymbol{z}\in\mathcal{R}}\left\|(\partial_i\tilde{\varphi}_i(\boldsymbol{z}))_{i\in\mathcal{I}}\right\|_2^2. \tag{32}$$

We bound these quantities as follows,

$$\tilde{G}^2 \leq \sum_{i\in\mathcal{I}}\frac{1}{a_i^2}G_i^2 \leq \sum_{i\in\mathcal{I}}\frac{4a_i^2}{a_i^2\epsilon^2} = \frac{4|\mathcal{I}|}{\epsilon^2}, \tag{33}$$

$$\frac{1}{\tilde{K}} = \max_i \tilde{\lambda}_i = \epsilon\max_i c_i, \quad \text{and } \tilde{\Omega} \leq \frac{1}{2\epsilon}\sum_{i\in\mathcal{I}} c_i^2. \tag{34}$$

Plugging these bounds in (30) yields the target result, which finishes the proof. ∎

While DAQ comes with standard no-regret guarantees and can therefore be viewed as a plausible behavioral model for repeated bidding, RMQ with adaptive step-sizes $(\eta_t^{(i)})_t$ does not generally enjoy regret guarantees. Nevertheless, RMQ acts as proxy to analyze the multi-agent behavior of DAQ when the step-sizes $\eta_t^{(i)}$ are time-invariant; in addition, RMQ can be interpreted as a distributed procedure for computing the unique Nash equilibrium of the stage game, as shown in Theorem 4.

The choice of $\eta_t^{(i)}$ optimizes the asymptotic dependency of the regret in the parameters problem when the budgets of other agents is unknown (see Remark 1). Under this choice of $\eta_t^{(i)}$, Theorem 4 provides an explicit finite-horizon bound of the deviation from equilibrium through the averaged gap $\overline{\mathrm{Gap}}_T(\mathrm{DAQ})$. The established bound highlights that convergence may deteriorate when the minimum admissible bid ϵ is small, since logarithmic utilities induce large gradients near 0. Finally, budget constraints introduce additional variability in the updates, which also contributes to slower convergence.

Now we study the case where all agents employ a myopic best response. This is a *simultaneous* best-response update: all agents revise in parallel from the last observed profile. Such parallel BR does not necessarily converge to a Nash equilibrium in general [32]. By contrast, in *unilateral* best-response updates, agents revise one at a time (cyclically or at random); in finite potential games, these dynamics converge to a pure Nash equilibrium [33].

Theorem 5. *If $V_i(\cdot) = a_i \ln(\cdot) + d_i$, $p_i(z) = z$, $\mathcal{A}_i = BR$ for all agents, and $\epsilon_i = \epsilon$ such that,*

$$\epsilon > \frac{1}{n-1}\left(\frac{(\sqrt{n}-1)^2}{\sqrt{n}} \max_{i\in\mathcal{I}} a_i - \delta\right). \tag{35}$$

then $(\boldsymbol{z}^{BR}(t))_t$ converges to the unique Nash equilibrium $\boldsymbol{z}^*$ linearly fast, i.e., $\exists \rho \in (0,1)$: $\|\boldsymbol{z}^{BR}(t) - \boldsymbol{z}^*\| \le \rho^t \|\boldsymbol{z}^{BR}(0) - \boldsymbol{z}^*\|$.

Proof: The best-response updates are fixed point iterations with the best-response operator $\boldsymbol{BR}$. We prove that this operator is a *contraction*, which yields convergence to the unique fixed point—which is also the NE of the stage game—at linear speed.

To prove that **BR** is a contraction, define $\widetilde{\mathrm{BR}}_i(s) = \left(-s + \sqrt{s^2 + 4a_i s}\right)/2$ and $\widetilde{\mathbf{BR}}(\boldsymbol{z}) \triangleq \left(\widetilde{\mathrm{BR}}_i(s_i(\boldsymbol{z}))\right)_{i\in\mathcal{I}}$, so that $\mathbf{BR}(\boldsymbol{z}) = \Pi_{\mathcal{R}}\left(\widetilde{\mathbf{BR}}(\boldsymbol{z})\right)$ (see (20)). Given that the projection map $\Pi_{\mathcal{R}}$ is 1-Lipschitz, $\widetilde{\mathrm{BR}}_i$ is smooth, and using the generalized mean-value theorem [34, Cor. 3.2], a sufficient condition for the best-response operator **BR** to be a contraction, is given by,

$$\sup_{z\in\mathcal{R}} \left\|\mathcal{J}_{\widetilde{\mathbf{BR}}}(\boldsymbol{z})\right\|_\infty < 1. \tag{36}$$

Direct calculations yield,

$$\left(\mathcal{J}_{\widetilde{\mathbf{BR}}}(\boldsymbol{z})\right)_{i,j} = \begin{cases}\zeta_i(s_i(\boldsymbol{z})), & j \neq i,\\ 0, & j = i.\end{cases} \tag{37}$$

where $\zeta_i(s) = -\frac{1}{2} + \frac{s+2a_i}{2\sqrt{s^2+4a_i s}}$, and $s_i(\boldsymbol{z}) = \sum_{j\neq i} z_j + \delta$. The function ζ_i is decreasing over $\mathbb{R}^+$, and thus,

$$\left\|\mathcal{J}_{\widetilde{\mathbf{BR}}}(\boldsymbol{z})\right\|_\infty = \max_{i\in\mathcal{I}} \sum_{j\neq i} \zeta_i(s_i(\boldsymbol{z})) = \max_{i\in\mathcal{I}} (n-1)\zeta_i(s_{\min}). \tag{38}$$

where $s_{\min} \triangleq (n-1)\epsilon + \delta$. Combining this with the fact that $(n-1)\zeta_i(s) < 1$ is satisfied whenever, $s > \frac{(\sqrt{n}-1)^2}{\sqrt{n}} a_i$ (see [27]), proves that (35) is indeed a sufficient condition for **BR** to be a contraction. This finishes the proof. ■

Note that the lower bound on the minimum bid in (35) scales inversely with the number of agents, i.e., $\epsilon = \mathcal{O}(1/n)$, and thus the convergence of best response dynamics is guaranteed with arbitrarily small minimum bid ϵ for large number of agents. Moreover, only $\mathcal{O}(\ln(1/p))$ iterations are needed to converge to a point whose distance from the NE is smaller than p.

TABLE I: Homogeneous dynamics: **Convergence iterations** in terms of the **fixed-point residual** r_t (threshold $< 10^{-5}$) under varying γ and n.

γ	n	BR	OGD$_V$	OGD$_F$	DAQ$_F$	DAQ$_V$	RRM$_V$
0	2	15	37	122	195	($r_T = 3.7\times10^{-2}$)	1682
	10	7	19	240	311	($r_T = 1.05$)	2291
	20	6	20	252	330	($r_T = 1.73$)	2361
5	2	15	111	127	184	($r_T = 1.0\times10^{-1}$)	504
	10	7	40	206	274	($r_T = 6.8\times10^{-1}$)	2182
	20	6	1811	221	289	($r_T = 7.5\times10^{-1}$)	2220
10	2	15	116	117	184	($r_T = 9.7\times10^{-2}$)	498
	10	8	533	185	253	($r_T = 4.5\times10^{-1}$)	1604
	20	8	533	194	259	($r_T = 4.5\times10^{-1}$)	2062

IV. Numerical Simulations

We consider the set of agents $\mathcal{I} = \{1, \dots, 10\}$. Each agent i's valuation function is $V_i(x) = a_i \ln x$, where $a_i > 0$ is an agent-specific parameter, and payment function $p_i(z) = z$. We set $\delta = 0.1$. Utilities heterogeneity is determined by γ by setting $a_i = \max(a - i\gamma, 1)$, with $\gamma \in \{0, 5, 10\}$, $a = 100$, and a budget constraint $c_i = c = 400$ and $\epsilon_i = \epsilon = 1$. The number of rounds in the repeated Kelly game is $T = 3000$, and results are averaged over 10 independent runs with random feasible initial bids. Agents follow one of the bidding algorithms described in Section III-C. Namely, OGD, DAQ, RRM, and BR. For OGD, DAQ, and RRM, we consider both fixed and time-varying learning rates: the fixed learning rate is $\eta^{(i)} = \frac{D_i}{G_i\sqrt{T}}$ and the time-varying one is $\eta_t^{(i)} = \frac{D_i}{G_i\sqrt{t}}$. We distinguish these variants by either adding the subscripts F for fixed or V for time-varying.

Our simulations include both *homogeneous dynamics* settings, where all agents follow the same update rule, and *heterogeneous dynamics* settings, where a fraction $\alpha_{\mathcal{A}_1}$ of agents use algorithm $\mathcal{A}_1$ while the remaining agents use $\mathcal{A}_2$.

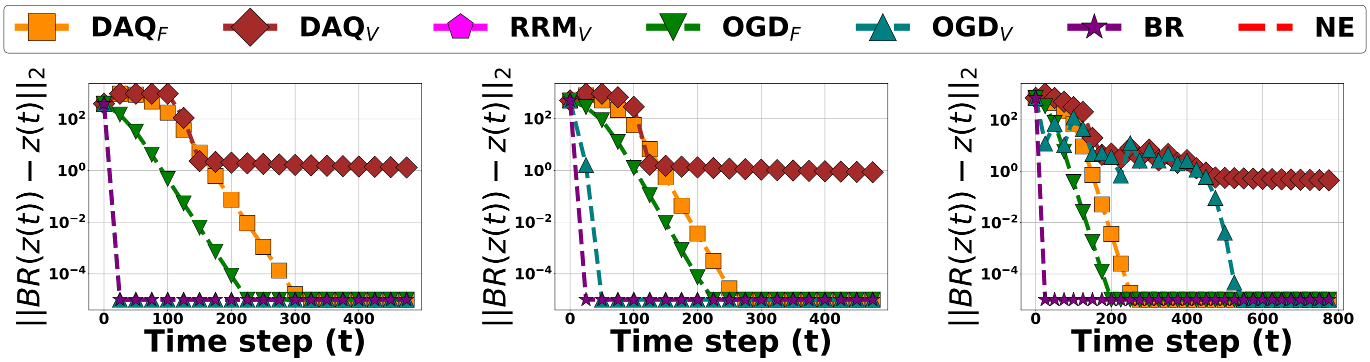
The objective is twofold: (1) to evaluate whether repeated play converges to the Nash equilibrium (NE) of the stage game while comparing convergence speeds; and (2) to compare the bidding algorithms in terms of time-average payoff.

To measure convergence to the NE, we use the metric r_t, defined as,

$$r_t \triangleq \|\mathbf{BR}(\boldsymbol{z}(t)) - \boldsymbol{z}(t)\|_2. \tag{39}$$

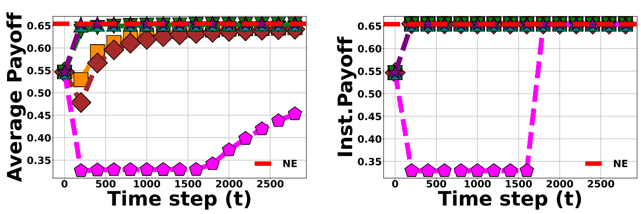
Indeed, the unique NE of the stage game is the fixed point of the best-response operator, i.e., $\mathbf{BR}(\boldsymbol{z}^*) = \boldsymbol{z}^*$. Accordingly, when the bid profile $\boldsymbol{z}(t)$ is near the NE, one expects $\mathbf{BR}(\boldsymbol{z}(t)) \approx \boldsymbol{z}(t)$.

To make payoffs comparable across agents, we normalize them to lie in $[0,1]$. Specifically, we define $\overline{\varphi_i}(\boldsymbol{z}) = \frac{\varphi_i(z)-\varphi_{\min}}{\varphi_{\max}-\varphi_{\min}}$ where $\varphi_{\min}$ and $\varphi_{\max}$ are the smallest and largest values of the payoff functions among all players and across all actions. We then compare the bidding algorithms using the time-average normalized payoff, i.e., $\frac{1}{T}\sum_{t=1}^{T} \overline{\varphi_i}(\boldsymbol{z}(t))$, which matches each player's objective of maximizing long-run payoff.



(a) $\gamma = 0$ (b) $\gamma = 5$ (c) $\gamma = 10$

Fig. 3: Convergence speed under homogeneous dynamics and varying payoff's heterogeneity levels.



(a) Time-average payoff. (b) Instantaneous payoff.

Fig. 4: Agent's payoff, $\gamma = 0$.

The rest of this section is organized as follows; Section IV-A addresses these questions under *homogeneous dynamics*, whereas Section IV-B addresses them under *heterogeneous dynamics*.

A. Homogeneous dynamics

Figure 3 reports the evolution of the metric r_t under homogeneous dynamics, for several values of γ. Repeated play with OGD, DAQ_F, or BR converges to the Nash equilibrium, which aligns with our theoretical results. In contrast, DAQ_V fails to converge with a similar precision. In terms of convergence speed, BR is the fastest, followed by OGD and DAQ_F, while OGD_V appears to be more sensitive to player heterogeneity (γ) than OGD_F.

Table I reports, for $n = |\mathcal{I}| \in \{2, 10, 20\}$, the minimum number of iterations needed to reach the threshold $r_t \leq 10^{-5}$. The results indicate that the convergence of BR becomes faster as the number of players increases, whereas the other bidding algorithms exhibit the opposite trend. This behavior is consistent with the $\mathcal{O}(n)$ convergence bound in Theorem 4. The method RRM converges significantly more slowly than the others. Finally, varying the heterogeneity level through γ has only a limited overall effect on convergence rates, though it slightly accelerates OGD_F, DAQ_F, and RRM.

Figure 4 reports, for a representative player, both the instantaneous payoff and the time-average payoff. Since $\gamma = 0$, all agents share the same payoff function, so the plotted curves are representative of any player. The figure shows that the ranking in payoff performance mirrors the ranking in convergence speed: best-response dynamics achieves the highest payoffs, followed by OGD, then DAQ, and finally RRM.

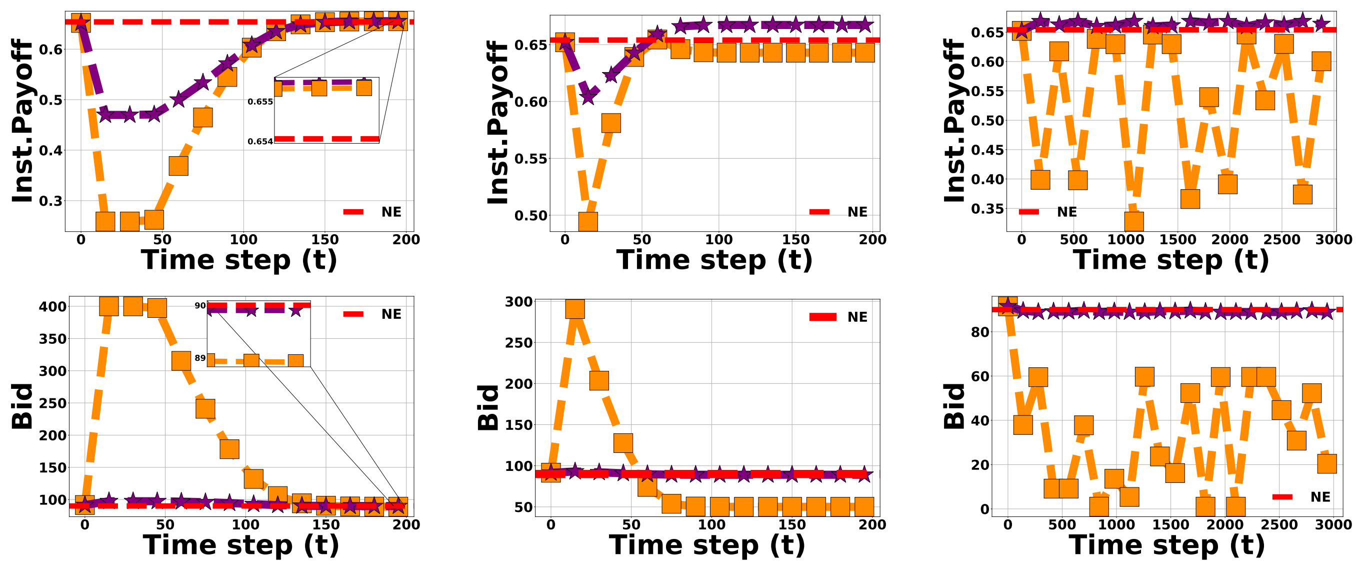
An additional observation is that, although DAQ_V does not attain the high-precision threshold in the residual metric r_t, it still achieves time-average payoffs comparable to its fixed-learning rate counterpart DAQ_F. Figures 4b and 3 suggest that, for both DAQ_V and RRM, reaching a moderately small r_t (e.g., $r_t < 1$) is already sufficient for the instantaneous payoff to be close to the Nash-equilibrium payoff; further reductions in r_t bring only marginal payoff improvements. This explains why DAQ_V performs well in terms of payoff despite not converging to very high precision. Finally, the poor payoff of RRM indicates that it is not an attractive update rule from a selfish perspective.

B. Heterogeneous dynamics

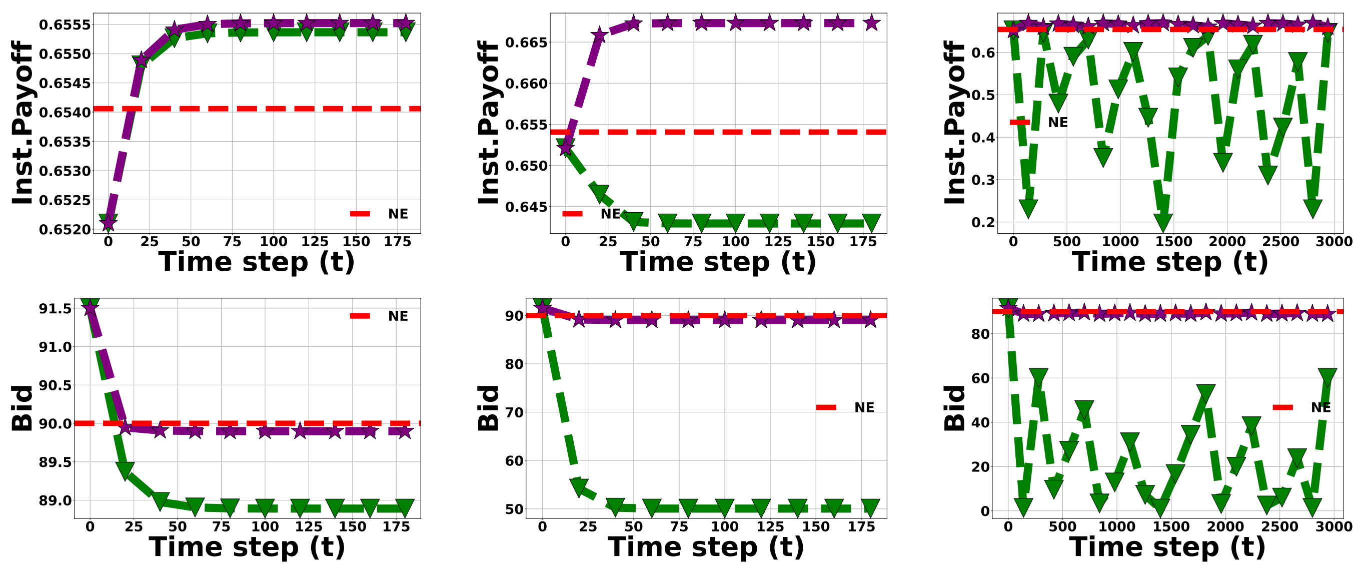
We consider now *heterogeneous dynamics* where a fraction $\alpha_{\mathcal{A}_1}$ of agents uses algorithm $\mathcal{A}_1$ while the remaining agents use $\mathcal{A}_2$ with $\gamma = 0$.

Figure 5 shows the evolution over time of the instantaneous bid and payoff of two representative agents—one using $\mathcal{A}_1$ and the other using $\mathcal{A}_2$—for $(\mathcal{A}_1, \mathcal{A}_2) \in \{(\text{BR}, \text{OGD}), (\text{BR}, \text{DAQ})\}$ and for $\alpha_{\text{BR}} \in \{10\%, 80\%, 90\%\}$. Similar results for the couple (OGD, DAQ) are available in the supplementary material.

Overall, heterogeneous dynamics dot not appear to converge to the stage-game NE. For $\alpha_{\text{BR}} \in \{10\%, 80\%\}$, the trajectories nevertheless settle to a steady regime. In the (BR, DAQ) regime, DAQ quickly saturates its budget constraint and exhibits a longer transient regime before stabilizing, whereas OGD adapts faster. This is consistent with DAQ aggregating gradients over time and OGD responding more strongly to recent feedback. For $\alpha_{\text{BR}} = 90\%$, both OGD and DAQ show persistent bid oscillations, while BR bids remain comparatively stable. Since BR agents form the large majority, the aggregate bid changes little so best responses vary only little because of the concavity of the payoff function (ln).



(a) BR vs. DAQ_F ($\alpha_{BR} = 10\%$)
(b) BR vs. DAQ_F ($\alpha_{BR} = 80\%$)
(c) BR vs. DAQ_F ($\alpha_{BR} = 90\%$)

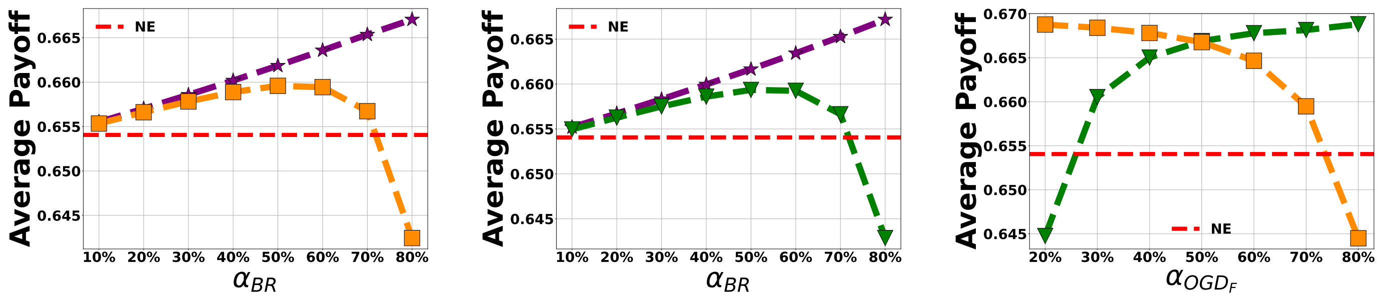


(d) BR vs. OGD_F ($\alpha_{BR} = 10\%$)
(e) BR vs. OGD_F ($\alpha_{BR} = 80\%$)
(f) BR vs. OGD_F ($\alpha_{BR} = 90\%$)

Fig. 5: Heterogeneous dynamics. In each sub-figure: instantaneous payoff (top) and bids (bottom).

In terms of instantaneous payoff, for $\alpha_{BR} \in \{10\%, 80\%\}$ the (BR, DAQ) configuration displays periods of low payoff for BR, but even lower for DAQ, due to the DAQ agents budget saturation. In contrast, with (BR, OGD) the system stabilizes more quickly, and these low payoff transients are much less pronounced. In both cases, the BR agent's payoff converges to a value close (and sometimes slightly above) the NE value, whereas the DAQ and OGD agents can fall below the NE payoff at $\alpha_{BR} = 80\%$. Finally, for $\alpha_{BR} = 90\%$, the oscillatory bids OGD and DAQ translate into large payoff fluctuations, with instantaneous payoffs repeatedly dropping from the NE-payoff to significanlty lower values.

Figure 6 reports the average payoff of two representative agents—one using $\mathcal{A}_1$ and the other using $\mathcal{A}_2$—for $(\mathcal{A}_1, \mathcal{A}_2) \in \{(BR, OGD), (BR, DAQ), (OGD, DAQ)\}$ and multiple values of $\alpha_{\mathcal{A}_1}$. Overall, heterogeneous play can yield time-average payoffs that are slightly above or below the stage-game NE payoff, sometimes benefiting one group more than the other. However, these deviations remain small: across all mixtures, observed payoffs lie in $[0.64, 0.67]$, while the NE payoff is roughly 0.655. Both OGD and DAQ exhibit similar trends: when they represent less than about 30% of the population, their time-average payoff fall below the NE payoff, whereas for larger fractions it can exceed it. In contrast, BR achieves payoffs consistently above the NE, with its highest values when α_{BR} is large. These results suggest that payoff differences across policies are small in this setting, although BR appears slightly preferable despite lacking no-regret



(a) $\mathtt{DAQ_F}$ vs. $\mathtt{BR}$ (b) $\mathtt{OGD_F}$ vs. $\mathtt{BR}$ (c) $\mathtt{DAQ_F}$ vs. $\mathtt{OGD_F}$

Fig. 6: Heterogeneous dynamics: Average payoff

guarantees.

V. Conclusion

In this paper, we studied the game induced by competition among agents in a proportional allocation auction. We derive a sufficient condition under which the game satisfies Rosen's *diagonal strict concavity* (DSC), and show that this condition holds in particular when agents have logarithmic utilities in their allocated share. We further relate these utilities to bandwidth allocation problems in which the objective is to balance fairness and throughput. We then consider the repeated version of the game, where all agents update their bids using either best-response dynamics or classical no-regret learning algorithms, namely Online Gradient Descent and Dual Averaging. Leveraging DSC, we establish convergence of these homogeneous dynamics to the stage-game Nash equilibrium.

As future work, we plan to characterize the system behavior under heterogeneous update rules. We also aim to extend our analysis to more general utility functions, such as α-fair utilities (with parameter α controlling the fairness-efficiency trade-off), and to settings with multiple heterogeneous resources.

References

[1] R. Johari, *Efficiency Loss in Market Mechanisms for Resource Allocation*. PhD thesis, Department of Electrical Engineering and Computer Science, 2004.

[2] R. T. Maheswaran and T. Basar, "Efficient signal proportional allocation (ESPA) mechanisms: decentralized social welfare maximization for divisible resources," *IEEE JSAC.*, vol. 24, no. 5, 2006.

[3] P. Maillé and B. Tuffin, "Multi-bid auctions for bandwidth allocation in communication networks," in *IEEE INFOCOM*, 2004.

[4] X. Wang *et al.*, "A distributed truthful auction mechanism for task allocation in mobile cloud computing," *IEEE Trans. Serv. Comput.*, no. 3, 2021.

[5] W. Saad *et al.*, "A noncooperative game for double auction-based energy trading between phevs and distribution grids," in *IEEE SmartGridComm*, 2011.

[6] L. Zheng *et al.*, "Auction-based order dispatch and pricing in ridesharing," in *35th IEEE ICDE*, IEEE, 2019.

[7] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, 1997.

[8] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Math. Oper. Res.*, 2004.

[9] S. Yang and B. E. Hajek, "Vcg-kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals," *IEEE J. Sel. Areas Commun.*, 2007.

[10] V. Syrgkanis and É. Tardos, "Composable and efficient mechanisms," in *STOC*, pp. 211–220, ACM, 2013.

[11] I. Caragiannis and A. A. Voudouris, "Welfare guarantees for proportional allocations," *Theory Comput. Syst.*, vol. 59, 2016.

[12] I. Caragiannis and A. A. Voudouris, "The efficiency of resource allocation mechanisms for budget-constrained users," in *EC*, pp. 681–698, ACM, 2018.

[13] E. Even-Dar, Y. Mansour, and U. Nadav, "On the convergence of regret minimization dynamics in concave games," in *STOC*, pp. 523–532, 2009.

[14] S. D'Oro *et al.*, "Auction-based resource allocation in openflow multi-tenant networks," *Comput. Networks*, 2017.

[15] E. Elkind, A. Ghosh, and P. W. Goldberg, "Continuous-time best-response and related dynamics in tullock contests with convex costs," *arXiv preprint arXiv:2402.08541*, 2024.

[16] J. D. Pérez-Castrillo and T. Verdier, "A general analysis of rent-seeking games," *Public choice*, vol. 73, no. 3, pp. 335–350, 1992.

[17] A. Héliou, J. Cohen, and P. Mertikopoulos, "Learning with bandit feedback in potential games," in *NIPS*, pp. 6369–6378, 2017.

[18] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave N-person games," *Econometrica*, vol. 33, July 1965.

[19] M. Datar, E. Altman, F. D. Pellegrini, R. E. Azouzi, and C. Touati, "A mechanism for price differentiation and slicing in wireless networks," in *WiOPT*, pp. 121–128, IEEE, 2020.

[20] H. Kim and Y. Han, "A proportional fair scheduling for multicarrier transmission systems," *IEEE Commun. Lett.*, 2005.

[21] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends® in Optimization*, 2016.

[22] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *ICML*, pp. 928–936, 2003.

[23] H. B. McMahan, "A survey of algorithms and analysis for adaptive online learning," *J. Mach. Learn. Res.*, vol. 18, 2017.

[24] P. Mertikopoulos and Z. Zhou, "Learning in games with continuous action sets and unknown payoff functions," *Math. Program.*, vol. 173, no. 1-2, pp. 465–507, 2019.

[25] P. Mertikopoulos *et al.*, "A unified stochastic approximation framework for learning in games," *Math. Program.*, 2024.

[26] L. Orseau, "A regret bound for online gradient descent with momentum," tech. rep., Personal Technical Report, June 2025. Available at: https://laurent-orseau.com/docs/OGDM_regret.pdf.

[27] C. M. Mboulou-Moutoubi, Y. Ben Mazziane, F. De Pellegrini, and E. Altman, "Best-response learning in budgeted α-fair kelly mechanisms," in *NETGCOOP*, pp. 90–99, 2025.

[28] Z. Zhou, P. Mertikopoulos, *et al.*, "Robust power management via learning and game design," *Oper. Res.*, vol. 69, no. 1, 2021.

[29] F. De Pellegrini, A. Massaro, L. Goratti, and R. El-Azouzi, "Bounded generalized Kelly mechanism for multi-tenant caching in mobile edge clouds," in *NetGCoop*, 2017.

[30] Y. Yang, R. T. B. Ma, and J. C. S. Lui, "Price differentiation and control in the kelly mechanism," *Perform. Evaluation*, 2013.

[31] Y. B. Mazziane, C.-m. Mboulou-Moutoubi, F. De Pellegrini, and E. Altman, "Learning to bid in proportional allocation auctions with budget constraints," in *2025 23rd WiOpt*, pp. 1–8, 2025.

[32] S. Hart and A. Mas-Colell, "Uncoupled dynamics do not lead to nash equilibrium," *American Economic Review*, 2003.

[33] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[34] R. Coleman, *Calculus on Normed Vector Spaces*. Springer, 2012.

VI. Supplementary Material

A. Proof of Theorem 2

Define the functions f_i, g_i, and $\psi_{\boldsymbol{r},\boldsymbol{V}}$ as

$$f_i(x) = (1-x)^2 \ddot{V}_i(x) - 2(1-x)\dot{V}_i(x), \tag{40}$$

$$g_i(x) = -x(1-x)\ddot{V}_i(x) + (2x-1)\dot{V}_i(x), \tag{41}$$

$$\psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) \triangleq \left(\sum_{i\in\mathcal{I}} \frac{r_i g_i(x_i)^2}{k_i(x_i)}\right)\left(\sum_{i\in\mathcal{I}} \frac{1}{r_i k_i(x_i)}\right), \tag{42}$$

where $k_i(x) = g_i(x) - f_i(x) + \delta^2 L_i$ and $L_i = \min_{z_i\in\mathcal{R}_i} \ddot{p}_i(z_i)$. Further define the set $\Delta = \{\boldsymbol{x} > \boldsymbol{0} : \sum_{i\in\mathcal{I}} x_i \leq \frac{\sum_{k\in\mathcal{I}} c_k}{\sum_{k\in\mathcal{I}} c_k + \delta}\}$.

We prove that,

$$\exists \boldsymbol{r} > \boldsymbol{0} : \sup_{\boldsymbol{x}\in\Delta} \psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) < 1 \implies \mathcal{G} \text{ is } \textbf{SDSC}(\boldsymbol{r}), \tag{43}$$

or equivalently we prove that the matrix $\boldsymbol{H_r}(\boldsymbol{z})$ is negative definite whenever $\sup_{\boldsymbol{x}\in\Delta} \psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) < 1$.

The second-order partial derivatives of the payoff functions $\partial^2_{i,j}\varphi_i$ can be written as

$$\partial^2_{i,j}\varphi_i(\boldsymbol{z}) = \partial^2_{i,j} U_i(\boldsymbol{z}) - \ddot{p}_i(z_i), \tag{44}$$

where $U_i(\boldsymbol{z}) = V_i(x_i(\boldsymbol{z}))$, and $\ddot{p}_i$ is the second derivative of p_i. Note that p_i is convex, which implies it contributes to the matrix $\boldsymbol{H_r}(\boldsymbol{z})$ as a diagonal matrix whose entries are $-2r_i\ddot{p}_i(z_i) \leq 0$ and thus is negative semi-definite. Consequently, any failure of $\boldsymbol{H_r}(\boldsymbol{z})$ to be negative definite would stem from the matrix formed by $\partial^2_{i,j} U_i(\boldsymbol{z})$. Lemma 1 shows that this matrix has a particular structure.

Lemma 1. *The partial derivatives of U_i verify,*

$$m(z)\partial^2_{i,j} U_i(\boldsymbol{z}) = \begin{cases} f_i(x_i(\boldsymbol{z})), & \textit{if } i = j, \\ g_i(x_i(\boldsymbol{z})), & \textit{otherwise.} \end{cases} \tag{45}$$

Proof: The calculation of the derivative of U_i with respect to z_i writes

$$\partial_i U_i(z_i) = \frac{\sum_{j\neq i}^N z_j + \delta}{\left(\sum_{l=1}^N z_l + \delta\right)^2} \dot{V}_i\left(\frac{z_i}{\sum_{l=1}^N z_l + \delta}\right). \tag{46}$$

The elements on the diagonal can be written as

$$\begin{aligned}
\partial^2_{i,i} U_i(\boldsymbol{z}) &= \frac{-2\sum_{l\neq i}^N z_i + \delta}{(\sum_{l=1}^N z_l + \delta)^3} \dot{V}_i\left(\frac{z_i}{\sum_{l=1}^N z_l + \delta}\right) \\
&\quad + \left(\frac{\sum_{l\neq i}^N z_l + \delta}{(\sum_{l=1}^N z_l + \delta)^2}\right)^2 \ddot{V}_i\left(\frac{z_i}{\sum_{l=1}^N z_l + \delta}\right) \\
&= \frac{\left[(1-x_i(\boldsymbol{z}))^2 \ddot{V}_i(x_i(\boldsymbol{z})) - 2(1-x_i(\boldsymbol{z}))\dot{V}_i(x_i(\boldsymbol{z}))\right]}{\left(\sum_{i=1}^N z_j + \delta\right)^2} \\
&= \frac{f_i(x_i(\boldsymbol{z}))}{m(z)^2}.
\end{aligned} \tag{47}$$

In a similar fashion we obtain the off-diagonal entries

$$\begin{aligned}
\partial^2_{i,j} U_i(\boldsymbol{z}) &= \frac{\sum_{l=1}^N z_l + \delta - 2\left(\sum_{l\neq i}^N z_l + \delta\right)}{\left(\sum_{l=1}^N z_l + \delta\right)^3} \dot{V}_i\left(\frac{z_i}{\sum_{l=1}^N z_l + \delta}\right) \\
&\quad - \frac{\sum_{l\neq i}^N z_l + \delta}{\left(\sum_{l=1}^N z_l + \delta\right)^4} \cdot z_i \cdot \ddot{V}_i\left(\frac{z_i}{\sum_{l=1}^N z_l + \delta}\right) \\
&= \frac{\left[-\ddot{V}_i(x_i(\boldsymbol{z}))x_i(\boldsymbol{z})(1-x_i(\boldsymbol{z})) + \dot{V}_i(x_i(\boldsymbol{z}))(2x_i(\boldsymbol{z})-1)\right]}{\left(\sum_{i=1}^N z_j + \delta\right)^2} \\
&= \frac{g_i(x_i(\boldsymbol{z}))}{m(z)^2}
\end{aligned} \tag{48}$$

which concludes the calculation. ■

Lemma 1 shows that the matrix of second partial derivatives of the functions U_i can be rewritten using a change of variable from $\boldsymbol{z} \in \mathcal{R}$ to $\boldsymbol{x}(\boldsymbol{z})$. Moreover, this reformulated matrix can be decomposed into the sum of a diagonal matrix, with entries $f_i(x_i) - g_i(x_i)$, and a rank-one matrix, whose entries are given by $g_i(x_i)$. Lemma 2 leverages this particular structure to prove (43).

Lemma 2. $\sup_{\boldsymbol{x}\in\Delta} \psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) < 1 \implies \boldsymbol{H_r}(\boldsymbol{z})$ *is negative definite for any* $\boldsymbol{z}$.

Proof: Let $\boldsymbol{r} > \boldsymbol{0}$. We assume that $\sup_{\boldsymbol{x}\in\Delta} \psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) < 1$ and prove that the matrix $\boldsymbol{H_r}(\boldsymbol{z})$ is negative definite for every $\boldsymbol{z} \in \mathcal{R}$. To this end, we introduce the following notation. We define the vectors $\boldsymbol{g} = (g_i(x_i(\boldsymbol{z})))_{i=1}^n$ and $\boldsymbol{f} = (f_i(x_i(\boldsymbol{z})))_{i=1}^n$, and set $\tilde{k}_i(\boldsymbol{z}) = g_i(x_i(\boldsymbol{z})) - f_i(x_i(\boldsymbol{z})) + m(\boldsymbol{z})\ddot{p}_i(z_i)$, so that $\tilde{\boldsymbol{k}} = (\tilde{k}_i(\boldsymbol{z}))_{i=1}^n$. Here, $\odot$ denotes the Hadamard (elementwise) product, $\boldsymbol{1}$ is the vector of ones, and $\Lambda(\boldsymbol{a})$ represents the diagonal matrix whose diagonal entries are the components of $\boldsymbol{a}$.

Thanks to Lemma 1, the matrix can be expressed as

$$m(\boldsymbol{z}) \cdot (\boldsymbol{H}_r(\boldsymbol{z}))_{i,j} = \begin{cases} 2r_i\Big(f_i(x_i(\boldsymbol{z})) - m(\boldsymbol{z})\,\ddot{p}_i(z_i)\Big), & \text{if } i = j, \\ r_i\, g_i(x_i(\boldsymbol{z})) + r_j\, g_j(x_j(\boldsymbol{z})), & \text{if } i \neq j. \end{cases} \tag{49}$$

Thus, the matrix $\boldsymbol{H}_r(\boldsymbol{z})$ can be written compactly as

$$\boldsymbol{H_r}(\boldsymbol{z}) = -2\,\Lambda\Big(\tilde{\boldsymbol{k}} \odot \boldsymbol{r}\Big) + \Big(\boldsymbol{g} \odot \boldsymbol{r}\Big)\boldsymbol{1}^\top + \boldsymbol{1}\Big(\boldsymbol{g} \odot \boldsymbol{r}\Big)^\top. \tag{50}$$

Notice that $\tilde{k}_i(\boldsymbol{z}) \geq k_i(x_i(\boldsymbol{z}))$. Moreover, replacing by the expressions of g_i and f_i in (24) and (23), we obtain,

$$k_i(x_i) = (x_i - 1)\,\ddot{V}_i(x_i) + \dot{V}_i(x_i) + \delta^2 \min_{z_i \in \mathcal{R}_i} \ddot{p}_i(z_i). \tag{51}$$

By Assumption 1, we have $\dot{V}_i > 0$, $\ddot{V}_i < 0$, and $\ddot{p}_i > 0$. Moreover, since $x_i(\boldsymbol{z}) < 1$ (as $\boldsymbol{x} \in \Delta$), it follows that $k_i(x_i(\boldsymbol{z})) > 0$. Define also the vector $\boldsymbol{k} = \big(k_i(x_i(\boldsymbol{z}))\big)_{i=1}^N$.

Let $\boldsymbol{v} \in \mathbb{R}^N$. Using the expression of $\boldsymbol{H}_r(\boldsymbol{z})$, we have

$$\begin{aligned} \boldsymbol{v}^T \boldsymbol{H_r}(\boldsymbol{z})\boldsymbol{v} &= -2\left(\sum_{i=1}^n v_i^2\, r_i\, \tilde{k}_i(\boldsymbol{z}) - \langle \boldsymbol{v}, \boldsymbol{g} \odot \boldsymbol{r}\rangle\langle \boldsymbol{v}, \boldsymbol{1}\rangle\right) \\ &\leq -2\left(\sum_{i=1}^n v_i^2\, r_i\, k_i(x_i(\boldsymbol{z})) - \langle \boldsymbol{v}, \boldsymbol{g} \odot \boldsymbol{r}\rangle\langle \boldsymbol{v}, \boldsymbol{1}\rangle\right), \end{aligned} \tag{52}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product.

We now rewrite the inner products in (52) by noting that, for any vector $\boldsymbol{a}$, the notation $\sqrt{\boldsymbol{a}}$ represents the vector obtained by taking the square root of each component of $\boldsymbol{a}$, and the ratio of two vectors is computed elementwise. In particular,

$$\langle \boldsymbol{v}, \boldsymbol{g} \odot \boldsymbol{r}\rangle = \left\langle \boldsymbol{v} \odot \sqrt{\boldsymbol{k} \odot \boldsymbol{r}},\, \frac{\sqrt{\boldsymbol{r}} \odot \boldsymbol{g}}{\sqrt{\boldsymbol{k}}} \right\rangle, \tag{53}$$

$$\langle \boldsymbol{v}, \boldsymbol{1}\rangle = \left\langle \boldsymbol{v} \odot \sqrt{\boldsymbol{k} \odot \boldsymbol{r}},\, \frac{\boldsymbol{1}}{\sqrt{\boldsymbol{k} \odot \boldsymbol{r}}} \right\rangle. \tag{54}$$

Applying the Cauchy-Schwarz inequality to these expressions and substituting into (52) yields

$$\begin{aligned} \boldsymbol{v}^T \boldsymbol{H_r}(\boldsymbol{z})\boldsymbol{v} &\leq -2\sum_{i=1}^n v_i^2\, r_i\, k_i(x_i(\boldsymbol{z})) \\ &+ 2\left(\sum_{i=1}^n v_i^2 r_i k_i(x_i(\boldsymbol{z}))\right)\sqrt{\sum_{i=1}^n \frac{r_i\, g_i(x_i(\boldsymbol{z}))^2}{k_i(x_i(\boldsymbol{z}))}}\sqrt{\sum_{i=1}^n \frac{1}{r_i k_i(x_i(\boldsymbol{z}))}} \\ &= 2\underbrace{\sum_{i=1}^n v_i^2 r_i k_i(x_i(\boldsymbol{z}))}_{>0}\left(-1 + \sqrt{\sum_{i=1}^n \frac{r_i g_i(x_i(\boldsymbol{z}))^2}{k_i(x_i(\boldsymbol{z}))} \sum_{i=1}^n \frac{1}{r_i k_i(x_i(\boldsymbol{z}))}}\right), \end{aligned}$$

which concludes the proof. ■

Assume that $V_i(x) = a_i \log(x) + d_i$, $\delta > 0$, and $r_i = a_i^{-1}$. Straightforward calculations show that $g_i(x) = a_i$ and $k_i(x) = a_i/x^2 + \delta^2 L_i$. Thus $\psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x})$ verifies,

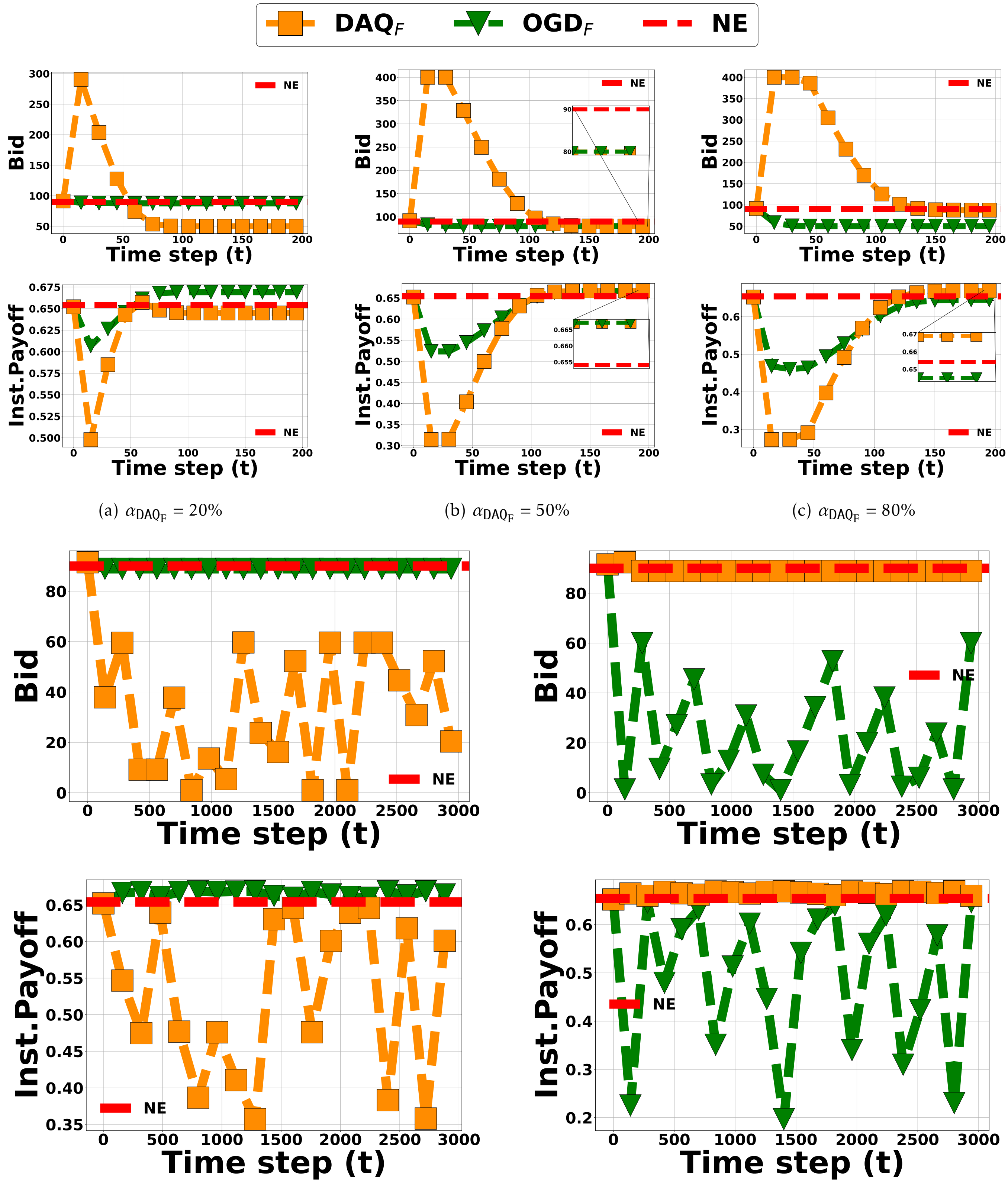
$$\psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) = \left(\sum_{i=1}^{n} \frac{a_i}{a_i/x_i^2 + \delta^2 L_i)}\right)\left(\sum_{i=1}^{n} \frac{1}{1/x_i^2 + \delta^2 a_i^{-1} L_i}\right) \tag{55}$$

$$\leq \left(\sum_{i=1}^{n} x_i^2\right)^2 \leq \left(\sum_{i=1}^{n} x_i\right)^4 \leq \left(\frac{C}{C+\delta}\right)^4 < 1. \tag{56}$$

Thus the condition $\sup_{\boldsymbol{x}\in\Delta} \psi_{\boldsymbol{r},\boldsymbol{V}}(\boldsymbol{x}) < 1$ is satisfied for logarithmic utilities. This finishes the proof.

B. Additional experiments

We consider now *heterogeneous dynamics* where a fraction $\alpha_{\texttt{DAQ}}$ of agents uses DAQ while the remaining agents use OGD with $\gamma = 0$. Figures 7 shows the evolution over time of the instantaneous bid and payoff of two representative agents—one using OGD and the other using DAQ—for $\alpha_{\texttt{DAQ}} \in \{10\%, 20\%, 50\%, 80\%, 90\%\}$. Convergence is observed only for $\alpha_{\texttt{DAQ}} \in [20\%, 80\%]$. In this regime, the algorithm used by the majority of agents achieves a utility above the NE-utility, while the minority attains a lower payoff. The dynamics are nearly symmetric for $\alpha_{\texttt{DAQ}} = 50\%$. On the other hand, for extreme splits $\alpha_{\mathcal{A}_1} \in \{10\%, 90\%\}$, we observe oscillations.



(a) $\alpha_{\mathtt{DAQ_F}} = 20\%$

(b) $\alpha_{\mathtt{DAQ_F}} = 50\%$

(c) $\alpha_{\mathtt{DAQ_F}} = 80\%$

(d) $\alpha_{\mathtt{DAQ_F}} = 10\%$

(e) $\alpha_{\mathtt{DAQ_F}} = 90\%$

Fig. 7: Heterogeneous dynamics: $\mathtt{DAQ_F}$ vs. $\mathtt{OGD_F}$. In each subfigure: bids (top) and instantaneous payoff (bottom).