Online Learning of Partitions in Additively Separable Hedonic Games

Saar Cohen and **Noa Agmon**
Department of Computer Science, Bar-Ilan University, Israel
saar30@gmail.com, agmon@cs.biu.ac.il

Abstract

Coalition formation involves partitioning agents into disjoint coalitions based on their preferences over other agents. In reality, agents may lack enough information to assess their preferences *before* interacting with others. This motivates us to initiate the research on coalition formation from the viewpoint of *online learning*. At each round, a possibly different subset of a given set of agents arrives, that a *learner* then partitions into coalitions. Only afterwards, the agents' preferences, which possibly change over time, are revealed. The *learner*'s goal is optimizing social cost by minimizing his (static or dynamic) regret. We show that even no-static regret is hard to approximate, and *constant* approximation in polynomial time is unattainable. Yet, for a *fractional* relaxation of our problem, we devise an algorithm that simultaneously gives the *optimal* static and dynamic regret. We then present a rounding scheme with an *optimal* dynamic regret, which converts our algorithm's output into a solution for our original problem.

1 Introduction

A set of computer science students initiates their Bachelor's degree, eager to enhance their learning experience and academic successes by forming study groups during the semester. Initially, with limited information about other students' skills and work habits, the students change their preferences through interactions, making informed decisions to join study groups aligned with their working styles and academic goals. This adaptive process enables them to optimize collaborative learning, leveraging the strengths of diverse groups. The composition of study groups may also evolve over time as students' preferences adapt and factors like schedules and health-related issues affect physical attendance at the university. Such situations and many other real-life scenarios noticed in our social, economic, and politic life, fall within the phenomenon of *coalition formation*, where *agents* perform activities in *coalitions* rather than on their own.

A popular framework for studying coalition formation is that of *hedonic games* [Drèze and Greenberg, 1980], which disregards *externalities*, i.e., agents' utilities solely depend on the coalition they are part of. The outcome of such games is a set of disjoint coalitions (hereafter, *partition*). Most of the hedonic games literature considers an *offline* setting, where a *single* game is fully available upfront. Yet, in many realistic cases as our study groups example (see, e.g., [Cohen and Agmon, 2023a]), not all agents are initially present, but different subsets of them arrive over time, while their preferences toward other agents are initially unknown and may change.

In this paper, we present and study a *new* model capturing such real-world scenarios in coalition formation from the perspective of *online learning*. At each round, a possibly different subset of a given set of agents arrives, that a *learner* then partitions into coalitions. Only afterwards, the agents' preferences, which possibly alter across time, are revealed. For performance evaluation, we focus on the classic metric of *regret* [Shalev-Shwartz, 2012]. Prior works on online learning typically consider the *static* regret [Hazan, 2016], evaluating the learner's decisions against an optimal fixed offline decision. Yet, it is not a suitable measure in *dynamic* settings such as ours, where the best decision may change frequently. As a remedy, recent studies focus on *dynamic regret* [Yang *et al.*, 2016], comparing the *learner*'s decisions against the decisions optimal for *individual* time steps. In our model, the *learner*'s goal is thus minimizing (*static* or *dynamic*) regret.

Our study aims to characterize what the *learner* can attain in terms of *static* and *dynamic* regret. First, we prove that no-*static* regret is hard to approximate. We show that, if we have an efficient learner, then we can devise an efficient algorithm that approximately solves an *offline* variant of our problem, proven to be hard to approximate due to Bilò *et al.* [2022]. Our result also indicates *constant* approximation to no-*static* regret is unattainable in polynomial time. Surprisingly, this result does *not* apply to the more stringent *dynamic* regret. To overcome our hardness result, we consider a *fractional* relaxation of our problem, where agents are *fractionally* assigned to multiple coalitions. We prove that such a relaxation allows us to obtain a best of both (regret) worlds algorithm that concurrently attains the ***optimal*** static and dynamic regret. Our algorithm is derived from a primal-dual formulation for our problem's *fractional* relaxation. Finally, we provide our main technical contribution: we supply a randomized rounding scheme proven to have an ***optimal*** dynamic regret, which converts the *fractional* solution produced by our above algorithm into an *integral* one. All omitted proofs can be found in

the supplementary materials [Cohen and Agmon, 2024].

2 Related Work

Hedonic games have been introduced by Drèze and Greenberg [1980], and later expanded to the study of various solution concepts such as stability, fairness, and optimality (see, e.g., [Aziz and Savani, 2016; Woeginger, 2013]). One major concern is designing computationally manageable classes of hedonic games, which led to an abundance of game representations. Some are *ordinal* and can *fully* express any preference over coalitions [Bouveret *et al.*, 2010; Elkind and Wooldridge, 2009], yet may require exponential space. In contrast, *cardinal* hedonic games, based on weighted graphs [Aziz *et al.*, 2019; Bogomolnaia and Jackson, 2002], are *not* fully expressive, but only require *polynomial* space for reasonable weights. Our work focuses on *additively separable hedonic games* [Bogomolnaia and Jackson, 2002], wherein a large body of research considers stability notions [Banerjee *et al.*, 2001; Bogomolnaia and Jackson, 2002; Aziz *et al.*, 2011; Ballester, 2004], while we regard economic efficiency [Aziz *et al.*, 2015; Bullinger, 2020; Elkind *et al.*, 2020].

Our work is closely related to the *online* version of hedonic games introduced by Flammini *et al.* [2021], where agents arrive one at a time and should be *immediately* and *irrevocably* assigned to coalitions with the goal of maximizing social cost. This problem was then studied by Bullinger and Romen [2023], for which they also recently considered various *stability* concepts [2024]. Notably, requiring that algorithms exactly know the agents' preferences *before* making decisions limits practicality as agents may lack enough information to assess their preferences before interacting with others as in our study groups example. In fact, social interactions are complex and relationships often require time to develop. Those works also make a *single*, *immediate* and *irrevocable* assignment for each agent, while assuming that agents' preferences do not change. In contrast, we consider realistic scenarios where preferences dynamically evolve based on interactions, allowing agents to adapt to changing situations. Our repeated game nature allows agents to learn the coalitions proven most relevant and effective for them.

We offer a novel framework that mitigates those issues by studying coalition formation from the viewpoint of *online learning*. Our study contributes to the research trend on online learning in combinatorial domains such as online task allocation [Cohen and Agmon, 2023b], submodular optimization [Hazan and Kale, 2012; Krause and Golovin, 2014], matching markets [Zhang *et al.*, 2022; Maheshwari *et al.*, 2022], clustering [Christou *et al.*, 2024; Fotakis *et al.*, 2021], bandits [Tekin and Van Der Schaar, 2015], online inventory problems [Hihat *et al.*, 2023] and many more. Our fractional relaxation technique is a common approach for handling impossibilities, proven fruitful for combinatorial online learning problems (see, e.g., [Gergatsouli and Tzamos, 2022]), and many other problems such as matchings [Aziz and Klaus, 2019] and fair division [Amanatidis *et al.*, 2023]. Our algorithm for this relaxation is also related to online learning primal-dual methods, which were applied in, e.g., bandits [Li *et al.*, 2021; Tirinzoni *et al.*, 2020]. Note that primal-dual analysis is also used in many other (online) domains beyond online learning (e.g., online matchings [Ekbatani *et al.*, 2023]). Yet, as far as we know, we are the *first* to study online learning in *hedonic games*, providing algorithms proven to have *optimal* regret.

In this context, most related to our work are *matching markets*, which can be seen as a *constrained* variant of hedonic games where coalitions are restricted to be of size at most 2. Unlike our work, the input instances are also limited to be *bipartite*, while either only *one* side must have preferences over the other one [Maheshwari *et al.*, 2022], or both sides have preferences over each other [Jagadeesan *et al.*, 2021; Zhang *et al.*, 2022]. Some works study learning matchings with monetary transfers [Liu *et al.*, 2020; Sankararaman *et al.*, 2021]. Yet, our setting is more challenging as it is *not* limited to matchings and monetary transfers are *not* available.

Those studies and ours are also connected to *online convex optimization* (see [Shalev-Shwartz, 2012; Hazan, 2016] for surveys). Its classic performance measure is the *static* regret, evaluating the learner's decisions against a *single* optimal decision in hindsight. Over the past decades, various algorithms, such as online gradient descent (OGD) [Zinkevich, 2003; Hazan *et al.*, 2007], have been proposed to yield (optimal) sub-linear regret under different scenarios. However, in real-life scenarios such as our *dynamic* context the optimal decision drifts over time. Hence, algorithms that guarantee decisions close to a *static* one may perform poorly in *dynamic* settings. As a remedy, the *dynamic* regret has recently become a popular metric [Zinkevich, 2003; Besbes *et al.*, 2015; Zhang *et al.*, 2017], comparing the learner's overall loss against a comparator sequence of solutions optimal for *individual* time steps. Due to the arbitrary fluctuation in the loss functions, it is well-known that the worst-case dynamic regret scales linearly in the number of rounds T [Yang *et al.*, 2016], unless some restrictions are imposed [Jadbabaie *et al.*, 2015].

Yet, the dynamic regret can be bounded by certain regularities of the comparator or function sequences. A natural regularity is the *path-length* [Zinkevich, 2003], denoted as $\mathcal{S}_\star^T$, reflecting the fluctuation of online optimal decisions. When known in advance, Zinkevich [2003] shows OGD has a dynamic regret of $O(\sqrt{T(1+\mathcal{S}_\star^T)})$ for convex functions, that can be improved to $O(\mathcal{S}_\star^T)$ for smooth functions that are either strongly convex [Mokhtari *et al.*, 2016], or convex with minimizers lying in the decision set's interior [Yang *et al.*, 2016]. For the latter, Yang *et al.* [2016] prove that $O(\mathcal{S}_\star^T)$ dynamic regret is *optimal*. Another regularity is the *temporal variability* [Besbes *et al.*, 2015; Baby and Wang, 2019], capturing the function values' variation.

In our work, we present the *interaction term*, a *novel* regularity measuring the variation in agents' interactions, which holds the potential of expanding the research on online learning within computational social choice in general. Further, our problem falls under the *learning from expert advice* problem, wherein the well-known multiplicative weights update (MWU) method is no-regret [Hazan, 2016]. However, MWU cannot be applied to our setting due to its demanding time and space requirements: MWU maintains different weights

for all possible partitions, which form a space exponential in the number of agents. This required us to devise new methods that can solve our problem efficiently.

3 Online Learning in Hedonic Games

We consider an *online learning* variant of additively separable hedonic games (ASHGs), where different subsets of a given agents set dynamically arrive across T rounds, while possibly changing their preferences over time. We study the most general setting, where an *online learning algorithm* (hereafter, *learner*) maintains a partition of the agents into disjoint subsets (i.e, *coalitions*) over time, while interacting with an *adversary* that controls both the agents' composition and their preferences across time without posing any restrictions on their behaviour. Formally, the input to our problem is given by a finite set $N = \{1, \dots, n\}$ of n agents. For $x \in \mathbb{N}$, we hereafter denote $[x] := \{1, \dots, x\}$ and $[0] = \{0\}$. At each time $t \in [T]$, the *learner* and the *adversary* interact according to the following protocol. The *adversary* first picks a subset $N^t \subseteq N$ of $n^t := |N^t|$ agents that are present in the game at time t, capturing (worst-case) adversarial arrivals. Without loss of generality, we assume that $N^t \neq \emptyset$ at any time t.

Afterwards, the *learner* selects a partition π^t of the agents in N^t, where we denote $|\pi^t|$ as the number of its coalitions and $\pi^t(i)$ as the coalition $C \in \pi^t$ such that $i \in C$. We focus on real-life scenarios where the number of coalitions and the size of each coalition are bounded. Fixing positive integers $\alpha \leq n$ and $k \leq n$ *before* any agent arrives, we assume that the *learner* chooses an (α, k)-partition π^t at each time t, i.e., $|\pi^t| \leq k$ and $|C| \leq \alpha$ for each coalition $C \in \pi^t$. We denote the collection of all (α, k)-partitions at time t as $\Pi^t_{\alpha,k}$.

Only after the *learner* picks a partition, the *adversary* reports the preferences of each agent $i \in N^t$. While her preferences are usually given as her *utility* from others, for the sake of the analysis we represent them by an analogous cardinal *disutility* $d^t_i : N^t \to \mathbb{R}$ with $d^t_i(i) = 0$, specifying that agent i assigns a disutility of $d^t_i(j) \in \mathbb{R}$ to any agent $i \neq j \in N^t$ which indicates the degree of her *dislike* for agent j. Essentially, it is the negation of agent i's *utility*. We denote the agents' *joint* disutility at time t as $\mathbf{d}^t = (d^t_i)_{i \in N^t}$.

Remark 1. (Valuations' Evolution) *For any pair of agents $i, j \in N^t$ that were also present during some time $t' < t$ (i.e., $i, j \in N^{t'}$), note that $d^t_i(j)$ may not be equal to $d^{t'}_i(j)$. This models scenarios where the agents' social interactions at different time instants may yield an (adversarial) alteration in their preferences. However, an agent's valuations may remain* unchanged *over time, i.e., each agent $i \in N$ may have a* fixed *disutility $d_i : N \to \mathbb{R}$ with $d_i(i) = 0$ such that her utility at time t for any agent $j \in N^t$ is simply $d^t_i(j) = d_i(j)$.*

For obtaining (cardinal) preferences over coalitions, individual disutilities are aggregated via their summation. That is, we denote by N^t_i the set of coalitions agent $i \in N^t$ belongs to, i.e., $N^t_i = \{C \subseteq N^t : i \in C\}$. Agent i's disutility can be then *additively* aggregated to preferences over each coalition $C \in N^t_i$ via $d^t_i(C) = \sum_{j \in C} d^t_i(j)$. We further denote by $d^t_i(\pi^t) = d^t_i(\pi^t(i))$ the disutility agent i receives from the partition π^t chosen by the *learner*, which induces the partition π^t's *disutility profile* $\mathbf{d}^t(\pi^t) := (d^t_i(\pi^t))_{i \in N^t}$.

After the agents' valuations are revealed, the *learner* incurs the *social cost* of the partition π^t, which is defined by the sum of all agents' disutilities for that partition. Intuitively, the social cost indicates the agents' overall *dissatisfaction* with the partition. Formally, for a coalition $C \subseteq N^t$, we let $\mathrm{sc}(C, \mathbf{d}^t) = \sum_{i \in C} d^t_i(C)$ be the *social cost of C w.r.t.* $\mathbf{d}^t$. The *social cost* of the partition π^t w.r.t. $\mathbf{d}^t$ is $\mathrm{sc}(\pi^t, \mathbf{d}^t) = \sum_{C \in \pi^t} \mathrm{sc}(C, \mathbf{d}^t) = \sum_{i \in N^t} d^t_i(\pi^t)$. Hence, a partition π of the agents in N^t is *cost-minimal* if it minimizes the social cost amongst all possible partitions, i.e., $\pi \in \arg\min_{\pi' \in \Pi^t_{\alpha,k}} \mathrm{sc}(\pi', \mathbf{d}^t)$. We can thus summarize the learner-adversary interaction protocol as follows:

1. At each time t, the *adversary* picks the subset of agents $N^t \subseteq N$ that are present in the game at time t.
2. The *learner* then selects an (α, k)-partition π^t of N^t.
3. Afterwards, the *adversary* reports the *joint* disutility profile $\mathbf{d}^t = (d^t_i)_{i \in N^t}$ for the agents in N^t.
4. The *learner* incurs a social cost of $\mathrm{sc}(\pi^t, \mathbf{d}^t)$.

We study an online learning problem, termed as (α, k)-*Online Partitions* ((α, k)-OP), which is defined as follows. At each time t, based on the agents' past joint disutilities $\{d^\tau\}_{\tau \in [t-1]}$, the *learner*'s goal is choosing an (α, k)-partition π^t at time t such that the *learner*'s cumulative social cost over time is close to the overall social cost of the cost-minimal partitions. The standard metric for online learning algorithms is the *static regret* [Shalev-Shwartz, 2012], which compares the learner's decisions against the *optimal* decision in hindsight. In our context, the optimal decision is a *single* partition of all n agents minimizing the *learner*'s cumulative social cost over time. Formally, we denote the set of all (α, k)-partitions of the agents in N as Π. Given a partition $\pi \in \Pi$, we can obtain an (α, k)-partition $\pi \cap N^t := (C \cap N^t)_{C \in \pi : C \cap N^t \neq \emptyset}$ of the agents present at time t (i.e., N^t). Note that $\pi \cap N^t$ is not necessarily cost-minimal at time t. Further, if the bounds on the number of coalitions and their size are *bounded* such that $\alpha \cdot k < n$, then not all the agents in N are assigned to a coalition in π. In this case, if the partition $\pi \cap N^t$ contains any empty coalition at time t, we use the convention that $d^t_i(\emptyset) = 0$ for each agent $i \in N^t$. As such, the learner's *single* optimal decision in hindsight is any partition $\pi \in \arg\min_{\pi' \in \Pi} \sum_{t=1}^{T} \mathrm{sc}(\pi' \cap N^t, \mathbf{d}^t)$.

Next, we present the formal definition of static regret. First, we denote the absolute maximum value of a non-zero single-agent disutility as $W := \max_{t \in [T], i, j \in N^t : d^t_i(j) \neq 0} |d^t_i(j)|$. If the learner's overall social cost is at most $c > 0$ times the total social cost of a *single* optimal (offline) partition up to an additive term that is sublinear in T, then the learner is said to be *no-c-regret*. Formally, the learner is *no-c-regret* ($c > 0$) if and only if, for any sequence of agent sets $\{N^t\}_{t \in [T]}$ and joint disutility profiles $\{\mathbf{d}^t\}_{t \in [T]}$ that are chosen by the adversary, the partitions $\{\pi^t\}_{t \in [T]}$ selected by the learner satisfy:

$$\sum_{t=1}^{T} \mathrm{sc}(\pi^t, \mathbf{d}^t) \leq c \cdot \min_{\pi \in \Pi} \sum_{t=1}^{T} \mathrm{sc}(\pi \cap N^t, \mathbf{d}^t) + \Theta(\mathrm{poly}(n, W) \cdot T^\delta) \quad (1)$$

where $\delta < 1$. If $c = 1$, then the learner is called *no-regret*. However, since the static regret benchmarks the learner with a

single fixed partition, it fails to accurately assess the quality of decisions in our *dynamic* setting. Namely, the minimum social cost is not static, but it changes dynamically as both the agents' composition and their preferences evolve over time. The stronger notion of *dynamic regret* naturally reflects this concept, allowing us to measure the performance difference between the learner and a set of partitions optimal for *individual* time instants in hindsight [Jadbabaie *et al.*, 2015]. Formally, for $c > 0$, the learner's *c-dynamic regret* for his selected partitions $\{\pi^t\}_{t\in[T]}$ is given by:

$$\mathcal{R}_c^T := \sum_{t=1}^T \mathrm{sc}(\pi^t, \mathbf{d}^t) - c \cdot \sum_{t=1}^T \mathrm{sc}(\pi_\star^t, \mathbf{d}^t) \qquad (2)$$

where $\pi_\star^t$ is a cost-minimal partition w.r.t. $\mathbf{d}^t$. For $c = 1$, we obtain the standard *dynamic regret* $\mathcal{R}^T := \mathcal{R}_1^T$. *No-c-dynamic regret* is defined similarly to (1).

4 Hardness of Being No-c-Regret

Though certain domains admit no-*static* regret algorithms (see, e.g., [Zinkevich, 2003]), we begin with proving that it is hard to approximate in our context, despite that it is weaker than *dynamic* regret. The key idea is that if we have a good efficient learner, then we can obtain an efficient algorithm that approximately solves an *offline* variant of our problem, proven to be hard to approximate due to Bilò *et al.* [2022]. Formally, our strong negative result holds even for a restricted version of (α, k)-OP termed as (α, k)-OP$^=$, where, at any time t, there are *exactly* $n^t = \alpha \cdot k$ agents and a partition must contain *exactly* k coalitions, each containing *exactly* α agents. We also consider the following *offline* variant, where the subsets of agents and their valuations are available upfront:

Definition 1. (Offline (α, k)-OP$^=$) *Given a sequence of B agent sets $\{N^b\}_{b\in[B]}$ with $|N^b| = \alpha k$ for any $b \in [B]$ and corresponding joint disutility profiles $\{\mathbf{d}^b\}_{b\in[B]}$, select an (α, k)-partition π of $N = \cup_{b\in[B]} N^b$ with* exactly *k coalitions, each containing* exactly *α agents, such that $\sum_{b=1}^B \mathrm{sc}(\pi \cap N^b, \mathbf{d}^b)$ is minimized. For any $b \in [B]$, note that $\pi \cap N^b$ is an (α, k)-partition, but it is* not *required to contain* exactly *k coalitions, each with* exactly *α agents. Denoting $\pi_\star$ as the* optimal *partition minimizing this objective, a c-*approximation algorithm *for offline (α, k)-OP$^=$ returns an (α, k)-partition π such that $\sum_{b=1}^B \mathrm{sc}(\pi \cap N^b, \mathbf{d}^b) \leq c \cdot \sum_{b=1}^B \mathrm{sc}(\pi_\star \cap N^b, \mathbf{d}^b)$ in polynomial time.*

To go from our learning problem to the combinatorial *offline* (α, k)-OP$^=$ problem, we perform the following approximation preserving reduction from (α, k)-OP$^=$ to *offline* (α, k)-OP$^=$ that will imply our main result:

Lemma 1. *Any (randomized or deterministic) no-c-regret algorithm for (α, k)-OP$^=$ yields a $(c+1)$-approximation algorithm for* offline *(α, k)-OP$^=$.*

Proof. As deterministic algorithms are a special case of randomized ones, we focus on the latter. Consider any randomized no-c-regret algorithm $\mathcal{A}$ for (α, k)-OP$^=$. For an instance of *offline* (α, k)-OP$^=$ given by a sequence of agent sets $\mathcal{N} := \{N^b\}_{b\in[B]}$ and their joint disutilities $\mathcal{D} := \{\mathbf{d}^b\}_{b\in[B]}$, we devise the algorithm $\mathcal{A}'$ that simulates this instance online over some $T > 0$ rounds using algorithm $\mathcal{A}$ as follows. First, $\mathcal{A}'$ picks $\tau \in [T]$ uniformly at random. Then, it runs an adversary that, at each time $t \in [T]$, selects *uniformly at random* a set of agents N^t and its corresponding joint disutility d^t. Let π^t and $\pi \in \Pi$ be the random partition returned by $\mathcal{A}$ at time t and the random offline optimal solution (resp.). The algorithm $\mathcal{A}'$ returns the partition π_τ. Since $\mathcal{A}$ is no-c-regret: $\sum_{t=1}^T \frac{1}{B} \sum_{b=1}^B \mathbb{E}[\mathrm{sc}(\pi^t, \mathbf{d}^b)] \leq c \cdot \frac{T}{B} \sum_{b=1}^B \mathbb{E}[\mathrm{sc}(\pi \cap N^b, \mathbf{d}^b)] + \Theta(\mathrm{poly}(n, W) \cdot T^\delta)$, where $\delta < 1$ and the expectation is over the randomness of $\mathcal{A}$ and the adversary. Thus, the expected social cost of $\mathcal{A}'$ is at most: $\frac{1}{T} \sum_{t=1}^T \sum_{b=1}^B \mathbb{E}[\mathrm{sc}(\pi^t, \mathbf{d}^t)] = \frac{B}{T} \sum_{t=1}^T \frac{1}{B} \sum_{b=1}^B \mathbb{E}[\mathrm{sc}(\pi^t, \mathbf{d}^t)] \leq \frac{cB}{T} \cdot \frac{T}{B} \sum_{b=1}^B \mathbb{E}[\mathrm{sc}(\pi \cap N^b, \mathbf{d}^t)] + \Theta(\mathrm{poly}(n, W) \cdot T^{\delta-1})$. As T was arbitrary, setting $T = \Theta(\mathrm{poly}(n, W)^{1/(1-\delta)})$ yields that $\mathcal{A}'$ is a $(c+1)$-approximation algorithm for *offline* (α, k)-OP$^=$. □

We can now obtain our computational hardness results. We remark that they hold even for *simple* games, where, at any time t, the disutility of each agent $i \in N^t$ for any other agent $i \neq j \in N^t$ satisfies $d_i^t(j) \in \{0, -1\}$.

Theorem 1. *When k is not constant, assuming the Exponential Time Hypothesis (ETH), there does not exist a no-$(n^{1/\log\log^\beta n} - 1)$-regret algorithm for (α, k)-OP$^=$, where $\beta > 0$ is a universal constant independent of n. Further, assuming that there is a constant $\varepsilon > 0$ s.t. no subexponential-time algorithm can distinguish between a satisfiable 3SAT formula and one which is only $(1-\varepsilon)$-satisfiable (also known as Gap-ETH), there does not exist a no-$(n^{f(n)} - 1)$-regret algorithm for (α, k)-OP$^=$ for any function $f \in o(1)$. Both results hold even for simple games.*

Proof. (*Sketch*) The proof readily follows via combining Lemma 1 with Theorem 11 by Bilò *et al.* [2022], and is thus deferred to Appendix A. First, we show that Theorem 11 by Bilò *et al.* [2022] can be applied to *offline* (α, k)-OP$^=$ with $B = 1$ by proving that it is a variant of the problem considered by Bilò *et al.* [2022]. Combining this observation with Lemma 1 establishes our hardness results. We also note that Theorem 11 by Bilò *et al.* [2022] is obtained by reduction from the Densest β-Subgraph (DβS) problem, known to be hard to approximate [Manurangsi, 2017]. For completeness, in Appendix A we supply our reduction from DβS to *offline* (α, k)-OP$^=$ that, combined with Lemma 1 and [Manurangsi, 2017, Corollary 1.3], gives an alternative proof. □

Remark 2. *As (α, k)-OP$^=$ is a subproblem of (α, k)-OP, Theorem 1 also holds for the latter. As it is well-known that the worst-case* dynamic *regret scales linearly in T, we show in Appendix B that a no-c-dynamic regret algorithm is unattainable for any $c > 0$. Hence, Theorem 1 does not apply to* dynamic *regret. Since we posed no restriction on the agents' disutilities, our result also holds when they remain* unchanged *over time (see Remark 1).*

This establishes the possible regret bounds and indicates that ***constant*** approximation to no-*static* regret cannot be attained in polynomial time for (α, k)-OP when α is *bounded*

(i.e., $\alpha < n^t$ at each time t). Thus, our result motivates us to assume the following most general setting:

Assumption 1. *The coalition size bound α is* unbounded *at each time t (i.e., $\alpha = n$), in which case we term the (α, k)-OP problem as k-OP for short. We also assume that $k \geq 2$ since the case of $k = 1$ is trivial.*

Next, we devise online learning schemes that attain non-trivial regret bounds in polynomial time. Initially, we consider a *fractional* relaxation of our problem, where agents are *fractionally* assigned to multiple coalitions (Section 5), for which we devise a no-*static* regret learner that simultaneously has *low dynamic* regret (Section 5.1). In fact, our algorithm is *best of both (regret) worlds*: it is ***optimal*** in terms of both static and dynamic regret. We then give an efficient randomized rounding scheme with ***optimal*** dynamic regret that converts the *fractional* partitions produced by our algorithm into *integral* ones (Section 6).

5 Fractional Online Partitions

In this section, we consider *fractional k-OP*, the *fractional* relaxation of our problem where agents are allowed to be part of several coalitions by being *fractionally* assigned to multiple coalitions. This approach is common in numerous related problems (e.g., matchings [Aziz and Klaus, 2019] and fair division [Amanatidis *et al.*, 2023]), and has been proven fruitful for overcoming impossibilities as the one in Theorem 1. This setting also has many real-life applications. For instance, a company manager may assign his employees to several project teams, where a fractional assignment models the amount of time an employee spends working on each project. Later, we show that such relaxation leads to a *no-regret* algorithm, proven to attain ***optimal*** static and dynamic regret.

Fractional k-OP (k-FOP) is defined as follows. First, we denote the k-dimensional simplex as $\Delta_k = \{\mathbf{y} \in \mathbb{R}^k_{\geq 0} : \sum_{\ell \in [k]} y_{i,\ell} = 1\}$. At each time t, after the *adversary* picks the subset of agents N^t, the *learner* selects a vector $\mathbf{y}_i^t = (y_{i,\ell}^t)_{\ell \in [k]} \in \Delta_k$ for any agent i (i.e., a *agent i's fractional assignment*), where $y_{i,\ell}^t$ stands for the fraction of agent i assigned to the ℓ-th (possible) coalition for each $\ell \in [k]$. We term $\mathbf{y}^t = (\mathbf{y}_i^t)_{i \in N^t}$ as the *fractional partition*. Afterwards, the *adversary* reports the *joint* disutility profile $\mathbf{d}^t = (d_i^t)_{i \in N^t}$ for the agents in N^t. Let $\mathbb{1}_{x_{i,\ell}^t > 0}$ be equal to 1 if $x_{i,\ell}^t > 0$, and 0 otherwise. The *learner* then incurs the *fractional social cost* $\mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t)$, which is the optimal value of the following convex linear program:

$$\begin{array}{ll} \min & \sum_{i \in N^t} \sum_{\ell \in [k]} \mathbb{1}_{x_{i,\ell}^t > 0} \sum_{i \neq j \in N^t} x_{j,\ell}^t d_i^t(j) \\ \text{s.t.} & \sum_{\ell \in [k]} x_{i,\ell}^t = 1, \forall i \in N^t \\ & x_{i,\ell}^t \geq 0, x_{i,\ell}^t \leq y_{i,\ell}^t, \forall i \in N^t, \ell \in [k] \end{array} \tag{3}$$

where the minimum is over the fractional partition $\mathbf{x}^t = ((x_{i,\ell}^t)_{\ell \in [k]})_{i \in N^t}$. The intuition behind problem (3)'s objective is that, if agent i is fractionally assigned to the ℓ-th coalition (i.e., $x_{i,\ell}^t > 0$), then she receives a disutility of $x_{j,\ell}^t d_i^t(j)$ from each agent $i \neq j \in N^t$. Thus, given the fractional partition $\mathbf{y}^t$ selected by the *learner* of the agents in N^t with the joint disutility $\mathbf{d}^t$ chosen by the *adversary*, problem (3) picks $\mathbf{x}^t$ for minimizing its objective such that $x_{i,\ell}^t \leq y_{i,\ell}^t$ for any agent $i \in N^t$ and $\ell \in [k]$. Since the *learner* chooses the fractional partition *before* observing the agents' valuations, he can solve (3) for obtaining the agents' fractional assignments only *after* the disutilities are revealed.

Once the agents' assignment vectors are *integral* (i.e., $x_{i,\ell}^t \in \{0, 1\}$), each agent can be assigned to at most one coalition. This reduces k-FOP to the original k-OP problem, in which case the *fractional* social cost equals to the *original* social cost. Thus, we infer the following relation:

Lemma 2. *For any sequence of agent sets $\{N^t\}_{t \in [T]}$ and joint disutility profiles $\{\mathbf{d}^t\}_{t \in [T]}$, the social cost of a cost-minimal partition $\pi_\star^t$ for k-OP at each time t is lower bounded by the* fractional *social cost of the optimal fractional partition $\mathbf{y}_\star^t$ for the corresponding k-FOP problem, i.e., $\mathrm{sc}(\pi_\star^t, \mathbf{d}^t) \geq \mathrm{fsc}(\mathbf{y}_\star^t, \mathbf{d}^t)$.*

Commonly used methods for solving problems such as k-FOP are based on *online gradient descent* (OGD) [Hazan, 2016]. Such algorithms use the *subgradients* of the problem's objective function, which are defined as follows:

Definition 2. (Subgradients) *For a function $g : \mathbb{R}^m \to \mathbb{R}$, a vector $\mathbf{z} \in \mathbb{R}^m$ is a* subgradient *of g at point $\mathbf{y} \in \mathbb{R}^m$ if and only if $g(\mathbf{y}') \geq g(\mathbf{y}) + \mathbf{z}^\top(\mathbf{y}' - \mathbf{y})$ for any $\mathbf{y}' \in \mathbb{R}^m$. The set of g's subgradients at point $\mathbf{y} \in \mathbb{R}^m$ is denoted as $\partial g(\mathbf{y})$.*

Generally, calculating subgradients is computationally challenging. However, we show that they can be easily attained for our problem (3) by solving a new primal-dual program, which we obtain in Lemma 3 by following the common approach for solving constrained optimization problems via their Lagrangian relaxed form.

Lemma 3. *Given the set N^t of n^t agents, their valuations $\mathbf{d}^t$, the fractional partition $\mathbf{y}^t$ and the fractional assignment $\mathbf{x}_i^t$ of each agent $i \in N^t$, the Lagrangian relaxation of* (3) *at time t gives the primal-dual optimization problem given by the following convex linear program:*

$$\begin{array}{ll} \max & \sum_{i \in N^t} \lambda_i^t - \sum_{i \in N^t} \sum_{\ell \in [k]} \mu_{i,\ell}^t \cdot y_{i,\ell}^t \\ & \sum_{i \neq j \in N^t : x_{j,\ell}^t > 0} d_j^t(i) + \mu_{i,\ell}^t = \lambda_i^t + \gamma_{i,\ell}^t, \\ & \lambda_i^t \geq 0, \mu_{i,\ell}^t \geq 0, \gamma_{i,\ell}^t \geq 0, \quad \forall i \in N^t, \ell \in [k] \end{array} \tag{4}$$

where the minimum is over the Lagrangian multipliers $\lambda^t \in \mathbb{R}^{n^t}_{\geq 0}$, $\mu^t \in \mathbb{R}^{n^t \times k}_{\geq 0}$ and $\gamma^t \in \mathbb{R}^{n^t \times k}_{\geq 0}$.

Proof. (*Sketch*) In Appendix C, we consider the Lagrangian function $\mathcal{L}(\mathbf{x}^t, \mathbf{y}^t, \lambda^t, \gamma^t, \mu^t)$ of (3) and the Lagrangian dual function $g(\lambda^t, \gamma^t, \mu^t) = \inf_{\mathbf{x}^t} \mathcal{L}(\mathbf{x}^t, \mathbf{y}^t, \lambda^t, \gamma^t, \mu^t)$. We find the supremum by computing the gradient $\nabla_{\mathbf{x}^t} \mathcal{L}(\mathbf{x}^t, \mathbf{y}^t, \lambda^t, \gamma^t, \mu^t)$ w.r.t. $\mathbf{y}^t$ and solving $\nabla_{\mathbf{x}^t} \mathcal{L}(\mathbf{x}^t, \mathbf{y}^t, \lambda^t, \gamma^t, \mu^t) = 0$. The latter gives us the constraints in (4), whose substitution to the Lagrangian function yields that the objective of our primal-dual problem is $g(\lambda^t, \gamma^t, \mu^t) = \sum_{i \in N^t} \lambda_i^t - \sum_{i \in N^t} \sum_{\ell \in [k]} \mu_{i,\ell}^t \cdot y_{i,\ell}^t$. □

Next, we show how the subgradients in $\partial\, \mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t)$ can be easily computed using the optimal solution to the optimization problem in (4). We also supply an upper bound on the subgradients, which depends on $m := \max_{t \in [T]} |N^t|$ (i.e.,

Algorithm 1 : **Online Gradient Descent for k-FOP**

Input: A constant step size $\eta > 0$

The *learner* initializes $\mathbf{z}_i^1$ as $z_{i,\ell}^1 = 1/k \; \forall i \in N, \ell \in [k]$.
for each time $t \in [T]$ **do**
  The *adversary* picks the agents N^t arriving at time t.
  The *learner* picks $y_{i,\ell}^t = z_{i,\ell}^t \; \forall i \in N^t, \ell \in [k]$.
  The *adversary* reports the agents' disutilities $\mathbf{d}^t$.
  The *learner* incurs $\mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t)$.
  The *learner* solves (3) and obtains $\mathbf{x}_{\star,i}^t \; \forall i \in N^t$.
  Given $\mathbf{x}_{\star,i}^t$ and $\mathbf{y}^t$, the *learner* solves (4) to attain $\mu_\star^t$.
  The *learner* sets $\mathbf{g}^t = ((-\mu_{\star,i,\ell}^t)_{\ell\in[k]})_{i\in N^t}$.
  The *learner* picks $\mathbf{y}^{t+1} = \mathcal{P}_{\Delta_k}(\mathbf{y}^t - \eta \mathbf{g}^t)$.
  The *learner* sets $\mathbf{z}_i^{t+1}$ as $\mathbf{y}_i^{t+1}$ if $i \in N^t$, and $\mathbf{y}_i^t$ o.w.

the maximum number of agents that are present at a single time instant), and the absolute maximum value W of a non-zero single-agent disutility (as defined before (2)).

Lemma 4. *Given a set N^t of n^t agents and their valuations $\mathbf{d}^t$ at time t, let $\lambda_\star^t, \mu_\star^t, \gamma_\star^t$ be the optimal solution of the problem* (5) *with respect to a fractional partition $\mathbf{y}^t$ and a fractional assignment $\mathbf{x}^t$. Then, for any fractional partition $\hat{\mathbf{y}}^t$ and fractional assignment $\hat{\mathbf{x}}^t$: $\mathrm{fsc}(\hat{\mathbf{y}}^t, \mathbf{d}^t) \geq \mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t) + \sum_{\ell\in[k]} \sum_{i\in N^t} (-\mu_{\star,i,\ell}^t) \cdot (\hat{y}_{i,\ell}^t - y_{i,\ell}^t)$. Further, the subgradients are upper bounded by: $|-\mu_{\star,i,\ell}^t| \leq (m-1)W$.*

5.1 A Best of Both (Regret) Worlds Algorithm

We are now ready to present our algorithm for k-FOP (Algorithm 1), which is *no-regret* while also providing *low* dynamic regret. Our algorithm provides the best of both (regret) worlds: it simultaneously achieves ***optimal*** static and dynamic regret. Algorithm 1 uses a modified variant of *online gradient descent* (OGD) [Zinkevich, 2003], and is parameterized by a constant $\eta > 0$. We will later provide the suitable selections of η that yields low regret. Algorithm 1 runs as follows. The *learner* first initializes a fractional partition $\mathbf{z}^1 = ((z_{i,\ell}^1)_{\ell\in[k]})_{i\in N}$ that splits each agent equally among the k possible coalitions (i.e., $z_{i,\ell}^1 = 1/k$ for each agent $i \in N$ and $\ell \in [k]$). At each time t, the *adversary* first selects the arriving agents N^t and the *learner* responds with $\mathbf{y}^t = ((z_{i,\ell}^t)_{\ell\in[k]})_{i\in N^t}$. The *adversary* then selects the agents' disutilities $\mathbf{d}^t$ and the *learner* incurs the fractional social cost $\mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t)$.

Given the *learner*'s obtained information, the *learner* can readily solve the *convex linear program* (3) using standard methods and obtain the optimal fractional assignment $\mathbf{x}_{\star,i}^t$ of each agent $i \in N^t$. The *learner* can then easily solve the *convex linear program* (4) via classic approaches and receive its optimal solution $\lambda_\star^t$, $\mu_\star^t$, $\gamma_\star^t$. Using Lemma 4, the *learner* computes the corresponding subgradient. Finally, the *learner* computes the fractional partition $\mathbf{y}^{t+1}$ at time $t+1$ by performing a gradient descent step w.r.t. the subgradient $\mathbf{g}^t = ((-\mu_{\star,i,\ell}^t)_{\ell\in[k]})_{i\in N^t}$ corresponding to the current solution $\mathbf{y}^t$, where $\mathcal{P}_{\Delta_k^n}(\cdot)$ denotes the projection onto the nearest point in Δ_k^n. Finally, for any agent $i \in N$, the *learner* updates agent i's $\mathbf{z}_i^{t+1}$ as $\mathbf{y}_i^{t+1}$ if $i \in N^t$, and $\mathbf{y}_i^t$ otherwise.

In Theorem 2, we show that Algorithm 1 is no-regret for a proper choice of $\eta > 0$. Note that an offline optimal decision for k-FOP is given by a fractional partition $\mathbf{y}$, from which we can derive a fractional partition $\mathbf{y}|_{N^t} := (\mathbf{y}_i)_{i\in N^t}$ at time t.

Remark 3. *At each time t, as $\mathbf{z}_i^t = \mathbf{y}_i^t$ for any $i \in N^t$, note that the fractional social cost induces a* loss function *$f_t : \Delta_k^n \to \mathbb{R}$ given by $f_t(\mathbf{z}^t) = \mathrm{fsc}(\mathbf{z}^t|_{N^t}, \mathbf{d}^t) = \mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t)$.*

Theorem 2. *For any sequence of agent sets $\{N^t\}_{t\in[T]}$ with joint disutilities $\{\mathbf{d}^t\}_{t\in[T]}$ that are chosen by the* adversary*, the* learner *is* no-regret *when producing the fractional partitions $\{\mathbf{y}^t\}_{t\in[T]}$ using Algorithm 1 with $\eta = \frac{\sqrt{\log(n)}}{(m-1)W\sqrt{2T}}$, and the* learner *also obtains a* **minimax optimal** *static regret bound of $O(2(m-1)W\sqrt{2T\log(n)})$.*

Proof. (*Sketch*) By Lemma 4 and using the triangle inequality, note that for any pair of fractional partitions $\mathbf{y}^t, \hat{\mathbf{y}}^t$: $|\mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t) - \mathrm{fsc}(\hat{\mathbf{y}}^t, \mathbf{d}^t)| \leq |\sum_{\ell\in[k]} \sum_{i\in N^t} \mu_{\star,i,\ell}^t (\hat{y}_{i,\ell}^t - y_{i,\ell}^t)| \leq \sum_{\ell\in[k]} \sum_{i\in N^t} |\mu_{\star,i,\ell}^t| \cdot |\hat{y}_{i,\ell}^t - y_{i,\ell}^t| \leq (m-1)W\|\hat{\mathbf{y}}^t - \mathbf{y}^t\|_1$, where $\|\cdot\|_1$ is the $\mathcal{L}_1$-norm. That is, $\mathrm{fsc}(\cdot, \mathbf{d}^t)$ is L-Lipschitz w.r.t. $\|\cdot\|_1$ for $L := (m-1)W$. In Appendix E, we show that this also applies to $f_t(\cdot)$ from Remark 3. Hence, setting $\eta = \frac{\sqrt{\log(n)}}{L\sqrt{2T}}$, Corollary 2.14 by Shalev-Shwartz [2012] yields the desired static regret bound when applying it to the sequence of loss functions $\{f_t\}_{t\in[T]}$.

Though no-static regret is hard to approximate for the original k-OP problem by Theorem 1, note that our obtained $O(\sqrt{T})$ static regret is **minimax optimal** due to Abernethy *et al.* [2008] (i.e., the minimal static regret that can be attained in the worst case for convex programs). □

Remark 4. *Observe that the bound's dependence on W indicates that, if the agents' disutilities fluctuate moderately over time (or even remain* unchanged *as discussed in Remark 1), then the static regret is reduced.*

Despite that we reached the minimax optimal *static* regret, recall that *dynamic* regret is still more suitable for evaluating the performance of the learner in our dynamic setting. Hence, we now focus on *dynamic* regret. It is well-known that the worst-case *dynamic* regret scales linearly in T since the agents' valuations arbitrarily fluctuate. Yet, it is possible to bound the dynamic regret in terms of certain regularities. We herein consider the *path-length* of the optimal sequence of fractional partitions [Zinkevich, 2003], given by $\mathcal{S}_\star^T := \sum_{t=2}^t \|\mathbf{y}_\star^t - \mathbf{y}_\star^{t-1}\|_2$, capturing the overall Euclidean norm of the difference between two successive solutions.

In Theorem 3, we show that Algorithm 1 is not only no-regret, but it also has an *optimal* dynamic regret.

Theorem 3. *For any sequence of agent sets $\{N^t\}_{t\in[T]}$ with joint disutilities $\{\mathbf{d}^t\}_{t\in[T]}$ that are chosen by the* adversary*, the fractional partitions $\{\mathbf{y}^t\}_{t\in[T]}$ produced by the* learner *using Algorithm 1 with $\eta = \frac{1}{2L}$ for any constant $L > 0$ yield an* **optimal** *dynamic regret bound of $O(\mathcal{S}_\star^T)$.*

Proof. (*Sketch*) In Appendix F, we first show that $\mathrm{fsc}(\mathbf{y}^t, \mathbf{d}^t)$ is convex as a function of $\mathbf{y}^t$ since it can be expressed as a

linear function of $\mathbf{y}^t$. Then, we prove that it is L-smooth for any constant $L > 0$, i.e., as it twice differentiable, we show that $\mathbf{z}^\top \nabla^2 \operatorname{fsc}(\mathbf{y}^t, \mathbf{d}^t)\mathbf{z} \leq L\|\mathbf{z}\|_2^2$ for any pair of fractional partitions $\mathbf{y}^t, \mathbf{z}$ and constant $L > 0$. Thus, $f_t(\cdot)$ from Remark 3 is also convex and L-smooth. By setting $\eta = \frac{1}{2L}$, the conditions of Theorem 3 by Yang *et al.* [2016] and we obtain the desired dynamic regret bound, when applying it to the sequence of loss functions $\{f_t\}_{t\in[T]}$. Since $\operatorname{fsc}(\mathbf{y}^t, \mathbf{d}^t)$ is a linear function of $\mathbf{y}^t$, then it is not *strongly* convex (see Appendix G for a proof). Due to Proposition 1 by Yang *et al.* [2016], our dynamic regret of $O(\mathcal{S}_\star^T)$ is ***optimal*** for smooth and (non-strongly) convex functions. □

Remark 5. *We conclude that, if the* learner *runs Algorithm 1 with η as in Theorem 2, then the* learner *simultaneously obtains the* **optimal** *static and dynamic regret since we can invoke Theorem 3 as $\eta = \frac{1}{2L}$ for $L = \frac{(m-1)W\sqrt{2T}}{2\sqrt{\log(n)}}$. Recall that no-dynamic regret is generally unattainable. However, if the optimal fractional partition only slightly changes between consecutive time steps, then our dynamic regret bound $O(\mathcal{S}_\star^T)$ can become smaller. For instance, if $\|\mathbf{y}_\star^t - \mathbf{y}_\star^{t-1}\| = \Omega(1/\sqrt{T})$ for each time t, then $\mathcal{S}_\star^T = \Omega(\sqrt{T})$.*

6 Randomized Rounding for k-OP

Recall that *static* regret not only fails to accurately reflect the learner's performance in *dynamic* environments, no-regret is also hard to approximate for k-OP due to Theorem 1. Hence, in this section we focus on the more stringent *dynamic* regret, which is also more suitable for our *dynamic* setting. Using standard randomized rounding techniques, we next show how the resulting *fractional* partition produced by Algorithm 1 can be converted into a randomized *integral* partition, while obtaining the *optimal* dynamic regret bound. Formally:

Mechanism 1. (Randomized Rounding) *Given $\eta > 0$ as input, run Algorithm 1 with η. At any time t, given the* fractional *partition $\mathbf{y}^t$, we can obtain an* integral *partition π^t via* randomized rounding*: If a fraction $y_{i,\ell}^t$ of agent i is assigned to the ℓ-th coalition, then agent i is assigned to the ℓ-th possible coalition with probability $y_{i,\ell}^t$ (Recall that $\sum_{\ell\in[k]} y_{i,\ell}^t = 1$ for any agent $i \in N^t$). Note that decisions are* independent *among different agents and different time instants.*

Our bound in Theorem 4 involves a *new* regularity metric for dynamic regret, called the *interaction term*, measuring the overall difference between the agents' disutilities from their interactions according to the fractional partition $\mathbf{y}^t$ and the respective optimal solution $\mathbf{x}^t$ of (3) at each time t. Namely, the *interaction term* is given as $\mathcal{I}^T = \sum_{t=1}^T \sum_{i\in N^t} \sum_{\ell\in[k]} \sum_{i\neq j\in N^t} |[y_{i,\ell}^t y_{j,\ell}^t - \mathbb{1}_{x_{i,\ell}^t>0} \cdot x_{j,\ell}^t]d_i^t(j)|$, which is at most $O(T)$ (in the worst-case) as $|y_{i,\ell}^t y_{j,\ell}^t - \mathbb{1}_{x_{i,\ell}^t>0} \cdot x_{j,\ell}^t| \leq 1$ and $|d_i^t(j)| \leq W$. Intuitively, at time t, $y_{i,\ell}^t y_{j,\ell}^t$ can be viewed as the probability that any pair of agents i, j are assigned to the ℓ-th possible coalition for any $\ell \in [k]$ (i.e., they form an *interaction*), where $y_{i,\ell}^t y_{j,\ell}^t d_i^t(j)$ is agent i's resulting fractional disutility from their interaction. Similarly, $\mathbb{1}_{x_{i,\ell}^t>0} \cdot x_{j,\ell}^t$ can be interpreted as the assignment probability of agent j to the ℓ-th coalition at time t, given that agent i is assigned to that coalition with a positive probability (i.e., $x_{i,\ell}^t > 0$), while agent i's resulting fractional disutility from their interaction is $\mathbb{1}_{x_{i,\ell}^t>0} \cdot x_{j,\ell}^t d_i^t(j)$.

Theorem 4. *For any sequence of agent sets $\{N^t\}_{t\in[T]}$ with joint disutilities $\{\mathbf{d}^t\}_{t\in[T]}$ that are chosen by the* adversary*, the fractional partitions $\{\mathbf{y}^t\}_{t\in[T]}$ produced by the* learner *using Mechanism 1 with $\eta = \frac{1}{2L}$ for any constant $L > 0$ yield an* **optimal** *dynamic regret bound of $\mathcal{I}^T + O(\mathcal{S}_\star^T)$.*

Proof. (*Sketch*) In Appendix H, we first prove $\sum_{t=1}^T \mathbb{E}[\operatorname{sc}(\pi^t, \mathbf{d}^t)] \leq \mathcal{I}^T + \sum_{t=1}^T \operatorname{fsc}(\mathbf{y}^t, \mathbf{d}^t)$, from which the bound is obtained by Theorem 3. For optimality, we show that a sublinear dynamic regret is unattainable if the the path length is unconstrained (i.e., it is upper bounded by $\Omega(T)$). However, if the path length is at most $o(T)$, then we prove that it is impossible to get a better dynamic regret bound of $\mathcal{I}^T + O((\mathcal{S}_\star^T)^\psi)$ for $\psi \in (0, 1)$. □

Remark 6. *We obtained the following groundbreaking result: while no-*static *regret is hard to approximate for* integral *partitions (but the* minimax optimal *static regret is attainable for* fractional *ones), the more stringent* dynamic *regret can be solved* **optimally** *in both the* fractional *and* integral *settings. Recall that no-dynamic regret is generally unfeasible. Yet, if the optimal fractional partition and the agents' interactions only slightly change between consecutive time steps, then our dynamic regret bound can become smaller. For instance, if $\|\mathbf{y}_\star^t - \mathbf{y}_\star^{t-1}\| = \Omega(1/\sqrt{T})$ and $|[y_{i,\ell}^t y_{j,\ell}^t - \mathbb{1}_{x_{i,\ell}^t>0} \cdot x_{j,\ell}^t]d_i^t(j)| = \Omega(1/m^2k\sqrt{T})$ for each time t, agents $i \neq j$ and $\ell \in [k]$, then $\mathcal{I}^T + \mathcal{S}_\star^T = \Omega(\sqrt{T})$.*

7 Conclusions and Future Work

In this paper, we presented a novel algorithmic framework for studying coalition formation in *dynamic* settings from the perspective of *online learning*. We characterized the *learner*'s capabilities in terms of both *static* and *dynamic* regret, obtaining the following remarkable results. While no-*static* regret is hard to approximate for *integral* partitions, *dynamic* regret is not. However, for a *fractional* relaxation of our problem, we devised an algorithm that concurrently has ***optimal*** static and dynamic regret, thus achieving the best of both (regret) worlds. Unlike *static* regret, the more stringent dynamic regret can be also solved ***optimally*** for *integral* partitions. Further, we presented the *interaction term*, a *new* regularity for bounding dynamic regret that reflects the variation in agents' interactions, and has the potential of expanding the research on online learning in computational social choice.

Our research paves the way for many future works. Immediate directions are the investigation of other classes of hedonic games and other adversaries. Further, as in many real-world domains preferences may be inherently *uncertain*, future work warrants developing algorithms for online learning in hedonic games under *partial* and possibly *noisy* feedback. Another intriguing direction is considering other measures of economic efficiency (e.g., Pareto-optimality) as well as other solution concepts, such as stability and fairness notions.

Acknowledgments

This research was funded in part by ISF grant 1563/22.

References

[Abernethy *et al.*, 2008] Jacob Abernethy, Peter L Bartlett, Alexander Rakhlin, and Ambuj Tewari. Optimal strategies and minimax lower bounds for online convex games. In *Proceedings of the 21th Annual conference on learning theory, COLT*, pages 415–424, 2008.

[Amanatidis *et al.*, 2023] Georgios Amanatidis, Haris Aziz, Georgios Birmpas, Aris Filos-Ratsikas, Bo Li, Hervé Moulin, Alexandros A. Voudouris, and Xiaowei Wu. Fair division of indivisible goods: Recent progress and open questions. *Artificial Intelligence*, 322:103965, 2023.

[Aziz and Klaus, 2019] Haris Aziz and Bettina Klaus. Random matching under priorities: stability and no envy concepts. *Social Choice and Welfare*, 53(2):213–259, 2019.

[Aziz and Savani, 2016] Haris Aziz and Rahul Savani. Hedonic games. In *Handbook of Computational Social Choice*, page 356–376. Cambridge University Press, 2016.

[Aziz *et al.*, 2011] Haris Aziz, Felix Brandt, and Hans Georg Seedig. Stable partitions in additively separable hedonic games. In *The 10th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, volume 1, pages 183–190, 2011.

[Aziz *et al.*, 2015] Haris Aziz, Serge Gaspers, Joachim Gudmundsson, Julián Mestre, and Hanjo Täubig. Welfare maximization in fractional hedonic games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI*, pages 461–467, 2015.

[Aziz *et al.*, 2019] Haris Aziz, Florian Brandl, Felix Brandt, Paul Harrenstein, Martin Olsen, and Dominik Peters. Fractional hedonic games. *ACM Transactions on Economics and Computation, TEAC*, 7(2):1–29, 2019.

[Baby and Wang, 2019] Dheeraj Baby and Yu-Xiang Wang. Online forecasting of total-variation-bounded sequences. *Advances in Neural Information Processing Systems, NeurIPS*, 32, 2019.

[Ballester, 2004] Coralio Ballester. Np-completeness in hedonic games. *Games and Economic Behavior*, 49(1):1–30, 2004.

[Banerjee *et al.*, 2001] Suryapratim Banerjee, Hideo Konishi, and Tayfun Sönmez. Core in a simple coalition formation game. *Social Choice and Welfare*, 18(1):135–153, 2001.

[Besbes *et al.*, 2015] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

[Bilò *et al.*, 2022] Vittorio Bilò, Gianpiero Monaco, and Luca Moscardelli. Hedonic games with fixed-size coalitions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9287–9295, 2022.

[Bogomolnaia and Jackson, 2002] Anna Bogomolnaia and Matthew O Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2002.

[Bouveret *et al.*, 2010] Sylvain Bouveret, Ulle Endriss, and Jérôme Lang. Fair division under ordinal preferences: Computing envy-free allocations of indivisible goods. In *Proceedings of the 19th European Conference on Artificial Intelligence, ECAI 2010*, pages 387–392, 2010.

[Bullinger and Romen, 2023] Martin Bullinger and René Romen. Online Coalition Formation Under Random Arrival or Coalition Dissolution. In *Proceedings of the 31st Annual European Symposium on Algorithms, ESA*, volume 274 of *LIPIcs*, pages 27:1–27:18, 2023.

[Bullinger and Romen, 2024] Martin Bullinger and René Romen. Stability in online coalition formation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9537–9545, 2024.

[Bullinger, 2020] Martin Bullinger. Pareto-optimality in cardinal hedonic games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems, AAMAS*, pages 213–221, 2020.

[Christou *et al.*, 2024] Dimitrios Christou, Stratis Skoulakis, and Volkan Cevher. Efficient online clustering with moving costs. *Advances in Neural Information Processing Systems, NeurIPS*, 36, 2024.

[Cohen and Agmon, 2023a] Saar Cohen and Noa Agmon. Complexity of probabilistic inference in random dichotomous hedonic games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5573–5581, 2023.

[Cohen and Agmon, 2023b] Saar Cohen and Noa Agmon. Online coalitional skill formation. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, pages 494–503, 2023.

[Cohen and Agmon, 2024] Saar Cohen and Noa Agmon. Online learning of partitions in additively separable hedonic games – supplementary materials. https://u.cs.biu.ac.il/~agmon/CohenIJCAI24Sup.pdf, 2024.

[Drèze and Greenberg, 1980] Jacques H Drèze and Joseph Greenberg. Hedonic coalitions: Optimality and stability. *Econometrica: Journal of the Econometric Society*, 48(4):987–1003, 1980.

[Ekbatani *et al.*, 2023] Farbod Ekbatani, Yiding Feng, and Rad Niazadeh. Online resource allocation with buyback: Optimal algorithms via primal-dual. In *Proceedings of the 24th ACM Conference on Economics and Computation, EC*, page 583, 2023.

[Elkind and Wooldridge, 2009] Edith Elkind and Michael Wooldridge. Hedonic coalition nets. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, volume 1, pages 417–424, 2009.

[Elkind *et al.*, 2020] Edith Elkind, Angelo Fanelli, and Michele Flammini. Price of pareto optimality in hedonic games. *Artificial Intelligence*, 288:103357, 2020.

[Flammini *et al.*, 2021] Michele Flammini, Gianpiero Monaco, Luca Moscardelli, Mordechai Shalom, and Shmuel Zaks. On the online coalition structure generation problem. *Journal of Artificial Intelligence Research, JAIR*, 72:1215–1250, 2021.

[Fotakis *et al.*, 2021] Dimitris Fotakis, Georgios Piliouras, and Stratis Skoulakis. Efficient online learning for dynamic k-clustering. In *International Conference on Machine Learning, ICML*, pages 3396–3406. PMLR, 2021.

[Gergatsouli and Tzamos, 2022] Evangelia Gergatsouli and Christos Tzamos. Online learning for min sum set cover and pandora's box. In *International Conference on Machine Learning, ICML*, pages 7382–7403. PMLR, 2022.

[Hazan and Kale, 2012] Elad Hazan and Satyen Kale. Online submodular minimization. *Journal of Machine Learning Research*, 13(10), 2012.

[Hazan *et al.*, 2007] Elad Hazan, Amit Agarwal, and Satyen Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69:169–192, 2007.

[Hazan, 2016] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[Hihat *et al.*, 2023] Massil Hihat, Stéphane Gaïffas, Guillaume Garrigos, and Simon Bussy. Online inventory problems: Beyond the iid setting with online convex optimization. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[Jadbabaie *et al.*, 2015] Ali Jadbabaie, Alexander Rakhlin, Shahin Shahrampour, and Karthik Sridharan. Online optimization: Competing with dynamic comparators. In *Artificial Intelligence and Statistics*, pages 398–406, 2015.

[Jagadeesan *et al.*, 2021] Meena Jagadeesan, Alexander Wei, Yixin Wang, Michael Jordan, and Jacob Steinhardt. Learning equilibria in matching markets from bandit feedback. *Advances in Neural Information Processing Systems, NeurIPS*, 34:3323–3335, 2021.

[Krause and Golovin, 2014] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability*, 3(71-104):3, 2014.

[Li *et al.*, 2021] Xiaocheng Li, Chunlin Sun, and Yinyu Ye. The symmetry between arms and knapsacks: A primal-dual approach for bandits with knapsacks. In *International Conference on Machine Learning, ICML*, pages 6483–6492. PMLR, 2021.

[Liu *et al.*, 2020] Lydia T Liu, Horia Mania, and Michael Jordan. Competing bandits in matching markets. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, AISTATS*, pages 1618–1628, 2020.

[Maheshwari *et al.*, 2022] Chinmay Maheshwari, Shankar Sastry, and Eric Mazumdar. Decentralized, communication-and coordination-free learning in structured matching markets. *Advances in Neural Information Processing Systems, NeurIPS*, 35:15081–15092, 2022.

[Manurangsi, 2017] Pasin Manurangsi. Almost-polynomial ratio eth-hardness of approximating densest k-subgraph. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 954–961, 2017.

[Mokhtari *et al.*, 2016] Aryan Mokhtari, Shahin Shahrampour, Ali Jadbabaie, and Alejandro Ribeiro. Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *IEEE 55th Conference on Decision and Control, CDC*, pages 7195–7201, 2016.

[Sankararaman *et al.*, 2021] Abishek Sankararaman, Soumya Basu, and Karthik Abinav Sankararaman. Dominate or delete: Decentralized competing bandits in serial dictatorship. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics, AISTATS*, pages 1252–1260, 2021.

[Shalev-Shwartz, 2012] Shai Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

[Tekin and Van Der Schaar, 2015] Cem Tekin and Mihaela Van Der Schaar. Distributed online learning via cooperative contextual bandits. *IEEE Transactions on Signal Processing*, 63(14):3700–3714, 2015.

[Tirinzoni *et al.*, 2020] Andrea Tirinzoni, Matteo Pirotta, Marcello Restelli, and Alessandro Lazaric. An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. *Advances in Neural Information Processing Systems, NeurIPS*, 33:1417–1427, 2020.

[Woeginger, 2013] GJ Woeginger. Core stability in hedonic coalition formation. In *39th International Conference on Current Trends in Theory and Practice of Computer Science, SOFSEM*, pages 33–50. Springer, 2013.

[Yang *et al.*, 2016] Tianbao Yang, Lijun Zhang, Rong Jin, and Jinfeng Yi. Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient. In *International Conference on Machine Learning, ICML*, pages 449–457. PMLR, 2016.

[Zhang *et al.*, 2017] Lijun Zhang, Tianbao Yang, Jinfeng Yi, Rong Jin, and Zhi-Hua Zhou. Improved dynamic regret for non-degenerate functions. *Advances in Neural Information Processing Systems, NeurIPS*, 30, 2017.

[Zhang *et al.*, 2022] Yirui Zhang, Siwei Wang, and Zhixuan Fang. Matching in multi-arm bandit with collision. *Advances in Neural Information Processing System, NeurIPS*, 35:9552–9563, 2022.

[Zinkevich, 2003] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning, ICML*, pages 928–936, 2003.