Learning the Pareto Set Under Incomplete Preferences: Pure Exploration in Vector Bandits

Efe Mert Karagözlü
Bilkent University, Ankara, Turkey

Yaşar Cahit Yıldırım
Bilkent University, Ankara, Turkey

Çağın Ararat
Bilkent University, Ankara, Turkey

Cem Tekin
Bilkent University, Ankara, Turkey

Abstract

We study pure exploration in bandit problems with vector-valued rewards, where the goal is to (approximately) identify the Pareto set of arms given incomplete preferences induced by a polyhedral convex cone. We address the open problem of designing sample-efficient learning algorithms for such problems. We propose Pareto Vector Bandits (PaVeBa), an adaptive elimination algorithm that nearly matches the gap-dependent and worst-case lower bounds on the sample complexity of (ϵ, δ)-PAC Pareto set identification. Finally, we provide an in-depth numerical investigation of PaVeBa and its heuristic variants by comparing them with the state-of-the-art multi-objective and vector optimization algorithms on several real-world datasets with conflicting objectives.

1 INTRODUCTION

Pure exploration seeks to identify the optimal arms through sequential interaction. Usually, the error upper bound for identification is given and one seeks to minimize the sampling budget. This approach is termed the *fixed confidence setting* (Karnin et al., 2013). Notable algorithms for this include Exponential Gap Elimination (Karnin et al., 2013) and Track-and-Stop strategy (Garivier and Kaufmann, 2016). A similar concept is the (ϵ, δ)-probably approximately correct (PAC) best arm identification introduced by Even-Dar et al. (2006), where the aim is to find an ϵ-optimal arm with $1-\delta$ confidence. At $\epsilon = 0$, it coincides with the fixed confidence setting.

Although traditional pure exploration approaches primarily focus on scalar rewards, many real-world exploration challenges present multiple competing objectives. For instance, a communication channel with a low error rate, high bit rate, and a narrow bandwidth tends to consume more power; and a more complex neural network tends to require more time, computation, and data to be trained. Therefore, single-objective approaches fall short for real problems with $D > 1$ objectives that cannot be simplified to scalar optimization. Such problems gained attention from bandit literature in applications like digital hardware design (Zuluaga et al., 2016) and treatment optimization (Lizotte and Laber, 2016). To handle these, one needs a vector-valued multi-armed bandit framework that extends scalar rewards. Although vector rewards can be scalarized by a weighted sum of individual objectives, this approach can be difficult for the practitioner since it requires choosing a weight vector. Furthermore, using weighted linear combinations is not the only way to scalarize the reward vectors, and each real-life problem may require its own wise choice of (possibly highly nonlinear) scalarization function; hence, scalarization becomes even harder, and it motivates the study of exploration among vector-valued bandits on its own.

To that end, some work on bandit literature focused on the identification of the Pareto set of arms, i.e., arms that are not dominated by any other arm in all objectives. Noteworthy contributions include algorithms for (ϵ, δ)-*PAC Pareto set* identification (Auer et al., 2016) defined similarly to the single-objective case, exploration using Gaussian processes in large datasets (Zuluaga et al., 2016; Shah and Ghahramani, 2016), feasible arm identification (Katz-Samuels and Scott, 2018),

Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238. Copyright 2024 by the author(s).

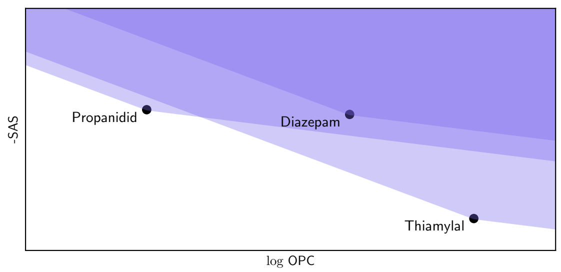


Figure 1: Three anesthetics ordered by a large polyhedral cone. The only Pareto optimal drug is Diazepam. (Data gathered from Wishart et al. (2018) and Huang et al. (2022).)

and algorithms using information-theoretic heuristics (Hernandez-Lobato et al., 2016; Belakaria et al., 2019; Tu et al., 2022). Additionally, regret minimization in multi-objective bandits is explored by Drugan and Nowe (2013) and Turgay et al. (2018), while multi-objective reinforcement learning has been delved into by Moffaert and Nowé (2014) and Hayes et al. (2022).

While multi-objective methods address vector-valued rewards, they use a specific domination concept, which we will call the standard componentwise order hereafter. The field of *vector optimization* (Jahn, 2011; Löhne, 2011) generalizes this by offering a more flexible dominance notion, a partial order $\preceq_C$ induced by a cone C defined as $\boldsymbol{\mu} \preceq_C \boldsymbol{\nu}$ if and only if (iff) $\boldsymbol{\nu} - \boldsymbol{\mu} \in C$. Choosing C as the positive orthant gives the standard approach, but the flexibility of the framework under different cones found diverse real-world applications since the choice of C allows users to reflect their incomplete (nontotal) preferences. For instance, *octanol-water partition coefficient* (OPC) and *synthetic accessibility score* (SAS) are both significant metrics in anesthetic design; the former has been shown to correlate well with anesthetic efficacy (Meyer, 1899; Overton, 1901), while the latter predicts the difficulty of synthesizing a molecule (Ertl and Schuffenhauer, 2009). To illustrate the problem with the standard componentwise order, Figure 1 shows the negative of SAS (ease of synthesis) and $\log \text{OPC}$ of three different anesthetics, all of which are Pareto optimal in the standard sense. However, Propanidid is at an absolute disadvantage in practice compared to Diazepam. This is because ease of synthesis has no value unless the drug is potent, and Diazepam is only marginally more difficult to synthesize but invaluably more effective in inducing anesthesia. By using a larger cone, as shown, the optimal molecule among the three can be reduced to Diazepam, eliminating the molecules that favor extreme trade-offs between accessibility and efficacy. On the other hand, in some scenarios in molecule design, where chemical properties are objectives, utilizing a smaller cone can be a better option. Initially, smaller cones aid cheap experiments to eliminate sub-optimal molecules (Jayatunga et al., 2022). Subsequently, detailed and costlier experiments focus on fewer molecules that are likely to be optimal in the standard sense.

Recently, Ararat and Tekin (2023) delved into vector optimization in the bandit context with sequential noisy samplings. While they identified key results and lower bounds on the sample complexity for (ϵ, δ)-PAC exploration, a matching sample-efficient algorithm remained elusive. In this paper, we explore the vector-valued bandit setting using a cone-induced order, following the vector optimization approach adopted by Jahn (2011) and Löhne (2011). We introduce *Pareto Vector Bandits* (PaVeBa) to bridge the gap in the literature, nearly matching the lower bounds identified by Ararat and Tekin (2023).

We also contribute to the bandit-based vector optimization theory developed by Ararat and Tekin (2023) by proving important properties about their fundamental *gap functions* $m(\cdot,\cdot)$ and $M(\cdot,\cdot)$, and *ordering complexity* β. In particular, we demonstrate that these gap functions are Lipschitz with constant β, emphasizing the role of β in dictating the complexity of the specific ordering. Additionally, we introduce a stricter success condition for (ϵ, δ)-PAC Pareto identification and present convex optimization formulations to perform Pareto identification tasks using ellipsoids from multi-objective Gaussian processes with correlated outputs.

Finally, we benchmark PaVeBa against state-of-the-art multi-objective algorithms and the Naïve Elimination algorithm by Ararat and Tekin (2023) for vector optimization. PaVeBa surpasses Naïve Elimination; its heuristic variants yield results akin to ϵ-PAL (Zuluaga et al., 2016), MESMO (Belakaria et al., 2019), and JES (Tu et al., 2022) in the standard multi-objective problem.

2 PROBLEM DEFINITION

Let $\mathcal{X}$ represent a finite set of arms. Each arm $x \in \mathcal{X}$ has a corresponding mean reward vector $\boldsymbol{f}(x) \in \mathbb{R}^D$, where $D \in \mathbb{N} := \{1, 2, \ldots\}$. We denote by $\|\cdot\|_2$ the Euclidean norm on $\mathbb{R}^D$ and by $B(\boldsymbol{\mu}, r)$ the closed ball in $\mathbb{R}^D$ with respect to $\|\cdot\|_2$ whose center is $\boldsymbol{\mu} \in \mathbb{R}^D$ and radius is $r \geq 0$. Given $\boldsymbol{\mu} \in \mathbb{R}^D$ and a nonempty set $A \subseteq \mathbb{R}^D$, we define $d(\boldsymbol{\mu}, A) := \inf_{\boldsymbol{\nu} \in A} \|\boldsymbol{\mu} - \boldsymbol{\nu}\|_2$. Let $t \in \mathbb{N}$ be an arbitrary index for rounds of sequential evaluations. If some agent decides to sample $x \in \mathcal{X}$ in round t (possibly along with other arms), they observe the random vector $\boldsymbol{y}_t(x) = \boldsymbol{f}(x) + \boldsymbol{\eta}_t(x)$, where

$\boldsymbol{\eta}_t(x)$ is the random noise vector at round t for arm x. We assume that the family $(\boldsymbol{\eta}_t(x))_{x\in\mathcal{X},t\in\mathbb{N}}$ consists of independent norm-subgaussian random vectors with a common parameter $\sigma > 0$. In particular,

$$\mathbb{P}\{\|\boldsymbol{\eta}_t(x)\|_2 \geq u\} \leq 2e^{-\frac{u^2}{2\sigma^2}}$$

for each $u \geq 0$ (Jin et al., 2019, Definition 3). For each $x \in \mathcal{X}$, let $N_t(x)$ denote the set of rounds by round t at which arm x is sampled, and let $n_t(x) = |N_t(x)|$.

We focus on pure exploration, where the goal is to find the set of "optimal" arms. Since rewards are multi-dimensional, one needs to define a notion of ranking between the reward vectors. Following Ararat and Tekin (2023), we use a convex polyhedral ordering cone C to partially order the reward vectors. We define $C \coloneqq \{\boldsymbol{\mu} \in \mathbb{R}^D \mid \boldsymbol{W}\boldsymbol{\mu} \geq \mathbf{0}\}$, where $\boldsymbol{W}$ is an $N \times D$ real matrix for some $N \in \mathbb{N}$ with rows $\boldsymbol{w}_1^\mathsf{T}, \ldots, \boldsymbol{w}_N^\mathsf{T}$. We assume that $\boldsymbol{W}$ has full row rank and $\|\boldsymbol{w}_n\|_2 = 1$ for each $n \in [N] \coloneqq \{1, \ldots, N\}$. Hence, the interior of C is $\mathrm{int}(C) = \{\boldsymbol{\mu} \in \mathbb{R}^D \mid \boldsymbol{W}\boldsymbol{\mu} > \mathbf{0}\}$ and the boundary of C is $\mathrm{bd}(C) = \{\boldsymbol{\mu} \in C \mid \exists n \in [N]\colon \boldsymbol{w}_n^\mathsf{T}\boldsymbol{\mu} = 0\}$. For each $n \in [N]$, we also introduce the constant $\alpha_n \coloneqq \sup_{\boldsymbol{u}\in B(\mathbf{0},1)\cap C} \boldsymbol{w}_n^\mathsf{T}\boldsymbol{u} \in (0, 1]$. Using C, we define a partial order $\preceq_C$ via $\boldsymbol{\mu} \preceq_C \boldsymbol{\nu}$ iff $\boldsymbol{\nu} - \boldsymbol{\mu} \in C$ for each $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^D$. In this case, we say that $\boldsymbol{\mu}$ is *weakly dominated* by $\boldsymbol{\nu}$. Similarly, we say that $\boldsymbol{\mu}$ is *strongly dominated* by $\boldsymbol{\nu}$, denoted by $\boldsymbol{\mu} \prec_C \boldsymbol{\nu}$, iff $\boldsymbol{\mu} - \boldsymbol{\nu} \in \mathrm{int}(C)$.

Previously, Ararat and Tekin (2023) identified two parameters associated with the difficulty of identifying if an arm is dominated by another arm, given as

$$\beta_1 \coloneqq \sup_{\boldsymbol{\mu}\notin C} \frac{d(\boldsymbol{\mu}, C \cap (\boldsymbol{\mu} + C))}{d(\boldsymbol{\mu}, C)}\ , \tag{1}$$

$$\beta_2 \coloneqq \sup_{\boldsymbol{\mu}\in\mathrm{int}(C)} \frac{d(\boldsymbol{\mu}, (\mathrm{int}(C))^c \cap (\boldsymbol{\mu} - C))}{d(\boldsymbol{\mu}, (\mathrm{int}(C))^c)}\ . \tag{2}$$

Theorem 2.4 of their work states that both of these constants are finite, and they call $\beta \coloneqq \max\{\beta_1, \beta_2\}$ the *ordering complexity* of cone C, which they associate with the difficulty of the vector optimization problem.

Having the basics established, we define our main goal next. Pure exploration in the vector bandits problem seeks to maximize the latent function $\boldsymbol{f}$ over $\mathcal{X}$ with respect to the partial order $\preceq_C$ with the minimum sampling budget. In other words, the objective is to find the set P^* of all arms yielding the maximal elements of $\{\boldsymbol{f}(x) \mid x \in \mathcal{X}\}$ with respect to $\preceq_C$, i.e.,

$$P^* = \{x \in \mathcal{X} \mid \nexists y \in \mathcal{X} \setminus \{x\}\colon \boldsymbol{f}(x) \preceq_C \boldsymbol{f}(y)\},$$

which is also referred to as the *Pareto set*.

To identify the Pareto set P^*, we employ two functions defined by Ararat and Tekin (2023) as generalizations of their multi-objective counterparts in the work of Auer et al. (2016). Different from the approach of Ararat and Tekin (2023), where these functions are defined on the set of arms, we generalize the use of these functions by defining them on arbitrary vectors $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^D$. First, we define $M(\boldsymbol{\mu}, \boldsymbol{\nu})$ as the minimum improvement along an arbitrary direction in C needed for $\boldsymbol{\nu}$ to dominate $\boldsymbol{\mu}$. Formally,

$$M(\boldsymbol{\mu}, \boldsymbol{\nu}) \coloneqq \inf\{s \geq 0 \mid \exists \boldsymbol{u} \in B(\mathbf{0}, 1) \cap C\colon \boldsymbol{\nu} + s\boldsymbol{u} \in \boldsymbol{\mu} + C\}. \tag{3}$$

Second, we define $m(\boldsymbol{\mu}, \boldsymbol{\nu})$ as the minimum improvement along an arbitrary direction in C needed for $\boldsymbol{\mu}$ not to be dominated by $\boldsymbol{\nu}$. Formally,

$$m(\boldsymbol{\mu}, \boldsymbol{\nu}) \coloneqq \inf\{s \geq 0 \mid \exists \boldsymbol{u} \in B(\mathbf{0}, 1) \cap C\colon \boldsymbol{\mu} + s\boldsymbol{u} \notin \boldsymbol{\nu} - \mathrm{int}(C)\}. \tag{4}$$

Some useful properties of these two functions that we have used for the sample complexity analysis of our algorithm were proven by Ararat and Tekin (2023) in their Propositions 4.2, 4.3; Corollary 4.5. In particular, it is sufficient to keep track of these two functions to identify the Pareto set because their signs reveal the dominance status of any two arms.

As mentioned in the introduction, many different objectives, such as fixed confidence and (ϵ, δ)-PAC identification, have been used in pure exploration. For the vector case, we use Definition 1.

Definition 1. *Let $\epsilon \geq 0, \delta \in (0, 1)$. A random set $P \subseteq \mathcal{X}$ is called an (ϵ, δ)-PAC Pareto set if the following success conditions hold at least with probability $1 - \delta$: (i) $P^* \subseteq P$; (ii) for every $x \in P \backslash P^*$, it holds $\Delta^*(x) \leq \epsilon$, where $\Delta^*(x) \coloneqq \max_{y\in P^*} m(\boldsymbol{f}(x), \boldsymbol{f}(y))$.*

Our success conditions are modified versions of the ones in Ararat and Tekin (2023). Specifically, we changed their first condition to a stricter one. Hence, every (ϵ, δ)-PAC Pareto set according to Definition 1 is also a one according to their Definition 4.6.

The sample complexity of any algorithm for (ϵ, δ)-PAC Pareto set identification will depend on how close the reward vectors are. To quantify the "closeness" of the arms to the Pareto set, we use the following definitions in Ararat and Tekin (2023), Auer et al. (2016): $\Delta^+(x) = \min_{y\in P^*\setminus\{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y))$ for $x \in P^*$ and $\Delta^+(x) = \max_{y\in P^*} m(\boldsymbol{f}(x), \boldsymbol{f}(y))$ for $x \in \mathcal{X} \setminus P^*$. So

$$\Delta^+(x) = \max\left\{\min_{y\in P^*\setminus\{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)), \max_{y\in P^*\setminus\{x\}} m(\boldsymbol{f}(x), \boldsymbol{f}(y))\right\}$$

since the product of the two gap functions is always zero by Corollary 4.5 of Ararat and Tekin (2023).

Inspired by the multi-objective framework of Auer et al. (2016), we further introduce an improved gap definition for Pareto arms. We need this new definition because of the design of our algorithm. In some cases, our algorithm continues to sample the arms even if it is almost sure that they are Pareto optimal. This is due to the fact that getting more sure about a Pareto arm can be useful in deciding on another arm's optimality. To quantify the contribution of such samplings to our sample complexity, for each $x \in P^*$, we define

$$\Delta(x) \coloneqq \min\Bigg\{ \min_{y \in P^* \setminus \{x\}} \{M(\boldsymbol{f}(x), \boldsymbol{f}(y)), M(\boldsymbol{f}(y), \boldsymbol{f}(x))\}, \min_{y \notin P^*} \left(M(\boldsymbol{f}(y), \boldsymbol{f}(x)) + 2\Delta^+(y) \right) \Bigg\}. \quad (5)$$

Finally, let $\widetilde{\Delta}^+_\epsilon(x) \coloneqq \max\{\Delta^+(x), \epsilon\}$, $\widetilde{\Delta}_\epsilon(x) \coloneqq \max\{\Delta(x), \epsilon\}$. The use of these ϵ-dependent gaps is for a tighter analysis of our sample complexity in case the algorithm terminates not because of its confidence 'beating' the natural gaps between the rewards of the arms but ϵ being so large that, even with low confidence on the reward vectors, our algorithm can be sure about the (ϵ, δ)-Pareto status of the arms.

3 PARETO VECTOR BANDITS

We present Pareto Vector Bandits (PaVeBa), an algorithm that solves the Pareto identification problem on the vector bandits (pseudocode in Algorithm 1).

At each round $t \in \mathbb{N}$, the algorithm keeps track of a set of undecided (or 'secret') arms $\mathcal{S}_t$, a set of Pareto arms $\mathcal{P}_t$, and a set of useful Pareto arms $\mathcal{U}_t$. It samples each 'active' arm in $\mathcal{A}_t \coloneqq \mathcal{S}_t \cup \mathcal{U}_t$ for judgment to get more confident as long as there are some undecided arms, i.e., $\mathcal{S}_t \neq \emptyset$. Then, it detects the set of arms that are unlikely to satisfy condition (i) in Definition 1 as

$$\mathcal{D}_t = \{x \in \mathcal{S}_t \mid \exists y \in \mathcal{A}_t \setminus \{x\} \colon \sup_{\substack{\boldsymbol{\mu} \in \mathcal{E}_t(x), \\ \boldsymbol{\nu} \in \mathcal{E}_t(y)}} M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0\}, \quad (6)$$

where $\mathcal{E}_t(x)$ is a high-probability confidence region for arm x at round t. After this, it removes $\mathcal{D}_t$ from $\mathcal{S}_t$ and $\mathcal{A}_t$ to find the intermediate sets $\overline{\mathcal{S}}_t$ and $\overline{\mathcal{A}}_t$. Similarly, it detects the set of arms that appear to satisfy condition (ii) in Definition 1 as

$$\overline{\mathcal{P}}_t = \{x \in \overline{\mathcal{S}}_t \mid \forall y \in \overline{\mathcal{A}}_t \setminus \{x\} \colon \sup_{\substack{\boldsymbol{\mu} \in \mathcal{E}_t(x), \\ \boldsymbol{\nu} \in \mathcal{E}_t(y)}} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon\}, \quad (7)$$

appends it to $\mathcal{P}_t$ to obtain $\mathcal{P}_{t+1}$, and removes it from $\mathcal{S}_t$ to obtain $\mathcal{S}_{t+1}$. It also finds a set of useful Pareto arms for the next round as

$$\mathcal{U}_{t+1} = \{y \in \mathcal{P}_{t+1} \mid \exists x \in \mathcal{S}_{t+1} \colon \sup_{\substack{\boldsymbol{\mu} \in \mathcal{E}_t(x), \\ \boldsymbol{\nu} \in \mathcal{E}_t(y)}} m(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \epsilon\}, \quad (8)$$

which consists of the arms in $\mathcal{P}_{t+1}$ that may help in determining the Pareto optimality of an arm in $\mathcal{S}_{t+1}$.

PaVeBa needs a confidence region for the reward of each arm to make decisions about its status. To that end, we define $\mathcal{B}_t(x) \coloneqq B(\boldsymbol{\mu}_t(x), r_t(x))$, where

$$r_t(x) \coloneqq \sqrt{\frac{8t\sigma^2}{n_t(x)^2} \log\left(\frac{\pi^2(D+1)|\mathcal{X}|t^2}{6\delta}\right)} \quad (9)$$

and $\boldsymbol{\mu}_t(x)$ is the sample mean of arm x at round t, i.e., $\boldsymbol{\mu}_t(x) = \frac{1}{n_t(x)} \sum_{\tau \in N_t(x)} \boldsymbol{y}_\tau(x)$. Further, we define $\mathcal{E}_t(x) \coloneqq \mathcal{E}_{t-1}(x) \cap \mathcal{B}_t(x)$ with $\mathcal{E}_1(x) \coloneqq \mathcal{B}_1(x)$ as the confidence region of x at round t.

Algorithm 1 Pareto Vector Bandits (PaVeBa)

```
Input: $\mathcal{X}$, $C$, $\epsilon$, $\delta$
Initialize: $\mathcal{S}_1 = \mathcal{X}$, $\mathcal{P}_1 = \emptyset$, $\mathcal{U}_1 = \emptyset$, $t = 1$;
while $\mathcal{S}_t \neq \emptyset$ do
    $\mathcal{A}_t = \mathcal{S}_t \cup \mathcal{U}_t$;
    for $x \in \mathcal{A}_t$ do
        Observe $\boldsymbol{y}_t(x)$, calculate $\mathcal{E}_t(x)$ using (9);
    end for
    Compute $\mathcal{D}_t$ using (6);
    $\overline{\mathcal{S}}_t = \mathcal{S}_t \setminus \mathcal{D}_t$, $\overline{\mathcal{A}}_t = \mathcal{A}_t \setminus \mathcal{D}_t$;
    Compute $\overline{\mathcal{P}}_t$ using (7);
    $\mathcal{P}_{t+1} = \mathcal{P}_t \cup \overline{\mathcal{P}}_t$, $\mathcal{S}_{t+1} = \overline{\mathcal{S}}_t \setminus \overline{\mathcal{P}}_t$;
    Compute $\mathcal{U}_{t+1}$ using (8);
    $t = t + 1$;
end while
return $\hat{P} = \mathcal{P}_t$
```

PaVeBa can be seen as a generalization of Algorithm 1 of Auer et al. (2016) for arbitrary ordering cones. However, it has some distinct features even when the ordering cone is the standard componentwise order. First, we assume norm-subgaussian noise, which results in spherical confidence regions, whereas Auer et al. (2016) assume subgaussian noise in each dimension, thus yielding hyperrectangular confidence regions. This assumption and the fact that they only consider the standard componentwise order reduce their discarding and Pareto identification operations to simple comparisons between the empirical means. In PaVeBa, we compare all vectors in the spherical confidence regions using the ordering cone by solving convex optimization problems. Second, some redundant comparisons in their work do not exist in ours. For instance, for the discarding operation, they search for arms to discard within $\mathcal{A}_t$ (the set A in their notation), whereas we only consider the arms in $\mathcal{S}_t$ for discarding as the other arms in $\mathcal{A}_t$ are likely to be Pareto optimal. Third, in their work, Pareto arms accumulate (in the set P in their notation) not when they are identified as Pareto optimal arms but when they become 'useless' in identifying the status of other arms.

This might not seem like a significant difference, but using the knowledge of Pareto arms, we can remove redundant comparisons in the algorithm, as exemplified by not seeking arms to discard among Pareto arms.

4 TECHNICAL ANALYSIS

Efficient Implementation of PaVeBa via Convex Programming. Computing the sets $\mathcal{D}_t$, $\overline{\mathcal{P}}_t$, $\mathcal{U}_{t+1}$ requires checking the validity of certain conditions for all possible choices of vectors in the confidence regions. To make our analysis compatible with the heuristic and future use of Gaussian processes, we assume that the confidence region of an arm $x \in \mathcal{X}$ at round $t \in \mathbb{N}$ is an intersection of the form $\mathcal{E}_t(x) = \bigcap_{\tau=1}^{t} \mathcal{B}_\tau(x)$, where $\mathcal{B}_\tau(x) = \{\boldsymbol{\nu} \in \mathbb{R}^D \mid (\boldsymbol{\nu} - \boldsymbol{\mu}_\tau(x))^\mathsf{T} \boldsymbol{\Sigma}_\tau^{-1}(x)(\boldsymbol{\nu} - \boldsymbol{\mu}_\tau(x)) \leq \alpha_\tau(x)\}$ is a generic ellipsoid with parameters $\boldsymbol{\Sigma}_\tau(x)$, a symmetric positive definite matrix, and $\alpha_\tau(x) > 0$ for each $\tau \in [t]$. (By choosing $\boldsymbol{\Sigma}_\tau(x)$ as the identity matrix and $\alpha_\tau(x) = r_\tau(x)$, we recover $\mathcal{B}_\tau(x) = B(\boldsymbol{\mu}_\tau(x), r_\tau(x))$ as in (9).) In the general case, Propositions 1 and 2, whose proofs are in the supplementary §1, provide methods for computing the sets of interest.

Proposition 1. *Let $x, y \in \mathcal{X}$ and $t \in \mathbb{N}$. For each $n \in [N]$, consider the convex optimization problem*

$$\textit{minimize} \quad \boldsymbol{w}_n^\mathsf{T}(\boldsymbol{\nu} - \boldsymbol{\mu}) \quad \textit{subject to:} \quad \boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^D,$$
$$(\boldsymbol{\mu} - \boldsymbol{\mu}_\tau(x))^\mathsf{T} \boldsymbol{\Sigma}_\tau^{-1}(x)(\boldsymbol{\mu} - \boldsymbol{\mu}_\tau(x)) \leq \alpha_\tau(x),\ \forall \tau \in [t],$$
$$(\boldsymbol{\nu} - \boldsymbol{\mu}_\tau(y))^\mathsf{T} \boldsymbol{\Sigma}_t^{-1}(y)(\boldsymbol{\nu} - \boldsymbol{\mu}_\tau(y)) \leq \alpha_\tau(y),\ \forall \tau \in [t].$$

Then, the optimal value of this problem is strictly negative for at least one $n \in [N]$ iff $\sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} M(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0$.

In view of Proposition 1, we can compute $\mathcal{D}_t$ by iterating first over $x \in \mathcal{S}_t$ and then over $y \in \mathcal{A}_t \setminus \{x\}$, and solving the convex optimization problem for each $n \in [N]$. If we cannot detect at least one y and n for which the optimal value is strictly negative, then we add x to $\mathcal{D}_t$.

Proposition 2. *Let $x, y \in \mathcal{X}$, $t \in \mathbb{N}$, and $\epsilon > 0$. Consider the convex feasibility problem*

$$\textit{minimize} \quad 0 \quad \textit{subject to:} \quad \boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^D,$$
$$\boldsymbol{w}_n^\mathsf{T}(\boldsymbol{\nu} - \boldsymbol{\mu}) \geq \epsilon \alpha_n,\ \forall n \in [N]$$
$$(\boldsymbol{\mu} - \boldsymbol{\mu}_\tau(x))^\mathsf{T} \boldsymbol{\Sigma}_\tau^{-1}(x)(\boldsymbol{\mu} - \boldsymbol{\mu}_\tau(x)) \leq \alpha_\tau(x),\ \forall \tau \in [t],$$
$$(\boldsymbol{\nu} - \boldsymbol{\mu}_\tau(y))^\mathsf{T} \boldsymbol{\Sigma}_\tau^{-1}(y)(\boldsymbol{\nu} - \boldsymbol{\mu}_\tau(y)) \leq \alpha_\tau(y),\ \forall \tau \in [t].$$

Then, this problem has at least one feasible solution iff $\sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \epsilon$.

In view of Proposition 2, to calculate $\overline{\mathcal{P}}_t$, we iterate first over $x \in \overline{\mathcal{S}}_t$ and then over $y \in \overline{\mathcal{A}}_t \setminus \{x\}$, and solve the convex feasibility problem. If the problem turns out to be infeasible for all y, then we add x to $\overline{\mathcal{P}}_t$.

In a similar way, we also calculate $\mathcal{U}_{t+1}$. In this case, we iterate first over $y \in \mathcal{P}_{t+1}$ and then over $x \in \mathcal{S}_{t+1}$, and solve the convex feasibility problem. If the problem has at least one feasible solution for at least one x, then we add y to $\mathcal{U}_{t+1}$ thanks to Proposition 2.

Note that in the convex programs of Propositions 1 and 2, the number of constraints grows linearly with time t. This is because we need to intersect the confidence regions $\mathcal{B}_\tau$ to ensure no new vector is introduced to a confidence region as time passes. This is to ensure the proofs are working, specifically Lemma 8. (see supplementary §2.4)

Remark 1. *When $\epsilon = 0$, Proposition 2 does not work. In this case, we suggest a slightly different algorithm with minor modifications in the definitions of $\overline{\mathcal{P}}_t$, $\mathcal{U}_{t+1}$. We provide another convex optimization formulation to calculate these sets and prove similar sample complexity bounds for that version in the supplementary §3.*

Sample Complexity of PaVeBa. In this part, we provide upper bounds for the sample complexity of the algorithm. The theoretical analysis follows from a series of lemmas, all of whose statements and proofs are provided in the supplementary §2.

We start by presenting the main results, namely, Proposition 3 and Proposition 4, which provide invaluable insights for understanding how the cone-dependent ordering complexity β plays a role in the vector bandits problem.

Proposition 3. *For every $\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\epsilon} \in \mathbb{R}^D$, we have $|M(\boldsymbol{\mu} + \boldsymbol{\epsilon}, \boldsymbol{\nu}) - M(\boldsymbol{\mu}, \boldsymbol{\nu})| \leq \beta_1 \|\boldsymbol{\epsilon}\|_2$.*

Proposition 3 follows from the definition of β_1 and a triangle inequality for $M(\cdot, \cdot)$ that extends Auer et al. (2016, Lemma 3) to the vector optimization setting.

Proposition 4. *For every $\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\epsilon} \in \mathbb{R}^D$, we have $|m(\boldsymbol{\mu} + \boldsymbol{\epsilon}, \boldsymbol{\nu}) - m(\boldsymbol{\mu}, \boldsymbol{\nu})| \leq \beta_2 \|\boldsymbol{\epsilon}\|_2$.*

The proof of Proposition 4 is more involved and consists of some novel geometric arguments. Together with Proposition 3, it establishes the Lipschitz continuity of our fundamental gap functions. Surprisingly, β_1 and β_2 turn out to be the corresponding Lipschitz constants. The importance of this can be seen by considering the vector $\boldsymbol{\epsilon}$ in the statements as an error between what an algorithm thinks the reward vector of an arm is and its actual value. Then, our two fundamental gap functions for Pareto identification are at most as erroneous as β_1 and β_2 times the norm of $\boldsymbol{\epsilon}$. Hence, it is natural to expect that the cone-dependence of an (ϵ, δ)-PAC Pareto set identifying algorithm making use of $m(\cdot, \cdot)$ and $M(\cdot, \cdot)$ is expressed in terms of β_1 and β_2. This is indeed the case for PaVeBa, as shown in Theorems 1 and 2, our final sample complexity bounds. Let $\log^+(\cdot) := \max\{\log(\cdot), 0\}$.

Theorem 1. *When PaVeBa is run on a finite set of arms $\mathcal{X}$, the maximum number of samples required for it to output an (ϵ, δ)-PAC Pareto set $\hat{P}$ is*

$$\frac{|\mathcal{X}| 512 \beta_2^2 \sigma^2}{\epsilon^2} \log^+ \left(\frac{256 \beta_2^2 \sigma^2}{\epsilon^2} \sqrt{\frac{\pi^2 (D+1) |\mathcal{X}|}{6\delta}} \right) + |\mathcal{X}| \ .$$

Our sample complexity bound in Theorem 1 does not depend on gaps between arms, and it nearly matches the worst-case lower bound $\Omega((|\mathcal{X}|\beta_2^2/\epsilon^2)\log(1/\delta))$ provided in Ararat and Tekin (2023, Theorem 5.3). The additional dependence on ϵ, β_2, $|\mathcal{X}|$ in the logarithmic term makes our sample complexity bound nearly-matching. Next, we present an improved result based on the individual gaps of arms.

Theorem 2. *When PaVeBa is run on a finite set of arms $\mathcal{X}$, the maximum number of samples required for it to output an (ϵ, δ)-PAC Pareto set $\hat{P}$ is*

$$|\mathcal{X}| + \sum_{x \in P^*} \frac{4608 \beta^2 \sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2} \log^+ \left(\frac{2304 \beta^2 \sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2} \sqrt{\frac{\pi^2 (D+1) |\mathcal{X}|}{6\delta}} \right)$$
$$+ \sum_{x \in \mathcal{X} \setminus P^*} \frac{512 \beta^2 \sigma^2}{\widetilde{\Delta}^+_{\epsilon}(x)^2} \log^+ \left(\frac{256 \beta^2 \sigma^2}{\widetilde{\Delta}^+_{\epsilon}(x)^2} \sqrt{\frac{\pi^2 (D+1) |\mathcal{X}|}{6\delta}} \right).$$

Here, our sample complexity bound nearly matches the gap-dependent lower bound $\Omega(\sum_{x\in\mathcal{X}}(1/\widetilde{\Delta}^\epsilon_x)\log(1/\delta))$ provided in Ararat and Tekin (2023, Theorem 5.1), where $\widetilde{\Delta}^\epsilon_x$ corresponds to $\widetilde{\Delta}^+_\epsilon(x)$ for suboptimal arms exactly and to $\widetilde{\Delta}_\epsilon(x)$ for Pareto arms except a little discrepancy explained in Remark 2. The additional dependence on the gaps in the logarithmic terms makes our sample complexity bound nearly-matching. To the best of our knowledge, PaVeBa is the first algorithm to nearly match the theoretical lower bounds on the sample complexity of the pure exploration problem for vector bandits. It is also worth noting that since our success condition is stricter than that of Ararat and Tekin (2023), the bounds we match are for an even easier problem.

Notice here that our worst-case sample complexity bound depends on β_2 (but not on β_1) while the gap-dependent one depends on β. This is because, in the worst-case gap configuration, PaVeBa eliminates arms not because it is sure about their actual Pareto status due to their confidence regions shrinking sufficiently to exploit the gaps but because arms are added to the Pareto set due to the ϵ-looseness in (7). Therefore, individual gaps of the arms do not matter. In addition, $m(\cdot,\cdot)$ becomes the main determining function in the algorithm since the discarding step requires comparisons between $M(\cdot,\cdot)$ and 0, that can be translated into comparisons using $m(\cdot,\cdot)$ due to Ararat and Tekin (2023, Corollary 4.5). These two causes make β_2 the only cone-dependent term appearing in the final result. However, in the gap-dependent analysis, $M(\cdot,\cdot)$ also plays a significant role since our gap definitions involve it, thus resulting in a β_1-dependence as well in the theorem.

We note that the exact dependence of the gap-dependent lower bound on β remains unresolved at this stage. Indeed, the gap-dependent lower bound in Ararat and Tekin (2023, Corollary 4.5) has a cone-dependent scalar k appearing in the proof. We think that β is highly related to the problem's difficulty since an adversarial placement of the mean vectors results in a β appearance in the sample complexity lower bound, and we match this in our worst-case bound. Further, we regress the average sample complexity under different cones on β^2 using the least squares approach in the experiments section (see Figure 4) and observe that β^2 precisely explains the variation in the sample complexity.

Remark 2. *For $x \in \mathcal{X} \setminus P^*$, the gap term $\widetilde{\Delta}^+_\epsilon(x)$ in Theorem 2 matches with the one in the gap-dependent lower bound (Theorem 5.1) of Ararat and Tekin (2023). For $x \in P^*$, the gap term $\widetilde{\Delta}_{3\epsilon}(x)$ in Theorem 2 may be different from the ones that appear on Theorem 5.1 of Ararat and Tekin (2023). One such case is when $\widetilde{\Delta}_{3\epsilon}(x) = M(\boldsymbol{f}(y), \boldsymbol{f}(x)) + 2\Delta^+(y)$ for some $y \notin P^*$. In this case, as in Auer et al. (2016), one can replace the gap term of x appearing in our upper bound with the gap term of y in the lower bound, still having a valid upper bound. Another case is when $\widetilde{\Delta}_{3\epsilon}(x) = M(\boldsymbol{f}(y), \boldsymbol{f}(x))$ for some $y \in P^* \setminus \{x\}$. This case can be handled by mapping x to some y' through a finite number of replacement iterations and replacing the gap term of x in our upper bound with the gap term of y' in the lower bound. When the number of arms $x \in P^*$ for which these cases happen is $O(1)$, the new upper bound matches with the lower bound in Ararat and Tekin (2023) in terms of the gaps up to logarithmic factors. The details of this argument can be found in the supplementary §2.7.*

When proving Theorems 1, 2, we begin by showing that, for all arms and rounds, the true reward vectors lie inside PaVeBa's confidence regions ($\mathcal{E}_t(x)$) around the empirical means at least with probability $1-\delta$. Then, we bound the variation between the latent values of functions $m(\cdot,\cdot)$, $M(\cdot,\cdot)$ and their estimates based on the empirical means, given that the actual values and estimates are close. To that end, we prove Propositions 3, 4 after a series of intermediate results about the nature of the vector bandits problem. Using these and the fact that confidence regions shrink arbitrarily, we show that each arm x gets eliminated from the set $\mathcal{S}_t$ after sufficiently many samplings dictated by its "gap" $\widetilde{\Delta}^+_\epsilon(x)$. Similarly, we show that arms get eliminated from $\mathcal{U}_t$ after a sampling number dictated by either $\widetilde{\Delta}^+_\epsilon(x)$ or $\widetilde{\Delta}_{3\epsilon}(x)$. In doing so, we follow a different approach from Auer et al. (2016) since our discarding and Pareto identification operations consider all vectors inside the confidence regions (not only empirical means) to work under arbitrary cones and intersections of ellipsoidal confidence regions. These results yield the sample complexity. As for the successful Pareto identification, given that confidence regions contain actual reward vectors, we show that true Pareto arms never get discarded, and the sets $\mathcal{P}_t$ never contain highly suboptimal arms that break the (ϵ, δ)-PAC-ness of the final set. Hence, as the rounds pass, a set $\hat{P}$ satisfying Definition 1 is accumulated.

Remark 3. *Different from our work, Katz-Samuels and Scott (2018) assume multi-dimensional subgaussian noise distribution. A σ-subgaussian D-dimensional distribution is $2\sqrt{2}\sigma\sqrt{D}$-norm-subgaussian, see Jin et al. (2019, Lemma 1). Auer et al. (2016) assume that the noise dimensions are marginally σ-subgaussian. Then, under the special case where noise dimensions are independent σ-subgaussian, it is known that the noise vector is σ-subgaussian, which is again $2\sqrt{2}\sigma\sqrt{D}$-norm-subgaussian. Hence, when the noise vector is σ-subgaussian, our sample complexity bounds in Theorems 1 and 2 will involve $8\sigma^2 D$ instead of σ^2.*

5 EXPERIMENTS

In this section, we assess the performance of PaVeBa and compare it with the state of the art. We introduce heuristic variants of PaVeBa that are tailored to achieve low sample complexity in problems with correlated arms. We keep this section concise due to space considerations. Further details can be found in the supplementary §5. Our implementation for PaVeBa can be found at `https://github.com/Bilkent-CYBORG/PaVeBa`.

Heuristic Variants. To improve PaVeBa's empirical performance in large, correlated experimental design problems, we introduce heuristic variants using Gaussian Processes (GPs). Consistent with common practice, each reward dimension is modeled by a distinct GP. We also employ correlated GP outputs, leading to ellipsoidal confidence regions. A derivation connecting this region's significance level to the confidence term δ is in the supplementary §4. PaVeBa variants utilizing independent and correlated GP outputs are termed PaVeBa-IH and PaVeBa-DE, with "IH" denoting "Independent and Hyperrectangular" and "DE" denoting "Dependent and Ellipsoidal" with the latter incorporating a Linear Model of Coregionalization.

5.1 Real-World Problems

SNW ($D = 2$, $|\mathcal{X}| = 206$): This dataset is derived from the domain of computational hardware design, specifically concerning the optimization of sorting network configurations (Zuluaga et al., 2012).

DB (Disc Brake, $D = 2$, $|\mathcal{X}| = 128$): This dataset addresses efficiency and safety in automotive engineering, presenting an optimization problem in disc brake manufacturing (Tanabe and Ishibuchi, 2020).

PK2 ($D = 2$, $|\mathcal{X}| = 500$): In the context of organic chemistry, this dataset aims at the optimization of the Paal-Knorr synthesis, a fundamental reaction for the synthesis of pyrroles and pyrrolidines (Moore and Jensen, 2012).

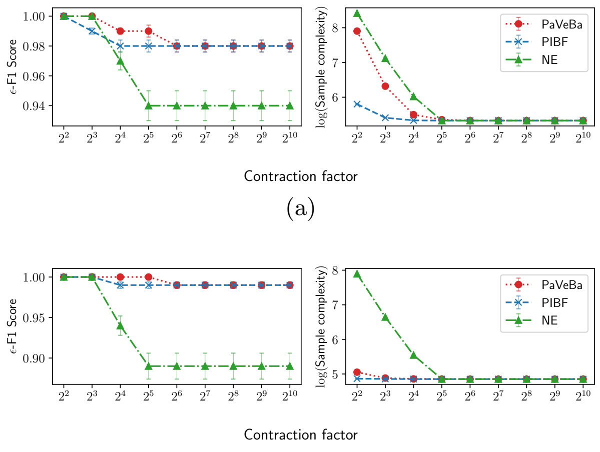
VC (Vehicle Crashworthiness, $D = 3$, $|\mathcal{X}| = 2000$): From the field of automotive safety, this dataset focuses on the optimization of vehicle structures to enhance crashworthiness (Tanabe and Ishibuchi, 2020; Liao et al., 2008). **VC1**: A smaller version of this dataset with $|\mathcal{X}| = 100$ that is used in compute-heavy settings.

MAR (Marine, $D = 4$, $|\mathcal{X}| = 500$): Within maritime engineering, this dataset is concerned with the optimization of bulk carrier designs to improve cargo transfer efficiency and maritime safety (Parsons and Scott, 2004; Tanabe and Ishibuchi, 2020).

5.2 Experimental Setup and Results

Before running our experiments, we min-max scale the inputs to unit intervals and standardize the outputs per usual. For all experiments, we employ a Gaussian noise as $\mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_D)$ where $\sigma_n = 0.1$, we set $\epsilon = 0.1$ and $\delta = 0.05$, unless stated otherwise. Note that when $D = 2$ this is both σ_n-subgaussian and σ_n-norm-subgaussian. To reduce the number of constraints in convex programs for PaVeBa from $O(t)$ to $O(1)$, we use $\mathcal{B}_t(\cdot)$ instead of $\mathcal{E}_t(\cdot)$ inside PaVeBa. For GP-based models, we use the RBF kernel, featuring automatic relevance determination. We assume we know the kernel parameters and choose the parameters according to maximum likelihood estimation from the dataset prior to optimization. We report ϵ-F1 scores as our accuracy metric (the higher, the better), on which further details are provided in the supplementary §5. All reported results are averaged over 50 runs.

Experiment 1. We assess the performance of PaVeBa in multi-objective optimization, a special case of vector optimization where the ordering cone is the positive orthant. PaVeBa is compared with Algorithm



(a)

(b)

Figure 2: Comparison with NE and PIBF on SNW (a) and DB (b) with $\sigma_n = \epsilon = 0.01$. The left figures show ϵ-F1 scores with different contraction factors, and the right figures show the sample complexities.

1 in Auer et al. (2016), denoted here as PIBF, and Naïve Elimination (NE) proposed in Ararat and Tekin (2023). PaVeBa and PIBF need wide confidence regions, and NE needs a high per-arm sampling budget (called L) for high-probability correct results. However, standard confidence parameters, being overly conservative, can hamper performance. To address this, we scale down the confidence regions given by PaVeBa and PIBF, i.e., $r_t(x)$, and the per-arm sampling budget calculated by NE, i.e., L. We term this scaling factor *contraction factor*, chosen as powers of 2 in $[2^2, 2^{10}]$. We set $\sigma_n = \epsilon = 0.01$ for this experiment due to high sample complexities. The results (Figure 2) reveal PaVeBa to be notably robust even with unmet confidence assumptions, maintaining similar sample complexities with PIBF at higher contraction factors while PaVeBa and PIBF outperform NE.

Experiment 2. We evaluate the sample complexity of PaVeBa under three different polyhedral ordering cones C and inspect the relation between the sample and ordering complexities. We take $C = C_\theta := \{\boldsymbol{x} \in \mathbb{R}^2 \mid \pi/4 - \theta/2 \leq \theta_{\boldsymbol{x}} \leq \pi/4 + \theta/2\}$, where $\theta_{\boldsymbol{x}}$ denotes the angle in the polar coordinates of a point $\boldsymbol{x} \in \mathbb{R}^2$ and $\theta \in \{\pi/4, \pi/2, 3\pi/4\}$. We also compare PaVeBa with NE, which is the only other algorithm that does vector optimization using ordering cones to the best of our knowledge. NE uses a per-arm sampling budget L, and for fairness, it is given the next multiple of $|\mathcal{X}|$ samples greater than what PaVeBa used. We use a contraction factor of 16 for PaVeBa. The results in Tables 1 and 2 show that PaVeBa works better under similar sample complexity.

		ϵ-F1 Score w.r.t. θ		
	ϵ	$\pi/4$	$\pi/2$	$3\pi/4$
NE	10^{-1}	**0.99**	0.97	0.97
	10^{-2}	0.93	0.87	0.78
PaVeBa	10^{-1}	**0.99**	**0.98**	**0.99**
	10^{-2}	**0.94**	**0.91**	**0.88**

(a)

NE	10^{-1}	**0.98**	0.94	**0.97**
	10^{-2}	0.93	0.84	0.78
PaVeBa	10^{-1}	**0.98**	**0.95**	**0.97**
	10^{-2}	**0.94**	**0.92**	**0.91**

(b)

Table 1: Results of NE and PaVeBa for different values of ϵ and θ on SNW (a) and DB (b) datasets.

		Sample Complexity w.r.t. θ		
	ϵ	$\pi/4$	$\pi/2$	$3\pi/4$
SNW	10^{-1}	654.50	378.68	272.44
	10^{-2}	10100.92	2594.00	789.96
DB	10^{-1}	304.86	167.58	141.78
	10^{-2}	2045.26	308.84	196.70

Table 2: Sample complexities of PaVeBa with different values of ϵ and θ on SNW and DB datasets.

Experiment 3. We compare the heuristic variants of PaVeBa against ϵ-PAL (Zuluaga et al., 2016), MESMO (Belakaria et al., 2019) and JES (Tu et al., 2022). Since these algorithms sample only one arm at each round, to ensure fairness, we modify PaVeBa-IH and PaVeBa-DE to have a batch size parameter K. The batch selection is done in a loop by choosing the arm with the maximum posterior covariance matrix trace and updating the posterior variances of arms until K arms have been chosen. While PaVeBa and ϵ-PAL work in the fixed confidence setting, others work with a fixed sampling budget, so in each run of the experiment, we run MESMO and JES algorithms with a budget that is equal to the number of samples taken by PaVeBa. We scale down $\alpha_\tau(x)$ by 64 for PaVeBa and β_t by 9 for ϵ-PAL, as in Zuluaga et al. (2016). To reduce the sample complexity of PaVeBa while maintaining ϵ correctness, we amend (6) to also include ϵ looseness in discarding. Further details are discussed in the supplementary §5.

First, we set $K = 1$. In the first simulation (Table 3), we compare the algorithms under the multi-objective cone. In the second simulation (Table 4), we focus on $D = 3$ and compare the algorithms under two other cones. For this purpose, we define two cones with matrices $\boldsymbol{W} = [[1, -2, 4], [4, 1, -2], [-2, 4, 1]]/\sqrt{21}$

Dataset	Algorithm	S.C.	ϵ-F1 Score
PK2	ϵ-PAL	178.42	1.00
	JES	57.62	0.86
	MESMO	57.62	0.86
	PaVeBa-IH	57.62	0.95
VC	ϵ-PAL	584.00	1.00
	JES	123.94	0.87
	MESMO	123.94	0.97
	PaVeBa-IH	123.94	0.97
MAR	ϵ-PAL	639.04	0.98
	JES	224.38	0.97
	MESMO	224.38	0.95
	PaVeBa-IH	224.38	0.97

Table 3: Comparison of PaVeBa with ϵ-PAL, MESMO, and JES under the multi-objective cone.

Cone	Algorithm	S.C.	ϵ-F1 Score
Acute	ϵ-PAL	88.08	0.98
	JES	91.94	0.81
	MESMO	91.94	0.97
	PaVeBa-DE	91.94	0.99
Obtuse	ϵ-PAL	88.08	1.00
	JES	27.58	0.86
	MESMO	27.58	0.85
	PaVeBa-DE	27.58	1.00

Table 4: Comparison of PaVeBa with ϵ-PAL, MESMO, and JES on VC1 dataset under acute and obtuse cones.

and $\boldsymbol{W} = [[1, 0.4, 1.6], [1.6, 1, 0.4], [0.4, 1.6, 1]]/\sqrt{3.72}$, called acute and obtuse cones, respectively. For algorithms other than PaVeBa, we calculate the cone-ordered Pareto set directly from the final posterior means of arms. The results show PaVeBa keeps up with state of the art on several datasets with different reward dimensions under all considered cones. While ϵ-PAL offers good ϵ-F1 scores, its sample complexity is generally higher. In the third simulation, we compare PaVeBa-IH with PaVeBa-DE under different K values. From Figure 3(Left), we observe (1) novel modeling of correlations between objectives can yield superior sample efficiency and (2) the negative effect of batch size on the sample complexity. From Figure 3(Right), we observe that the ϵ-F1 scores of two different variants of PaVeBa do not vary meaningfully, indicating that the key benefit of utilizing ellipsoidal confidence regions may lie in reducing sample complexities.

Experiment 4. To relate the empirical sample complexity with the ordering complexity, we do a regression analysis on DB dataset. We use the same cone definition as in Experiment 2. We regress the average sample complexity on β^2 using the least squares approach. In Figure 4, it can be seen that β^2 matches well with the empirical observations.

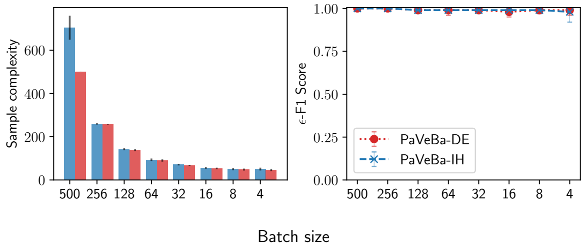


Figure 3: The sample complexities (Left) and ϵ-F1 scores (Right) of PaVeBa-IH and PaVeBa-DE on PK2.

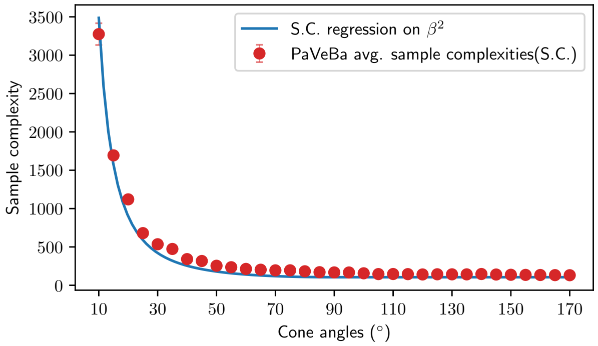


Figure 4: Least squares fit of the average sample complexity for different cone angles. Angles are sampled with step size 5° from the range $[10°, 170°]$.

6 CONCLUSION

We studied the vector bandits problem and proposed PaVeBa, the first algorithm to our knowledge that nearly matches the lower bounds on the sample complexity of the problem. It is based on a simple round-based heuristic and performs very well in the experiments. We also reinforce the existing theory by proving further results that establish the link between the ordering complexity and two fundamental gap functions, $M(\cdot, \cdot)$ and $m(\cdot, \cdot)$. Designing algorithms based on different heuristics, such as entropy search, studying regret minimization in vector bandits with arbitrary cones, or studying partial observations where only a subset of reward dimensions are available can be valuable future research direction.

Acknowledgements

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant 121E159. Yaşar Cahit Yıldırım was supported by Turk Telekom as part of 5G and Beyond Joint Graduate Support Program coordinated by Information and Communication Technologies Authority. Cem Tekin acknowledges the support by the Turkish Academy of Sciences Distinguished Young Scientist Award Program (TÜBA-GEBİP-2023).

References

Ararat, Ç. and Tekin, C. (2023). Vector optimization with stochastic bandit feedback. In *Proc. 26th International Conference on Artificial Intelligence and Statistics*, pages 2165–2190.

Auer, P., Chiang, C.-K., Ortner, R., and Drugan, M. (2016). Pareto front identification from stochastic bandit feedback. In *Proc. 19th International Conference on Artificial Intelligence and Statistics*, pages 939–947.

Belakaria, S., Deshwal, A., and Doppa, J. R. (2019). Max-value entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 32.

Drugan, M. M. and Nowe, A. (2013). Designing multi-objective multi-armed bandits algorithms: A study. In *The International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Ertl, P. and Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, 1(8):1–11.

Even-Dar, E., Mannor, S., and Mansour, Y. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(39):1079–1105.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Proc. Conference on Learning Theory*, pages 998–1027.

Hayes, C. F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L. M., Dazeley, R., Heintz, F., Howley, E., Irissappane, A. A., Mannion, P., Nowé, A., Ramos, G., Restelli, M., Vamplew, P., and Roijers, D. M. (2022). A practical guide to multi-objective reinforcement learning and planning. *Autonomous Agents and Multi-Agent Systems*, 36(26):1–59.

Hernandez-Lobato, D., Hernandez-Lobato, J., Shah, A., and Adams, R. (2016). Predictive entropy search for multi-objective bayesian optimization. In *Proc. International Conference on Machine Learning*, pages 1492–1501.

Huang, K., Fu, T., Gao, W., Zhao, Y., Roohani, Y., Leskovec, J., Coley, C. W., Xiao, C., Sun, J., and Zitnik, M. (2022). Artificial intelligence foundation for therapeutic science. *Nature Chemical Biology*, 18(10):1033–1036.

Jahn, J. (2011). *Vector Optimization Theory: Applications, and Extensions*. Springer, 2nd edition.

Jayatunga, M. K., Xie, W., Ruder, L., Schulze, U., and Meier, C. (2022). AI in small-molecule drug discovery: A coming wave? *Nature Reviews Drug Discovery*, 21(3):175–176.

Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., and Jordan, M. I. (2019). A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736*.

Karnin, Z., Koren, T., and Somekh, O. (2013). Almost optimal exploration in multi-armed bandits. In *Proc. International Conference on Machine Learning*, pages 1238–1246.

Katz-Samuels, J. and Scott, C. (2018). Feasible arm identification. In *Proc. International Conference on Machine Learning*, pages 2535–2543.

Liao, X., Li, Q., Yang, X., Zhang, W., and Li, W. (2008). Multiobjective optimization for crash safety design of vehicles using stepwise regression model. *Structural and Multidisciplinary Optimization*, 35:561–569.

Lizotte, D. J. and Laber, E. B. (2016). Multi-objective markov decision processes for data-driven decision support. *Journal of Machine Learning Research*, 17(210):1–28.

Löhne, A. (2011). *Vector optimization with infimum and Supremum*. Springer.

Meyer, H. (1899). Zur theorie der alkoholnarkose: Erste mittheilung. welche eigenschaft der anästhetica bedingt ihre narkotische wirkung? *Archiv für experimentelle Pathologie und Pharmakologie*, 42:109–118.

Moffaert, K. V. and Nowé, A. (2014). Multi-objective reinforcement learning using sets of pareto dominating policies. *Journal of Machine Learning Research*, 15(107):3663–3692.

Moore, J. S. and Jensen, K. F. (2012). Automated multitrajectory method for reaction optimization in a microfluidic system using online IR analysis. *Organic Process Research & Development*, 16(8):1409–1415.

Overton, C. E. (1901). *Studien über die Narkose: zugleich ein Beitrag zur allgemeinen Pharmakologie*. G. Fischer.

Parsons, M. G. and Scott, R. L. (2004). Formulation of multicriterion design optimization problems for solution with scalar numerical optimization methods. *Journal of Ship Research*, 48(01):61–76.

Shah, A. and Ghahramani, Z. (2016). Pareto frontier learning with expensive correlated objectives. In *Proc. International Conference on Machine Learning*, pages 1919–1927.

Tanabe, R. and Ishibuchi, H. (2020). An easy-to-use real-world multi-objective optimization problem suite. *Applied Soft Computing*, 89:106078.

Tu, B., Gandy, A., Kantas, N., and Shafei, B. (2022). Joint entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 35, pages 9922–9938.

Turgay, E., Oner, D., and Tekin, C. (2018). Multi-objective contextual bandit problem with similarity information. In *Proc. International Conference on Artificial Intelligence and Statistics*, pages 1673–1681.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Research*, 46(D1):D1074–D1082.

Zuluaga, M., Krause, A., and Püschel, M. (2016). ε-PAL: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research*, 17(104):1–32.

Zuluaga, M., Milder, P., and Püschel, M. (2012). Computer generation of streaming sorting networks. In *DAC Design Automation Conference*, pages 1241–1249.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes
   (b) Complete proofs of all theoretical results. Yes
   (c) Clear explanations of any assumptions. Yes

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). Yes
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Yes
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Yes
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). No

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator if your work uses existing assets. Yes
   (b) The license information of the assets, if applicable. Not Applicable
   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable
   (d) Information about consent from data providers/curators. Not Applicable
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. Not Applicable
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not applicable
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

Learning the Pareto Set Under Incomplete Preferences: Pure Exploration in Vector Bandits: Supplementary Materials

1 PROOFS FOR THE CONVEX PROGRAMMING FORMULATIONS

In this section, we provide the proofs of Propositions 1 and 2.

1.1 Proof of Proposition 1

Note that $\mathcal{E}_t(x)$, $\mathcal{E}_t(y)$ are compact sets and $M(\cdot,\cdot)$ is a continuous function. Hence, the supremum under consideration is indeed a maximum. Then, $\sup_{\boldsymbol{\mu}\in\mathcal{E}_t(x),\boldsymbol{\nu}\in\mathcal{E}_t(y)} M(\boldsymbol{\mu},\boldsymbol{\nu}) > 0$ iff there exist $\boldsymbol{\mu}\in\mathcal{E}_t(x)$, $\boldsymbol{\nu}\in\mathcal{E}_t(y)$ such that $\boldsymbol{\nu}-\boldsymbol{\mu}\in C^c$ by Ararat and Tekin (2023, Corollary 4.5). By the definition of cone C, this is possible iff $\boldsymbol{w}_n^{\mathsf{T}}(\boldsymbol{\nu}-\boldsymbol{\mu}) < 0$ for some $n\in[N]$, $\boldsymbol{\mu}\in\mathcal{E}_t(x)$, and $\boldsymbol{\nu}\in\mathcal{E}_t(y)$. In other words, for at least one $n\in[N]$, we have $\min_{\boldsymbol{\mu}\in\mathcal{E}_t(x),\boldsymbol{\nu}\in\mathcal{E}_t(y)} \boldsymbol{w}_n^{\mathsf{T}}(\boldsymbol{\nu}-\boldsymbol{\mu}) < 0$, which corresponds precisely to the problem defined in Proposition 1. □

1.2 Proof of Proposition 2

Note that $\mathcal{E}_t(x)$, $\mathcal{E}_t(y)$ are compact sets and $m(\cdot,\cdot)$ is a continuous function. Hence, the supremum under consideration is indeed a maximum. Then, $\sup_{\boldsymbol{\mu}\in\mathcal{E}_t(x),\boldsymbol{\nu}\in\mathcal{E}_t(y)} m(\boldsymbol{\mu},\boldsymbol{\nu}) \geq \epsilon$ iff there exist $\boldsymbol{\mu}\in\mathcal{E}_t(x)$, $\boldsymbol{\nu}\in\mathcal{E}_t(y)$ such that $m(\boldsymbol{\mu},\boldsymbol{\nu})\geq\epsilon$. By Ararat and Tekin (2023, Proposition 4.2(iv)), we have

$$m(\boldsymbol{\mu},\boldsymbol{\nu}) = \min_{n\in[N]} \frac{(\boldsymbol{w}_n^{\mathsf{T}}(\boldsymbol{\nu}-\boldsymbol{\mu}))^+}{\alpha_n}.$$

Then,

$$m(\boldsymbol{\mu},\boldsymbol{\nu})\geq\epsilon \quad\Leftrightarrow\quad \forall n\in[N]\colon (\boldsymbol{w}_n^{\mathsf{T}}(\boldsymbol{\nu}-\boldsymbol{\mu}))^+ \geq \epsilon\alpha_n.$$

Since $\epsilon\alpha_n > 0$ for each $n\in[N]$, we can drop the positive part function and obtain

$$m(\boldsymbol{\mu},\boldsymbol{\nu})\geq\epsilon \quad\Leftrightarrow\quad \forall n\in[N]\colon \boldsymbol{w}_n^{\mathsf{T}}(\boldsymbol{\nu}-\boldsymbol{\mu}) \geq \epsilon\alpha_n.$$

Therefore, $\sup_{\boldsymbol{\mu}\in\mathcal{E}_t(x),\boldsymbol{\nu}\in\mathcal{E}_t(y)} m(\boldsymbol{\mu},\boldsymbol{\nu}) \geq \epsilon$ iff there exist $\boldsymbol{\mu}\in\mathcal{E}_t(x)$, $\boldsymbol{\nu}\in\mathcal{E}_t(y)$ such that $\boldsymbol{w}_n^{\mathsf{T}}(\boldsymbol{\nu}-\boldsymbol{\mu}) \geq \epsilon\alpha_n$ holds for each $n\in[N]$. These conditions are precisely the constraints of the feasibility problem in the proposition. Hence, the result follows. □

2 PROOFS FOR THE SAMPLE COMPLEXITY BOUNDS

In this section, we provide the proofs of the results concerning the sample complexity of PaVeBa.

2.1 Preliminary Results on Gap Functions

We start with a remark that is useful in proving Propositions 3 and 4.

Remark 4. *The gap functions $m(\cdot,\cdot)$ and $M(\cdot,\cdot)$ depend on their arguments only through the difference between the two arguments. Hence, by a slight abuse of notation, we could alternatively define them via*

$$m(\boldsymbol{\xi}) = d(\boldsymbol{\xi},(\mathrm{int}(C))^c\cap(\boldsymbol{\xi}-C)),\quad M(\boldsymbol{\xi}) = d(\boldsymbol{\xi},C\cap(\boldsymbol{\xi}+C))$$

for each $\boldsymbol{\xi}\in\mathbb{R}^D$.

Next, we present several lemmata that will lead to Theorems 1 and 2. The first one states the triangle inequality for $M(\cdot,\cdot)$.

Lemma 1. *For each $\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\xi} \in \mathbb{R}^D$, we have $M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq M(\boldsymbol{\mu}, \boldsymbol{\xi}) + M(\boldsymbol{\xi}, \boldsymbol{\nu})$.*

Proof. Let $\varepsilon > 0$. Then, by (3) and the definition of infimum, there exist $s_1 \leq M(\boldsymbol{\mu}, \boldsymbol{\xi}) + \frac{\varepsilon}{2}$, $s_2 \leq M(\boldsymbol{\xi}, \boldsymbol{\nu}) + \frac{\varepsilon}{2}$ and $\boldsymbol{u_1}, \boldsymbol{u_2} \in B(\mathbf{0}, 1) \cap C$ such that $\boldsymbol{\mu} \preceq_C \boldsymbol{\xi} + s_1 \boldsymbol{u_1}$ and $\boldsymbol{\xi} \preceq_C \boldsymbol{\nu} + s_2 \boldsymbol{u_2}$. By adding these two inequalities, we get $\boldsymbol{\mu} + \boldsymbol{\xi} \preceq_C \boldsymbol{\xi} + \boldsymbol{\nu} + s_1 \boldsymbol{u_1} + s_2 \boldsymbol{u_2}$ and by canceling out terms, we get

$$\boldsymbol{\mu} \preceq_C \boldsymbol{\nu} + s\boldsymbol{u},$$

where $s := \|s_1 \boldsymbol{u}_1 + s_2 \boldsymbol{u}_2\|_2$; $\boldsymbol{u} := \frac{1}{s}(s_1 \boldsymbol{u_1} + s_2 \boldsymbol{u_2})$ if $s > 0$ and $\boldsymbol{u} := \mathbf{0}$ if $s = 0$ (for definiteness). Moreover, triangle inequality yields

$$s = \|s_1 \boldsymbol{u_1} + s_2 \boldsymbol{u_2}\|_2 \leq \|s_1 \boldsymbol{u_1}\|_2 + \|s_2 \boldsymbol{u_2}\|_2 \leq s_1 + s_2, \tag{S.1}$$

where the last step is due to $\boldsymbol{u_1}, \boldsymbol{u_2} \in B(\mathbf{0}, 1)$. Also, note that $s_1 \boldsymbol{u_1} + s_2 \boldsymbol{u_2} \in C$ so that $\boldsymbol{u} \in C$. Then, by the definition of $M(\boldsymbol{\mu}, \boldsymbol{\nu})$ and (S.1), we have $M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq s_1 + s_2$. In particular,

$$M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq M(\boldsymbol{\mu}, \boldsymbol{\xi}) + M(\boldsymbol{\xi}, \boldsymbol{\nu}) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we obtain the desired triangle inequality. □

The next three results are concerned with the Lipschitz-continuity of the gap functions. We show that $M(\cdot, \cdot)$ is β_1-Lipschitz and $m(\cdot, \cdot)$ is β_2-Lipschitz when seen as functions of the difference vectors, see Remark 4.

Lemma 2. *For every $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^D$, we have $M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \beta_1 \|\boldsymbol{\mu} - \boldsymbol{\nu}\|_2$ and $m(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \beta_2 \|\boldsymbol{\mu} - \boldsymbol{\nu}\|_2$.*

Proof. We consider the following three cases for $\boldsymbol{\nu} - \boldsymbol{\mu}$.

Case 1: Suppose that $\boldsymbol{\nu} - \boldsymbol{\mu} \in \operatorname{int}(C)$. By Ararat and Tekin (2023, Corollary 4.5(i)), we have $M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ so that the inequality for $M(\boldsymbol{\mu}, \boldsymbol{\nu})$ becomes trivial. Furthermore, by (2), we have

$$\frac{d(\boldsymbol{\nu} - \boldsymbol{\mu}, (\operatorname{int}(C))^c \cap (\boldsymbol{\nu} - \boldsymbol{\mu} - C))}{d(\boldsymbol{\nu} - \boldsymbol{\mu}, (\operatorname{int}(C))^c)} \leq \beta_2.$$

By Ararat and Tekin (2023, Proposition 4.2(ii)), this directly implies that

$$m(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \beta_2 d(\boldsymbol{\nu} - \boldsymbol{\mu}, (\operatorname{int}(C))^c) \leq \beta_2 \|\boldsymbol{\mu} - \boldsymbol{\nu}\|_2,$$

where the last step follows since $\mathbf{0} \in (\operatorname{int}(C))^c$.

Case 2: Suppose that $\boldsymbol{\nu} - \boldsymbol{\mu} \in \operatorname{bd}(C)$. By Ararat and Tekin (2023, Corollary 4.5(ii)), we have $m(\boldsymbol{\mu}, \boldsymbol{\nu}) = M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$. Hence, both inequalities become trivial.

Case 3: Suppose that $\boldsymbol{\nu} - \boldsymbol{\mu} \in C^c$. By Ararat and Tekin (2023, Corollary 4.5(iii)), we have $m(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$. Hence, the inequality for $m(\boldsymbol{\mu}, \boldsymbol{\nu})$ becomes trivial. Furthermore, by (1), we have

$$\frac{d(\boldsymbol{\nu} - \boldsymbol{\mu}, (C \cap (\boldsymbol{\nu} - \boldsymbol{\mu} + C)))}{d(\boldsymbol{\nu} - \boldsymbol{\mu}, C)} \leq \beta_1.$$

By Ararat and Tekin (2023, Proposition 4.3(ii)), this directly implies that

$$M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq \beta_1 d(\boldsymbol{\nu} - \boldsymbol{\mu}, C) \leq \beta_1 \|\boldsymbol{\mu} - \boldsymbol{\nu}\|_2,$$

where the last step follows since $\mathbf{0} \in C$.

Hence, we have the desired inequalities for all cases. □

2.2 Proof of Proposition 3

By Lemma 1, we have $M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq M(\boldsymbol{\mu}, \boldsymbol{\mu} + \boldsymbol{\epsilon}) + M(\boldsymbol{\mu} + \boldsymbol{\epsilon}, \boldsymbol{\nu})$. Hence, Lemma 2 yields

$$M(\boldsymbol{\mu}, \boldsymbol{\nu}) - M(\boldsymbol{\mu} + \boldsymbol{\epsilon}, \boldsymbol{\nu}) \leq M(\boldsymbol{\mu}, \boldsymbol{\mu} + \boldsymbol{\epsilon}) \leq \beta_1 \|\boldsymbol{\epsilon}\|_2.$$

Similarly,

$$M(\boldsymbol{\mu} + \boldsymbol{\epsilon}, \boldsymbol{\nu}) - M(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq M(\boldsymbol{\mu} + \boldsymbol{\epsilon}, \boldsymbol{\mu}) \leq \beta_1 \|\boldsymbol{\epsilon}\|_2.$$

Therefore, the desired inequality follows. □

2.3 Proof of Proposition 4

We first prove the statement under the assumption that $m(\boldsymbol{\mu},\boldsymbol{\nu}) \geq m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu})$. We consider the following cases. If $m(\boldsymbol{\mu},\boldsymbol{\nu}) = m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu})$, then the desired inequality follows trivially.

For the rest of the proof, let us suppose that $m(\boldsymbol{\mu},\boldsymbol{\nu}) > m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu})$. Let

$$\varepsilon \in \left(0, \frac{1}{2}(m(\boldsymbol{\mu},\boldsymbol{\nu}) - m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}))\right) \tag{S.2}$$

be arbitrarily chosen. Then, by (4) and the definition of infimum, there exist $s \leq m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) + \varepsilon$ and $\boldsymbol{u} \in B(\boldsymbol{0},1)\cap C$ such that $\boldsymbol{\mu}+\boldsymbol{\epsilon}+s\boldsymbol{u} \notin \boldsymbol{\nu} - \operatorname{int}(C)$, i.e., $\boldsymbol{\nu}-\boldsymbol{\mu}-\boldsymbol{\epsilon}-s\boldsymbol{u} \in (\operatorname{int}(C))^c$. In particular, (S.2) ensures that

$$s < m(\boldsymbol{\mu},\boldsymbol{\nu}). \tag{S.3}$$

Let $h \coloneqq d(\boldsymbol{\nu}-\boldsymbol{\mu}-s\boldsymbol{u}, (\operatorname{int}(C))^c)$. Then, we immediately have

$$h \leq \|(\boldsymbol{\nu}-\boldsymbol{\mu}-s\boldsymbol{u}) - (\boldsymbol{\nu}-\boldsymbol{\mu}-\boldsymbol{\epsilon}-s\boldsymbol{u})\|_2 = \|\boldsymbol{\epsilon}\|_2.$$

We claim that $\boldsymbol{\nu}-\boldsymbol{\mu}-s\boldsymbol{u} \in \operatorname{int}(C)$. Suppose that this is not the case, i.e., $\boldsymbol{\mu}+s\boldsymbol{u} \notin \boldsymbol{\nu}-\operatorname{int}(C)$. By the definition of $m(\boldsymbol{\mu},\boldsymbol{\nu})$, this implies that $m(\boldsymbol{\mu},\boldsymbol{\nu}) \leq s$, which is a contradiction to (S.3). Hence, the claim holds, and we have $h > 0$. As an immediate consequence, by the definition of β_2 and Ararat and Tekin (2023, Proposition 4.2(ii)), we also get

$$\begin{aligned}\beta_2 &= \sup_{\boldsymbol{x}\in\operatorname{int}(C)} \frac{d(\boldsymbol{x},(\operatorname{int}(C))^c \cap (\boldsymbol{x}-C))}{d(\boldsymbol{x},(\operatorname{int}(C))^c)} \\ &\geq \frac{d(\boldsymbol{\nu}-\boldsymbol{\mu}-s\boldsymbol{u},(\operatorname{int}(C))^c \cap (\boldsymbol{\nu}-\boldsymbol{\mu}-s\boldsymbol{u}-C))}{d(\boldsymbol{\nu}-\boldsymbol{\mu}-s\boldsymbol{u},(\operatorname{int}(C))^c)} = \frac{m(\boldsymbol{\mu}+s\boldsymbol{u},\boldsymbol{\nu})}{h}.\end{aligned}$$

Next, we claim that $m(\boldsymbol{\mu}+s\boldsymbol{u},\boldsymbol{\nu}) + s \geq m(\boldsymbol{\mu},\boldsymbol{\nu})$. Indeed, let $r \geq 0$ be such that $\boldsymbol{\mu}+s\boldsymbol{u}+r\boldsymbol{v} \notin \boldsymbol{\nu}-\operatorname{int}(C)$ for some $\boldsymbol{v} \in B(\boldsymbol{0},1)\cap C$. Then, $s\boldsymbol{u}+r\boldsymbol{v} \in C$ since C is a convex cone. Hence, $m(\boldsymbol{\mu},\boldsymbol{\nu}) \leq ||s\boldsymbol{u}+r\boldsymbol{v}||_2 \leq s+r$. Then, taking infimum over all choices of r yields $m(\boldsymbol{\mu},\boldsymbol{\nu}) \leq s + m(\boldsymbol{\mu}+s\boldsymbol{u},\boldsymbol{\nu})$. Therefore,

$$\beta_2\|\boldsymbol{\epsilon}\|_2 \geq m(\boldsymbol{\mu}+s\boldsymbol{u},\boldsymbol{\nu}) \geq m(\boldsymbol{\mu},\boldsymbol{\nu}) - s.$$

In particular, by taking $s = m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) + \varepsilon$, we get

$$\beta_2\|\boldsymbol{\epsilon}\|_2 \geq m(\boldsymbol{\mu},\boldsymbol{\nu}) - m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) - \varepsilon.$$

Finally, since ε can be chosen arbitrarily small, we conclude that

$$\beta_2\|\boldsymbol{\epsilon}\|_2 \geq m(\boldsymbol{\mu},\boldsymbol{\nu}) - m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) = |m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) - m(\boldsymbol{\mu},\boldsymbol{\nu})|,$$

which completes the proof when $m(\boldsymbol{\mu},\boldsymbol{\nu}) \geq m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu})$ holds.

Next, suppose that $m(\boldsymbol{\mu},\boldsymbol{\nu}) < m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu})$. In this case, let us define $\tilde{\boldsymbol{\mu}} \coloneqq \boldsymbol{\mu}+\boldsymbol{\epsilon}$ and $\tilde{\boldsymbol{\epsilon}} \coloneqq -\boldsymbol{\epsilon}$. Then, we have

$$m(\tilde{\boldsymbol{\mu}}+\tilde{\boldsymbol{\epsilon}},\boldsymbol{\nu}) = m(\boldsymbol{\mu},\boldsymbol{\nu}) < m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) = m(\tilde{\boldsymbol{\mu}},\boldsymbol{\nu}).$$

Hence, by the result of the previous step, we have

$$|m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu}) - m(\boldsymbol{\mu},\boldsymbol{\nu})| = |m(\boldsymbol{\mu},\boldsymbol{\nu}) - m(\boldsymbol{\mu}+\boldsymbol{\epsilon},\boldsymbol{\nu})| = |m(\tilde{\boldsymbol{\mu}}+\tilde{\boldsymbol{\epsilon}},\boldsymbol{\nu}) - m(\tilde{\boldsymbol{\mu}},\boldsymbol{\nu})| \leq \beta_2\|\tilde{\boldsymbol{\epsilon}}\|_2 = \beta_2\|\boldsymbol{\epsilon}\|_2,$$

which completes the proof.

□

2.4 Norm-Subgaussian Concentration

The aim of this subsection is to prove a concentration inequality for a random sum of norm-subgaussian random vectors to be used in the proof of Lemma 5 below. As a preparation, we first provide a moment inequality for a single norm-subgaussian random vector. The inequality is also given in Jin et al. (2019, Lemma 2) without specifying the constant multiplied by the bounding term $\sigma\sqrt{p}$. Since we need the exact value of this constant for our purposes, we re-derive the inequality.

Lemma 3. *Let $\boldsymbol{y}$ be D-dimensional norm-subgaussian random vector with parameter σ.*

(i) For each $p \in \mathbb{N}$,

$$\mathbb{E}[\|\boldsymbol{y}\|_2^{2p}] \leq 2p!\,(2\sigma^2)^p.$$

(ii) Define a $(D+1)\times(D+1)$-dimensional random matrix by

$$\tilde{\boldsymbol{y}} \coloneqq \begin{pmatrix} 0 & \boldsymbol{y}^\mathsf{T} \\ \boldsymbol{y} & \boldsymbol{0} \end{pmatrix}.$$

Then, for each $\theta \in \mathbb{R}$,

$$\mathbb{E}[e^{\theta\tilde{\boldsymbol{y}}}] \preceq e^{2\theta^2\sigma^2}\boldsymbol{I},$$

where $e^{\boldsymbol{A}}$ denotes the matrix exponential of a square matrix $\boldsymbol{A}$, $\boldsymbol{I}$ denotes the $(D+1)\times(D+1)$-dimensional identity matrix, and $\preceq$ denotes the positive semidefinite (Loewner) order on the space of symmetric matrices.

Proof. (i) Let $p \geq 1$. Since $\|\boldsymbol{y}\|_2^{2p}$ is a nonnegative random variable, we have

$$\mathbb{E}[\|\boldsymbol{y}\|_2^{2p}] = \int_0^\infty \mathbb{P}\{\|\boldsymbol{y}\|_2^{2p} > u\}du = \int_0^\infty \mathbb{P}\{\|\boldsymbol{y}\|_2 > u^{\frac{1}{2p}}\}du = 2p\int_0^\infty \mathbb{P}\{\|\boldsymbol{y}\|_2 > t\}t^{2p-1}dt,$$

where the last equality is a result of the substitution $u^{\frac{1}{2p}} = t$. Since $\boldsymbol{y}$ is a norm-subgaussian random vector with parameter σ, we have

$$\begin{aligned}\mathbb{E}[\|\boldsymbol{y}\|_2^{2p}] \leq 4p\int_0^\infty e^{-\frac{t^2}{2\sigma^2}}t^{2p-1}dt &= 4p\int_0^\infty e^{-\frac{t^2}{2\sigma^2}}(t^2)^{p-1}tdt \\ &= 4p\int_0^\infty e^{-v}(2\sigma^2 v)^{p-1}\sigma^2 dv \\ &= 4p2^{p-1}(\sigma^2)^p(p-1)!\int_0^\infty \frac{e^{-v}v^{p-1}}{(p-1)!}dv \\ &= 2p!\,(2\sigma^2)^p,\end{aligned}$$

where the last integral equals 1 as the integral of the probability density function of the gamma distribution with shape index p and scale parameter 1.

(ii) The proof mainly follows the arguments as in the proof of Jin et al. (2019, Lemma 4), but has some additional tweaks that lead to a tighter upper bound. Note that $\tilde{\boldsymbol{y}}$ is a symmetric matrix of rank 2 whose eigenvalues are $\|\boldsymbol{y}\|_2$ and $-\|\boldsymbol{y}\|_2$. Let $\|\tilde{\boldsymbol{y}}\|$ denote the matrix (operator) norm of $\tilde{\boldsymbol{y}}$. Then, we have $\|\tilde{\boldsymbol{y}}\| = \|\boldsymbol{y}\|_2$. Using this, we obtain

$$\begin{aligned}
\mathbb{E}[e^{\theta\tilde{\boldsymbol{y}}}] = \mathbb{E}\left[\sum_{p=0}^{\infty}\frac{(\theta\tilde{\boldsymbol{y}})^p}{p!}\right] &= \frac{\mathbb{E}[\tilde{\boldsymbol{y}}^0]}{0!} + \sum_{p\in\{1,3,\dots\}}\frac{\theta^p\mathbb{E}[\tilde{\boldsymbol{y}}^p]}{p!} + \sum_{p\in\{2,4,\dots\}}\frac{\theta^p\mathbb{E}[\tilde{\boldsymbol{y}}^p]}{p!}\\
&= \boldsymbol{I} + \sum_{p\in\{2,4,\dots\}}\frac{\theta^p\mathbb{E}[\tilde{\boldsymbol{y}}^p]}{p!}\\
&= \boldsymbol{I} + \sum_{p=1}^{\infty}\frac{\theta^{2p}\mathbb{E}[\tilde{\boldsymbol{y}}^{2p}]}{(2p)!}\\
&\preceq \left(1 + \sum_{p=1}^{\infty}\frac{\theta^{2p}\mathbb{E}[\|\tilde{\boldsymbol{y}}\|^{2p}]}{(2p)!}\right)\boldsymbol{I}\\
&= \left(1 + \sum_{p=1}^{\infty}\frac{\theta^{2p}\mathbb{E}[\|\boldsymbol{y}\|_2^{2p}]}{(2p)!}\right)\boldsymbol{I}\\
&\preceq \left(1 + \sum_{p=1}^{\infty}\frac{\theta^{2p}2p!\,(2\sigma^2)^p)}{(2p)!}\right)\boldsymbol{I} \quad \text{due to (i)}\\
&= \left(1 + \sum_{p=1}^{\infty}\frac{(2\theta^2\sigma^2)^p 2p!}{(2p)!}\right)\boldsymbol{I}\\
&\preceq \left(1 + \sum_{p=1}^{\infty}\frac{(2\theta^2\sigma^2)^p}{p!}\right)\boldsymbol{I} \quad \text{due to } \frac{2p!}{(2p)!} \leq \frac{1}{p!} \text{ for } p \in \mathbb{N}\\
&= \boldsymbol{I} + \sum_{p=1}^{\infty}\frac{(2\theta^2\sigma^2\boldsymbol{I})^p}{p!} = \sum_{p=0}^{\infty}\frac{(2\theta^2\sigma^2\boldsymbol{I})^p}{p!} = e^{2\theta^2\sigma^2\boldsymbol{I}}\ .
\end{aligned}$$

Here, the inequality $\frac{2p!}{(2p)!} \leq \frac{1}{p!}$ for $p \geq 1$ can be proven by induction.

□

The next lemma is a 'random sum' version of Jin et al. (2019, Lemma 6) within our setting.

Lemma 4. *Let $x \in \mathcal{X}$ and $t \in \mathbb{N}$. Let $\delta_t \in (0,1)$. Then, for each $\theta > 0$, it holds*

$$\mathbb{P}\left\{\left\|\sum_{\tau\in N_t(x)}\boldsymbol{\eta}_\tau(x)\right\|_2 > 2\theta n_t(x)\sigma^2 + \frac{1}{\theta}\log\left(\frac{D+1}{\delta_t}\right)\right\} \leq \delta_t\ .$$

In particular,

$$\mathbb{P}\left\{\left\|\sum_{\tau\in N_t(x)}\boldsymbol{\eta}_\tau(x)\right\|_2 > \sqrt{8t\sigma^2\log\left(\frac{D+1}{\delta_t}\right)}\right\} \leq \delta_t\ . \tag{S.4}$$

Proof. For each $t \in \mathbb{N}$, let $\mathcal{F}_t$ denote the σ-algebra corresponding to the information accumulated by round t. Let $\mathcal{F}_0$ be the trivial σ-algebra. As in Lemma 3, we denote by $\tilde{\boldsymbol{\eta}}_t(x)$ the random matrix corresponding to $\boldsymbol{\eta}_t(x)$ for each $t \in \mathbb{N}$. Let us fix $t \in \mathbb{N}$. Let $\theta > 0$. Note that

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)\right]\\
&\quad= \mathbb{E}\left[\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)(1_{\{t\in N_t(x)\}} + 1_{\{t\notin N_t(x)\}}) \mid \mathcal{F}_{t-1}\right]\right],
\end{aligned}$$

where $\operatorname{tr}\boldsymbol{A}$ denotes the trace of a square matrix $\boldsymbol{A}$ and 1_E denotes the indicator random variable of an event E. By the design of PaVeBa, $N_t(x)$ is an $\mathcal{F}_{t-1}$-measurable random set. In particular, $n_t(x)$ is $\mathcal{F}_{t-1}$-measurable. Moreover, $\tilde{\boldsymbol{\eta}}_t(x)$ is independent of $\mathcal{F}_{t-1}$ and we have $\mathbb{E}[e^{\theta\tilde{\boldsymbol{\eta}}_t(x)}] \preceq e^{2\theta^2\sigma^2}\boldsymbol{I}$ by Lemma 3(ii).

Under the event $\{t \in N_t(x)\}$, note that we have $N_t(x) \setminus \{t\} = N_{t-1}(x)$ and $n_t(x) - 1 = n_{t-1}(x)$. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)1_{\{t\in N_t(x)\}} \mid \mathcal{F}_{t-1}\right] \\
&\leq \operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)\setminus\{t\}}\tilde{\boldsymbol{\eta}}_\tau(x) + \log\mathbb{E}[e^{\theta\tilde{\boldsymbol{\eta}}_t(x)}]\right)1_{\{t\in N_t(x)\}} \\
&\leq \operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)\setminus\{t\}}\tilde{\boldsymbol{\eta}}_\tau(x) + 4\theta^2\sigma^2\boldsymbol{I}\right)1_{\{t\in N_t(x)\}} \\
&= \operatorname{tr}\exp\left(-2\theta^2 (n_t(x)-1)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)\setminus\{t\}}\tilde{\boldsymbol{\eta}}_\tau(x)\right)1_{\{t\in N_t(x)\}} \\
&= \operatorname{tr}\exp\left(-2\theta^2 n_{t-1}(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_{t-1}(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)1_{\{t\in N_t(x)\}},
\end{aligned}$$

where $\log\boldsymbol{A}$ denotes the matrix logarithm of a square matrix $\boldsymbol{A}$.

On the other hand, under the event $\{t \notin N_t(x)\}$, we have $N_t(x) = N_{t-1}(x)$ and $n_t(x) = n_{t-1}(x)$. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)1_{\{t\notin N_t(x)\}} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_{t-1}(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_{t-1}(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)1_{\{t\notin N_t(x)\}} \mid \mathcal{F}_{t-1}\right] \\
&= \operatorname{tr}\exp\left(-2\theta^2 n_{t-1}(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_{t-1}(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)1_{\{t\notin N_t(x)\}}.
\end{aligned}$$

Combining these gives

$$\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)\right] \leq \mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_{t-1}(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_{t-1}(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)\right].$$

Iterating this inductively yields

$$\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2\boldsymbol{I} + \theta\sum_{\tau\in N_t(x)}\tilde{\boldsymbol{\eta}}_\tau(x)\right)\right] \leq \operatorname{tr}\exp(0\boldsymbol{I}) = D+1.$$

Finally, for each $c_t > 0$, arguing as in the proof of Jin et al. (2019, Lemma 6) via Markov's inequality, we get

$$\begin{aligned}
&\mathbb{P}\left\{\left\|\sum_{\tau\in N_t(x)} \boldsymbol{\eta}_\tau(x)\right\|_2 > 2\theta n_t(x)\sigma^2 + \frac{c_t}{\theta}\right\} \\
&= \mathbb{P}\left\{\left\|\sum_{\tau\in N_t(x)} \tilde{\boldsymbol{\eta}}_\tau(x)\right\| > 2\theta n_t(x)\sigma^2 + \frac{c_t}{\theta}\right\} \\
&= \mathbb{P}\left\{\lambda_{\max}\left(\sum_{\tau\in N_t(x)} \tilde{\boldsymbol{\eta}}_\tau(x)\right) > 2\theta n_t(x)\sigma^2 + \frac{c_t}{\theta}\right\} \\
&= \mathbb{P}\left\{\lambda_{\max}\left(e^{\theta\sum_{\tau\in N_t(x)} \tilde{\boldsymbol{\eta}}_\tau(x)}\right) > e^{2\theta^2 n_t(x)\sigma^2 + c_t}\right\} \\
&= \mathbb{P}\left\{\operatorname{tr}\left(e^{\theta\sum_{\tau\in N_t(x)} \tilde{\boldsymbol{\eta}}_\tau(x)}\right) > e^{2\theta^2 n_t(x)\sigma^2 + c_t}\right\} \\
&\le e^{-c_t}\mathbb{E}\left[\operatorname{tr}\exp\left(-2\theta^2 n_t(x)\sigma^2 \boldsymbol{I} + \theta\sum_{\tau\in N_t(x)} \tilde{\boldsymbol{\eta}}_\tau(x)\right)\right] \le e^{-c_t}(D+1),
\end{aligned}$$

where $\lambda_{\max}(\boldsymbol{A})$ denotes the maximum eigenvalue of a square matrix $\boldsymbol{A}$. Here, since $\sum_{\tau\in N_t(x)} \tilde{\boldsymbol{\eta}}_{\boldsymbol{\tau}}$ is a symmetric matrix of rank 2 with eigenvalues $\|\sum_{\tau\in N_t(x)} \boldsymbol{\eta}_\tau\|_2$ and $-\|\sum_{\tau\in N_t(x)} \boldsymbol{\eta}_\tau\|_2$, the second equality follows. Hence, setting $c_t \coloneqq \log(\frac{D+1}{\delta_t})$ yields $e^{-c_t}(D+1) = \delta_t$, which completes the proof of the main inequality.

Since $n_t(x) \le t$ due to the construction of PaVeBa, we get

$$\mathbb{P}\left\{\left\|\sum_{\tau\in N_t(x)} \boldsymbol{\eta}_\tau(x)\right\|_2 > 2\theta t\sigma^2 + \frac{1}{\theta}\log\left(\frac{D+1}{\delta_t}\right)\right\} \le \delta_t$$

by a monotonicity bound. In particular, choosing $\theta > 0$ optimally as $\theta = \sqrt{\frac{1}{2t\sigma^2}\log(\frac{D+1}{\delta_t})}$ yields

$$2\theta t\sigma^2 + \frac{1}{\theta}\log\left(\frac{D+1}{\delta_t}\right) = \sqrt{8t\sigma^2\log\left(\frac{D+1}{\delta_t}\right)}$$

so that (S.4) follows. □

2.5 Lemmata for the Proof of Theorem 1

The next lemma is the main result on concentration. It provides a "good event" that holds with high probability. The results to follow will make some statements that hold under this event. For each $x \in \mathcal{X}$ and $t \in \mathbb{N}$, recall that $N_t(x)$ denotes the set of rounds by round t (including round t) at which arm x is sampled and $n_t(x) = |N_t(x)|$.

Take

$$r_t(x) \coloneqq \sqrt{\frac{8t\sigma^2}{n_t(x)^2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|t^2}{6\delta}\right)}.$$

Lemma 5. *We have* $\mathbb{P}\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{E}_t(x)\} \ge 1 - \delta$.

Proof. Since $\mathcal{E}_t(x) = \bigcap_{\tau=1}^{t} \mathcal{B}_\tau(x)$ for each $x \in \mathcal{X}$ and $t \in \mathbb{N}$, we have

$$\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{E}_t(x)\} = \{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{B}_t(x)\}.$$

Fix $x \in \mathcal{X}$ and $t \in \mathbb{N}$. Then,

$$\|\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)\|_2 = \left\|\sum_{\tau\in N_t(x)} \frac{\boldsymbol{y}_\tau(x) - \boldsymbol{f}(x)}{n_t(x)}\right\|_2 = \frac{1}{n_t(x)}\left\|\sum_{\tau\in N_t(x)} \boldsymbol{\eta}_\tau(x)\right\|_2.$$

Given $\delta_t \in (0,1)$, by (S.4) in Lemma 4, we have

$$\begin{aligned}
&\mathbb{P}\left\{\|\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)\|_2 \geq \sqrt{\frac{8t\sigma^2}{n_t(x)^2}\log\left(\frac{D+1}{\delta_t}\right)}\right\} \\
&= \mathbb{P}\left\{\left\|\sum_{\tau \in N_t(x)} \boldsymbol{\eta}_\tau(x)\right\|_2 \geq \sqrt{8t\sigma^2 \log\left(\frac{D+1}{\delta_t}\right)}\right\} \leq \delta_t.
\end{aligned}$$

Let us set $\delta_t \coloneqq \frac{6\delta}{\pi^2 t^2 |\mathcal{X}|}$. Then, a union bound gives

$$\mathbb{P}\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N} \colon \boldsymbol{f}(x) \in \mathcal{B}_t(x)\} \geq 1 - \sum_{x \in \mathcal{X}} \sum_{t \in \mathbb{N}} \frac{6\delta}{\pi^2 t^2 |\mathcal{X}|} = 1 - \sum_{x \in \mathcal{X}} \frac{\delta}{|\mathcal{X}|} \geq 1 - \delta$$

since, for each $x \in \mathcal{X}$ and $t \in \mathbb{N}$, we have

$$\sqrt{\frac{8t\sigma^2}{n_t(x)^2}\log\left(\frac{D+1}{\delta_t}\right)} = \sqrt{\frac{8t\sigma^2}{n_t(x)^2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|t^2}{6\delta}\right)} = r_t(x),$$

which completes the proof. □

The next result ensures a low uncertainty radius when an arm is sampled sufficiently many times.

The following lemma combines the Lipschitz-continuity results with the concentration lemma.

Lemma 6. *Under the event in Lemma 5, the following statements hold for every $x, y \in \mathcal{X}$ and $t \in \mathbb{N}$:*

1. $\|(\boldsymbol{\mu}_t(x) - \boldsymbol{\mu}_t(y)) - (\boldsymbol{f}(x) - \boldsymbol{f}(y))\|_2 \leq r_t(x) + r_t(y)$.
2. $|m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - m(\boldsymbol{f}(x), \boldsymbol{f}(y))| \leq \beta_2\,(r_t(x) + r_t(y))$.
3. $|M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - M(\boldsymbol{f}(x), \boldsymbol{f}(y))| \leq \beta_1\,(r_t(x) + r_t(y))$.

Proof. Under the event in Lemma 5, we have $\|\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)\|_2 \leq r_t(x)$ for every $x \in \mathcal{X}$ and $t \in [T]$.

1. We have

$$\begin{aligned}
\|(\boldsymbol{\mu}_t(x) - \boldsymbol{\mu}_t(y)) - (\boldsymbol{f}(x) - \boldsymbol{f}(y))\|_2 &= \|(\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)) + (\boldsymbol{f}(y) - \boldsymbol{\mu}_t(y))\|_2 \\
&\leq \|\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)\|_2 + \|\boldsymbol{f}(y) - \boldsymbol{\mu}_t(y)\|_2 \\
&\leq r_t(x) + r_t(y).
\end{aligned}$$

2. Note that $m(\boldsymbol{f}(x), \boldsymbol{f}(y)) = m(\boldsymbol{f}(x) - \boldsymbol{f}(y), \boldsymbol{0})$ and $m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) = m(\boldsymbol{\mu}_t(x) - \boldsymbol{\mu}_t(y), \boldsymbol{0})$. Hence,

$$\begin{aligned}
|m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - m(\boldsymbol{f}(x), \boldsymbol{f}(y))| &= |m(\boldsymbol{f}(x) - \boldsymbol{f}(y), \boldsymbol{0}) - m(\boldsymbol{\mu}_t(x) - \boldsymbol{\mu}_t(y), \boldsymbol{0})| \\
&\leq \beta_2 \|(\boldsymbol{\mu}_t(y) - \boldsymbol{f}(y)) - (\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x))\|_2 \\
&\leq \beta_2\left(\|\boldsymbol{\mu}_t(y) - \boldsymbol{f}(y)\|_2 + \|\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)\|_2\right) \\
&\leq \beta_2(r_t(x) + r_t(y)),
\end{aligned}$$

   where the passage to the second line is due to Proposition 4.

3. By Lemma 1, Proposition 3, and Lemma 2, we have

$$\begin{aligned}
&M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - M(\boldsymbol{f}(x), \boldsymbol{f}(y)) \\
&\leq M(\boldsymbol{\mu}_t(x), \boldsymbol{f}(y)) + M(\boldsymbol{f}(y), \boldsymbol{\mu}_t(y)) - M(\boldsymbol{f}(x), \boldsymbol{f}(y)) \\
&\leq |M(\boldsymbol{\mu}_t(x), \boldsymbol{f}(y)) - M(\boldsymbol{f}(x), \boldsymbol{f}(y))| + M(\boldsymbol{f}(y), \boldsymbol{\mu}_t(y)) \\
&\leq \beta_1 \|\boldsymbol{\mu}_t(x) - \boldsymbol{f}(x)\|_2 + \beta_1 \|\boldsymbol{f}(y) - \boldsymbol{\mu}_t(y)\|_2 \\
&\leq \beta_1(r_t(x) + r_t(y)).
\end{aligned}$$

   By symmetry, $M(\boldsymbol{f}(x), \boldsymbol{f}(y)) - M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) \leq \beta_1(r_t(x) + r_t(y))$ holds as well.

□

Lemma 7. *Under the event in Lemma 5, we have $P^* \subseteq \mathcal{P}_t \cup \mathcal{S}_t$ for every $t \in \mathbb{N}$.*

Proof. Clearly, at $t = 1$, $P^* \subseteq \mathcal{P}_t \cup \mathcal{S}_t$ holds. Note that, as t increases, arms get removed from $\mathcal{P}_t \cup \mathcal{S}_t$ only in the discarding step (line 9 of PaVeBa). Let $x \in \mathcal{X}$ be an arm that is discarded at some round $t \geq 1$. Hence, $x \in \mathcal{D}_t$, i.e., there exists $y \in \mathcal{A}_t \setminus \{x\}$ such that

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0.$$

This implies that $M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ for every $\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)$. Moreover, by Lemma 5, with probability at least $1 - \delta$, we have $\boldsymbol{f}(x) \in \mathcal{E}_t(x)$ and $\boldsymbol{f}(y) \in \mathcal{E}_t(y)$; in particular, we have $M(\boldsymbol{f}(x), \boldsymbol{f}(y)) = 0$, which implies that $x \preceq_C y$ by Ararat and Tekin (2023, Proposition 4.3(iii)). Hence, $x \notin P^*$ under this event, which completes the proof. □

Lemma 8. *Under the event in Lemma 5, for every $t \in \mathbb{N}$, the set $\mathcal{P}_t$ is an 'ϵ-accurate' Pareto set, that is, for every $x \in \mathcal{P}_t \setminus P^*$, we have*

$$\Delta^*(x) = \max_{y \in P^*} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) \leq \epsilon.$$

Proof. Obviously, the result holds for $t = 1$. Note that, as t increases, arms get added to $\mathcal{P}_t$ only at the Pareto selection step (line 11 of PaVeBa). Let $x \in \mathcal{S}_t$ be an arm that is added at the Pareto selection step at some round $t \geq 1$. Hence, $x \in \overline{\mathcal{P}}_t$, i.e., for every $y \in \overline{\mathcal{A}}_t \setminus \{x\}$, we have

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon. \tag{S.5}$$

By Lemma 5, we have $\boldsymbol{f}(x) \in \mathcal{E}_t(x)$ and $\boldsymbol{f}(y) \in \mathcal{E}_t(y)$ with probability at least $1 - \delta$. Hence,

$$m(\boldsymbol{f}(x), \boldsymbol{f}(y)) < \epsilon \tag{S.6}$$

for every $y \in \overline{\mathcal{A}}_t = \mathcal{A}_t \setminus \mathcal{D}_t$ under this event.

Moreover, we claim that

$$m(\boldsymbol{f}(x), \boldsymbol{f}(y)) < \epsilon \tag{S.7}$$

for every $y \in \mathcal{P}_t \setminus \mathcal{U}_t$. To see this, assume otherwise that there exists $y \in \mathcal{P}_t \setminus \mathcal{U}_t$ such that

$$m(\boldsymbol{f}(x), \boldsymbol{f}(y)) \geq \epsilon. \tag{S.8}$$

Hence,

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_{t-1}(x), \boldsymbol{\nu} \in \mathcal{E}_{t-1}(y)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \epsilon, \tag{S.9}$$

which implies that $y \in \mathcal{U}_t$ and we reach a contradiction.

Therefore, we have

$$\max_{y \in \mathcal{P}_t \cup (\mathcal{S}_t \setminus \mathcal{D}_t)} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) = \max_{y \in \overline{\mathcal{A}}_t \cup (\mathcal{P}_t \setminus \mathcal{U}_t)} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) < \epsilon. \tag{S.10}$$

Finally, by Lemma 7, under the event in Lemma 5, we have $P^* \subseteq \mathcal{P}_{t+1} \cup \mathcal{S}_{t+1} = (\mathcal{P}_t \cup \mathcal{S}_t) \setminus \mathcal{D}_t$. This implies

$$\Delta^*(x) = \max_{y \in P^*} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) \leq \max_{y \in \mathcal{P}_{t+1} \cup \mathcal{S}_{t+1}} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) \leq \epsilon,$$

which completes the proof. □

Lemma 9. *For every $t \in \mathbb{N}$ and $x \in \mathcal{A}_t$, we have $n_t(x) = t$.*

Proof. Let $t \in \mathbb{N}$. We claim that $\mathcal{A}_{t+1} \subseteq \mathcal{A}_t$. To get a contradiction, suppose that there exists $x \in \mathcal{A}_{t+1}$ such that $x \notin \mathcal{A}_t$. Hence, $x \in \mathcal{S}_{t+1}$ or $x \in \mathcal{U}_{t+1}$. The former is not possible since $x \notin \mathcal{A}_t$ implies $x \notin \mathcal{S}_t$ and $\mathcal{S}_{t+1} \subseteq \mathcal{S}_t$. Therefore, $x \in \mathcal{U}_{t+1}$. We consider the following two cases.

Case 1: Suppose that $x \notin \mathcal{P}_t$. Since $x \in \mathcal{U}_{t+1}$, we have $x \in \mathcal{P}_{t+1}$. Together with $x \notin \mathcal{P}_t$, this implies that $x \in \overline{\mathcal{P}}_t$; hence, $x \in \mathcal{S}_t$. But we showed that this is a contradiction.

Case 2: Suppose that $x \in \mathcal{P}_t$. Note that $x \notin \mathcal{A}_t$ implies $x \notin \mathcal{U}_t$. Hence, for every $y \in \mathcal{S}_t$, we have

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_{t-1}(y), \boldsymbol{\nu} \in \mathcal{E}_{t-1}(x)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon. \tag{S.11}$$

However, since $x \in \mathcal{U}_{t+1}$, there exists $\bar{y} \in \mathcal{S}_{t+1}$ such that

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_{t}(\bar{y}), \boldsymbol{\nu} \in \mathcal{E}_{t}(x)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \epsilon. \tag{S.12}$$

This is a contradiction since $\mathcal{S}_{t+1} \subseteq \mathcal{S}_t$ and we have $\mathcal{E}_t(\bar{y}) \subseteq \mathcal{E}_{t-1}(\bar{y})$, $\mathcal{E}_t(x) \subseteq \mathcal{E}_{t-1}(x)$. This completes the proof of the claim.

By an inductive argument, it follows that $\mathcal{A}_t \subseteq \ldots \subseteq \mathcal{A}_1$. Since all the arms in $\mathcal{A}_\tau$ get sampled once at round τ for each $\tau \in \{1, \ldots, t\}$, we have $n_t(x) = t$ for every $x \in \mathcal{A}_t$. □

Lemma 10. *Let $t \in \mathbb{N}$ and define $R_t := \max_{x \in \mathcal{A}_t} r_t(x)$. Under the event in Lemma 5, if $R_t < \frac{\epsilon}{4\beta_2}$, then the algorithm terminates at round t.*

Proof. Let $x \in \mathcal{S}_t$. First, suppose that there exists $y \in \mathcal{A}_t \setminus \{x\}$ such that $m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) > \frac{\epsilon}{2}$. Let $\boldsymbol{\mu} \in \mathcal{E}_t(x)$, $\boldsymbol{\nu} \in \mathcal{E}_t(y)$. Since $||\boldsymbol{\mu} - \boldsymbol{\mu}_t(x)||_2 \leq r_t(x)$ and $||\boldsymbol{\nu} - \boldsymbol{\mu}_t(y)||_2 \leq r_t(y)$, by Proposition 4, we have

$$\begin{aligned}
\frac{\epsilon}{2} - m(\boldsymbol{\mu}, \boldsymbol{\nu}) &< |m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - m(\boldsymbol{\mu}, \boldsymbol{\nu})| \\
&= |m(\boldsymbol{\mu}_t(x) - \boldsymbol{\mu}_t(y), \mathbf{0}) - m(\boldsymbol{\mu} - \boldsymbol{\nu}, \mathbf{0})| \\
&\leq \beta_2 \|(\boldsymbol{\mu}_t(x) - \boldsymbol{\mu}_t(y)) - (\boldsymbol{\mu} - \boldsymbol{\nu})\|_2 \\
&\leq \beta_2 (r_t(x) + r_t(y)) \leq 2\beta_2 R_t < \frac{\epsilon}{2}.
\end{aligned}$$

It follows that $m(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0$. Hence, arm x gets discarded at round t.

Second, suppose that $x \notin \mathcal{D}_t$ and for every $y \in \mathcal{A}_t \setminus \{x\}$ we have $m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) \leq \frac{\epsilon}{2}$. Let $\boldsymbol{\mu} \in \mathcal{E}_t(x)$, $\boldsymbol{\nu} \in \mathcal{E}_t(y)$. Since $||\boldsymbol{\mu} - \boldsymbol{\mu}_t(x)||_2 \leq r_t(x)$ and $||\boldsymbol{\nu} - \boldsymbol{\mu}_t(y)||_2 \leq r_t(y)$, by Proposition 4, we have

$$m(\boldsymbol{\mu}, \boldsymbol{\nu}) - \frac{\epsilon}{2} \leq |m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - m(\boldsymbol{\mu}, \boldsymbol{\nu})| \leq \beta_2 (r_t(x) + r_t(y)) \leq 2\beta_2 R_t < \frac{\epsilon}{2}.$$

It follows that $m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon$. Hence, arm x is added to the returned Pareto set at round t. Therefore, each arm gets eliminated from $\mathcal{S}_t$, hence the algorithm stops. □

2.5.1 Proof of Theorem 1

Let t be the round after which PaVeBa stops. Assume that $t > 1$; otherwise the result is trivial. We have $\mathcal{S}_t \neq \emptyset$ and $\mathcal{S}_{t+1} = \emptyset$. Then, for all $x \in \mathcal{A}_t$, we must have $r_{t-1}(x) \geq \frac{\epsilon}{4\beta_2}$; otherwise, by Lemma 10, the algorithm will terminate in round $t-1$ since we will have $r_{t-1}(y) < \frac{\epsilon}{4\beta_2}$ for all $y \in \mathcal{A}_{t-1}$. To see this, note that by Lemma 9, $\mathcal{A}_t \subseteq \mathcal{A}_{t-1}$, and by the sampling rule of PaVeBa (line 6), all arms in $\mathcal{A}_{t-1}$ are equally sampled. Let $\tau = t-1$.

We have for all $x \in \mathcal{A}_t$,

$$\begin{aligned} r_{t-1}^2(x) \geq \frac{\epsilon^2}{16\beta_2^2} &\Leftrightarrow \frac{8\sigma^2(t-1)}{n_{t-1}(x)^2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|(t-1)^2}{6\delta}\right) \geq \frac{\epsilon^2}{16\beta_2^2} \\ &\Leftrightarrow \frac{8\sigma^2}{\tau}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|\tau^2}{6\delta}\right) \geq \frac{\epsilon^2}{16\beta_2^2} \\ &\Leftrightarrow \log\left(\frac{\pi^2(D+1)|\mathcal{X}|\tau^2}{6\delta}\right) \geq \frac{\epsilon^2}{128\beta_2^2\sigma^2}\tau \\ &\Leftrightarrow \log\left(\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}\right) + 2\log(\tau) \geq \frac{\epsilon^2}{128\beta_2^2\sigma^2}\tau \\ &\Leftrightarrow \log(\tau) \geq \frac{\epsilon^2}{256\beta_2^2\sigma^2}\tau - \frac{1}{2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}\right), \end{aligned}$$

where Lemma 9 is used for the second line. By Antos et al. (2010, Lemma 8), the above display implies that

$$\begin{aligned} \tau &< \frac{512\beta_2^2\sigma^2}{\epsilon^2}\left(\log\left(\frac{256\beta_2^2\sigma^2}{\epsilon^2}\right) + \log\left(\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ \\ &= \frac{512\beta_2^2\sigma^2}{\epsilon^2}\left(\log\left(\frac{256\beta_2^2\sigma^2}{\epsilon^2}\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ , \end{aligned}$$

where $(\cdot)^+ := \max\{\cdot, 0\}$. Therefore, an arm in $\mathcal{A}_t$ is sampled no more than

$$\frac{512\beta_2^2\sigma^2}{\epsilon^2}\left(\log\left(\frac{256\beta_2^2\sigma^2}{\epsilon^2}\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ + 1 .$$

times by PaVeBa. Multiplying this number by the cardinality of $\mathcal{X}$ yields the sample complexity upper bound.

2.6 Lemmata for the Proof of Theorem 2

Lemma 11. *Let $t \in \mathbb{N}$ and $x \in \mathcal{S}_t$. Under the event in Lemma 5, if $x \notin P^*$ and $R_t = \max_{y\in\mathcal{A}_t} r_t(y) < \frac{1}{4\beta}\widetilde{\Delta}_\epsilon^+(x)$, or if $x \in P^*$ and $R_t = \max_{y\in\mathcal{A}_t} r_t(y) < \frac{1}{12\beta}\widetilde{\Delta}_{3\epsilon}(x)$, then x will be removed from the undecided set, i.e., $x \notin \mathcal{S}_{t+1}$.*

Proof. Assume that the given condition holds for x. We consider the following three cases.

Case 1: Suppose that $x \notin P^*$ and $\Delta^+(x) = \max_{y\in P^*\setminus\{x\}} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) > \epsilon$. Hence, the condition in the lemma statement implies

$$R_t < \frac{1}{4\beta}\max_{y\in P^*\setminus\{x\}} m(\boldsymbol{f}(x), \boldsymbol{f}(y)).$$

Let $y^* \in P^* \setminus \{x\}$ be such that

$$m(\boldsymbol{f}(x), \boldsymbol{f}(y^*)) = \max_{y\in P^*\setminus\{x\}} m(\boldsymbol{f}(x), \boldsymbol{f}(y)).$$

Note that, by Lemma 7, $y^* \in \mathcal{P}_t \cup \mathcal{S}_t$ under the event in Lemma 5. We claim that $y^* \in \mathcal{A}_t = \mathcal{S}_t \cup \mathcal{U}_t$. Indeed, assume otherwise. Then, we must have $y^* \in \mathcal{P}_t \setminus \mathcal{U}_t$. In particular, for every $z \in \mathcal{S}_t$, we have

$$\sup_{\boldsymbol{\mu}\in\mathcal{E}_{t-1}(z),\boldsymbol{\nu}\in\mathcal{E}_{t-1}(y^*)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon. \tag{S.13}$$

This is a contradiction, since, for $(z, \boldsymbol{\mu}, \boldsymbol{\nu}) = (x, \boldsymbol{f}(x), \boldsymbol{f}(y^*))$, we have $m(\boldsymbol{\mu}, \boldsymbol{\nu}) = m(\boldsymbol{f}(x), \boldsymbol{f}(y^*)) = \Delta^+(x) > \epsilon$. Hence, $y^* \in \mathcal{A}_t$.

Next, notice that

$$2\beta R_t < m(\boldsymbol{f}(x), \boldsymbol{f}(y^*)) - 2\beta R_t \leq m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y^*)) ,$$

where the last step is by Lemma 6 under the event in Lemma 5. Let $\boldsymbol{\mu} \in \mathcal{E}_t(x)$ and $\boldsymbol{\nu} \in \mathcal{E}_t(y^*)$. Since $||\boldsymbol{\mu} - \boldsymbol{\mu}_t(x)||_2 \leq r_t(x)$ and $||\boldsymbol{\nu} - \boldsymbol{\mu}_t(y^*)||_2 \leq r_t(y^*)$, by Proposition 4, we have

$$|m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y^*)) - m(\boldsymbol{\mu}, \boldsymbol{\nu})| \leq \beta(r_t(x) + r_t(y^*)) \leq 2\beta R_t < m(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y^*)),$$

which implies that $m(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0$. Hence, $M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ by Ararat and Tekin (2023, Corollary 4.5). Since $\sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y^*)} M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$, arm x gets discarded at round t.

Case 2: Suppose that $x \in P^*$ and $\Delta(x) > 3\epsilon$. In particular, we have $\min_{y \notin P^*} (M(\boldsymbol{f}(y), \boldsymbol{f}(x)) + 2\Delta^+(y)) > 3\epsilon$ and $\min_{y \in P^* \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)) > 3\epsilon > \epsilon$. Hence, the condition in the lemma statement implies

$$R_t < \frac{1}{12\beta} \min_{y \in P^* \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)) < \frac{1}{4\beta} \min_{y \in P^* \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)).$$

We claim that

$$\min_{y \in P^* \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)) \leq \min_{y \in \overline{\mathcal{A}}_t \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)). \tag{S.14}$$

To see this, note that we have

$$\forall y \in \overline{\mathcal{A}}_t \setminus P^* \; \exists y^* \in P^* \colon M(\boldsymbol{f}(x), \boldsymbol{f}(y^*)) \leq M(\boldsymbol{f}(x), \boldsymbol{f}(y)).$$

Indeed, each dominated arm $y \in \mathcal{X} \setminus P^*$ has a nonempty set of dominating arms, which has a maximal point $y^* \in P^*$ as a finite partially ordered set. Then, by Lemma 1, we have

$$M(\boldsymbol{f}(x), \boldsymbol{f}(y^*)) \leq M(\boldsymbol{f}(x), \boldsymbol{f}(y)) + M(\boldsymbol{f}(y), \boldsymbol{f}(y^*)) = M(\boldsymbol{f}(x), \boldsymbol{f}(y)),$$

where the last step is due to Ararat and Tekin (2023, Corollary 4.5).

Finally, to prove our claim, we need $y^* \neq x$. Assume otherwise; then, we have $(M(\boldsymbol{f}(y), \boldsymbol{f}(x)) + 2\Delta^+(y)) = 2\Delta^+(y) > 3\epsilon$. Also, note that by the condition of the lemma statement, we have $12\beta R_t < \max\{\Delta(x), 3\epsilon\} \leq \max\{2\Delta^+(y), 3\epsilon\}$. This means we have $6\beta R_t < \Delta^+(y)$ or $4\beta R_t < \epsilon$. If the latter is true, the algorithm stops by Lemma 10. The former cannot be true since $\Delta^+(y) > 6\beta R_t > 4\beta R_t$ implies that y will be discarded in round t by case 1 and $y \in \overline{\mathcal{A}}_t$ cannot be true.

Hence, the claim follows, and we get

$$R_t < \frac{1}{4\beta} \min_{y \in \overline{\mathcal{A}}_t \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)).$$

Next, we have for all $y \in \overline{\mathcal{A}}_t \setminus \{x\}$

$$2\beta R_t < \min_{y' \in \overline{\mathcal{A}}_t \setminus \{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y')) - 2\beta R_t \leq M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)),$$

where the last step is by Lemma 6 and taking minimum over the set $\overline{\mathcal{A}}_t \setminus \{x\}$. Let $\boldsymbol{\mu} \in \mathcal{E}_t(x)$ and $\boldsymbol{\nu} \in \mathcal{E}_t(y)$. Since $||\boldsymbol{\mu} - \boldsymbol{\mu}_t(x)||_2 \leq r_t(x)$ and $||\boldsymbol{\nu} - \boldsymbol{\mu}_t(y)||_2 \leq r_t(y)$, for every $y \in \overline{\mathcal{A}}_t \setminus \{x\}$, we have

$$|M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - M(\boldsymbol{\mu}, \boldsymbol{\nu})| \leq \beta(r_t(x) + r_t(y)) \leq 2\beta R_t < M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)),$$

which implies that $M(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0$. For each $y \in \overline{\mathcal{A}}_t$, since $\inf_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} M(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0$, we have

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0 < \epsilon \tag{S.15}$$

by Ararat and Tekin (2023, Corollary 4.5). Hence, arm x is added to the returned Pareto set at round t.

Case 3: Suppose that $x \notin P^*$ and $\Delta^+(x) \leq \epsilon$. Hence, the condition in the lemma statement implies

$$R_t < \frac{\epsilon}{4\beta}.$$

Then, by Lemma 10, the algorithm terminates at round t. In particular, x is eliminated either by being discarded or by being added to the returned Pareto set.

Case 4: Suppose that $x \in P^*$ and $\Delta(x) \leq 3\epsilon$. Hence, the condition in the lemma statement implies

$$R_t < \frac{3\epsilon}{12\beta} = \frac{\epsilon}{4\beta}.$$

Then, by Lemma 10, the algorithm terminates at round t. In particular, x is being added to the returned Pareto set due to Lemma 7. □

Lemma 12. *Under the event in Lemma 5, let $y \in P^*$ be such that $R_t = \max_{x\in\mathcal{A}_t} r_t(x) < \frac{1}{12\beta}\widetilde{\Delta}_{3\epsilon}(y)$. Then, we have $y \notin \mathcal{U}_{t+1}$.*

Proof. Case 1: Suppose that $R_t < \frac{\epsilon}{4\beta}$. Then, the algorithm terminates by Lemma 10 so that $\mathcal{S}_{t+1} = \emptyset$. This implies that $y \notin \mathcal{U}_{t+1}$.

Case 2: Suppose that $R_t \geq \frac{\epsilon}{4\beta}$. Since $R_t < \frac{1}{12\beta}\widetilde{\Delta}_{3\epsilon}(y)$, we must have $3\epsilon < \Delta(y)$ and $R_t < \frac{\Delta(y)}{12\beta}$.

To get a contradiction, suppose that $y \in \mathcal{U}_{t+1}$. Hence, there exist $x \in \mathcal{S}_{t+1} \subseteq \mathcal{S}_t$, $\boldsymbol{\mu} \in \mathcal{E}_t(x)$, and $\boldsymbol{\nu} \in \mathcal{E}_t(y)$ such that

$$m(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq \epsilon > 0, \tag{S.16}$$

which implies that $M(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ by Ararat and Tekin (2023, Corollary 4.5). Then, by Lemma 6,

$$\begin{aligned} M(\boldsymbol{f}(x), \boldsymbol{f}(y)) &\leq |M(\boldsymbol{f}(x), \boldsymbol{f}(y)) - M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y))| + |M(\boldsymbol{\mu}_t(x), \boldsymbol{\mu}_t(y)) - M(\boldsymbol{\mu}, \boldsymbol{\nu})| \\ &\leq 2\beta R_t + 2\beta R_t \leq 4\beta R_t. \end{aligned}$$

We continue by splitting Case 2 into two sub-cases and analyzing them separately.

Case 2.1: Suppose that $x \in P^*$. Then, we have $\Delta(y) \leq M(\boldsymbol{f}(x), \boldsymbol{f}(y)) \leq 4\beta R_t < 12\beta R_t$, which contradicts with $R_t < \frac{\Delta(y)}{12\beta}$. Hence, $y \notin \mathcal{U}_{t+1}$.

Case 2.2: Suppose that $x \notin P^*$. As a first case, if $\Delta^+(x) \leq \epsilon$, then we have $\Delta^+(x) \leq \epsilon \leq 4\beta R_t$ by supposition. As a second case, assume that $\Delta^+(x) > \epsilon$. Since $x \notin P^*$, there exists $x^* \in P^*$ such that $\Delta^+(x) = m(\boldsymbol{f}(x), \boldsymbol{f}(x^*))$; see Section 2. Then, by Lemma 7, we have $x^* \in \mathcal{S}_t \cup \mathcal{P}_t$. Indeed, we claim that $x^* \in \mathcal{A}_t$. Otherwise, we would have $x^* \in \mathcal{P}_t \setminus \mathcal{U}_t$ so that for every $\bar{x} \in \mathcal{S}_t$, $\boldsymbol{\mu} \in \mathcal{E}_{t-1}(\bar{x})$, and $\boldsymbol{\nu} \in \mathcal{E}_{t-1}(x^*)$, we would have

$$m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon. \tag{S.17}$$

But choosing $(\bar{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) = (x, \boldsymbol{f}(x), \boldsymbol{f}(x^*))$ contradicts this. Hence, $x^* \in \mathcal{A}_t$ follows. But since x is not discarded by x^* yet, we have $x \notin \mathcal{D}_t$ so that $M(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0$ for some $\boldsymbol{\mu} \in \mathcal{E}_t(x)$ and $\boldsymbol{\nu} \in \mathcal{E}_t(x^*)$. Then, by Ararat and Tekin (2023, Corollary 4.5), we have

$$m(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0.$$

Therefore, $\Delta^+(x) = m(\boldsymbol{f}(x), \boldsymbol{f}(x^*)) \leq 4\beta R_t$ by Lemma 6. In either case, we have $\Delta^+(x) \leq 4\beta R_t$.

Finally, we have

$$\Delta(y) \leq M(\boldsymbol{f}(x), \boldsymbol{f}(y)) + 2\Delta^+(x) \leq 4\beta R_t + 8\beta R_t = 12\beta R_t,$$

which contradicts with $R_t < \frac{\Delta(y)}{12\beta}$. Hence, $y \notin \mathcal{U}_{t+1}$.

□

Lemma 13. *Let $t \in \mathbb{N}$. Let $x \notin P^*$ be such that $x \in \mathcal{U}_t$. Under the event in Lemma 5, if $R_t = \max_{y\in\mathcal{A}_t} r_t(y) < \frac{1}{4\beta}\widetilde{\Delta}^+_\epsilon(x)$, then $x \notin \mathcal{U}_{t+1}$.*

Proof. Suppose that $R_t < \frac{1}{4\beta}\widetilde{\Delta}^+_\epsilon(x)$. As $x \notin P^*$, we have $\Delta^+(x) = \max_{y\in P^*\setminus\{x\}} m(\boldsymbol{f}(x), \boldsymbol{f}(y))$; see Section 2. We claim that $\Delta^+(x) \leq \epsilon$. Assume otherwise. Since $x \in \mathcal{U}_t \subseteq \mathcal{P}_t$, there exists some round $\tau \in \{1, \ldots, t\}$ such that $x \in \overline{\mathcal{P}}_\tau$. Let $y^* \in P^* \setminus \{x\}$ be such that

$$\Delta^+(x) = \max_{y\in P^*\setminus\{x\}} m(\boldsymbol{f}(x), \boldsymbol{f}(y)) = m(\boldsymbol{f}(x), \boldsymbol{f}(y^*)).$$

Note that, by Lemma 7, under the event in Lemma 5, we have $y^* \in \mathcal{S}_\tau \cup \mathcal{P}_\tau$. Indeed, we have $y^* \in \mathcal{A}_\tau$. Otherwise, we would have $y^* \in \mathcal{P}_\tau \setminus \mathcal{U}_\tau$ so that, for every $z \in \mathcal{S}_\tau$, we would have

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_{\tau-1}(z), \boldsymbol{\nu} \in \mathcal{E}_{\tau-1}(y^*)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon; \tag{S.18}$$

then, choosing $(z, \boldsymbol{\mu}, \boldsymbol{\nu}) = (x, \boldsymbol{f}(x), \boldsymbol{f}(y^*))$ would contradict this.

Going back, since $x \in \overline{\mathcal{P}}_\tau$, we have

$$\sup_{\boldsymbol{\mu} \in \mathcal{E}_{\tau}(x), \boldsymbol{\nu} \in \mathcal{E}_{\tau}(y)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < \epsilon \tag{S.19}$$

for every $y \in \overline{\mathcal{A}}_\tau \setminus \{x\}$. Then, choosing $(y, \boldsymbol{\mu}, \boldsymbol{\nu}) = (y^*, \boldsymbol{f}(x), \boldsymbol{f}(y^*))$ gives $m(\boldsymbol{\mu}, \boldsymbol{\nu}) = \Delta^+(x) > \epsilon$, which is a contradiction.

Since $\Delta^+(x) \leq \epsilon$, we have $\widetilde{\Delta}^+_\epsilon(x) = \epsilon$. Hence, we have $R_t < \frac{\epsilon}{4\beta}$. By Lemma 10, the algorithm terminates at round t. Therefore, $x \notin \mathcal{U}_{t+1}$. □

2.6.1 Proof of Theorem 2

First, we upper bound the number of times each arm in $\mathcal{X}$ gets sampled by PaVeBa under the event in Lemma 5.

Case 1: Fix an arm $x \in \mathcal{X} \setminus P^*$. Let t_x be the round in which x is sampled for the last time by PaVeBa; hence $x \in \mathcal{A}_{t_x}$ and $x \notin \mathcal{A}_{t_x+1}$. Assume that $t_x > 1$; otherwise the result is trivial. Then, for all $y \in \mathcal{A}_t$, we must have $r_{t_x-1}(y) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)}{4\beta}$; otherwise, by Lemmata 11 and 13, x will be eliminated in round $t_x - 1$ since we will have $r_{t_x-1}(y) < \frac{\widetilde{\Delta}^+_\epsilon(x)}{4\beta}$ for all $y \in \mathcal{A}_{t_x-1}$. To see this, note that by Lemma 9, $\mathcal{A}_t \subseteq \mathcal{A}_{t-1}$, and by the sampling rule of PaVeBa (line 6), all arms in $\mathcal{A}_{t-1}$ are equally sampled. Let $\tau = t_x - 1$. We have for all $y \in \mathcal{A}_t$,

$$\begin{aligned}
r^2_{t_x-1}(x) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)^2}{16\beta^2} &\Leftrightarrow \frac{8\sigma^2(t_x-1)}{n_{t_x-1}(x)^2} \log\left(\frac{\pi^2(D+1)|\mathcal{X}|(t_x-1)^2}{6\delta}\right) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)^2}{16\beta^2} \\
&\Leftrightarrow \frac{8\sigma^2}{\tau} \log\left(\frac{\pi^2(D+1)|\mathcal{X}|\tau^2}{6\delta}\right) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)^2}{16\beta^2} \\
&\Leftrightarrow \log\left(\frac{\pi^2(D+1)|\mathcal{X}|\tau^2}{6\delta}\right) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)^2}{128\beta^2\sigma^2}\tau \\
&\Leftrightarrow \log\left(\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}\right) + 2\log(\tau) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)^2}{128\beta^2\sigma^2}\tau \\
&\Leftrightarrow \log(\tau) \geq \frac{\widetilde{\Delta}^+_\epsilon(x)^2}{256\beta^2\sigma^2}\tau - \frac{1}{2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}\right),
\end{aligned}$$

where Lemma 9 is used for the second line. By Antos et al. (2010, Lemma 8), the above display implies that

$$\begin{aligned}
\tau &< \frac{512\beta^2\sigma^2}{\widetilde{\Delta}^+_\epsilon(x)^2}\left(\log\left(\frac{256\beta^2\sigma^2}{\widetilde{\Delta}^+_\epsilon(x)^2}\right) + \log\left(\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ \\
&= \frac{512\beta^2\sigma^2}{\widetilde{\Delta}^+_\epsilon(x)^2}\left(\log\left(\frac{256\beta^2\sigma^2}{\widetilde{\Delta}^+_\epsilon(x)^2}\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+,
\end{aligned}$$

where $(\cdot)^+ := \max\{\cdot, 0\}$. Therefore, arm x is sampled no more than

$$\frac{512\beta^2\sigma^2}{\widetilde{\Delta}^+_\epsilon(x)^2}\left(\log\left(\frac{256\beta^2\sigma^2}{\widetilde{\Delta}^+_\epsilon(x)^2}\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ + 1 \tag{S.20}$$

times by PaVeBa. Summing (S.20) over $x \in \mathcal{X} \setminus P^*$ yields the second term in the sample complexity upper bound.

Case 2: Fix an arm $x \in P^*$. Let t_x be the round in which x is sampled for the last time by PaVeBa; hence $x \in \mathcal{A}_{t_x}$ and $x \notin \mathcal{A}_{t_x+1}$. Assume that $t_x > 1$; otherwise the result is trivial. Then, similar to Case 1, for all $y \in \mathcal{A}_t$, we must have $r_{t_x-1}(y) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)}{12\beta}$; otherwise, by Lemmata 11 and 12, x will be eliminated in round $t_x - 1$ since we will have $r_{t_x-1}(y) < \frac{\widetilde{\Delta}_{3\epsilon}(x)}{12\beta}$ for all $y \in \mathcal{A}_{t_x-1}$. Let $\tau = t_x - 1$. We have for all $y \in \mathcal{A}_t$,

$$
\begin{aligned}
r^2_{t_x-1}(x) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)^2}{144\beta^2} &\Leftrightarrow \frac{8\sigma^2(t_x-1)}{n_{t_x-1}(x)^2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|(t_x-1)^2}{6\delta}\right) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)^2}{144\beta^2} \\
&\Leftrightarrow \frac{8\sigma^2}{\tau}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|\tau^2}{6\delta}\right) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)^2}{144\beta^2} \\
&\Leftrightarrow \log\left(\frac{\pi^2(D+1)|\mathcal{X}|\tau^2}{6\delta}\right) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)^2}{1152\beta^2\sigma^2}\tau \\
&\Leftrightarrow \log\left(\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}\right) + 2\log(\tau) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)^2}{1152\beta^2\sigma^2}\tau \\
&\Leftrightarrow \log(\tau) \geq \frac{\widetilde{\Delta}_{3\epsilon}(x)^2}{2304\beta^2\sigma^2}\tau - \frac{1}{2}\log\left(\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}\right),
\end{aligned}
$$

where Lemma 9 is used for the second line. By Antos et al. (2010, Lemma 8), the above display implies that

$$
\begin{aligned}
\tau &< \frac{4608\beta^2\sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2}\left(\log\left(\frac{2304\beta^2\sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2}\right) + \log\left(\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ \\
&= \frac{4608\beta^2\sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2}\left(\log\left(\frac{2304\beta^2\sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2}\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ .
\end{aligned}
$$

Therefore, arm x is sampled no more than

$$
\frac{4608\beta^2\sigma^2}{\widetilde{\Delta}_{3\epsilon}(x)^2}\left(\log\left(\frac{2304\beta^2\sigma^2}{\widetilde{\Delta}_{3\epsilon}(y)^2}\sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}}\right)\right)^+ + 1 \tag{S.21}
$$

times by PaVeBa. Summing (S.21) over $x \in P^*$ yields the first term in the sample complexity upper bound.

As the final step, we show that the set returned by PaVeBa, i.e., $\hat{P}$, is an (ϵ, δ)-PAC Pareto set according to Definition 1. Due to Lemma 7, $P^* \subseteq \hat{P}$, hence the first condition for being an (ϵ, δ)-PAC Pareto set in Definition 1 holds. The second condition in Definition 1 holds by Lemma 8. Hence, when PaVeBa stops, $\hat{P}$ is an (ϵ, δ)-PAC Pareto set.

2.7 Proof of Remark 2

We compare the gaps appearing at the sample complexity upper bound of PaVeBa with the gaps appearing at the gap-dependent sample complexity lower bound established in Ararat and Tekin (2023) on an arm-by-arm basis. We show that, for many cases, the gaps for an arm $x \in \mathcal{X}$ match either exactly or up to a problem-independent constant. For such cases, we say that arm x maps to itself. We also show that for some $x \in P^*$, the gaps may not match. Then, we define a procedure that maps x to another arm $y \in \mathcal{X}$ such that the gap of arm x ($\widetilde{\Delta}_{3\epsilon}(x)$) in PaVeBa's sample complexity upper bound in Theorem 2 is lower bounded by the gap of arm y (given in Ararat and Tekin (2023, Theorem 5.1), denoted here by $\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(y)$) in the sample complexity lower bound. Let $c(y)$ represent the number of arms $x \in \mathcal{X}$ that are mapped to arm y (including arm y if it maps to itself). This mapping allows us to obtain the following upper bound on the sample complexity of PaVeBa, ignoring constant and logarithmic factors.

$$
\begin{aligned}
\text{(Theorem 2) } \tilde{O}\left(\sum_{x\in P^*}\frac{1}{(\widetilde{\Delta}_{3\epsilon}(x))^2} + \sum_{x\in\mathcal{X}\setminus P^*}\frac{1}{(\widetilde{\Delta}^+_{\epsilon}(x))^2}\right) &\leq \tilde{O}\left(\sum_{x\in\mathcal{X}}\frac{c(x)}{(\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(x))^2}\right) \\
&\leq \tilde{O}\left((\max_{x\in\mathcal{X}} c(x))\left(\sum_{x\in\mathcal{X}}\frac{1}{(\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(x))^2}\right)\right) ,
\end{aligned}
$$

where $\tilde{O}\left(\frac{1}{(\tilde{\Delta}^{\epsilon,\mathrm{AT}}(x))^2}\right)$ represents the sample complexity lower bound. Without further restriction on the bandit environment class, there exist instances for which $\max_{c\in\mathcal{X}} c(x) = O(|\mathcal{X}|)$.

Let $c_* \geq 0$ be an integer. We define the norm subgaussian bandit environment class $\mathcal{E}(c_*)$ as the set of all norm subgaussian bandits defined over arm set $\mathcal{X}$ for which $\max_{x\in\mathcal{X}} c(x) \leq c_*$. Clearly, this environment class is a subset of norm subgaussian bandits studied here in our work and in Ararat and Tekin (2023); therefore, the sample complexity lower bound of Ararat and Tekin (2023) is still valid even though there might exist a smaller lower bound due to the restriction put on the environment class. For $\mathcal{E}(c_*)$, we can treat $\max_{x\in\mathcal{X}} c(x) \leq c_* = O(1)$, matching the sample complexity lower bound up to constant factors.

Next, we describe how to map each $x \in \mathcal{X}$. First, suppose that $x \in \mathcal{X} \setminus P^*$. For x, the lower bound in Ararat and Tekin (2023, Theorem 5.1) involves

$$\frac{1}{(\max\{\max_{y\in P^*} m(\boldsymbol{f}(x), \boldsymbol{f}(y)), \epsilon\})^2}.$$

This term matches with our upper bound, which involves

$$\frac{1}{(\tilde{\Delta}^+_\epsilon(x))^2} = \frac{1}{(\max\{\Delta^+(x), \epsilon\})^2} = \frac{1}{(\max\{\max_{y\in P^*} m(\boldsymbol{f}(x), \boldsymbol{f}(y)), \epsilon\})^2}.$$

Hence, x maps to itself in this case.

Next, suppose that $x \in P^*$. We define the following terms:

$$\Delta^{+,\mathrm{AT}}(x) \coloneqq \min_{y\in P^*\setminus\{x\}} M(\boldsymbol{f}(x), \boldsymbol{f}(y)),$$

$$\Delta^{+,\mathrm{NEW}}(x) \coloneqq \min_{y\in P^*\setminus\{x\}} M(\boldsymbol{f}(y), \boldsymbol{f}(x)),$$

$$\Delta^{-}(x) = \min_{y\notin P^*} M(\boldsymbol{f}(y), \boldsymbol{f}(x)) + 2\Delta^+(y).$$

Case 1: $\tilde{\Delta}_{3\epsilon}(x) = 3\epsilon$.

In this case, we have $\min\{\Delta^{+,\mathrm{AT}}(x), \Delta^{+,\mathrm{NEW}}(x), \Delta^-(x)\} \leq 3\epsilon$ and the upper bound in PaVeBa involves

$$\frac{1}{\tilde{\Delta}_{3\epsilon}(x)^2} = \frac{1}{9\epsilon^2}.$$

We have at least one of the three subcases true:

Case 1.1: $(\tilde{\Delta}_{3\epsilon}(x) = 3\epsilon) \wedge (\Delta^{+,\mathrm{AT}}(x) \leq 3\epsilon)$.

In this subcase, the lower bound in Ararat and Tekin (2023) involves $\frac{1}{(\max\{\Delta^{+,\mathrm{AT}}(x),\epsilon\})^2}$. Moreover, we have $\epsilon \leq \max\{\Delta^{+,\mathrm{AT}}(x), \epsilon\} \leq 3\epsilon$. Therefore,

$$\frac{1}{9\epsilon^2} \leq \frac{1}{(\max\{\Delta^{+,\mathrm{AT}}(x), \epsilon\})^2} \leq \frac{1}{\epsilon^2}.$$

Again, we match the lower bound up to a constant. Hence, x maps to itself in this subcase.

Case 1.2: $(\tilde{\Delta}_{3\epsilon}(x) = 3\epsilon) \wedge (\Delta^{+,\mathrm{AT}}(x) > 3\epsilon) \wedge (\Delta^{+,\mathrm{NEW}}(x) \leq 3\epsilon)$.

In this subcase, let $y^* \in P^* \setminus \{x\}$ be such that $\Delta^{+,\mathrm{NEW}}(x) = M(\boldsymbol{f}(y^*), \boldsymbol{f}(x))$. Then, we have $\Delta^{+,\mathrm{AT}}(y^*) \leq M(\boldsymbol{f}(y^*), \boldsymbol{f}(x)) = \Delta^{+,\mathrm{NEW}}(x) \leq 3\epsilon$. Moreover, the lower bound for y^* in Ararat and Tekin (2023) involves $\frac{1}{(\max\{\Delta^{+,\mathrm{AT}}(y^*),\epsilon\})^2}$. Note that we have $\epsilon \leq \max\{\Delta^{+,\mathrm{AT}}(y^*), \epsilon\} \leq 3\epsilon$. Then,

$$\frac{1}{9\epsilon^2} \leq \frac{1}{(\max\{\Delta^{+,\mathrm{AT}}(y^*), \epsilon\})^2} \leq \frac{1}{\epsilon^2}.$$

Therefore, our upper bound for x matches with the lower bound for $y^* \in P^*$ up to a constant. Hence, x maps to y^* in this subcase.

Case 1.3: $(\tilde{\Delta}_{3\epsilon}(x) = 3\epsilon) \wedge (\Delta^{+,\mathrm{AT}}(x) > 3\epsilon) \wedge (\Delta^{+,\mathrm{NEW}}(x) > 3\epsilon) \wedge (\Delta^{-}(x) \leq 3\epsilon)$.

In this subcase, let $y^* \notin P^*$ be such that $\Delta^-(x) = M(\boldsymbol{f}(y^*), \boldsymbol{f}(x)) + 2\Delta^+(y^*)$. Then, we have $\Delta^+(y^*) \leq \frac{3\epsilon}{2}$. Moreover, the lower bound for y^* in Ararat and Tekin (2023) involves

$$\frac{1}{(\max\{\max_{y\in P^*} m(\boldsymbol{f}(y^*), \boldsymbol{f}(y)), \epsilon\})^2} = \frac{1}{(\max\{\Delta^+(y^*), \epsilon\})^2}.$$

Note that we have $\epsilon \leq \max\{\Delta^+(y^*), \epsilon\} \leq \frac{3\epsilon}{2}$. Then,

$$\frac{4}{9\epsilon^2} \leq \frac{1}{(\max\{\Delta^+(y^*), \epsilon\})^2} = \frac{1}{(\Delta^+(y^*))^2} \leq \frac{1}{\epsilon^2}.$$

Therefore, our upper bound for x matches with the lower bound for $y^* \notin P^*$ up to a constant. Hence, x maps to y^* in this subcase.

Case 2: $(\tilde{\Delta}_{3\epsilon}(x) > 3\epsilon) \wedge (\tilde{\Delta}_{3\epsilon}(x) = \Delta^{+,\mathrm{AT}}(x))$.

In this case, we have $\Delta^{+,\mathrm{AT}}(x) > \epsilon$. Moreover, the upper bound in PaVeBa involves

$$\frac{1}{\tilde{\Delta}_{3\epsilon}(x)^2} = \frac{1}{(\Delta^{+,\mathrm{AT}}(x))^2}$$

while the lower bound in Ararat and Tekin (2023) involves

$$\frac{1}{(\max\{\Delta^{+,\mathrm{AT}}(x), \epsilon\})^2} = \frac{1}{(\Delta^{+,\mathrm{AT}}(x))^2}.$$

Therefore, we match the lower bound. Hence, x maps to itself in this case.

Case 3: $(\tilde{\Delta}_{3\epsilon}(x) > 3\epsilon) \wedge (\tilde{\Delta}_{3\epsilon}(x) < \Delta^{+,\mathrm{AT}}(x)) \wedge (\tilde{\Delta}_{3\epsilon}(x) = \Delta^{-}(x))$.

In this case, we have $\widetilde{\Delta}_{3\epsilon}(x) = M(\boldsymbol{f}(y), \boldsymbol{f}(x)) + 2\Delta^+(y) \geq \Delta^+(y)$ for some $y \notin P^*$. Thus, the upper bound in PaVeBa includes $\frac{1}{(\Delta^-(x))^2}$ while the term for y in the lower bound includes

$$\frac{1}{(\Delta^-(x))^2} \leq \frac{1}{(\Delta^+(y))^2}.$$

Hence, we match the term for y in the lower bound if $\max\{\Delta^+(y), \epsilon\} \neq \epsilon$. If not, then we still have $\Delta^-(x) > 3\epsilon$ so that

$$\frac{1}{(\Delta^-(x))^2} < \frac{1}{9\epsilon^2} = \frac{1}{9\max\{\Delta^+(y), \epsilon\}^2}.$$

Hence, x maps to y in this case.

Case 4: $(\tilde{\Delta}_{3\epsilon}(x) > 3\epsilon) \wedge (\tilde{\Delta}_{3\epsilon}(x) < \Delta^{+,\mathrm{AT}}(x)) \wedge (\tilde{\Delta}_{3\epsilon}(x) < \Delta^{-}(x))$.

In this case, $\tilde{\Delta}_{3\epsilon}(x) = \Delta^{+,\mathrm{NEW}}(x)$. Let x_1^* be such that $M(\boldsymbol{f}(x_1^*), \boldsymbol{f}(x)) = \min_{y\in P^*\setminus\{x\}} M(\boldsymbol{f}(y), \boldsymbol{f}(x)) = \Delta^{+,\mathrm{NEW}}(x)$. Then, we have

$$\Delta(x_1^*) \leq \Delta^{+,\mathrm{AT}}(x_1^*) \leq M(\boldsymbol{f}(x_1^*), \boldsymbol{f}(x)) = \Delta^{+,\mathrm{NEW}}(x). \tag{S.22}$$

If $(\tilde{\Delta}_{3\epsilon}(x_1^*) \leq 3\epsilon) \vee (\tilde{\Delta}_{3\epsilon}(x_1^*) \geq \Delta^{+,\mathrm{AT}}(x_1^*)) \vee (\tilde{\Delta}_{3\epsilon}(x_1^*) \geq \Delta^{-}(x_1^*))$, then by one of the three cases above, x_1^* maps to an arm x_2^* such that

$$\frac{1}{(\tilde{\Delta}_{3\epsilon}(x))^2} = \frac{1}{(\Delta^{+,\mathrm{NEW}}(x))^2} \leq \frac{1}{(\Delta(x_1^*))^2} = O\left(\frac{1}{(\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(x_2^*))^2}\right).$$

Hence, x maps to x_2^* in this case.

On the other hand, if $(\tilde{\Delta}_{3\epsilon}(x_1^*) > 3\epsilon) \wedge (\tilde{\Delta}_{3\epsilon}(x_1^*) < \Delta^{+,\mathrm{AT}}(x_1^*)) \wedge (\tilde{\Delta}_{3\epsilon}(x_1^*) < \Delta^{-}(x_1^*))$, then we are in Case 4 for x_1^*. In this case, (S.22) implies the existence of $x_2^* \in \mathcal{P} \setminus \{x_1^*\}$ such that $\Delta(x_2^*) \leq \Delta^{+,\mathrm{NEW}}(x_1^*)$. The arguments above will be repeated for x_2^* until an arm x_n^* for which one of the first three cases holds is found. If such an arm is found, then,

$$\frac{1}{(\tilde{\Delta}_{3\epsilon}(x))^2} = \frac{1}{(\Delta^{+,\mathrm{NEW}}(x))^2} \leq \frac{1}{(\Delta(x_1^*))^2} \leq \frac{1}{(\Delta(x_2^*))^2} \leq \cdots \leq \frac{1}{(\Delta(x_{n-1}^*))^2} = O\left(\frac{1}{(\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(x_n^*))^2}\right).$$

We will conclude by proving that this procedure terminates at some x_n^* by showing that the same arm will not be visited more than once. In this case, x maps to x_n^*. Let $x_0^* = x$ and x_i^* be the ith arm found in the ith iteration of the procedure above.

Case 4.1: Assume that the procedure has not terminated at the end of iteration n, yielding the sequence of arms $x_0^*, \ldots, x_n^*$ such that $x_i^* \neq x_j^*$ for all $i, j \in \{0, \ldots, n\}$, $i \neq j$. If $n \geq |\mathcal{X}|$, we get a contradiction. This means that the procedure should terminate after less than $|\mathcal{X}|$ iterations yielding an arm x^* such that

$$\frac{1}{(\tilde{\Delta}_{3\epsilon}(x))^2} \leq O\left(\frac{1}{(\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(x^*))^2}\right).$$

Case 4.2: Contrary to Case 4.1, assume that the procedure has not terminated at the end of iteration n, yielding the sequence of arms $x_0^*, \ldots, x_n^*$ such that there exist indices $a < b$, $a, b \in \{0, \ldots, n\}$ for which $x_a^* = x_b^*$. Then, for all x_i^*, $0 \leq i \leq n$, the statement at the beginning of Case 4 holds implying that $\Delta^{+,\mathrm{NEW}}(x_i^*) < \Delta^{+,\mathrm{AT}}(x_i^*)$. Now, for all j such that $a + 1 \leq j \leq b$, we have

$$M(\boldsymbol{f}(x_j^*), \boldsymbol{f}(x_{j-1}^*)) = \Delta^{+,\mathrm{NEW}}(x_{j-1}^*) < \Delta^{+,\mathrm{AT}}(x_{j-1}^*) = \min_{x_j \in P^* \setminus \{x_{j-1}^*\}} M(\boldsymbol{f}(x_{j-1}^*), \boldsymbol{f}(x_j)).$$

Using these inequalities constructively, beginning with $j = b - 1$ and using the fact that $x_b^* = x_a^*$, we have

$$\begin{aligned} M(\boldsymbol{f}(x_b^*), \boldsymbol{f}(x_{b-1}^*)) &= M(\boldsymbol{f}(x_a^*), \boldsymbol{f}(x_{b-1}^*)) \\ &< \min_{x_a \in P^* \setminus \{x_{b-1}^*\}} M(\boldsymbol{f}(x_{b-1}^*), \boldsymbol{f}(x_a)) \\ &\leq M(\boldsymbol{f}(x_{b-1}^*), \boldsymbol{f}(x_{b-2}^*)) \\ &< \min_{x_{b-1} \in P^* \setminus \{x_{b-2}^*\}} M(\boldsymbol{f}(x_{b-2}^*), \boldsymbol{f}(x_{b-1})) \\ &< \ldots < \min_{x_{a+1} \in P^* \setminus \{x_a^*\}} M(\boldsymbol{f}(x_a^*), \boldsymbol{f}(x_{a+1})) \leq M(\boldsymbol{f}(x_a^*), \boldsymbol{f}(x_{b-1}^*)), \end{aligned}$$

which is a contradiction. Hence, such a loop cannot exist. Therefore, we must have $x_i^* \neq x_j^*$ for all $i, j \in \{0, \ldots, n\}$, $i \neq j$, which by Case 4.1 implies that the procedure should terminate after less than $|\mathcal{X}|$ iterations yielding an arm x^* such that

$$\frac{1}{(\tilde{\Delta}_{3\epsilon}(x))^2} \leq O\left(\frac{1}{(\widetilde{\Delta}^{\epsilon,\mathrm{AT}}(x^*))^2}\right).$$

3 IMPLEMENTATION AND SAMPLE COMPLEXITY OF PaVeBa WHEN $\epsilon = 0$

The problem with PaVeBa when we have $\epsilon = 0$ is twofold: (i) As pointed out in Remark 1, Proposition 2 does not work when $\epsilon = 0$ due to the positive part for the computation of $m(\cdot, \cdot)$ in Ararat and Tekin (2023, Proposition 4.2) making all $m(\cdot, \cdot)$ nonnegative, and (ii) from the definitions in (7) and (8), we would have

$$\begin{gathered} \overline{\mathcal{P}}_t = \{x \in \overline{\mathcal{S}}_t \,|\, \forall y \in \overline{\mathcal{A}}_t \setminus \{x\}\colon \sup_{\boldsymbol{\mu} \in \mathcal{E}_t(x), \boldsymbol{\nu} \in \mathcal{E}_t(y)} m(\boldsymbol{\mu}, \boldsymbol{\nu}) < 0\} = \emptyset \\ \implies \mathcal{P}_{t+1} = \mathcal{P}_t,\ \mathcal{S}_{t+1} = \overline{\mathcal{S}}_t = \mathcal{S}_t \setminus \mathcal{D}_t \text{ (PaVeBa lines 11 and 9).} \end{gathered}$$

Since $\mathcal{P}_1 = \emptyset$, the above iteration will result in $\mathcal{P}_t = \emptyset$ for all t. Therefore,

$$\mathcal{U}_{t+1} = \{y \in \mathcal{P}_{t+1} \,|\, \exists x \in \mathcal{S}_{t+1}\colon \sup_{\substack{\boldsymbol{\mu} \in \mathcal{E}_t(x), \\ \boldsymbol{\nu} \in \mathcal{E}_t(y)}} m(\boldsymbol{\mu}, \boldsymbol{\nu}) \geq 0\} = \emptyset, \forall t\ .$$

Then, by line 4 of PaveBa, $\mathcal{A}_t = \mathcal{S}_t$ for all t.

To conclude, PaVeBa samples all undecided arms in all rounds. PaVeBa can only discard arms, but it will never declare an arm as Pareto optimal. Since there are Pareto optimal arms in the undecided set at the beginning, it will never terminate.

To circumvent these, we offer a slight change in the definitions of $\overline{\mathcal{P}}_t$ and $\mathcal{U}_{t+1}$ as follows:

$$\overline{\mathcal{P}}_t = \{x \in \overline{\mathcal{S}}_t \,|\, \forall y \in \overline{\mathcal{A}}_t \setminus \{x\}\colon \sup_{\substack{\boldsymbol{\mu} \in \mathcal{E}_t(x), \\ \boldsymbol{\nu} \in \mathcal{E}_t(y)}} m(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq 0\}, \tag{S.23}$$

$$\mathcal{U}_{t+1} = \{y \in \mathcal{P}_{t+1} \,|\, \exists x \in \mathcal{S}_{t+1}\colon \sup_{\substack{\boldsymbol{\mu} \in \mathcal{E}_t(x), \\ \boldsymbol{\nu} \in \mathcal{E}_t(y)}} m(\boldsymbol{\mu}, \boldsymbol{\nu}) > 0\}. \tag{S.24}$$

Let the new variant of PaVeBa using the sets defined in (S.23) and (S.24) be PaVeBa-0. Note that all the algorithm steps remain the same with PaVeBa except these two slight changes in the definitions of $\overline{\mathcal{P}}_t$ and $\mathcal{U}_{t+1}$. We now present the sample complexity analysis for PaVeBa-0.

3.1 Sample Complexity of PaVeBa-0

Throughout this subsection, we assume that there exists no pair of distinct arms $x, y \in \mathcal{X}$, $x \neq y$ such that $\boldsymbol{f}(x) - \boldsymbol{f}(y) \in \mathrm{bd}(C)$. Consider P in Definition 1. When $\epsilon = 0$, condition (ii) becomes $\Delta^*(x) = 0$ for every $x \in P \setminus P^*$. Under the assumption above, by Ararat and Tekin (2023, Corollary 4.5) $\Delta^*(x) = 0$ implies that $M(\boldsymbol{f}(x), \boldsymbol{f}(y)) > 0$ for all $y \in P^*$, and hence, $x \not\preceq_C y$ for all $y \in P^*$. Then, by definition of P^*, such an x cannot be in $\mathcal{X} \setminus P^*$, resulting in $P \setminus P^* = \emptyset$. Due to this, the success conditions in Definition 1 reduces to the following when $\epsilon = 0$.

Definition 2. *Let $\delta \in (0,1)$. A random set $P \subseteq \mathcal{X}$ is called a δ-Probably Correct (PC) Pareto set if $P = P^*$ with probability at least $1 - \delta$.*

Theorem 3. *Assume that there exists no pair of distinct arms $x, y \in \mathcal{X}$, $x \neq y$ such that $\boldsymbol{f}(x) - \boldsymbol{f}(y) \in \mathrm{bd}(C)$. When PaVeBa-0 is run, the maximum number of samples required for it to output a δ-PC Pareto set $\hat{P}$ is*

$$\begin{aligned}
&\sum_{x \in P^*} \frac{4608\beta^2\sigma^2}{\Delta(x)^2} \log^+ \left(\frac{2304\beta^2\sigma^2}{\Delta(x)^2} \sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}} \right) \\
&+ \sum_{x \in \mathcal{X} \setminus P^*} \frac{512\beta^2\sigma^2}{\Delta^+(x)^2} \log^+ \left(\frac{256\beta^2\sigma^2}{\Delta^+(x)^2} \sqrt{\frac{\pi^2(D+1)|\mathcal{X}|}{6\delta}} \right) \\
&+ |\mathcal{X}|.
\end{aligned}$$

Proof. Proofs presented in the supplemental Section 2 are still valid for PaVeBa-0 except for some minor changes in the inclusion cases of inequalities.

In particular, the statement of Lemma 7 remains unchanged. The statement of Lemma 8 becomes "Under the event in Lemma 5, for every $t \in \mathbb{N}$ if $x \in \mathcal{X} \setminus P^*$, then $x \notin \mathcal{P}_t$". The statement of Lemma 9 remains unchanged. The statement of Lemma 11 becomes "Let $t \in \mathbb{N}$ and $x \in \mathcal{S}_t$. Under the event in Lemma 5, if $x \notin P^*$ and $R_t = \max_{y \in \mathcal{A}_t} r_t(y) < \frac{1}{4\beta}\Delta^+(x)$, or if $x \in P^*$ and $R_t = \max_{y \in \mathcal{A}_t} r_t(y) < \frac{1}{12\beta}\Delta(x)$, then x will be removed from the undecided set, i.e., $x \notin \mathcal{S}_{t+1}$". The statement of Lemma 12 becomes "Under the event in Lemma 5, let $y \in P^*$ be such that $R_t = \max_{x \in \mathcal{A}_t} r_t(x) < \frac{1}{12\beta}\Delta(y)$. Then, we have $y \notin \mathcal{U}_{t+1}$". The case in Lemma 13 never happens under the event in Lemma 5 thanks to the revised version of Lemma 8.

The sample complexity upper bound is found by following the same reasoning as in the proof of Theorem 2.

Due to Lemma 7, $P^* \subseteq \hat{P}$. Due to revised Lemma 8, if $x \in \mathcal{X} \setminus P^*$ then $x \notin \hat{P}$. Therefore, $\hat{P} = P^*$ Hence, when PaVeBa-0 stops, $\hat{P}$ is a δ-PC Pareto set. □

3.2 Efficient Implementation of PaVeBa-0 via Convex Programming

We now need to find another convex programming method to compute the updated sets in (S.23) and (S.24).

Proposition 5. *Given a pair of arms x, y and associated confidence regions $\mathcal{E}_t(x), \mathcal{E}_t(y)$; $\sup_{\boldsymbol{\mu}\in\mathcal{E}_t(x),\boldsymbol{\nu}\in\mathcal{E}_t(y)} m(\boldsymbol{\mu},\boldsymbol{\nu}) \leq 0$ if and only if the optimal value of the following feasibility problem is $+\infty$, i.e., it yields infeasibility:*

$$\begin{aligned} \textit{minimize} \quad & 0 \\ \textit{subject to} \quad & \boldsymbol{w}_n^\mathsf{T}(\boldsymbol{\mu}_y - \boldsymbol{\mu}) \geq 0,\ \boldsymbol{w}_n^\mathsf{T}(\boldsymbol{\mu} - \boldsymbol{\mu}_x) \geq 0, \forall n \in [N], \\ & (\boldsymbol{\mu}_x - \boldsymbol{\mu}_\tau(x))^\mathsf{T} \boldsymbol{\Sigma}_\tau^{-1}(x)(\boldsymbol{\mu}_x - \boldsymbol{\mu}_\tau(x)) \leq \alpha_t, \forall \tau \in [t], \\ & (\boldsymbol{\mu}_y - \boldsymbol{\mu}_\tau(y))^\mathsf{T} \boldsymbol{\Sigma}_\tau^{-1}(y)(\boldsymbol{\mu}_y - \boldsymbol{\mu}_\tau(y)) \leq \alpha_t, \forall \tau \in [t], \\ & \boldsymbol{\mu}, \boldsymbol{\mu}_x, \boldsymbol{\mu}_y \in \mathbb{R}^D. \end{aligned}$$

Proof. Given two vectors $\boldsymbol{\mu}_x, \boldsymbol{\mu}_y \in \mathbb{R}^D$, observe that

$$\boldsymbol{\mu}_y - \boldsymbol{\mu}_x \notin C \quad \Leftrightarrow \quad \boldsymbol{\mu}_y \notin \boldsymbol{\mu}_x + C \quad \Leftrightarrow \quad (\boldsymbol{\mu}_y - C) \cap (\boldsymbol{\mu}_x + C) = \emptyset.$$

Then, we may write

$$\begin{aligned} & \{\forall \boldsymbol{\mu}_x \in \mathcal{E}_t(x),\ \forall \boldsymbol{\mu}_y \in \mathcal{E}_t(y) \colon \boldsymbol{\mu}_y - \boldsymbol{\mu}_x \notin C\} \\ & = \{\forall \boldsymbol{\mu}_x \in \mathcal{E}_t(x),\ \forall \boldsymbol{\mu}_y \in \mathcal{E}_t(y) \colon (\boldsymbol{\mu}_y - C) \cap (\boldsymbol{\mu}_x + C) = \emptyset\} \\ & = \left\{ \left[\bigcup_{\boldsymbol{\mu}_x \in \mathcal{E}_t(x)} (\boldsymbol{\mu}_y - C) \right] \cap \left[\bigcup_{\boldsymbol{\mu}_y \in \mathcal{E}_t(y)} (\boldsymbol{\mu}_x + C) \right] = \emptyset \right\} \\ & = \{(\mathcal{E}_t(y) - C) \cap (\mathcal{E}_t(x) + C) = \emptyset\} \ . \end{aligned}$$

Let $x, y \in \mathcal{X}$ with $x \neq y$. In order to check if $(\mathcal{E}_t(y) - C) \cap (\mathcal{E}_t(x) + C) = \emptyset$, let us consider the following feasibility problem expressed in the form of a mathematical program:

$$\begin{aligned} \text{minimize} \quad & 0 \\ \text{subject to} \quad & \boldsymbol{\mu} \in \mathcal{E}_t(y) - C, \quad \boldsymbol{\mu} \in \mathcal{E}_t(x) + C, \quad \boldsymbol{\mu} \in \mathbb{R}^D. \end{aligned}$$

This problem can be rewritten more explicitly as

$$\begin{aligned} \text{minimize} \quad & 0 \\ \text{subject to} \quad & \boldsymbol{\mu}_y - \boldsymbol{\mu} \in C, \quad \boldsymbol{\mu} - \boldsymbol{\mu}_x \in C, \\ & \boldsymbol{\mu}_x \in \mathcal{E}_t(x), \quad \boldsymbol{\mu}_y \in \mathcal{E}_t(y), \\ & \boldsymbol{\mu}, \boldsymbol{\mu}_x, \boldsymbol{\mu}_y \in \mathbb{R}^D, \end{aligned}$$

which is equivalent to

$$\begin{aligned} \text{minimize} \quad & 0 \\ \text{subject to} \quad & \boldsymbol{w}_n^\mathsf{T}(\boldsymbol{\mu}_y - \boldsymbol{\mu}) \geq 0, \quad \boldsymbol{w}_n^\mathsf{T}(\boldsymbol{\mu} - \boldsymbol{\mu}_x) \geq 0, \quad \forall n \in [N], \\ & (\boldsymbol{\mu}_x - \boldsymbol{\mu}_t(x))^\mathsf{T} \boldsymbol{\Sigma}_t^{-1}(x)(\boldsymbol{\mu}_x - \boldsymbol{\mu}_t(x)) \leq \alpha_t, \forall \tau \in [t], \\ & (\boldsymbol{\mu}_y - \boldsymbol{\mu}_t(y))^\mathsf{T} \boldsymbol{\Sigma}_t^{-1}(y)(\boldsymbol{\mu}_y - \boldsymbol{\mu}_t(y)) \leq \alpha_t, \forall \tau \in [t], \\ & \boldsymbol{\mu}, \boldsymbol{\mu}_x, \boldsymbol{\mu}_y \in \mathbb{R}^D. \end{aligned}$$

This is a convex optimization (feasibility) problem with affine and quadratic constraints. We have $(\mathcal{E}_t(y) - C) \cap (\mathcal{E}_t(x) + C) = \emptyset$ if and only if the problem yields $+\infty$ as its optimal value, i.e., the problem is infeasible. Otherwise, the optimal value of the problem is zero, and the empty intersection property does not hold for x, y.

□

4 DERIVATION OF CONFIDENCE REGIONS FOR GAUSSIAN PROCESSES

We include this analysis to make our algorithm compatible with the use of GPs for improved regression in the experiments. We assume a multi-output GP with possibly correlated objectives. Then, we take our single-round confidence regions to be

$$\mathcal{B}_t(x) = \{\boldsymbol{\nu} : (\boldsymbol{\nu} - \boldsymbol{\mu}_t(x))^\mathsf{T} \boldsymbol{\Sigma}_t^{-1}(x)(\boldsymbol{\nu} - \boldsymbol{\mu}_t(x)) \leq \alpha_t\}, \tag{S.25}$$

where $\alpha_t = 8D\log(12) + 4\log(\frac{\pi^2 t^2 |\mathcal{X}|}{6\delta})$, $\boldsymbol{\mu}_t(x)$ is the posterior mean of the GP of arm x at round t, and $\boldsymbol{\Sigma}_t(x)$ is the posterior covariance matrix of the GP for arm x at round t. Further, we define

$$\mathcal{E}_t(x) \coloneqq \mathcal{E}_{t-1}(x) \cap \mathcal{B}_t(x) \tag{S.26}$$

with $\mathcal{E}_1(x) \coloneqq \mathcal{B}_1(x)$ as the confidence region of x at round t. (Note that the intersection of confidence regions is a technical detail for the proofs to work in the sample complexity analysis. However, it happens rarely in practice for $\mathcal{B}_{t+1}(x)$ not to be a subset of $\mathcal{B}_t(x)$; hence, we dropped intersecting regions $\mathcal{B}_\tau(x)$ for $\tau \leq t$ in the experiments where heuristic variants are used.)

We start with a covering lemma that is a refinement of Lattimore and Szepesvári (2020, Lemma 20.1).

Lemma 14. *For every $\varepsilon \in (0, 2)$, there exists a set $\mathcal{C}_\varepsilon \subseteq \mathbb{S}^{D-1}$ such that $|\mathcal{C}_\varepsilon| \leq (\frac{6}{\varepsilon})^D$ and*

$$\forall \boldsymbol{w} \in \mathbb{S}^{D-1}\ \exists \tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon \colon \|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \leq \varepsilon.$$

Proof. By Lattimore and Szepesvári (2020, Lemma 20.1), for every $\varepsilon > 0$, there exists a set $\tilde{\mathcal{C}}_\varepsilon \subseteq \mathbb{R}^D$ such that $|\tilde{\mathcal{C}}_\varepsilon| \leq (\frac{3}{\varepsilon})^D$ and

$$\forall \boldsymbol{w} \in \mathbb{S}^{D-1}\ \exists \tilde{\boldsymbol{w}} \in \tilde{\mathcal{C}}_\varepsilon \colon \|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \leq \varepsilon.$$

Moreover, without loss of generality, we assume that

$$\forall \tilde{\boldsymbol{w}} \in \tilde{\mathcal{C}}_\varepsilon\ \exists \boldsymbol{w} \in \mathbb{S}^{D-1} \colon \|\boldsymbol{w} - \tilde{\boldsymbol{w}}\| \leq \varepsilon.$$

as otherwise one can remove from $\tilde{\mathcal{C}}_\varepsilon$ the elements $\tilde{\boldsymbol{w}}$ for which there is no $\boldsymbol{w} \in \mathbb{S}^{D-1}$ with $\|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \leq \varepsilon$ and the new set still satisfies the two conditions that $\tilde{\mathcal{C}}_\varepsilon$ satisfies.

We claim that $\tilde{\mathcal{C}}_\varepsilon \subseteq B(\mathbf{0}, 1+\varepsilon) \setminus \operatorname{int} B(\mathbf{0}, 1-\varepsilon)$ whenever $\varepsilon \in (0, 1)$. Let $\tilde{\boldsymbol{w}} \in \tilde{\mathcal{C}}_\varepsilon$. By the assumption above, there exists $\boldsymbol{w} \in \mathbb{S}^{D-1}$ such that $\|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \leq \varepsilon$. Then, by triangle inequality, we have

$$\|\tilde{\boldsymbol{w}}\|_2 \leq \|\tilde{\boldsymbol{w}} - \boldsymbol{w}\|_2 + \|\boldsymbol{w}\|_2 \leq \varepsilon + 1.$$

Hence, $\tilde{\boldsymbol{w}} \in B(\mathbf{0}, 1+\varepsilon)$. Moreover, by reverse triangle inequality, we have

$$\|\tilde{\boldsymbol{w}}\|_2 = \|(\tilde{\boldsymbol{w}} - \boldsymbol{w}) - (-\boldsymbol{w})\|_2 \geq |\|\tilde{\boldsymbol{w}} - \boldsymbol{w}\|_2 - \|-\boldsymbol{w}\|_2| = |\|\tilde{\boldsymbol{w}} - \boldsymbol{w}\|_2 - 1| \geq 1 - \|\tilde{\boldsymbol{w}} - \boldsymbol{w}\|_2 \geq 1 - \varepsilon.$$

Hence, $\tilde{\boldsymbol{w}} \notin \operatorname{int} B(\mathbf{0}, 1-\varepsilon)$, which completes the proof of the claim.

Let us fix $\varepsilon \in (0, 2)$ and define

$$\mathcal{C}_\varepsilon \coloneqq \left\{ \frac{\tilde{\boldsymbol{w}}}{\|\tilde{\boldsymbol{w}}\|_2} \colon \tilde{\boldsymbol{w}} \in \tilde{\mathcal{C}}_{\frac{\varepsilon}{2}} \right\} \subseteq \mathbb{S}^{D-1}.$$

Note that $|\mathcal{C}_\varepsilon| \leq |\tilde{\mathcal{C}}_{\frac{\varepsilon}{2}}| \leq (\frac{6}{\varepsilon})^D$. Let $\boldsymbol{w} \in \mathbb{S}^{D-1}$. Then, there exists $\tilde{\boldsymbol{w}} \in \tilde{\mathcal{C}}_{\frac{\varepsilon}{2}}$ such that

$$\|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \leq \frac{\varepsilon}{2}.$$

Then, we have

$$\left\| \boldsymbol{w} - \frac{\tilde{\boldsymbol{w}}}{\|\tilde{\boldsymbol{w}}\|_2} \right\|_2 \leq \|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 + \left\| \tilde{\boldsymbol{w}} - \frac{\tilde{\boldsymbol{w}}}{\|\tilde{\boldsymbol{w}}\|_2} \right\|_2 \leq \frac{\varepsilon}{2} + |1 - \|\tilde{\boldsymbol{w}}\|_2|.$$

Moreover, since $\tilde{\boldsymbol{w}} \in B(\mathbf{0}, 1 + \frac{\varepsilon}{2}) \setminus \operatorname{int} B(\mathbf{0}, 1 - \frac{\varepsilon}{2})$, we have

$$-\frac{\varepsilon}{2} \leq 1 - \|\tilde{\boldsymbol{w}}\|_2 \leq \frac{\varepsilon}{2},$$

that is, $|1-\|\tilde{\boldsymbol{w}}\|_2| \leq \frac{\varepsilon}{2}$. It follows that

$$\left\|\boldsymbol{w}-\frac{\tilde{\boldsymbol{w}}}{\|\tilde{\boldsymbol{w}}\|_2}\right\|_2 \leq \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

This completes the proof since $\frac{\tilde{\boldsymbol{w}}}{\|\tilde{\boldsymbol{w}}\|_2} \in \mathcal{C}_\varepsilon$. □

Now, we present a lemma showing our choice of confidence regions indeed gives at least $1-\delta$ confidence.

Lemma 15. *Using the definitions in* (S.25) *and* (S.26), *we have* $\mathbb{P}\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{E}_t(x)\} \geq 1-\delta$.

Proof. Since $\mathcal{E}_t(x) = \bigcap_{\tau=1}^{t} \mathcal{B}_\tau(x)$ for each $t \in [T]$, we have

$$\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{E}_t(x)\} = \{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{B}_t(x)\}.$$

So, instead, we prove $\mathbb{P}\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N}\colon \boldsymbol{f}(x) \in \mathcal{B}_t(x)\} \geq 1-\delta$.

Let $\mathcal{F}_t$ be the information at round t. The conditional distribution of $\boldsymbol{f}(x)$ given $\mathcal{F}_t$ is the Gaussian distribution with mean vector $\boldsymbol{\mu}_t(x)$ and covariance matrix $\boldsymbol{\Sigma}_t(x)$. Thus, for each $\boldsymbol{w} \in \mathbb{R}^D \setminus \{\boldsymbol{0}\}$, the conditional distribution of $\boldsymbol{w}^\mathsf{T}\boldsymbol{f}(x)$ is the Gaussian distribution with mean $\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}_t(x)$ and variance $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_t(x)\boldsymbol{w}$. In particular, applying Gaussian concentration gives for $\tilde{\alpha} > 0$

$$\mathbb{P}\left\{\boldsymbol{w}^\mathsf{T}\boldsymbol{f}(x) > \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}_t(x) + \sqrt{\tilde{\alpha}\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_t(x)\boldsymbol{w}}\right\} = \mathbb{P}\left\{\frac{\boldsymbol{w}^\mathsf{T}(\boldsymbol{f}(x)-\boldsymbol{\mu}_t(x))}{\sqrt{\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_t(x)\boldsymbol{w}}} > \sqrt{\tilde{\alpha}}\right\} \leq e^{-\frac{\tilde{\alpha}}{2}}. \tag{S.27}$$

Moreover, the ellipsoid $\mathcal{B}_t(x)$ is a closed convex set. Hence, an application of separation theorem in convex analysis yields

$$\mathbb{P}\{\boldsymbol{f}(x) \notin \mathcal{B}_t(x)\} = \mathbb{P}\left\{\exists \boldsymbol{w} \in \mathbb{S}^{D-1}\colon \boldsymbol{w}^\mathsf{T}\boldsymbol{f}(x) > \sup_{\boldsymbol{\mu}_x \in \mathcal{B}_t(x)} \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}_x\right\}.$$

Here, note that

$$\begin{aligned}\mathcal{B}_t(x) &= \{\boldsymbol{\mu}_x \in \mathbb{R}^D \mid (\boldsymbol{\mu}_x - \boldsymbol{\mu}_t(x))^\mathsf{T}(\alpha_t\boldsymbol{\Sigma}_t(x))^{-1}(\boldsymbol{\mu}_x - \boldsymbol{\mu}_t(x)) \leq 1\} \\ &= \boldsymbol{\mu}_t(x) + \sqrt{\alpha_t}\boldsymbol{\Sigma}_t^{1/2}(x)B(\boldsymbol{0},1)\ .\end{aligned}$$

Hence,

$$\sup_{\boldsymbol{\mu}_x \in \mathcal{B}_t(x)} \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}_x = \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}_t(x) + \sqrt{\alpha_t\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_t(x)\boldsymbol{w}}\ . \tag{S.28}$$

Using (S.28) we get

$$\mathbb{P}\{\boldsymbol{f}(x) \notin \mathcal{B}_t(x)\} = \mathbb{P}\left\{\exists \boldsymbol{w} \in \mathbb{S}^{D-1}\colon \boldsymbol{w}^\mathsf{T}\boldsymbol{f}(x) > \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}_t(x) + \sqrt{\alpha_t\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}_t(x)\boldsymbol{w}}\right\}. \tag{S.29}$$

Let $\varepsilon \in (0,2)$ and $\tilde{\alpha} > 0$. By Lemma 14, there exists a set $\mathcal{C}_\varepsilon \subseteq \mathbb{S}^{D-1}$ such that $|\mathcal{C}_\varepsilon| \leq (\frac{6}{\varepsilon})^D$ and

$$\forall \boldsymbol{w} \in \mathbb{S}^{D-1}\ \exists \tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon\colon \|\boldsymbol{w}-\tilde{\boldsymbol{w}}\|_2 \leq \varepsilon. \tag{S.30}$$

For each $\tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon$, we may take $\boldsymbol{w} = \boldsymbol{\Sigma}_t^{-1/2}(x)\tilde{\boldsymbol{w}}$ in (S.27), which gives

$$\mathbb{P}\left\{\tilde{\boldsymbol{w}}^\mathsf{T}\boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x)-\boldsymbol{\mu}_t(x)) > \sqrt{\tilde{\alpha}}\right\} \leq e^{-\frac{\tilde{\alpha}}{2}}$$

since $\|\tilde{\boldsymbol{w}}\|_2 = 1$. After applying a union bound, we obtain

$$\mathbb{P}\{\exists \tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon \colon \tilde{\boldsymbol{w}}^\top \boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)) > \sqrt{\tilde{\alpha}}\} \leq \left(\frac{6}{\varepsilon}\right)^D e^{-\frac{\tilde{\alpha}}{2}}. \tag{S.31}$$

It remains to make the connection between (S.31) and (S.29). To that end, let us introduce the notation $\|\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}} = \sqrt{\boldsymbol{\mu}^\top \boldsymbol{\Sigma} \boldsymbol{\mu}}$ for a $D \times D$ symmetric positive definite matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu} \in \mathbb{R}^D$. Note that $\|\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}} = \|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\mu}\|_2$. Also recall that the ℓ^2 norm has the variational characterization $\|\boldsymbol{\mu}\|_2 = \max_{\boldsymbol{w} \in \mathbb{S}^{D-1}} \boldsymbol{w}^\top \boldsymbol{\mu}$. Then, under the complement of the event in (S.31), we have

$$\begin{aligned}
\|\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)\|_{\boldsymbol{\Sigma}_t^{-1}(x)} &= \left\|\boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x))\right\|_2 \\
&= \max_{\boldsymbol{w} \in \mathbb{S}^{D-1}} \boldsymbol{w}^\top \boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)) \\
&= \max_{\boldsymbol{w} \in \mathbb{S}^{D-1}} \min_{\tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon} \left[(\boldsymbol{w} - \tilde{\boldsymbol{w}})^\top \boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)) + \tilde{\boldsymbol{w}}^\top \boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x))\right] \\
&\leq \sup_{\boldsymbol{w} \in \mathbb{S}^{D-1}} \min_{\tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon} \left[\|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \, \|\boldsymbol{\Sigma}_t^{-1/2}(x)(\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x))\|_2 + \sqrt{\tilde{\alpha}}\right] \\
&= \sup_{\boldsymbol{w} \in \mathbb{S}^{D-1}} \min_{\tilde{\boldsymbol{w}} \in \mathcal{C}_\varepsilon} \left[\|\boldsymbol{w} - \tilde{\boldsymbol{w}}\|_2 \, \|\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)\|_{\boldsymbol{\Sigma}_t^{-1}(x)} + \sqrt{\tilde{\alpha}}\right] \\
&\leq \varepsilon \|\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)\|_{\boldsymbol{\Sigma}_t^{-1}(x)} + \sqrt{\tilde{\alpha}},
\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz-Bunyakovski inequality and the definition of the event, the second inequality follows from the covering property in (S.30). Rearranging the terms gives

$$\|\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)\|_{\boldsymbol{\Sigma}_t^{-1}(x)} \leq \frac{\sqrt{\tilde{\alpha}}}{1 - \varepsilon}.$$

Note that

$$\boldsymbol{f}(x) \in \mathcal{B}_t(x) \;\Leftrightarrow\; \|\boldsymbol{f}(x) - \boldsymbol{\mu}_t(x)\|^2_{\boldsymbol{\Sigma}_t^{-1}(x)} \leq \alpha_t.$$

Hence, for a given probability level $\delta_t' \in (0,1)$, in order to ensure that $\mathbb{P}\{\boldsymbol{f} \notin \mathcal{B}_t(x)\} \leq \delta_t'$, it suffices to choose $\varepsilon, \tilde{\alpha}$ such that

$$\left(\frac{6}{\varepsilon}\right)^D e^{-\frac{\tilde{\alpha}}{2}} = \delta_t', \quad \frac{\sqrt{\tilde{\alpha}}}{1 - \varepsilon} \leq \sqrt{\alpha_t}.$$

Hence, we take

$$\tilde{\alpha} = 2\log\left(\frac{\left(\frac{6}{\varepsilon}\right)^D}{\delta_t'}\right) = 2D\log\left(\frac{6}{\epsilon}\right) + 2\log\left(\frac{1}{\delta_t'}\right).$$

Then, choosing $\varepsilon = \frac{1}{2}$ imposes the following condition on α_t:

$$\frac{\sqrt{\tilde{\alpha}}}{1 - \varepsilon} = 2\sqrt{2(D\log(12) + \log(1/\delta_t'))} \leq \sqrt{\alpha_t} \;\Leftrightarrow\; 8D\log(12) + 4\log(1/\delta_t') \leq \alpha_t.$$

for every $\delta_t' \in (0,1)$. Let us set $\delta_t' \coloneqq \frac{6\delta}{\pi^2 t^2 |\mathcal{X}|}$. Then, a union bound gives

$$\mathbb{P}\{\forall x \in \mathcal{X}, \forall t \in \mathbb{N} \colon \boldsymbol{f}(x) \in \mathcal{B}_t(x)\} \geq 1 - \sum_{x \in \mathcal{X}} \sum_{t \in \mathbb{N}} \frac{6\delta}{\pi^2 t^2 |\mathcal{X}|} = 1 - \sum_{x \in \mathcal{X}} \frac{\delta}{|\mathcal{X}|} \geq 1 - \delta,$$

where $\mathcal{B}_t(x) = \{\boldsymbol{\nu} : (\boldsymbol{\nu} - \boldsymbol{\mu}_t(x))^\top \boldsymbol{\Sigma}_t^{-1}(x)(\boldsymbol{\nu} - \boldsymbol{\mu}_t(x)) \leq \alpha_t\}$ where $\alpha_t = 8D\log(12) + 4\log(\frac{\pi^2 t^2 |\mathcal{X}|}{6\delta})$.

This finishes the proof. □

5 FURTHER DETAILS OF EXPERIMENTS

5.1 Libraries

We use CVXPY (Diamond and Boyd, 2016; Agrawal et al., 2018) for solving convex optimization problems. For Gaussian Process modeling, we use GPyTorch (Gardner et al., 2018).

5.2 Heuristic Variants

PaVeBa-IH: We use the batch-independent multi-output GP formulation from GPyTorch to implement PaVeBa-IH. This approach involves having an independent GP for each reward dimension with a likelihood that allows for a correlated noise structure, though we only use the global noise and assume i.i.d. noise for each reward dimension. We then employ hyperrectangles that encompass the confidence ball in (S.25), with α_t scaled down by a contraction factor of 64.

PaVeBa-DE: We use the multitask GP formulation from GPyTorch for modeling reward dimensions with cross-correlations. It utilizes a formulation equivalent to the Linear Model of Coregionalization. We apply it with full-rank inter-task covariance. Given the cross-task covariance, we directly employ the hyperellipsoidal region that the GP posterior provides in (S.25), with α_t scaled down by a contraction factor of 64.

For batch selection in both variants, we do the following: we select the arm with maximum posterior covariance trace, we update all posterior covariances of active arms as if we played the selected arm (*fantasy update*), and loop until K arms are selected. We can do fantasy updates since variance update in GP formulation does not depend on observed reward (Contal et al., 2013).

5.3 Real-World Problems

Since we work in a finite arm setting, all datasets are generated with Sobol samples taken from the domain of the problem (except for SNW, which is already finite). We used the implementations available in the library Botorch (Balandat et al., 2020) for DB and VC. We used implementations provided with Tu et al. (2022) for PK2 and MAR.

SNW ($D = 2$, $|\mathcal{X}| = 206$): This dataset is derived from the domain of computational hardware design, specifically concerning the optimization of sorting network configurations. The reward vector reflects the trade-off between the network's throughput (speed of sorting) and the physical area required by the synthesized hardware, crucial factors in efficient hardware design (Zuluaga et al., 2012).

DB (Disc Brake, $D = 2$, $|\mathcal{X}| = 128$): This dataset addresses efficiency and safety in automotive engineering, presenting an optimization problem in disc brake manufacturing. The reward vector accounts for the brake's mass (affecting vehicle efficiency and brake performance) and the vehicle's stopping time (critical for safety). The original problem formulated in Tanabe and Ishibuchi (2020) also includes constraints that can be integrated as a third objective, but we work with two objectives instead.

PK2 ($D = 2$, $|\mathcal{X}| = 500$): In the context of organic chemistry, this dataset aims at the optimization of the Paal-Knorr synthesis, a fundamental reaction for the synthesis of pyrroles and pyrrolidines. The reward vector considers the yield of the pyrrolidine ring and the space-time yield, representing the efficiency of this organic synthesis process (Moore and Jensen, 2012). The implementation follows `https://github.com/adamc1994/MultiChem`.

VC (Vehicle Crashworthiness, $D = 3$, $|\mathcal{X}| = 2000$): From the field of automotive safety, this dataset focuses on the optimization of vehicle structures to enhance crashworthiness. The reward vector integrates factors like weight (impacting vehicle performance and fuel efficiency), acceleration characteristics (indicating potential safety performance), and toe-board intrusion levels (measuring the deformation of the vehicle structure in crashes) (Tanabe and Ishibuchi, 2020; Liao et al., 2008).

VC1 (Vehicle Crashworthiness, $D = 3$, $|\mathcal{X}| = 100$): Smaller version of VC dataset.

MAR (Marine, $D = 4$, $|\mathcal{X}| = 500$): Within maritime engineering, this dataset is concerned with the optimization of bulk carrier designs to improve cargo transfer efficiency and maritime safety. The reward vector assesses a range of factors, including transportation cost (vital for economic efficiency), weight (related to fuel consumption and vessel stability), annual cargo capacity (determining operational efficiency), and compliance with design constraints (ensuring safety and regulatory adherence) (Parsons and Scott, 2004; Tanabe and Ishibuchi, 2020).

5.4 Implementations of Other Methods

NE: We use the implementation of naïve elimination provided in Ararat and Tekin (2023).

PIBF: Since there are no implementations available that we could find, we implemented the algorithm with great care to Auer et al. (2016). Since we use PaVeBa's theoretical confidence regions to compare with PIBF, we also implement the theoretical confidence regions for PIBF. We then employ *contraction factor*s that scales down PaVeBa's $r_t(x)$ and PIBF's β_i's.

ϵ-PAL: We implement ϵ-PAL (Zuluaga et al., 2016) in Python since the published code for it is in MATLAB. We use the paper and the original MATLAB code for guidance while implementing. We scale down β_t by 9 for ϵ-PAL, as in Zuluaga et al. (2016).

MESMO and JES: Since both MESMO (Belakaria et al., 2019) and JES (Tu et al., 2022) operate in the continuous domain, they are at a disadvantage in finding the ϵ-Pareto set in finite arm setting. So, to ensure fairness, we modify them by calculating their acquisition functions only for the available arms and calculating the Pareto fronts from their final posterior means of arms. Moreover, we tweak the original code of MESMO to accommodate a noise parameter.

Cone ordering for other methods: To accommodate cones C that are different than the multi-objective cone $C = \mathbb{R}^D_+$ in the experiments that include other GP-based algorithms, i.e., ϵ-PAL, MESMO and JES, we run the algorithms, then calculate the Pareto front using the cone ordering from the final posterior means of arms.

5.5 Performance Metrics

ϵ-F1 Score: Given an input space $\mathcal{X}$ and parameter ϵ, define the positive arms set Π_ϵ such that

$$\Pi_\epsilon = \{x \in \mathcal{X} : \Delta^*(x) \leq \epsilon\},$$

where $\Delta^*(x)$ is defined in Definition 1. Let an algorithm for (ϵ, δ)-PAC Pareto set identification return $\hat{P}$ as the predicted Pareto set. Then, the ϵ-F1 Score for that algorithm is defined as

$$\epsilon\text{-F1} = \frac{2|\Pi_\epsilon \cap \hat{P}|}{2|\Pi_\epsilon \cap \hat{P}| + |\Pi_{\epsilon \setminus \hat{P}}| + |\hat{P} \setminus \Pi_\epsilon|},$$

where $\Pi_{\epsilon \setminus \hat{P}}$ is the set of Pareto optimal arms that is not covered by $\hat{P}$ where covering is defined in the part (i) of Definition 4.6 of Ararat and Tekin (2023).

ϵ-F1 Score is a loose F1-Score designed for ϵ approximate identification, where the Pareto identification problem is seen as a classification problem with a positive class of ϵ-accurate Pareto optimal arms. We note that

- $P^* \subseteq \Pi_\epsilon$.
- If an algorithm satisfies success condition (i) in Definition 1 ($P^* \subseteq \hat{P}$), then we have $\Pi_{\epsilon \setminus \hat{P}} = \emptyset$. If an algorithm does not satisfy success condition (i) in Definition 1 but still if all Pareto optimal arms are ϵ-covered (success condition (i) in Definition 4.6 of Ararat and Tekin (2023)), then we have $\Pi_{\epsilon \setminus \hat{P}} = \emptyset$.
- If an algorithm satisfies success condition (ii) in Definition 1, then we have $\hat{P} \setminus \Pi_\epsilon = \emptyset$.
- ϵ-F1 $= 1$ iff an algorithm satisfies Definition 4.6 of Ararat and Tekin (2023).

We chose this as our main accuracy metric to follow the literature on classification tasks. We always compute the ϵ-F1 score of the algorithms with the same ϵ fed into the algorithm for (ϵ, δ)-PAC Pareto identification.

Sample Complexity: Given a Pareto front identification algorithm, this is simply the number of evaluations the algorithm performs.

5.6 Further Results

Here, we provide error bars and some new results for the experiments in the main paper. Tables 5, 6, 7, 8 correspond to Tables 1, 2, 3, 4 respectively.

Experiment 2: We add two new cone angles $\theta = \pi/3$ and $2\pi/3$ and the standard deviations to Tables 1 and 2 in the main paper, resulting in Tables 5 and 6.

Experiment 3: We add standard deviations to Tables 3 and 4 to get Tables 7 and 8, respectively.

		ϵ-F1Came Score w.r.t. θ				
	ϵ	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	$3\pi/4$
NE	10^{-1}	**0.99 ± .01**	0.99 ± .01	0.97 ± .03	0.96 ± .04	0.97 ± .04
	10^{-2}	0.93 ± .02	0.91 ± .03	0.87 ± .04	0.82 ± .07	0.78 ± .10
PaVeBa	10^{-1}	**0.99 ± .01**	**1.00 ± .00**	**0.98 ± .01**	**0.99 ± .02**	**0.98 ± .02**
	10^{-2}	**0.94 ± .02**	**0.93 ± .03**	**0.91 ± .04**	**0.88 ± .05**	**0.88 ± .08**

(a)

NE	10^{-1}	**0.98 ± .02**	0.96 ± .04	0.94 ± .06	**0.95 ± .06**	**0.97 ± .06**
	10^{-2}	0.93 ± .04	0.89 ± .05	0.84 ± .09	0.93 ± .07	0.78 ± .17
PaVeBa	10^{-1}	**0.98 ± .02**	**0.97 ± .03**	**0.95 ± .05**	**0.95 ± .07**	**0.97 ± .06**
	10^{-2}	**0.94 ± .03**	**0.94 ± .04**	**0.92 ± .07**	**0.94 ± .08**	**0.91 ± .14**

(b)

Table 5: Results of NE and PaVeBa for different values of ϵ and θ on SNW (a) and DB (b) datasets. The best results are in bold.

		Sample Complexity w.r.t. θ				
	ϵ	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	$3\pi/4$
SNW	10^{-1}	654.50 ± 50.93	499.12 ± 42.35	378.68 ± 28.15	306.88 ± 21.73	272.44 ± 17.22
	10^{-2}	10100.92 ± 3315.00	6053.92 ± 1479.97	2594.00 ± 982.45	1162.86 ± 683.95	789.96 ± 442.67
DB	10^{-1}	304.86 ± 38.71	218.76 ± 21.11	167.58 ± 10.44	145.00 ± 9.69	141.78 ± 7.31
	10^{-2}	2045.26 ± 1915.25	780.00 ± 561.32	308.84 ± 163.49	473.82 ± 603.64	196.70 ± 67.16

Table 6: Sample complexities of PaVeBa with different values of ϵ and θ on SNW and DB datasets.

Dataset	$\|\mathcal{X}\|$	Algorithm	Sample Complexity	ϵ-F1 Score
PK2	500	ϵ-PAL	178.4 ± 55.9	1.00 ± .01
		JES	57.62 ± 8.12	0.86 ± .06
		MESMO	57.62 ± 8.12	0.86 ± .05
		PaVeBa-IH	57.62 ± 8.12	0.95 ± .08
VC	2000	ϵ-PAL	584.0 ± 61.0	1.00 ± .00
		JES	123.94 ± 14.80	0.87 ± .04
		MESMO	123.94 ± 14.80	0.97 ± .02
		PaVeBa-IH	123.94 ± 14.80	0.97 ± .03
MAR	500	ϵ-PAL	639.04 ± 152.63	0.98 ± .01
		JES	224.38 ± 11.76	0.97 ± .01
		MESMO	224.38 ± 11.76	0.95 ± .03
		PaVeBa-IH	224.38 ± 11.76	0.97 ± .01

Table 7: Comparison of PaVeBa with ϵ-PAL, MESMO, and JES under the multi-objective cone.

Cone	Algorithm	Sample Complexity	ϵ-F1 Score
Acute	ϵ-PAL	88.08 ± 28.06	0.98 ± .02
	JES	91.94 ± 15.33	0.81 ± .04
	MESMO	91.94 ± 15.33	0.97 ± .02
	PaVeBa-DE	91.94 ± 15.33	0.99 ± .02
Obtuse	ϵ-PAL	88.08 ± 28.06	1.00 ± .00
	JES	27.58 ± 5.68	0.86 ± .07
	MESMO	27.58 ± 5.68	0.85 ± .11
	PaVeBa-DE	27.58 ± 5.68	1.00 ± .03

Table 8: Comparison of PaVeBa with ϵ-PAL, MESMO, and JES on VC1 dataset under acute and obtuse cones.

References

Agrawal, A., Verschueren, R., Diamond, S., and Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60.

Antos, A., Grover, V., and Szepesvári, C. (2010). Active learning in heteroscedastic noise. *Theoretical Computer Science*, 411(29-30):2712–2728.

Ararat, Ç. and Tekin, C. (2023). Vector optimization with stochastic bandit feedback. In *Proc. 26th International Conference on Artificial Intelligence and Statistics*, pages 2165–2190.

Auer, P., Chiang, C.-K., Ortner, R., and Drugan, M. (2016). Pareto front identification from stochastic bandit feedback. In *Proc. 19th International Conference on Artificial Intelligence and Statistics*, pages 939–947.

Balandat, M., Karrer, B., Jiang, D. R., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E. (2020). Botorch: A framework for efficient Monte-Carlo Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, pages 21524–21538.

Belakaria, S., Deshwal, A., and Doppa, J. R. (2019). Max-value entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 32.

Contal, E., Buffoni, D., Robicquet, A., and Vayatis, N. (2013). Parallel Gaussian process optimization with upper confidence bound and pure exploration. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 225–240.

Diamond, S. and Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5.

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., and Wilson, A. G. (2018). GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems*, volume 31.

Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., and Jordan, M. I. (2019). A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736*.

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

Liao, X., Li, Q., Yang, X., Zhang, W., and Li, W. (2008). Multiobjective optimization for crash safety design of vehicles using stepwise regression model. *Structural and Multidisciplinary Optimization*, 35:561–569.

Moore, J. S. and Jensen, K. F. (2012). Automated multitrajectory method for reaction optimization in a microfluidic system using online IR analysis. *Organic Process Research & Development*, 16(8):1409–1415.

Parsons, M. G. and Scott, R. L. (2004). Formulation of multicriterion design optimization problems for solution with scalar numerical optimization methods. *Journal of Ship Research*, 48(01):61–76.

Tanabe, R. and Ishibuchi, H. (2020). An easy-to-use real-world multi-objective optimization problem suite. *Applied Soft Computing*, 89:106078.

Tu, B., Gandy, A., Kantas, N., and Shafei, B. (2022). Joint entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 35, pages 9922–9938.

Zuluaga, M., Krause, A., and Püschel, M. (2016). ε-PAL: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research*, 17(104):1–32.

Zuluaga, M., Milder, P., and Püschel, M. (2012). Computer generation of streaming sorting networks. In *DAC Design Automation Conference*, pages 1241–1249.