Université de Lille, faculté des Sciences et Technologies

École Doctorale MADIS

Efficient Multi-Objective Pure Exploration: Pareto Set Identification and Applications to Clinical Trials

Exploration Pure à Objectifs Multiples : Identification du Front de Pareto et Applications aux Essais Cliniques de Phase Précoce

Présentée par

CYRILLE KONE

Thèse de Doctorat

Spécialité **Informatique**

sous la direction d'**Émilie Kaufmann** et de **Laura Richert**

Soutenue publiquement à Villeneuve d'Ascq, le 09/12/2025 devant un jury composé de

M. Vianney **Perchet**	Professeur, CREST/Ensae Paris	Rapporteur
M. Wouter M. **Koolen**	Professeur, CWI	Rapporteur
M. Antoine **Chambaz**	Professeur, Université Paris Cité	Président du jury
M. Peter **Auer**	Professeur, University of Leoben	Examinateur
M. Clayton **Scott**	Professeur, University of Michigan	Examinateur
M^{me} Sarah **Zohar**	Directrice de recherche, INSERM	Invitée
M^{me} Émilie **Kaufmann**	Chargée de recherche, CNRS	Directrice de thèse
M^{me} Laura **Richert**	Professeure, Université de Bordeaux	Co-directrice de thèse

Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL),
UMR 9189 Équipe Scool, 59650, Villeneuve d'Ascq, France

Remerciements

I would like to express my sincere and profound gratitude to my advisors, Émilie and Laura, for their guidance throughout my doctoral studies. Their intellectual rigour, availability, and steady support have had a decisive influence on this work. I am particularly grateful for their high standards, careful feedback, and continued encouragement, which were essential to both the direction and the completion of this thesis.

À toi, Émilie, merci pour ta générosité et pour la confiance que tu m'as accordée. Travailler à tes côtés au quotidien a été un vrai plaisir. Au fil des années, j'ai énormément appris auprès de toi, tant sur le plan scientifique que dans la manière de faire de la recherche avec rigueur. Ta maîtrise des sujets, ainsi que ton aisance à les rendre accessibles, m'ont permis de développer une compréhension plus profonde et une exigence accrue dans mon travail. Merci aussi pour ta bienveillance, tes conseils, et pour l'autonomie que tu m'as progressivement donnée.

À toi, Laura, merci pour tous ces séjours à Bordeaux et pour ton accueil toujours chaleureux. Tu as su rendre accessibles des notions clés liées aux essais cliniques, tout en nous partageant avec clarté les défis concrets qui ont nourri nos problématiques en bandits. Je te suis très reconnaissant pour cette expérience pluridisciplinaire particulièrement enrichissante, et pour la qualité de nos échanges, à la fois exigeants, ouverts et stimulants. De surcroît, vous avez été toutes deux d'un immense apport dans l'élaboration de mon projet post-thèse. Merci pour votre implication constante, vos conseils, et votre soutien dans cette étape importante. Tout cela montre que vous êtes non seulement d'excellentes scientifiques, mais aussi de superbes encadrantes.

I would also like to thank the reviewers of this thesis, Prof. Vianney Perchet and Prof. Wouter Koolen for the time they devoted to evaluating my work. I am grateful for their careful reading, insightful comments, and constructive suggestions, which helped improve both the manuscript and its presentation. I also thank all the members of the examination committee for their participation and for the stimulating discussions.

I am grateful to my colleagues and friends at Inria, whose presence and collaboration made this PhD an enriching scientific and human experience: Hector, Brahim, Waris, Marc, Achraf, Riad, Timothée, Adrienne, Sabrine, Rémy, Tuan, Yann, Mahdi, Udvas, Thomas, Sumit, Fabien, Patrick, Dorian, Matheus, Ayoub, Anthony, Deb, Julien, Amélie, Odalric, Naafi, Alena, Hernan, Riccardo, Mohamed Yacine, Reda, Nathan, Redouane, Nicolas, Hadrien, Adrien, Juliette, Lorenzo, Hatim, Guillaume, Tanguy, Brell, Mericel, Mickaël, and Philippe. I thank them for the many discussions, their support, and the constructive working environment that accompanied this research. Beyond the scientific exchanges, I particularly appreciated the Scool team's atmosphere: a rare combination of intellectual rigour, generosity, and genuine camaraderie. Thank you for the unforgettable pétanque parties, badminton sessions, and football matches, which provided welcome opportunities to unwind and recharge during these years. I feel fortunate to have been surrounded by such talented researchers and colleagues; your curiosity, energy, and professionalism have contributed greatly to my development throughout this journey. Finally, I would like to thank Philippe, as team

director, for fostering and sustaining this exceptional environment.

I would like to offer special thanks to Riad, Brahim, Waris, Hector, and Tuan for the many moments shared outside the laboratory, including our culinary explorations in Lille, which remain among my most memorable experiences during these years.

Je suis également reconnaissant aux membres de l'équipe SISTM à Bordeaux. Merci en particulier à Rodolphe pour son implication constante et la qualité de nos échanges lors de mes séjours à Bordeaux. Merci aussi à Quentin, Mélanie, Boris, Myrtille et Linda pour ces discussions stimulantes et enrichissantes autour des problématiques liées aux essais cliniques, qui ont nourri ma compréhension de ces sujets et donné une profondeur supplémentaire à ce travail.

I would also like to thank Arnab, Nicolas and Kevin at the University of Washington for the many fruitful exchanges during my visit to Seattle.

À mes amis de longue date — Abdoulaye, Ibrahim, Paul, Arthur, Milacet, Christian et Simplice — je vous remercie chaleureusement pour votre présence constante, vos encouragements et votre soutien. Je repense avec gratitude à tous les beaux moments qui ont ponctué ces trois années de thèse: vos messages, nos discussions, les retrouvailles, les rires, et cette manière si naturelle que vous avez de remettre les choses en perspective. Votre amitié m'a apporté stabilité et réconfort dans les périodes exigeantes, et a rendu ce parcours bien plus léger et plus joyeux.

À ma famille, je tiens à exprimer ma profonde reconnaissance, en particulier envers mes frères et sœurs — Julie, Estelle, Anicet — pour leur affection, leur patience et leurs encouragements tout au long de ce parcours.

Je souhaite adresser un remerciement tout particulier à mon oncle Nanontia Koné, dont le soutien indéfectible, depuis mon enfance, a été déterminant. Cette thèse n'aurait pas été possible sans sa générosité, sa confiance et sa présence constante. C'est grâce à l'ordinateur familial qu'il a acquis en 2008 que j'ai découvert l'informatique ; depuis, mon attachement à cette discipline n'a jamais cessé de grandir. Je lui en suis infiniment reconnaissant.

Enfin, je souhaite rendre hommage à mes parents, trop tôt disparus, Émilie Touré et Koné Niéna. Leur mémoire a été, durant toutes ces années, une source de motivation, et je leur dédie ce travail. Je ne pourrais ouvrir ce manuscrit sans rappeler combien je le dois à ma mère. Elle m'a transmis très tôt le goût des études et la fierté d'apprendre. Je me souviens avec émotion de la joie qui était la sienne lorsque je rapportais mes premiers résultats à l'école primaire ; c'est le souvenir de cette joie qui m'a porté tout au long de ce parcours, dans les moments d'élan comme dans les périodes de doute. Et malgré mille détours, les épreuves et les turbulences, je garde l'espoir que ce travail soit, à sa manière, le prolongement de ce que vous avez semé en moi.

Abstract

In this dissertation, we study a class of *multi-objective pure exploration problems* in stochastic bandits. Our goal is to design algorithms that identify the set of *Pareto-optimal* arms—those that cannot be improved in one objective without deteriorating another—while minimizing the number of samples required to achieve a prescribed level of confidence.

We begin by reviewing the literature on classical and multi-objective bandits, outlining the key theoretical tools that motivate the study of *Pareto Set Identification* (PSI). Throughout the thesis, we progressively move from simple algorithms based on confidence sequences with finite-time guarantees to more advanced approaches that achieve asymptotic optimality, while maintaining computational tractability. All proposed methods are systematically evaluated on synthetic and real-world–inspired datasets to provide a comprehensive understanding of their empirical performance.

In the fixed-budget regime, we develop EGE (*Empirical Gap Elimination*), the first algorithm for PSI in this setting. In the fixed-confidence regime, we introduce APE, a δ-correct procedure based on adaptive confidence intervals. Applications to multi-endpoint vaccine trials illustrate how PSI can guide early-stage clinical decision-making by identifying treatment strategies that jointly optimize multiple immunological indicators.

Chapter 4 extends PSI to the *structured* case, where arm means are linearly related through an unknown parameter matrix θ and known feature representations. We derive problem-dependent lower bounds on the sample complexity of adaptive algorithms and we propose efficient algorithms that exploit this structure to significantly reduce the number of required samples.

Chapter 5 focuses on algorithms that achieve the *information-theoretic lower bound*. We introduce PSIPS (*Pareto Set Identification with Posterior Sampling*), the first computationally efficient and asymptotically optimal algorithm for PSI with correlated objectives. PSIPS uses posterior sampling for both the sampling and stopping rules, eliminating the need to predefine the confidence parameter δ as a parameter of the sampling rule. It achieves theoretical optimality and demonstrates excellent empirical performance in extensive experiments, including on the *COV-BOOST* vaccine dataset, while being several orders of magnitude faster than gradient-based approaches.

Finally, Chapter 6 explores PSI under *linear constraints*, where the goal is to identify the set of Pareto-optimal arms among those that satisfy feasibility conditions. We propose e-cAPE, an extension of APE, and characterize the information-theoretic complexity of constrained PSI.

Overall, this work advances the theoretical and algorithmic foundations of *multi-objective pure exploration*. By bridging sequential decision-making, information theory, optimal experimental design, and Bayesian principles, it demonstrates that one can achieve both theoretical optimality and computational efficiency. Beyond bandits, the insights and methods developed here open promising avenues toward a general theory of *multi-objective sequential learning*, with applications to adaptive clinical trials, reinforcement learning, and multi-criteria decision-making under uncertainty.

Keywords: Multi-armed bandits, Pure exploration, Multi-objective, Pareto, Sequential learning, Clinical trials

Résumé

Cette thèse porte sur l'exploration pure à objectifs multiples dans le cadre des bandits stochastiques. L'objectif est de concevoir des algorithmes capables d'identifier l'ensemble des bras Pareto-optimaux; ceux qu'il est impossible d'améliorer sur un objectif sans en détériorer un autre, tout en minimisant le nombre d'échantillons nécessaires pour atteindre un niveau de confiance donné.

Après une revue de la littérature sur les bandits classiques et multiobjectifs, nous introduisons le problème d'Identification de l'Ensemble de Pareto (IEP) et les outils théoriques qui le sous-tendent. Nous progressons des méthodes basées sur des intervalles de confiance offrant des garanties en temps fini vers des approches asymptotiquement optimales, tout en préservant l'efficacité computationnelle. Les algorithmes proposés sont évalués de manière systématique sur des jeux de données synthétiques et réels, afin d'analyser finement leurs performances empiriques.

La première partie présente l'algorithme EGE, qui identifie efficacement l'ensemble de Pareto lorsque le budget d'échantillons est fixé. Revisitant les travaux sur l'IEP à confiance fixée, nous introduisons APE, une procédure adaptative δ-correcte reposant sur des intervalles de confiance. APE fournit de fortes garanties en nombre fini d'échantillons et de très bonnes performances pratiques. Des applications à des jeux de données issus d'essais vaccinaux multicritères illustrent comment EGE et APE peuvent guider la sélection de stratégies thérapeutiques optimisant simultanément plusieurs marqueurs immunologiques.

Nous étendons ensuite l'IEP au cas structuré, où les moyennes des bras dépendent linéairement de descripteurs connus via une matrice de régression inconnue. L'algorithme GEGE exploite cette structure pour réduire la complexité d'échantillonnage et atteint des bornes inférieures spécifiques à l'instance de bandit.

Le chapitre suivant identifie les limites de ces approches et introduit PSIPS, premier algorithme à la fois asymptotiquement optimal et à faible coût de calcul pour l'IEP avec objectifs corrélés. Basé sur un échantillonnage a posteriori pour le choix des bras et la règle d'arrêt, PSIPS élimine la nécessité de fixer à l'avance le paramètre de confiance δ. Il atteint la borne inférieure théorique sur la complexité d'échantillonnage et se montre très compétitif, y compris sur des données réelles comme COV-BOOST, tout en étant bien plus rapide que les méthodes à base de gradient.

Enfin, le dernier chapitre traite du cas contraint, où seuls les bras respectant certains critères doivent être considérés pour le calcul de l'ensemble de Pareto. Un algorithme doit donc apprendre simultanément les bras viables et déterminer leur optimalité. Nous proposons e-cAPE, un algorithme qui généralise celui du Chapitre 3 et établit des bornes théoriques pour l'IEP sous contraintes.

Dans l'ensemble, cette thèse fait progresser les fondements théoriques et algorithmiques de l'exploration pure à objectifs multiples. En combinant les outils issus de statistiques, apprentissage séquentiel, conception expérimentale optimale et principes bayésiens, elle montre qu'il est possible de concilier optimalité théorique et efficacité pratique. Les idées développées ouvrent la voie à une théorie générale de l'apprentissage séquentiel à objec-

tifs multiples, avec des applications aux essais cliniques adaptatifs, à l'apprentissage par renforcement et plus globalement à l'apprentissage multicritère sous incertitude.

Mots-clés : Bandits manchots, Exploration pure, Multi-objectif, Pareto, Apprentissage séquentiel, Essais cliniques

Table of Contents

Chapter 1

Introduction

This doctoral research was conducted between November 2022 and October 2025 within the Scool team at Inria Lille, hosted by Inria and the CRIStAL Computer Science Laboratory of the University of Lille. It was carried out under the supervision of Dr. Émilie Kaufmann (CNRS/Inria Lille) and Prof. Laura Richert (Inserm Bordeaux / University of Bordeaux), thanks to joint Inria–Inserm funding. As part of this collaboration, I made regular visits to Prof. Richert's group in Bordeaux, which provided valuable opportunities for interdisciplinary exchanges between machine learning and clinical-trial design.

The manuscript is organized as follows. Each chapter studies a particular aspect of Pareto set identification, beginning with confidence-based algorithms offering finite-time guarantees, and gradually extending to structured and constrained formulations. The work culminates in the design of asymptotically optimal algorithms based on posterior sampling, combining theoretical guarantees with computational efficiency. Finally, the empirical sections validate these approaches on both synthetic and real-world-inspired datasets, and the thesis concludes with perspectives on future research in multi-objective pure exploration.

1.1 Bandit Algorithms for Early-Stage Clinical Trials

Vaccination remains one of the most effective strategies to prevent and control infectious diseases. However, the clinical development of vaccines—like that of most experimental drugs—remains a long, costly, and uncertain process. The accelerated development of first-generation SARS-CoV-2 vaccines during the COVID-19 pandemic was an exceptional case of global mobilization and investment; in most contexts, vaccine development proceeds under much tighter resource and time constraints. At the beginning of a vaccine's clinical development plan, early to intermediate trials (phases I and II) are conducted to evaluate safety and explore immunogenicity in a limited number of participants. These early-stage trials play a crucial role: they must identify, among a potentially large set of vaccine candidates, those most promising for further efficacy testing in large phase IIb/III studies, which are extremely costly and logistically demanding.

From clinical motivation to bandit models. Adaptive clinical trial designs are one of the most promising approaches to accelerate and optimize vaccine development, as they offer a higher degree of adaptivity and statistical efficiency than traditional methods. The stochastic multi-armed bandit (MAB) problem was introduced as a stylized abstraction of an *adaptive clinical trial* (Thompson 1933; H. E. Robbins 1952), where a physician must sequentially allocate patients to competing treatments while learning from observed outcomes. Each treatment, or *arm*, may correspond, for instance, to a particular combination of dose, adjuvant, injection interval, or delivery platform. Patient responses are modeled as independent and identically distributed (*i.i.d.*) samples from the arm received by the patient.

Upon observing the responses of treated participants, the algorithm adaptively updates its allocation strategy, typically to balance exploration of uncertain options with exploitation of the most promising ones. This abstraction directly mirrors adaptive trial design: the aim is to identify promising arms as efficiently as possible, while respecting sample size and ethical constraints.

Initially, this abstraction was developed for *phase III* clinical trials, where treatments are tested for efficacy on cohorts of patients, often including patients affected by the target disease. In such settings, ethical considerations are paramount: since enrolled patients may derive therapeutic benefit from the experimental intervention, one seeks to minimize the number of individuals exposed to suboptimal treatments. This perspective aligns naturally with the *regret-minimization* formulation in bandit theory (H. E. Robbins 1952; T. Lai & H. Robbins 1985; Auer 2003; Garivier & Cappé 2011; Agrawal & Goyal 2012; Kaufmann, Korda, et al. 2012; Lattimore & Szepesvari 2020; Russo et al. 2020), where the goal is to balance learning (reducing uncertainty about treatment effects) and earning (maximizing the number of patients cured).

Early-stage vaccine development. In contrast, *early-phase trials* (typically phases I and II) pursue a different goal: rather than maximizing immediate patient outcomes, they aim to

Figure 1.1: Illustration of the multi-arm trial model

collect enough evidence to assess the safety, tolerability, or immunogenicity of candidate vaccines or treatments, often on healthy volunteers. Maximizing cumulative outcomes during the trial is secondary, and the focus shifts to accurate final recommendations. These trials involve small cohorts and limited budgets.

Adaptive allocation is particularly useful in this setting as it allows learning from early immunological responses to guide subsequent patient assignments, improving both statistical efficiency and ethical acceptability. This dissertation focuses on this side of the problem, which corresponds to the *pure exploration* setting in the literature on multi-armed bandits (Bubeck, Munos, et al. 2009; Degenne 2019). Here, treatments are allocated to maximize information gain rather than patient outcomes, and success is measured by the probability of correctly identifying effective treatments at the end of the trial.

Moreover, vaccine strategies for novel pathogens often yield a large number of experimental arms: multiple formulations, dose levels, injection intervals, booster strategies, and adjuvant combinations may all be evaluated concurrently. The COVID-19 booster-dose study (Munro et al. 2021), for example, simultaneously compared several booster vaccines across hundreds of participants—a design that naturally resembles a large-scale multi-armed bandit problem.

The multi-objective challenge. Most adaptive algorithms and trial designs treat treatment quality as one-dimensional. However, at the early stages of development, no validated *correlate of protection*, that is a single predictive immune marker of efficacy, is typically available (Plotkin & Gilbert 2012). Instead, several immunological readouts are monitored simultaneously (*e.g.*, antibody titers, neutralizing activity, T-cell response, and reactogenicity), and these may not align. In such settings, the notion of a single "best" treatment is ill-defined: one candidate may induce stronger antibodies but higher reactogenicity (Mark C et al. 2013), while another elicits broader T-cell responses but weaker neutralizing titers (Munro et al. 2021). Such trade-offs imply that no single vaccine candidate dominates all others. Instead, the goal we pursue in this dissertation is to identify the subset of *non-dominated* treatments forming the *Pareto front* or *Pareto set*: those that are not strictly worse across all measured criteria.

Formally, if $\mu_i \coloneqq (\mu_i^1, \ldots, \mu_i^d)$ denotes the mean response vector for arm i, we say that arm i is *dominated* by j (denoted $i \preceq j$ or $\mu_i \preceq \mu_j$) if $\mu_i^c \leqslant \mu_j^c$ for all $c \in \{1, 2, \ldots, d\}$, with strict inequality in at least one dimension. The goal is then to identify the set of treatments that are not dominated by any other: the *Pareto-optimal set*.

The COVID-19 example. For instance, in the early stages of COVID-19 vaccine development, several immune correlates—such as binding antibody titers, neutralizing activity, and T-cell responses—were simultaneously monitored to characterize immunogenicity before a definitive correlate of protection was established (Munro et al. 2021). The study of Munro et al. 2021 simultaneously evaluated multiple COVID-19 booster strategies, while measuring several immunological endpoints. In such cases, each vaccine candidate is associated with multiple response dimensions, and no single metric fully captures efficacy. Such designs naturally give rise to many arms (representing distinct vaccine strategies), a limited number of participants, and multiple partially conflicting measures of efficacy. This example illustrates precisely the kind of design challenges that motivate this work.

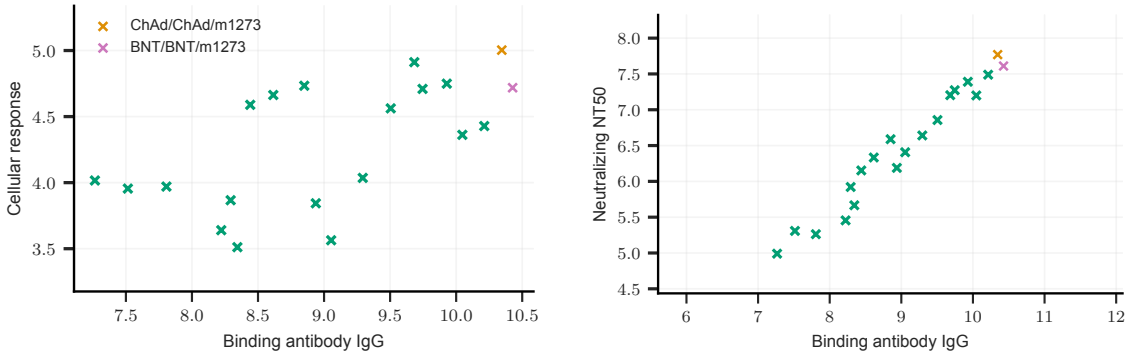


Figure 1.2: Empirical (arithmetic) average of the log-transformed immune response for three immunogenicity indicators reported by Munro et al. 2021. Each acronym corresponds to a vaccine. There are two groups of arms corresponding to the first 2 doses: one with prime BNT/BNT (BNT as first and second dose) and the second with prime ChAd/ChAd (ChAd as first and second dose). Combining the first two doses received and the third dose investigated in the trial, there were $K = 20$ arms.

Formal problem statement. We model this setting as a stochastic multi-armed bandit with vector-valued outcomes. At each time $t = 1, 2, \ldots, T$, a treatment A_t is adaptively assigned to the arriving participant, and a random vector $Z_t = (Z_t^1, \ldots, Z_t^d)$ is observed. Each arm k corresponds to an unknown distribution ν_k over $\mathbb{R}^d$ with mean vector $\mu_k \in \mathbb{R}^d$; the outcome Z_t is modeled as a sample from the distribution ν_{A_t}. These distributions belong to some family $\mathcal{P}$. The goal is to identify the Pareto-optimal subset $\mathcal{S}^\star(\mu) := \{i \in [K] : \nexists\ j,\ \mu_i \preceq \mu_j\}$ as efficiently as possible: for example by minimizing the probability of misidentification of optimal vaccine strategies using only a few patients (formalized in Section 1.2). In the example of Figure 1.2, two vaccine strategies are Pareto-optimal.

Dissertation scope and objectives. This dissertation develops adaptive algorithms tailored to the design of early-stage, multi-objective clinical trials. The focus is on the *pure exploration* setting, where the goal is to identify the best (or non-dominated) treatments at the end of the trial rather than to maximize patient benefit during the trial. Such designs are especially relevant to vaccine development, where the number of participants is small, the arms are many, and multiple endpoints must be considered simultaneously.

Throughout the manuscript, in particular in the experiments, we will frequently refer to the COVID-19 booster-dose trial (Munro et al. 2021) as a motivating application. It epitomizes the challenges addressed in this work: large numbers of candidate arms, multiple immunological endpoints, limited resources. We build upon and extend recent advances in the theory of bandit algorithms to handle:

- multi-dimensional outcomes and Pareto set identification,
- linear or feasibility constraints between arms, and
- Bayesian formulations and randomized algorithms based on posterior sampling.

The resulting algorithms provide statistical guarantees (*e.g.*, controlled error probability, sample efficiency) and their relevance is shown through simulations, datasets inspired by real-world vaccine studies and other benchmarks from the literature on multi-objective bandits.

Positioning with respect to the literature. For phase IIb/III trials where the patients should be cured during the trial, the field of regret minimization in bandits is now well developed, with extensions to non-stationary environments, contextual and adversarial settings, and structured action spaces (Lattimore & Szepesvari 2020). Upper Confidence Bound (UCB) algorithms, introduced by Auer 2003, and their variants, KL-UCB (Garivier & Cappé 2011), as well as randomized Bayesian algorithms such as Thompson Sampling (Thompson 1933), were shown to combine strong empirical performance with rigorous guarantees (Agrawal & Goyal 2012; Kaufmann, Korda, et al. 2012; Russo et al. 2020).

For early vaccine trials that focus on identifying the most immunogenic or balanced strategies, existing work on bandit algorithms provides a rich theoretical foundation, ranging from fixed-confidence best-arm identification (Jamieson & Nowak 2014; Garivier & Kaufmann 2016; Shang et al. 2020; Jourdan, Degenne, Baudry, et al. 2022) to fixed-budget pure exploration (Audibert & Bubeck 2010; Karnin et al. 2013; Locatelli et al. 2016). Yet, few studies have addressed the case of vector-valued outcomes or Pareto set identification, besides the work of (Auer et al. 2016). This dissertation builds precisely on that gap by proposing algorithms tailored to multi-objective exploration problems.

The next section introduces the general *pure exploration* framework underlying this work, detailing sampling, stopping, and recommendation principles.

1.2 Active Testing (Pure Exploration)

Pure exploration, also called *active testing*, models learning scenarios in which an algorithm is judged only by its final recommendation (Bubeck, Munos, et al. 2009; Degenne & W. Koolen 2019). Unlike regret minimization, interim outcomes are irrelevant; the goal is to collect just enough information to answer a prespecified query with high confidence.

Bandit model and query task. A bandit instance is described by $\nu := (\nu_1, \dots, \nu_K)$ with unknown mean vectors $\mu := (\mu_1, \dots, \mu_K) \in \mathcal{I}^K$ (scalar outcomes: $d = 1$; multi-objective: $\mu_k \in \mathbb{R}^d$). A *query task* is a (possibly set-valued) mapping

$$\mathcal{S}^\star : \mathcal{I}^K \to 2^{[K]},$$

whose value $\mathcal{S}^\star(\mu)$ collects all acceptable answers (*e.g.*, $\{\arg\max_k \mu_k\}$ in BAI; the Pareto set in PSI). Any element of $\mathcal{S}^\star(\mu)$ is correct.

Example 1.1 (Best Arm Identification). For best arm identification, the means are scalar-valued, $\mu = (\mu_1, \dots, \mu_K) \in \mathbb{R}^K$, and the query task associates each $\mu \in \mathbb{R}^K$ to the index of the best arm: $\mathcal{S}^\star(\mu) = \{\{k\} : k \in [K] \text{ and } \mu_k \geqslant \max_{i \neq k} \mu_i\}$.

Example 1.2 (Pareto Set Identification). For Pareto set identification, the mean of each arm is vector-valued, $\mu = (\mu_1, \dots, \mu_K) \in (\mathbb{R}^d)^K$, and the goal is to identify the set of arms that are not uniformly dominated on all d objectives by another arm. Formally, $\mathcal{S}^\star(\mu) = \{k \in [K] : \nexists\, i \in [K] \backslash \{k\},\ \mu_k \preceq \mu_i\}$, where $\mu_i \preceq \mu_j$ means that for all objectives $c \in \{1, \dots, d\}$, $\mu_i^c \leqslant \mu_j^c$, with strict inequality for at least one objective. This is further illustrated in Figure 1.3.

In this section, we focus on problems with a unique correct answer, *i.e.*, for a given μ, the task operator $\mathcal{S}^\star(\mu)$ is a singleton-valued (its unique element may itself be a set, as in PSI). We then write $\mathcal{S}^\star(\mu) \neq S$ to denote without ambiguity $S \notin \mathcal{S}^\star(\mu)$, and similarly, we write $S = \mathcal{S}^\star(\mu)$ to denote the unique element of $\mathcal{S}^\star(\mu)$.

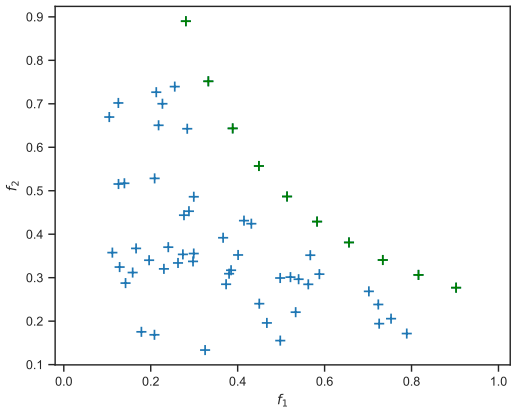


Figure 1.3: Pareto set in a bi-objective setting. Green points are Pareto-optimal, while blue points are sub-optimal.

At round $t \geqslant 1$, the learner chooses an arm $A_t \in [K]$, observes $Z_t \sim \nu_{A_t}$, and updates based on the history $\mathcal{H}_t = \sigma(A_1, Z_1, \dots, A_t, Z_t)$.

Policies and performance criteria. A pure exploration policy (or learner, or agent, or algorithm) has three main components: (i) a *sampling rule* $(A_t)_{t \geqslant 1}$ that dictates the arm to sample at each round, (ii) a *recommendation rule* $\widehat{S}_t$ (a random element or subset of $[K]$) representing a guess for the correct answer, and (iii) a *stopping rule* τ (a stopping time *w.r.t.*

$\mathcal{H}_t$), which dictates when to stop collecting new samples from the environment. The final output is $\widehat{S}_\tau$.

For risk $\delta \in (0,1)$, a policy is δ-PAC (or δ-*correct*) if, for every instance with mean μ,

$$\mathbb{P}_\nu\big(\tau < \infty,\ \widehat{S}_\tau \neq \mathcal{S}^\star(\mu)\big) \ \leqslant\ \delta. \tag{1.1}$$

Two classical formulations exist:

- **Fixed-confidence (FC)**: the goal is to minimize the *sample complexity*, *i.e.*, the (instance-specific) number of samples used before stopping, while satisfying (1.1) (Even-Dar et al. 2002; Kalyanakrishnan et al. 2012; Jamieson & Nowak 2014; Degenne, Ménard, et al. 2020). When the algorithm draws exactly one sample per round (or a fixed number per round), this is equivalent to minimizing the stopping time τ (or its expected value $\mathbb{E}[\tau]$).

- **Fixed-budget (FB)**: Audibert & Bubeck 2010; Locatelli et al. 2016; Katz-Samuels & Scott 2018; Roy Chaudhuri & Kalyanakrishnan 2019 for a given horizon T, stop at $\tau := T$ and minimize $e_T(\nu) := \mathbb{P}_\nu(\widehat{S}_T \neq \mathcal{S}^\star(\mu))$; interest centers on the decay rate of $e_T(\nu)$.

In short, the fixed-budget setting emphasizes optimal *allocation* (stopping is fixed), whereas the fixed-confidence setting couples *adaptive sampling* with *adaptive stopping*.

Algorithm 1.1: Generic Pure Exploration Bandit Algorithm

Require: Arms $[K] = \{1, \dots, K\}$; risk δ (FC) or budget T (FB)

Initialize counts $N_k \leftarrow 0$, for all $k \in [K]$; $t \leftarrow 0$.
repeat
  $t \leftarrow t + 1$.
  `// Sampling rule (policy)`
  Select arm $A_t \leftarrow$ **SamplingRule**$(\mathcal{H}_{t-1})$.
  Observe outcome $Z_t \sim \nu_{A_t}$.
  Update $N_{A_t} \leftarrow N_{A_t} + 1$ and estimates $\hat{\mu}_{t,A_t}$ (and confidence intervals/posterior as needed).
  `// Recommendation rule (current guess)`
  $\widehat{S}_t \leftarrow$ **RecommendationRule**$(\{\hat{\mu}_{t,k}\}_{k=1}^K, \mathcal{H}_t)$ (*e.g.*empirical best, empirical Pareto set, etc.).
until ***StoppingRule***$(\mathcal{H}_t, \delta)$ *is true* ***(FC)*** **or** $t = T$ ***(FB)***
return final recommendation $\widehat{S}_\tau$.

1.2.1 Stopping and Recommendation Rules

The recommendation rule maps the observed data to an element of $\{\mathcal{S}^\star(\lambda) : \lambda \in \mathcal{I}^K\}$. The most frequent recommendation rules are:

- Plug-in estimate: use empirical data to estimate the answer, *e.g.*, $\widehat{S}_t = \arg\max_k \hat{\mu}_{t,k}$ in BAI, or $\widehat{S}_t = \{i : \nexists j,\ \hat{\mu}_{t,i} \preceq \hat{\mu}_{t,j}\}$ in PSI.
- Confidence-based: return the most "optimistic" answer *e.g.*, the arm with the largest upper confidence bound (Gabillon et al. 2012).
- Bayesian: greedy response *w.r.t.* posterior sample, most probable answer as measured by the posterior (Russo 2016; Shang et al. 2020)

In practice, the recommendation is coupled with stopping: the algorithm halts only when the chosen recommendation is certified correct by the stopping rule.

Stopping rules are (Markov) stopping times *w.r.t.* the filtration $\{\mathcal{H}_t\}_{t=1}^\infty$ and determine when the accumulated evidence suffices to recommend a correct answer with the desired confidence. Designing efficient and computationally tractable *stopping rules* is a central challenge in the construction of bandit algorithms. Below, we unify frequentist and Bayesian constructions. Before introducing them, let $\mathrm{Alt} : 2^{[K]} \to (\mathbb{R}^d)^K$ defined as

$$\mathrm{Alt}(\mathcal{S}^\star(\mu)) \coloneqq \{\lambda \in \mathcal{I}^K : \mathcal{S}^\star(\lambda) \neq \mathcal{S}^\star(\mu)\}$$

be the set of parameters that have a different correct response than μ.

An *anytime confidence region* $(\mathcal{R}_t(\delta))_{t\geqslant 1}$ at level δ satisfies $\mathbb{P}(\forall t \geqslant 1,\ \mu \in \mathcal{R}_t(\delta)) \geqslant 1-\delta$. With an estimator $\hat{\mu}_t$ of μ, consider the generic stopping rule

$$\tau \coloneqq \inf\big\{t \geqslant 1\ \ \mathcal{R}_t(\delta) \cap \mathrm{Alt}(\widehat{S}_t) = \emptyset\big\}. \tag{1.2}$$

This yields δ-PAC correctness, since

$$\begin{aligned}\mathbb{P}\big(\tau < \infty,\ \widehat{S}_\tau \neq \mathcal{S}^\star(\mu)\big) &\leqslant \mathbb{P}\big(\exists t\colon \mu \in \mathrm{Alt}(\widehat{S}_t),\ \mathcal{R}_t(\delta) \cap \mathrm{Alt}(\widehat{S}_t) = \emptyset\big)\\ &\leqslant \mathbb{P}\big(\exists t\colon \mu \notin \mathcal{R}_t(\delta)\big) \leqslant \delta.\end{aligned}$$

In words, on the event $\{\forall t \geqslant 1,\ \mu \in \mathcal{R}_t(\delta)\}$, the true parameter is never excluded before stopping, and at τ all parameters consistent with data agree on the same answer.

Depending on the geometry of the confidence regions, various forms of stopping rules can be derived from Eq. (1.2).

Individual confidence sequences. For scalar-valued problems, if each arm distribution ν_k is σ-subgaussian, *i.e.*, $\log \mathbb{E}_{X\sim\nu_k}[e^{\lambda(X-\mu_k)}] \leqslant \sigma^2\lambda^2/2$ for all $\lambda \in \mathbb{R}$, then the empirical mean $\hat{\mu}_{t,k}$ of arm k after $N_{t,k}$ samples admits an anytime *Hoeffding-type* confidence interval:

$$[L_k(t,\delta), U_k(t,\delta)] = \left[\hat{\mu}_{t,k} - \sqrt{\tfrac{2\sigma^2\log(Ct^2K/\delta)}{N_{t,k}}},\ \hat{\mu}_{t,k} + \sqrt{\tfrac{2\sigma^2\log(Ct^2K/\delta)}{N_{t,k}}}\right],$$

valid with probability at least $1-\delta/K$ for a universal constant $C > 0$. The Cartesian product $\mathcal{R}_t(\delta) = \prod_{k=1}^{K}[L_k(t,\delta), U_k(t,\delta)]$ defines an *anytime confidence region* for μ.

For the best-arm identification (BAI) query, substituting $\mathcal{R}_t$ into the generic stopping rule (1.2) yields the classical LUCB-style stopping time:

$$\tau_\delta = \inf\Big\{t \geqslant 1 :\ L_{\widehat{S}_t}(t,\delta) \geqslant U_k(t,\delta),\ \forall k \neq \widehat{S}_t\Big\},$$

where $\widehat{S}_t = \arg\max_k \hat{\mu}_{t,k}$ is the empirical best arm. At stopping, every $\lambda \in \mathcal{R}_t(\delta)$ satisfies $\lambda_{\widehat{S}_t} \geqslant \max_k \lambda_k$, ensuring that the empirical best arm is also the true best arm for all mean vectors consistent with the collected data.

When the arms belong to a one-parameter exponential family, tighter bounds are obtained by replacing Hoeffding radii with Kullback-Leibler (KL) constraints. Let $d(x,y)$ denote the KL divergence between ν^x and ν^y for distributions in a parametric family indexed by x, y. The *KL-based confidence interval* for an arm k is defined implicitly by

$$\begin{aligned} U_k(t,\delta) &= \sup\{q \geqslant \hat{\mu}_{t,k} : N_{t,k}\, d(\hat{\mu}_{t,k}, q) \leqslant f(t,\delta)\}, \\ L_k(t,\delta) &= \inf\{q \leqslant \hat{\mu}_{t,k} : N_{t,k}\, d(\hat{\mu}_{t,k}, q) \leqslant f(t,\delta)\}\ , \end{aligned}$$

where $f(t,\delta)$ is a time-dependent exploration function of order $\log(Ct/\delta)$. The corresponding product region $\mathcal{R}_t(\delta) = \prod_k [L_k(t,\delta), U_k(t,\delta)]$ plugged into (1.2) yields KL-LUCB-type stopping rules (Kaufmann & Kalyanakrishnan 2013), offering sharper, distribution-dependent confidence bounds and improved finite-time performance.

More refined confidence regions can be constructed from sums of KL or likelihood bounds, yielding Chernoff or GLR-type stopping rules such as those used in Garivier & Kaufmann 2016.

Joint KL confidence region/Chernoff stopping. A global (joint) confidence region aggregates evidence across arms:

$$\mathcal{R}_t(\delta) \ = \ \Big\{\lambda \in \mathcal{I}^K :\ \sum_{k=1}^{K} N_{t,k}\, d\big(\hat{\mu}_{t,k}, \lambda_k\big) \leqslant \beta(t,\delta)\Big\},$$

with β chosen such that $\mathbb{P}(\forall t,\ \mu \in \mathcal{R}_t(\delta)) \geqslant 1-\delta$. This region captures the collective statistical evidence about the parameter μ. Then Eq. (1.2) becomes

$$\begin{aligned} \tau &= \inf\left\{t \geqslant 1 : \forall \lambda \in \mathrm{Alt}(\widehat{S}_t),\ \sum_{k=1}^{K} N_{t,k}\, d\big(\hat{\mu}_{t,k}, \lambda_k\big) > \beta(t,\delta)\right\} \\ &= \inf\left\{t \geqslant 1 :\ \inf_{\lambda \in \mathrm{Alt}(\widehat{S}_t)} \left[\sum_{k=1}^{K} N_{t,k}\, d\big(\hat{\mu}_{t,k}, \lambda_k\big)\right] > \beta(t,\delta)\right\}. \end{aligned}$$

The latter form corresponds to the GLR stopping rule or *Chernoff stopping*, which can also be derived from a hypothesis-testing perspective (Garivier & Kaufmann 2021), and has been at the core of the most recent algorithms for pure exploration (Garivier & Kaufmann 2016; Degenne, Ménard, et al. 2020).

Remark. Confidence-sequence and generalized likelihood-ratio (GLR) constructions are two sides of the same coin: modern results (Ramdas et al. 2022) show equivalences between anytime valid confidence sets and likelihood-based tests (such as Chernoff stopping), which explains the parallel forms of the rules above. Deriving tight confidence sequences has played a central role in the design of bandit algorithms (T. L. Lai 1976; Howard et al. 2018; Kaufmann & W.-M. Koolen 2021; Emmenegger et al. 2023; Kirschner et al. 2025).

Bayesian stopping. Let P_0 be a prior over μ and P_t the posterior at round t. A basic principle is to stop when the posterior probability of error is small:

$$e_t(q) \coloneqq \mathbb{P}_{\lambda\sim P_t}\big(\mathcal{S}^\star(\lambda) \neq q \mid \mathcal{H}_{t-1}\big),\ \widehat{S}_t \in \arg\min_q e_t(q),\ \tau_\delta = \inf\{t \geqslant 1 :\ e_t(\widehat{S}_t) \leqslant f(t,\delta)\}.$$

This stopping time triggers when the posterior mass on incorrect answers falls below δ. Proposed by Russo 2016 and analyzed by Shang et al. 2020, these rules yield near-optimal sample complexity but are often computationally demanding in complex, structured, or multi-dimensional settings.

Summary. Across all approaches, the core idea remains identical: stop as soon as the collected evidence is sufficient to rule out all alternative hypotheses with high confidence. Stopping rules derived from Hoeffding-type confidence regions are simple and offer finite-time guarantees but can be conservative, while GLR-based rules achieve asymptotic optimality but are much more involved to compute. A bandit policy must then pull arms in order to reduce uncertainty as fast as possible.

1.2.2 Sampling Rules

Sampling determines where information is gathered. The most frequent sampling rules belong to the following generic families:

- Non-adaptive allocations (uniform, stratified) are simple and analytically convenient, but can be wasteful when the number of arms is large (this will be illustrated throughout the manuscript). They are often coupled with adaptive eliminations, for example, when some arms can be confidently ruled out as sub-optimal.
- Uncertainty-based (racing, LUCB-style) prioritizes arms whose status (optimal vs. suboptimal; feasible vs. infeasible) is most uncertain under current confidence sets. A typical example is the OFU (*Optimism in the Face of Uncertainty*) principle, which, as employed in Jamieson, Malloy, et al. 2014, pulls the arm with the largest optimistic estimate of its mean.
- Tracking aims at targeting instance-optimal proportions (solutions to max-min formulations of instance-dependent lower bounds) via tracking or posterior sampling (*e.g.*, Top-Two variants (Russo 2016), Track-and-Stop (Garivier & Kaufmann 2016)). These naturally pair with GLR or Bayesian stopping to approach information-theoretic lower bounds.

1.2.3 Lower Bounds

At a deeper level, the limits of adaptivity are governed by *information-theoretic lower bounds*, which formalize the minimal number of samples required for reliable identification. These results show that, regardless of the sampling strategy, no algorithm can beat a threshold determined by the intrinsic difficulty of the problem instance. A key ingredient in such analyses is the *information contraction lemma*, popularized by Kaufmann, Cappé, et al. 2016 and later refined in Garivier, Ménard, et al. 2019; Garivier & Kaufmann 2021. Informally, it states that if two bandit models ν and ν' generate similar distributions over observation histories (in the sense of KL divergence), then no algorithm can reliably distinguish them without collecting a sufficient number of samples. This principle underlies all instance-dependent lower bounds developed in this dissertation.

Lemma 1.2.1 (Information contraction, Kaufmann, Cappé, et al. 2016; Garivier, Ménard, et al. 2019)**.** *Let ν and ν' be two bandit models, and let $\mathbb{P}_\nu$ and $\mathbb{P}_{\nu'}$ denote the distributions over histories induced by an arbitrary (possibly adaptive) strategy up to stopping time τ. Then for any event $\mathcal{E}$ measurable with respect to $\mathcal{H}_\tau$,*

$$\sum_{a=1}^{K} \mathbb{E}_\nu[N_{\tau,a}]\ \mathrm{KL}(\nu_a \| \nu'_a)\ \geqslant\ \mathrm{kl}\big(\mathbb{P}_\nu(\mathcal{E}),\, \mathbb{P}_{\nu'}(\mathcal{E})\big),$$

where $\mathrm{kl}(\cdot,\cdot)$ denotes the binary relative entropy and $N_{\tau,a}$ is the number of samples drawn from arm a before stopping.

This result shows that the ability of an algorithm to discriminate between events under two instances is fully determined by its sampling frequencies and the KL divergence between arm distributions. It forms the cornerstone of all lower bound arguments used later in this work.

In the next sections, we instantiate these generic components for the multi-objective setting and, in particular, Pareto set identification.

1.3 Background

Originally inspired by adaptive clinical trial design, where treatments are sequentially allocated to balance learning and patient welfare, the bandit framework has since evolved into a general paradigm for sequential decision-making under uncertainty.

This chapter reviews prior work, with a focus on the transition from single–to multi-objective formulations. We first revisit core results in *Best Arm Identification* (BAI) and related single-objective problems, which form the methodological foundation for our study. We then turn to the multi-objective setting, reviewing existing approaches to *Pareto Set Identification* (PSI) and its related problems.

1.3.1 Best Arm Identification and Beyond

The rich literature on BAI has laid the theoretical foundation for pure exploration in bandits, with sophisticated tools such as change-of-measure arguments (Kaufmann, Cappé, et al. 2016), law of iterated logarithm (LIL) confidence bounds (Jamieson, Malloy, et al. 2014), and generalized likelihood ratio stopping rules (Garivier & Kaufmann 2021). These methods will play a central role as we extend pure exploration to multi-objective settings, the focus of this dissertation.

Understanding the complexity of pure exploration, particularly the BAI problem, has long been a central question in the stochastic bandit literature. Given K arms with means $\mu = (\mu_1, \dots, \mu_K)$, the learner sequentially samples arms until a stopping time τ, and must recommend $\widehat{S}_\tau \in [K]$ such that $\mathbb{P}(\widehat{S}_\tau \neq \mathcal{S}^\star) \leqslant \delta$, where $\mathcal{S}^\star = \operatorname{argmax}_a \mu_a$. The design of efficient algorithms depends on both the setting (fixed-confidence or fixed-budget) and the statistical assumptions on arm distributions.

From gap-based heuristics to information-theoretic analysis. Early fixed-confidence algorithms, such as *Successive Elimination*, *LUCB*, and *LIL-UCB* (Even-Dar et al. 2002; Gabillon et al. 2012; Kalyanakrishnan et al. 2012; Karnin et al. 2013; Jamieson, Malloy, et al. 2014), relied on subgaussian concentration inequalities to derive confidence intervals around empirical means. Their performance was characterized by the sub-optimality gaps, $\Delta_i = \mu_{\mathcal{S}^\star} - \mu_i$ for $i \neq \mathcal{S}^\star$ and $\Delta_{\mathcal{S}^\star} = \min_{i \neq \mathcal{S}^\star} \Delta_i$, leading to sample complexity guarantees of the form

$$\mathbb{E}_\nu[\tau] \leqslant \mathcal{O}\big(H(\nu) \log(H(\nu)/\delta)\big), \qquad H(\nu) := \sum_{i=1}^{K} \frac{1}{\Delta_i^2}.$$

While these results captured key scaling laws, the guarantees were not instance-optimal, an important mismatch was observed between best known lower and upper bounds (Jamieson, Malloy, et al. 2014; Kaufmann, Cappé, et al. 2016).

A major refinement came from replacing subgaussian confidence bounds with information-theoretic quantities such as Kullback-Leibler divergences. Kaufmann & Kalyanakrishnan 2013 introduced *KL-LUCB*, which achieves sharper confidence intervals and tighter sample complexity bounds governed by the Chernoff information rather than mean gaps. The theoretical shift from difference-based to KL-based analysis eventually culminated in the *game-value characterization* of bandit complexity, introduced by Garivier & Kaufmann 2016 and defined as:

$$T^\star(\nu)^{-1} = \max_{w \in \triangle_K} \inf_{\lambda \in \mathrm{Alt}(\mathcal{S}^\star(\mu))} \sum_{i=1}^{K} w_i \, \mathrm{KL}(\nu_a \| \lambda_a), \tag{1.3}$$

where $\triangle_K := \{w \in \mathbb{R}_+^K : \sum_{i=1}^K w_i = 1\}$ is the probability simplex, ν is a bandit with means μ, and $\mathrm{Alt}(\mathcal{S}^\star(\mu))$ denotes the set of alternative bandit instances with a different best arm. The quantity $T^\star(\nu)$ is called the *characteristic time* of instance ν and, essentially $T^\star(\nu) \log \frac{1}{\delta}$ represents the optimal sample complexity for which an algorithm can distinguish ν from its alternatives, in particular when δ is small.

We say that an algorithm is *asymptotically optimal* if its sample complexity matches that lower bound, namely if $\limsup_{\delta \to 0} \mathbb{E}_\nu[\tau_\delta] / \log \frac{1}{\delta} \leqslant T^\star(\nu)$.

Asymptotically optimal algorithms. Garivier & Kaufmann 2016 established the first asymptotically optimal BAI algorithm, *Track-and-Stop*, achieving $\mathbb{E}_\nu[\tau_\delta] \sim_{\delta\to 0} T^\star(\nu)\log\frac{1}{\delta}$ for any bandit with distributions in one-parameter exponential families. They were the first to exactly match the lower bound, in the asymptotic regime ($\delta \to 0$).

The idea of *Track-and-Stop* algorithms is first, to have a forced exploration procedure to ensure that at least a sublinear fraction of samples is allocated to every arm, *e.g.*, ensure that at round t each arm is sampled at least $\sqrt{t}$ times. This ensures, by the law of large numbers, that empirical estimates converge to the true means almost surely. Then, at each round, *Track-and-Stop* algorithms compute the $\hat{w}^\star(\hat{\mu}_t) \in \triangle_K$, a maximizer of Eq. (1.3) for the empirical means, and following a tracking procedure, they will select the next arm to sample. This sampling rule is coupled with a generalized likelihood ratio (GLR) stopping rule ensuring δ-PAC correctness. Similar algorithmic designs further appeared in more general identification problems, linear bandits (Ménard 2019; Degenne, Ménard, et al. 2020; Jedra & Proutiere 2020a).

In parallel, Bayesian approaches such as *Top-Two Thompson Sampling* (Russo 2016) and related variants showed that posterior sampling can asymptotically realize the same optimal allocation without explicitly solving the optimization problem in Eq. (1.3), while admitting elegant analyses in terms of posterior contraction rates (Shang et al. 2020; Jourdan, Degenne, Baudry, et al. 2022; You et al. 2023).

Fixed-budget formulations. The fixed-budget regime, formalized by Audibert & Bubeck 2010, considers a fixed sampling horizon T and seeks to minimize the error probability $e_T(\nu) := \mathbb{P}(\widehat{S}_T \neq \mathcal{S}^\star)$. Two influential algorithms, *Successive Rejects* (SR) and *Sequential Halving* (SH), were proposed by Audibert & Bubeck 2010 and Karnin et al. 2013, respectively.

These algorithms allocate the budget in phases, discarding empirically suboptimal arms. Their analysis shows that the error probability decreases at an exponential rate

$$e_T(\nu) \leqslant \exp\left(-c\,\frac{T}{H(\nu)\log K}\right),$$

and lower bounds (Carpentier & Locatelli 2016; Degenne 2023) show that these rates are essentially unimprovable up to logarithmic factors.

Beyond BAI. Beyond standard Best Arm Identification, several extensions have broadened the scope of pure exploration in single-objective settings. These include *thresholding bandits* (Locatelli et al. 2016), where the goal is to classify arms as above or below a fixed level.

Other extensions include structured formulations such as *(transductive) linear* and *combinatorial* bandits (Chen et al. 2014; Soare et al. 2014; Fiez et al. 2019; Degenne, Ménard, et al. 2020). Other formulations consider *multi-fidelity* (Poiani et al. 2024) or *cost-sensitive* sampling (Kanarios et al. 2024), where arms differ in evaluation cost or reliability, as well as *safe* or *constrained* bandits (Camilleri, A. Wagenmaker, et al. 2022), where exploration must respect safety or feasibility conditions throughout the learning process; good arm identification (Roy Chaudhuri & Kalyanakrishnan 2019), a relaxation of *thresholding bandits* where the goal is to identify an arm whose mean is above a given threshold.

Summary. The evolution from gap-based to information-theoretic designs has led to a unified theoretical understanding of pure exploration in single-objective bandits. Fixed-budget algorithms optimize allocation under a deterministic horizon, while fixed-confidence algorithms balance adaptive sampling and stopping. The resulting principles: optimal allocation, information-theoretic stopping, and posterior contraction, form the methodological backbone for the multi-objective extensions developed in this dissertation.

1.3.2 Multi-objective Pure Exploration

The extension of pure exploration to multi-objective settings, where arms are evaluated along multiple potentially conflicting criteria, introduces fundamental challenges absent from single-objective problems. Rather than identifying a single best arm, the learner must characterize a set of arms representing different trade-offs–such as the Pareto set of non-dominated alternatives, or arms satisfying feasibility and ranking constraints. This section surveys the landscape of multi-objective pure exploration, reviewing theoretical results and algorithmic approaches for Pareto set identification and related multi-objective formulations.

Pareto Set Identification (PSI). Given a multi-armed bandit with K arms, where each arm $i \in [K]$ has a mean vector $\mu_i \in \mathbb{R}^d$, we say that arm i is *Pareto dominated* by arm j (denoted $\mu_i \preceq \mu_j$) if $\mu_i^c \leqslant \mu_j^c$ for all objectives $c \in [d]$ with strict inequality for at least one objective. The strict domination (strict inequality for all objectives) is denoted $\mu_i \prec \mu_j$. The *Pareto set* is the set of non-dominated arms:

$$\mathcal{S}^\star(\mu) = \{i \in [K] : \nexists j \in [K] \setminus \{i\},\ \mu_i \preceq \mu_j\}.$$

The goal of Pareto set identification is to identify $\mathcal{S}^\star(\mu)$ with high confidence. When $d = 1$, this reduces to the classical best-arm identification problem.

Early studies of PSI focused on uniform sampling and elimination schemes (Zuluaga, Sergent, et al. 2013; Auer et al. 2016; Zuluaga, Krause, et al. 2016), formulated under the fixed-confidence setting. The seminal contribution of Auer et al. 2016 provided the first formal instance-dependent analysis of PSI algorithms, introducing key instance-dependent quantities to characterize problem difficulty and establishing worst-case matching upper and lower bounds on their sample complexity.

To quantify the hardness of PSI, Auer et al. 2016 defined arm-specific *suboptimality gaps* that measures how difficult it is to determine whether an arm is Pareto-optimal. For arms $i, j \in [K]$, define[1]

$$\begin{aligned} \mathrm{m}(i,j) &\coloneqq \min_{c\in[d]} \left[\mu_j^c - \mu_i^c\right], \\ \mathrm{M}(i,j) &\coloneqq \max_{c\in[d]} \left[\mu_i^c - \mu_j^c\right]. \end{aligned} \tag{1.4}$$

[1]The initial definition was $\widetilde{\mathrm{m}}(i,j) = \max\{\mathrm{m}(i,j), 0\}$ and $\widetilde{\mathrm{M}}(i,j) = \max\{\mathrm{M}(i,j), 0\}$. Some algebra shows that these expressions can be simplified to Eq. (1.4) while keeping the gaps' value.

If $\mu_i \not\prec \mu_j$, then $\mathrm{M}(i,j)$ is the smallest uniform increase of j required to dominate i, while, if $\mu_i \prec \mu_j$, then $\mathrm{m}(i,j)$ is the smallest coordinate-wise increase of i needed to make it non-strictly-dominated by j. From these primitives, one defines for each arm i the suboptimality gap

$$\Delta_i := \begin{cases} \max_{j \in \mathcal{S}^\star} \mathrm{m}(i,j), & \text{if } i \notin \mathcal{S}^\star, \\ \min\{\delta_i^+, \delta_i^-\}, & \text{if } i \in \mathcal{S}^\star, \end{cases} \tag{1.5}$$

where

$$\begin{aligned} \delta_i^+ &:= \min_{j \in \mathcal{S}^\star \setminus \{i\}} \min\{\mathrm{M}(i,j), \mathrm{M}(j,i)\}, \\ \delta_i^- &:= \min_{j \notin \mathcal{S}^\star} \big[(\mathrm{M}(j,i))_+ + \Delta_j\big]. \end{aligned}$$

These quantities capture how close each arm lies to the Pareto frontier. The overall instance complexity is then measured by

$$H(\nu) := H(\mu) := \sum_{i=1}^{K} \frac{1}{\Delta_i^2}\,. \tag{1.6}$$

This term was shown by Auer et al. 2016 to govern the finite-time instance-dependent sample complexity of fixed-confidence PSI algorithms through an upper bound on the sample complexity of their algorithm of order

$$\mathcal{O}\big(H(\mu)\log(dH(\mu)/\delta)\big),$$

combined with a lower bound, also scaling with $H(\mu)$ (recalled in Theorem 1.3.1). This result establishes a nearly-matching information-theoretic lower bound, confirming (up to logarithmic factors) the dependence of the sample complexity on $H(\mu)$ for some instances.

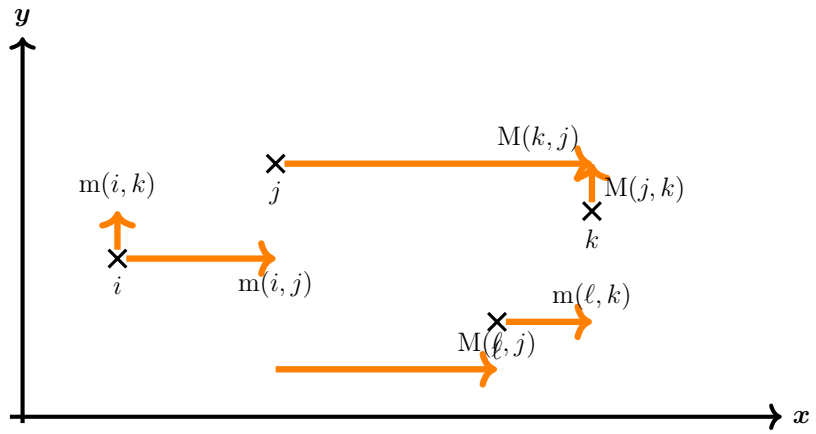


Figure 1.4: PSI gaps and distances

Theorem 1.3.1 (Theorem 17 of Auer et al. 2016). *For any collection of mean vectors $\mu_i \in [1/4, 3/4]^d$, there exist distributions $(\mathcal{D}_i)_{1 \leqslant i \leqslant K}$ such that, with probability at least $1-\delta$, any δ-correct algorithm for PSI requires at least*

$$\Omega\left(\sum_{i=1}^{K} \frac{1}{\widetilde{\Delta}_i^2} \log\frac{1}{\delta}\right)$$

samples to identify the Pareto set, where for suboptimal arms $\widetilde{\Delta}_i := \Delta_i$ and for optimal arms $\widetilde{\Delta}_i := \delta_i^+$.

From an application perspective, PSI provides a rigorous abstraction of multi-criteria decision-making in which several performance indicators must be considered simultaneously. In clinical research, this is particularly relevant to early-phase vaccine or treatment studies, as motivated in Section 1.1.

In what follows, we review several extensions of PSI proposed to handle more general forms of partial orders, feasibility constraints, and structured models.

Cone-Ordered Pareto Set Identification. A generalization of the Pareto set identification problem was proposed by Ararat & Tekin 2023, who extended the notion of dominance to arbitrary partial orders induced by convex cones. In this framework, a cone $\mathcal{C} \subset \mathbb{R}^d$ defines a preference relation between mean vectors, where $\mu_i \prec_{\mathcal{C}} \mu_j$ if and only if $\mu_j - \mu_i \in \mathcal{C}$. The standard Pareto order corresponds to the positive orthant $\mathbb{R}^d_+$, but other cones can encode domain-specific trade-offs between objectives. Another example of a cone is $\mathcal{C}_\theta = \{x : \theta^\intercal x \geqslant 0\}$, for a direction $\theta \in \mathbb{R}^d$, which results in a scalarization of the multi-objective problem (Miettinen 1998).

This cone-based formulation unifies PSI and several related problems, such as constrained optimization and scalarization-based selection, within a single mathematical framework. Ararat & Tekin 2023 analyzed this problem in the fixed-confidence setting and proposed a general elimination algorithm whose sample complexity depends on cone-specific quantities generalizing the suboptimality gaps of Auer et al. 2016. Their work showed that, although the underlying geometry of the cone affects the difficulty of the problem, similar logarithmic sample complexity dependencies can be achieved, extending PSI theory beyond the axis-aligned Pareto order. This line of work was later studied by Shukla & Basu 2024; Das et al. 2025 where, instead of identifying optimal arms, the goal is rather to identify a policy supported on optimal arms. The authors derived an information-theoretic lower bound and a Track-and-Stop style algorithm.

Feasible Arm Identification. The *Feasible Arm Identification* problem was introduced by Katz-Samuels & Scott 2018 in the fixed-budget setting. In this formulation, the learner is given a convex polyhedron $P \subset \mathbb{R}^d, d \geqslant 1$ and by interacting with a multi-objective bandit model, must identify the set of arms whose mean vectors lie within P, given a prescribed exploration budget. This setting can be viewed as a generalization of the classical *bandit thresholding problem* (Locatelli et al. 2016), extended to multi-dimensional objectives. The

authors proposed MD-APT (Multi-Dimensional Anytime Parameter-Free Thresholding), an algorithm that achieves a near-optimal fixed-budget guarantee:

$$e_T \leqslant \mathcal{O}\left(e^{-c\frac{T}{H(\mu)}}\right),$$

where c is an absolute constant, $H(\mu) \coloneqq \sum_i \Delta_i^{-2}$ is a complexity measure based on the arm-dependent gaps

$$\Delta_i \coloneqq \begin{cases} \mathrm{dist}(\mu_i, P^{\mathsf{c}}), & \text{if } \mu_i \in P, \\ \mathrm{dist}(\mu_i, \partial P), & \text{otherwise.} \end{cases}$$

The authors further showed that this rate is essentially optimal by establishing instance-dependent lower bounds for specific problem configurations. Beyond its theoretical appeal, the feasible-arm formulation has natural connections to adaptive clinical trial design: the polyhedron P can represent eligibility or safety constraints-such as acceptable toxicity-efficacy trade-offs, and identifying feasible arms corresponds to discovering treatments that satisfy predefined clinical criteria under limited experimental budget.

Top Feasible Arm Identification. The *Top Feasible Arm Identification* problem (Katz-Samuels & Scott 2019) extends the feasible-arm framework by incorporating preference rankings over multiple objectives.

In this setting, the learner is given a convex polyhedron $P \subset \mathbb{R}^d$, a weighting vector $w \in \mathbb{R}^d$, and an integer parameter $m \leqslant K$. Letting $F(\mu) \coloneqq \{k : \mu_k \in P\}$ denote the set of feasible arms, the learner's goal is to identify

$$\mathcal{O}^\star \coloneqq \left\{k : \mu_k \in P \text{ and } w^\top \mu_k \geqslant \overset{m}{\max}_{i \in F(\mu)} w^\top \mu_i\right\},$$

that is, the m best feasible arms according to the scalarized objective $k \mapsto w^\top \mu_k$. The authors formulated a fixed-confidence version of the problem, called δ-*PAC-explanatory*, where the learner outputs a triplet $(\widehat{O}, \widehat{S}, \widehat{I})$ forming a partition of $[K]$ such that

$$\widehat{O} = \mathcal{O}^\star, \qquad \widehat{I} \subset F(\mu)^{\mathsf{c}}, \qquad \widehat{S} \subset \{k : w^\top \mu_k \leqslant \max_{i \in \mathcal{O}^\star} w^\top \mu_i\}.$$

This formulation captures both the feasibility constraint and the ranking objective within a unified PAC framework, and provides statistical guarantees on the correctness of the identified sets.

From a clinical trial perspective, this problem naturally models the selection of the most promising treatments among those satisfying pre-defined safety or eligibility criteria. The weight vector w encodes the relative importance of various clinical endpoints-such as immunogenicity, safety, or tolerability-while the polyhedron P specifies admissible safety-efficacy trade-offs. Hence, identifying $\mathcal{O}^\star$ corresponds to finding the m best treatment candidates that are both clinically acceptable and statistically supported within a fixed-budget or fixed-confidence experiment.

Constrained Bi-Objective Identification. A closely related special case of multi-objective pure exploration with constraints was introduced by Faizal & Nair 2022. The authors considered a bi-objective ($d = 2$) bandit model where each arm $k \in [K]$ is associated with an unknown mean vector $\mu_k = (\mu_k^{(1)}, \mu_k^{(2)}) \in [0,1]^2$. The learner aims to identify the arm maximizing the first objective $\mu_k^{(1)}$ under a feasibility constraint on the second, namely

$$\mathcal{S}^\star_{\text{feas}} \;\coloneqq\; \Big\{ k \in [K] : \mu_k^{(2)} \leqslant b \text{ and } \mu_k^{(1)} = \max_{i:\, \mu_i^{(2)} \leqslant b} \mu_i^{(1)} \Big\},$$

where b is a known threshold. This setting can be interpreted as a variant of *Top Feasible Arm Identification*, where the feasibility region is $\{\mu : \mu^{(2)} \leqslant b\}$.

In the fixed-budget regime, the authors proposed a Successive-Rejects-type algorithm, C-SR, which alternates between (i) feasibility testing to discard arms likely to violate the constraint and (ii) dominance elimination among the remaining feasible arms. They showed that the algorithm achieves an exponential rate

$$e_T(\nu) \;\leqslant\; \mathcal{O}\Big(\exp\Big(-\, c \frac{T}{H_{\text{feas}}(\mu) \log K} \Big) \Big),$$

for some constant $c > 0$, where the instance complexity is governed by the composite measure

$$H_{\text{feas}}(\mu) \;\coloneqq\; \sum_{i=1}^{K} \frac{1}{\widetilde{\Delta}_i^2},$$

where for $i \neq \mathcal{S}^\star$, and assuming the feasible set is non-empty, $\Delta_i = \min(b - \mu^{(2)}_{\mathcal{S}^\star}, \tilde{\delta}_i)$ with

$$\tilde{\delta}_i \;\coloneqq\; \begin{cases} \mu^{(1)}_{\mathcal{S}^\star} - \mu^{(1)}_i & \text{if } \mu_i^{(2)} \leqslant b \\ \mu_i^{(2)} - b & \text{if } \mu_i^{(2)} > b \text{ and } \mu_i^{(1)} \geqslant \mu^{(1)}_{\mathcal{S}^\star} \\ \max(\mu_i^{(2)} - b, \mu^{(1)}_{\mathcal{S}^\star} - \mu^{(1)}_i) & \text{else.} \end{cases}$$

This expression highlights the two competing sources of difficulty: arms close to the feasibility boundary (small $|\mu_k^{(2)} - b|$) and arms close to optimality within the feasible region (small $\mu^{(1)}_{\mathcal{S}^\star} - \mu_k^{(1)}$). Although limited to $d = 2$, this work formally connected feasibility constraints and multi-objective identification, and provided one of the first fixed-budget guarantees for a constrained pure exploration problem. Such a setting is directly relevant to early-phase clinical trials, where treatments must simultaneously achieve sufficient efficacy and remain below pre-specified toxicity or reactogenicity thresholds.

Summary and open challenges. The studies reviewed above reveal that, before the onset of this work, the theory of multi-objective pure exploration in bandits remained largely fragmented. Existing algorithms were either restricted to low-dimensional cases (*e.g.*, two-objective feasibility problems), limited to fixed-confidence formulations, or relied on uniform sampling strategies with conservative performance guarantees. Most of the existing algorithms were designed under the assumption of uncorrelated objectives, or were not able to exploit potential correlation that may arise for example when both toxicity and efficacy are recorded.

In contrast to the well-understood single-objective setting, several fundamental questions were still open: How to extend instance-dependent analyses to the fixed-budget regime? How to adaptively allocate samples across arms to solve large-scale, high-dimensional Pareto problems? How to design an asymptotically optimal and computationally-efficient algorithm for PSI? How to handle additional structural or feasibility constraints that naturally arise in clinical and engineering applications?

The next section introduces the methodological and theoretical contributions of this dissertation, which aim to fill these gaps by developing adaptive, efficient, and interpretable algorithms for multi-objective exploration.

1.4 Contributions of This Dissertation

Building on the gaps identified in the previous section, this chapter presents the main contributions of the dissertation. It develops new theoretical and algorithmic tools for *Pareto Set Identification* (PSI) and related *pure exploration* problems in stochastic bandits, with a focus on multi-objective and constrained formulations motivated by early-stage clinical trials. The section is organized around three broad themes: (i) the fixed-budget setting, where the sampling budget is prescribed in advance, (ii) the fixed-confidence setting, where sampling continues until sufficient evidence is collected, and (iii) structured and constrained extensions of PSI. Each of these directions addresses an open question and is detailed in the following paragraphs.

1.4.1 Fixed-Budget PSI

This section summarizes our work on the fixed-budget formulation of Pareto Set Identification (PSI) presented in Kone, Kaufmann, et al. 2024. This constitutes the first theoretical and algorithmic analysis of PSI when the total number of samples is fixed in advance.

In many experimental settings, particularly in early-stage vaccine trials, the total number of enrolled subjects is predetermined. The learner must therefore allocate a fixed budget T across K arms and recommend an estimated Pareto set $\widehat{S}_T$ after exactly T samples. The goal is to minimize the probability of misidentification

$$e_T(\nu) = \mathbb{P}_\nu(\widehat{S}_T \neq \mathcal{S}^\star(\mu)) \,,$$

and to characterize how quickly $e_T(\nu)$ decays with T.

While the fixed-confidence regime had been studied, *e.g.*, in Zuluaga, Sergent, et al. 2013; Auer et al. 2016; Zuluaga, Krause, et al. 2016, no results were available in the fixed-budget setting. To bridge this gap, we propose *Empirical Gap Elimination* (EGE), a family of elimination-based algorithms that generalizes classical strategies such as Successive Rejects and Successive Halving to the multi-objective setting.

The design of EGE relies on an empirical reformulation of the suboptimality gaps introduced by Auer et al. 2016. Recall that these gaps depend on the true Pareto set $\mathcal{S}^\star(\mu)$ and therefore

were difficult to estimate directly. We show that these gaps (as recalled in Eq. (1.5)) can be equivalently rewritten as

$$\Delta_i = \max\{\Delta_i^\star, \delta_i^\star\} \,,$$

where we have

$$\begin{aligned} \Delta_i^\star &\coloneqq \max_{j\in[K]} \mathrm{m}(i,j) \\ \delta_i^\star &\coloneqq \min_{j\in[K]\setminus\{i\}} \left[\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta_j^\star)_+)\right] \,, \end{aligned}$$

which makes them amenable to empirical estimation without prior knowledge of $\mathcal{S}^\star(\mu)$. EGE maintains at each phase r an active set of arms $\mathcal{A}_r$ and empirical means $(\hat{\mu}_{i,r})_{i\in\mathcal{A}_r}$, and computes the empirical gaps $\widehat{\Delta}_{i,r}$ obtained by replacing μ with $\hat{\mu}$ and $[K]$ with $\mathcal{A}_r$. Arms with large empirical gaps are progressively eliminated with a particular tie-breaking rule.

The budget T is split across successive elimination rounds. In EGE-SR, the sampling budget is divided into $K-1$ phases of increasing length, and one arm is removed per phase. In EGE-SH, the budget is divided geometrically and approximately half of the remaining arms are eliminated at each round. Both strategies share the same elimination criterion but differ in their allocation schedule.

Our analysis establishes the first information-theoretic lower bound for PSI in the fixed-budget setting. For any adaptive algorithm, there exists a universal constant $c > 0$ and a class of instances $\widetilde{\mathcal{D}}^K$ such that

$$\forall\, \nu \,\in\, \widetilde{\mathcal{D}}^K \,, \liminf_{T\to\infty} -\frac{1}{T}\log e_T(\nu) \;\leqslant\; \frac{c}{H_2(\mu)},$$

where

$$H_2(\nu) \;\coloneqq\; H_2(\mu) \;\coloneqq\; \max_{k\in[K]} k\,\Delta_{(k)}^{-2}, \tag{1.7}$$

and $k \mapsto (k)$ is the permutation such that $\Delta_{(1)} \leqslant \Delta_{(2)} \leqslant \ldots \leqslant \Delta_{(K)}$. We further show that both EGE-SR and EGE-SH achieve

$$e_T(\nu) \;\leqslant\; \mathcal{O}\left(\exp\left(-c\frac{T}{H_2(\mu)\log K}\right)\right) \,,$$

which nearly matches the lower bound up to logarithmic factors, thereby characterizing the fundamental error exponent for PSI under a fixed sampling budget.

Empirical evaluations confirm the theoretical predictions: both EGE-SR and EGE-SH substantially outperform uniform allocation, particularly on unbalanced instances where dominated arms are easily identifiable.

In practice, they offer principled and adaptive allocation strategies for early-phase clinical studies, where limited sample budgets must be distributed across multiple candidate treatments and efficacy indicators.

1.4.2 Adaptive Fixed-Confidence Algorithms for PSI

In this contribution, we introduce the first fully adaptive, LUCB-style algorithm for fixed-confidence Pareto Set Identification (PSI) with finite-time guarantees. While prior works such as Auer et al. 2016 established PAC guarantees under uniform sampling, our goal is to design fully adaptive algorithms that pull one or two informative arms per round.

Formally, given K arms with d-dimensional mean vectors $\mu_1, \dots, \mu_K$, the learner must identify the Pareto set $\mathcal{S}^\star = \{i \in [K] : \nexists j,\ \mu_i \prec \mu_j\}$ with probability at least $1-\delta$, while minimizing the expected stopping time $\mathbb{E}[\tau]$. The central difficulty lies in the fact that dominance relations depend on all objectives simultaneously, making the problem inherently multi-dimensional and asymmetric. To capture this structure, we introduce an adaptive gap-based framework that extends the LUCB (Lower-Upper Confidence Bound) principle from Best Arm Identification to multi-objective bandits.

At each round t, the algorithm maintains $(1-\delta)$-level confidence boxes for each arm,

$$\mathcal{R}_{t,i}(\delta) = \prod_{c=1}^{d} \big[\hat{\mu}^c_{t,i} - \beta_i(t,\delta), \hat{\mu}^c_{t,i} + \beta_i(t,\delta)\big],$$

where $\hat{\mu}^c_{t,i}$ denotes the empirical mean on objective c, and $\beta_i(t,\delta)$ is a confidence radius calibrated via a law-of-iterated-logarithm bound. Arms whose confidence boxes are entirely dominated by those of others are eliminated, while sampling concentrates on ambiguous pairs (i,j) where the dominance relation remains uncertain. This adaptive allocation rule implicitly tracks the "most confusing" arms, in analogy to classical LUCB algorithms for single-objective identification.

We derive finite-time, instance-dependent guarantees showing that the proposed algorithms achieve near-optimal sample complexity:

$$\mathbb{E}_\nu[\tau_\delta] \ \leqslant\ \mathcal{O}\big(H(\mu)\ \log(dH(\mu)/\delta)\big),$$

where $H(\nu) \coloneqq \sum_{i=1}^{K} \Delta_i^{-2}$ is the complexity term defined from the Pareto gaps in (1.6). This matches the information-theoretic lower bound up to logarithmic factors, establishing the first finite-time optimality guarantees for PSI. Our analysis further characterizes the adaptive evolution of the active set and provides confidence-based stopping rules that ensure δ-PAC correctness without knowledge of instance parameters.

In addition, we introduce a relaxed formulation called the *k-Pareto set identification* problem, in which the learner returns a set $\widehat{S}$ of k elements satisfying $\widehat{S} \subseteq \mathcal{S}^\star$ (or the full Pareto set if $k \geqslant |\mathcal{S}^\star|$).

This relaxation handles practical situations where only a subset of the Pareto set needs to be identified. It is particularly meaningful in clinical studies, where one may only need to shortlist a few optimal strategies for further testing. We show that its complexity $H^k(\nu) \leqslant H(\nu)$ yields provably faster identification.

Empirically, our algorithms substantially outperform uniform and non-adaptive baselines, particularly when the number of arms is large. These results establish the first fully adaptive

algorithm for fixed-confidence PSI, bridging the methodological gap between classical LUCB-style exploration and PSI.

1.4.3 PSI with Linear Structure

In this contribution, we introduce and analyze *Pareto Set Identification (PSI) with linear structure*. This setting generalizes classical (unstructured) PSI by assuming that all arms share a low-dimensional parameterization: each arm $a \in [K]$ is associated with a known feature vector $x_a \in \mathbb{R}^h$, and its expected outcome vector is linear in an unknown parameter $\theta \in \mathbb{R}^{h \times d}$, that is

$$\mu_a = \theta^\intercal x_a, \qquad \theta \in \mathbb{R}^{h \times d} \text{ and } x_a \in \mathbb{R}^h,$$

where each arm a has σ-subgaussian marginals. This structural assumption induces correlations between arms, allowing information gathered from one arm to benefit all others.

The goal remains to identify the Pareto-optimal set $\mathcal{S}^\star(\mu)$ with probability at least $1 - \delta$, while minimizing the number of samples. In contrast to previous unstructured algorithms that sample arms uniformly or according to empirical dominance relations, we design a fully *structure-aware* sampling strategy guided by principles of optimal experimental design (Soare et al. 2014).

At the heart of our approach lies a connection between PSI estimation and G-optimal design. Let $V(w)$ denote the matrix associated with an allocation $w \in \triangle_K$, defined as

$$V(w) = \sum_{i=1}^{K} w_i \, x_i x_i^\top.$$

The key observation is that if arms are pulled according to a static allocation w, and an empirical least-squares estimate of θ is built, the uncertainty (covariance) on each arm's estimated mean $\hat{\mu}_a = \hat{\theta}^\top x_a$ is governed by $\|x_a\|^2_{V(w)^{-1}} \cdot \Sigma$, where Σ is the (assumed common) covariance matrix of each arm's marginals. An allocation is thus informative if it minimizes the worst-case prediction variance across the set of relevant arms, that is,

$$\rho^\star = \operatorname*{argmin}_{w \in \triangle_K} \max_{a \in [K]} \|x_a\|^2_{V(w)^{-1}}.$$

This optimization corresponds exactly to the classical G-optimal design criterion (Fiez et al. 2019).

Adaptive algorithm and analysis. We design an elimination-style algorithm that estimates θ using a G-optimal design, and then estimate the empirical PSI gaps similarly to the technique used in Section 1.4.1. We show that the resulting algorithm enjoys instance-dependent guarantees: given $\delta \in (0, 1)$, its sample complexity is upper bounded by

$$\mathcal{O}\left(\inf_{w \in \triangle_K} \max_{k \in [K]} \frac{\|x_k\|^2_{V(w)^{-1}}}{\Delta_k^2} \log \frac{Kd}{\delta \Delta_{\min}} \cdot \log \frac{1}{\Delta_{\min}} \right)$$

with probability at least $1-\delta$. Further refinement allows one to upper bound the sample complexity by

$$\log_2(2/\Delta_{(1)}) + \mathcal{O}\left(\sum_{i=2}^{h} \frac{\sigma^2}{\Delta_{(i)}^2} \log\left(\frac{Kd}{\delta}\log_2\left(\frac{2}{\Delta_{(i)}}\right)\right)\right),$$

where $(\cdot)$ is a re-indexing such that $\Delta_{\min} := \Delta_{(1)} \leqslant \Delta_{(2)} \leqslant \ldots \leqslant \Delta_{(K)}$. Thus, the sample complexity only depends on the h-smallest gaps, which is a major improvement when $h \ll K$.

In terms of fixed-budget guarantees, we prove that the probability of misidentification of the Pareto set for the proposed algorithm is upper-bounded by

$$\exp\left(-\frac{T}{c\sigma^2 H_{2,\text{lin}}(\nu)\lceil\log_2 h\rceil} + \log C(h,d,K)\right),$$

where $C(h,d,K) = 2d\,(K+h+\lceil\log_2 h\rceil)$, c is an absolute constant and

$$H_{2,\text{lin}}(\nu) \;:=\; \max_{k\in[h]} k\,\Delta_{(k)}^{-2},$$

which is a major improvement over Eq. (1.7), again when $h \ll K$.

From an application perspective, this structured approach provides a natural way to share information across correlated treatments or dosages, *e.g.*, where key components of those treatments are correlated (the vaccine platform, dosages, schedule, or formulation, etc.).

1.4.4 Asymptotically Optimal Algorithms for Fixed-Confidence PSI

The first algorithms introduced for fixed-confidence PSI provided only near-optimal sample complexity guarantees, with non-negligible gaps between upper and lower bounds. In contrast, the Best Arm Identification (BAI) literature has established sharp asymptotic characterizations: information-theoretic lower bounds (Garivier & Kaufmann 2016) and matching algorithms such as Track-and-Stop (Garivier & Kaufmann 2016) or Bayesian approaches (Russo 2016; Jourdan, Degenne & Kaufmann 2023) achieve the optimal expected sample complexity

$$\mathbb{E}_\nu[\tau_\delta] = T^\star(\nu)\,\log(1/\delta)\,(1+o(1)).$$

Reaching a similar level of optimality in PSI is significantly more challenging, since the learner must classify all arms simultaneously and the space of alternative hypotheses is combinatorial. Hence, unlike the single-objective case, no algorithm was known to achieve *asymptotic instance-optimality* for PSI.

Information-theoretic lower bound. Let $\mathcal{I} \subset \mathbb{R}^d$ denote the set of possible mean vectors, and $\mathcal{S}^\star(\lambda)$ the Pareto set associated with the collection $\{\lambda_1,\ldots,\lambda_K\}$. To guarantee

δ-correctness, an algorithm must statistically distinguish the true parameter μ from every alternative $\lambda \in \mathcal{I}^K$ whose Pareto set differs from $\mathcal{S}^\star(\mu)$. Define the set of such alternatives as

$$\mathrm{Alt}(\mathcal{S}^\star(\mu)) \;=\; \big\{\lambda \in \mathcal{I}^K : \mathcal{S}^\star(\lambda) \neq \mathcal{S}^\star(\mu)\big\}.$$

Using information-divergence arguments analogous to those of Garivier & Kaufmann 2016, we obtain the following lower bound on the expected stopping time of any δ-PAC PSI algorithm.

Theorem 1.4.1 (Fixed-confidence lower bound). *Let $\mathcal{D} := \{\mathcal{N}(\mu, \Sigma) : \mu \in \mathcal{I}\}$ denote the set of d-dimensional multivariate Gaussian arms with known covariance Σ and mean vectors in $\mathcal{I}$. Any algorithm that is δ-correct on all bandit instances in $\mathcal{D}^K$ satisfies, for all ν parameterized by $\mu \in \mathcal{I}^K$,*

$$\mathbb{E}_\nu[\tau_\delta] \;\geqslant\; T^\star(\nu)\, \log\Big(\tfrac{1}{2.4\delta}\Big),$$

where the characteristic time $T^\star(\nu)$ *is defined by*

$$T^\star(\nu)^{-1} \;:=\; \sup_{w \in \triangle_K} \inf_{\lambda \in Alt(\mathcal{S}^\star(\mu))} \left[\sum_{i=1}^K \frac{w_i}{2} \big\|\mu_i - \lambda_i\big\|^2_{\Sigma^{-1}}\right].$$

An algorithm is said to be *asymptotically optimal* if

$$\lim_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log(1/\delta)} = T^\star(\nu).$$

This bound characterizes the minimal information required to distinguish the true Pareto set from all incorrect ones. In BAI, algorithms achieving such tight guarantees typically rely on *generalized likelihood-ratio* (GLR) stopping rules, which admit closed-form expressions for exponential-family models. For PSI, this stopping rule generalizes to

$$\tau \;=\; \inf\left\{ t \geqslant 1 : \inf_{\lambda \in \mathrm{Alt}(\widehat{S}_t)} \sum_{k=1}^K N_{t,k} \left\|\hat{\mu}_{t,k} - \lambda_k\right\|^2_{\Sigma^{-1}} > \beta(t,\delta) \right\},$$

but the inner minimization over $\lambda \in \mathrm{Alt}(\widehat{S}_t)$ has no known closed-form solution, making direct implementation intractable for PSI, due to its combinatorial and non-convex structure.

Posterior-resampling stopping rule. To circumvent this limitation, we propose a posterior-based alternative inspired by Bayesian sequential testing. At time t, let P_t denote the posterior distribution over θ given the observed history $\mathcal{H}_{t-1}$, and let $\theta_t^1, \theta_t^2, \ldots$ be *i.i.d.* samples from P_t. Fix a recommendation $\widehat{S}_t = \mathcal{S}^\star(\hat{\mu}_t)$ and a threshold $M(t,\delta)$. We define the *posterior geometric stopping rule* as

$$\tau \;=\; \inf\Big\{ t \geqslant 1 : \forall\, m \leqslant M(t,\delta),\; \mathcal{S}^\star(\theta_t^m) = \widehat{S}_t \Big\}, \tag{1.8}$$

that is, we stop once $M(t,\delta)$ consecutive posterior samples yield the same optimal set as the current recommendation.

By proving novel anti-concentration results for Gaussian vectors, we calibrate the threshold $M(t, \delta)$ to ensure δ-correctness.

This rule is conceptually close to Chernoff's GLR stopping: both trigger when the posterior (or likelihood) ratio between the true and alternative hypotheses becomes sufficiently large. However, unlike GLR-based tests, it does not require explicit evaluation of the posterior probabilities–only the ability to draw posterior samples, which is computationally straightforward for standard conjugate models. This makes it particularly attractive for high-dimensional PSI problems where $\mathrm{Alt}(\mathcal{S}^\star(\mu))$ has a complex combinatorial structure.

Main result. We combine this stopping rule with a randomized adaptive allocation strategy based on posterior samples. We prove that the resulting algorithm is asymptotically instance-optimal:

$$\mathbb{E}_\nu[\tau_\delta] \;=\; T^\star(\nu)\,\log(1/\delta)\,\big(1+o(1)\big),$$

and achieves optimal posterior contraction rates. This constitutes the first successful application of posterior sampling to multi-objective pure exploration. Beyond theory, the algorithm is computationally practical and statistically efficient, providing a randomized and interpretable design principle for early-phase clinical trials where decisions must be both adaptive and provably optimal.

1.4.5 PSI with Linear Feasibility Constraints

In many real-world applications, not all arms are admissible: a vaccine dose may need to satisfy toxicity thresholds, a hardware configuration must respect power or latency limits, or a treatment must achieve a minimum efficacy level before being considered. Such constraints are often linear in nature, motivating the definition of the feasible set

$$P = \{x \in \mathbb{R}^d : Ax \leqslant b\},$$

with $A \in \mathbb{R}^{m\times d}$ and $b \in \mathbb{R}^m$. Given this set, the learner must identify the subset of arms that are both feasible and Pareto-optimal among feasible ones:

$$\mathcal{S}^\star_{\text{feas}}(\mu) = \Big\{ i \in [K] : \mu_i \in P \text{ and } \nexists j \in [K],\ \mu_j \in P,\ \mu_i \prec \mu_j \Big\}.$$

Compared to unconstrained PSI, this setting introduces a dual challenge: each arm must be simultaneously classified by both *feasibility* and *dominance*. The sample complexity is thus driven by two quantities: the Pareto gap Δ_i, characterizing the uncertainty about optimality, and the feasibility margin

$$\eta_i \;:=\; \begin{cases} \mathrm{dist}(\mu_k, P^c), & \text{if } \mu_k \in P, \\ \mathrm{dist}(\mu_k, \partial P), & \text{otherwise.} \end{cases},$$

which measures how close an arm's mean lies to the constraint boundary.

We propose the *explainable constrained Adaptive Pareto Exploration* (e-cAPE) algorithm, which generalizes the APE algorithm to constrained environments.

The algorithm recommends a partition $R_\tau = (O_\tau, S_\tau, I_\tau)$ such that

$$i)\ O_\tau = \mathcal{S}_{\text{feas}}(\mu), \quad ii)\ S_\tau \subseteq \text{SubOpt}(\mu) \text{ and } I_\tau \subseteq \{k : \mu_k \notin P\} \tag{1.9}$$

holds at stopping time τ, where (i) $\mathcal{S}_{\text{feas}}(\mu)$ is the Pareto set of $\{\mu_k : \mu_k \in P, k \in [K]\}$, $\text{SubOpt}(\mu) := \{i \in [K] \mid \exists j \text{ such that } \mu_i \prec \mu_j \text{ and } \mu_j \in P\}$ be the set of arms that are dominated by a feasible arm.

Since some arms can be both infeasible and dominated by a feasible arm, such arms could be correctly classified either in S_τ or I_τ. Therefore, multiple correct ways to partition $[K]$ may exist, so multiple correct partitions (O_τ, S_τ, I_τ). Given the polyhedron P, we define the set of valid answers for (S_τ, I_τ) as $\mathcal{M}(P, \mu)$.

The learner maintains empirical estimates of both feasibility and Pareto gaps, and dynamically allocates samples according to their estimated classification difficulty. At each iteration, arms near the Pareto-feasibility boundary are targeted. This mechanism achieves a near-optimal balance between exploration of uncertain arms and exploitation of structural information encoded by P.

We proved that e-cAPE correctly identifies a partition in the fixed-confidence setting, with sample complexity characterized by a constrained instance complexity term

$$\mathbb{E}_\nu[\tau_\delta] \lessapprox \mathcal{O}(C^\star_{\mathcal{M}}(\nu) \log(KdC^\star_{\mathcal{M}}(\nu)/\delta)), \quad \text{(for small } \delta\text{)},$$

where

$$C(\nu, S, I) := \sum_{i \in \mathcal{S}_{\text{feas}}} \frac{1}{\min(\Delta_i^2(\mathcal{S}_{\text{feas}} \cup S), \eta_i^2)} + \sum_{i \in S} \frac{1}{\Delta_i^2(\mathcal{S}_{\text{feas}} \cup S)} + \sum_{i \in I} \frac{1}{\eta_i^2}, \tag{1.10}$$

and, for $\mathcal{X} \subseteq [K]$, $\Delta_i^2(\mathcal{X})$ denotes the PSI gaps computed in the $|\mathcal{X}|$-armed bandit with means $\{\mu_k : k \in \mathcal{X}\}$. $C^\star_{\mathcal{M}}$ is the complexity of the *easiest to verify* partition:

$$C^\star_{\mathcal{M}}(\nu) := \min_{(O,S,I) \in \mathcal{M}(P,\mu)} C(\nu, S, I).$$

We further establish an information-theoretic lower bound of the same order, proving that the dependence on $C^\star_{\mathcal{M}}$ is unavoidable.

Finally, experiments on synthetic and realistic multi-objective datasets (Mark C et al. 2013) confirm that e-cAPE significantly outperforms uniform and unconstrained sampling, especially when feasible arms lie near the boundary of P. In practice, such constrained exploration directly models the setting of early-stage vaccine development, where identifying acceptable trade-offs between immunogenicity and safety is more realistic than seeking a single optimal dose.

1.5 Outline of the Dissertation

The dissertation is organized into five technical chapters plus a conclusion chapter. Each technical chapter addresses a distinct aspect of Pareto set identification in multi-objective bandits. The progression follows a natural path, from gap-based and finite-sample analyses to asymptotically optimal, constrained, and structured extensions.

- **Chapter 2** introduces the *fixed-budget* formulation of PSI, relevant when the sampling horizon is predetermined. It presents the *Empirical Gap Elimination* (EGE) framework, extending Successive Rejects and Sequential Halving to multi-objective settings, and establishes the first information-theoretic lower bound for PSI under fixed budget.
- **Chapter 3** addresses PSI in the *fixed-confidence* regime. It develops adaptive elimination algorithms with near-optimal instance-dependent sample complexity and discusses relaxations such as k-Pareto set identification, where the learner outputs a manageable representative subset of the Pareto-optimal arms.
- **Chapter 4** extends PSI to *multi-output linear models*, where arm means depend linearly on known features. Exploiting linear structure, we design sample-efficient algorithms for structured action spaces, connecting PSI estimation with optimal experimental design.
- **Chapter 5** studies *Bayesian algorithms for PSI*, based on posterior sampling and geometric stopping rules. These algorithms achieve asymptotic instance-optimality while remaining computationally efficient, offering a Bayesian randomized algorithm for PSI.
- **Chapter 6** introduces *constrained PSI*, where feasible arms must satisfy linear inequalities (*e.g.*, safety or resource constraints). We develop adaptive algorithms that jointly assess feasibility and dominance and derive complexity measures that couple feasibility margins with Pareto gaps.
- **Chapter 7** concludes the dissertation, outlining future research directions, including large-scale PSI, preference-driven objectives, and extensions to reinforcement learning and other sequential decision frameworks.

Each chapter is largely self-contained, but contributes to a unified view of multi-objective pure exploration, bridging theoretical foundations, algorithmic design, and experiments toward applications to adaptive clinical settings such as vaccine trials.

1.6 Publications

The chapters of this dissertation are primarily based on peer-reviewed papers I co-authored during this doctoral work. Each contribution has been published in international machine learning conferences, reflecting different facets of the Pareto set identification problem.

Conference Papers

- **C. Kone**, E. Kaufmann, L. Richert.
  Adaptive Algorithms for Relaxed Pareto Set Identification.
  Proceedings of the Thirty-Seventh Conference on Neural Information Processing Systems (NeurIPS 2023).

- **C. Kone**, E. Kaufmann, L. Richert.
  Bandit Pareto Set Identification: The Fixed-Budget Setting.
  Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS 2024).

- **C. Kone**, E. Kaufmann, L. Richert.
  Bandit Pareto Set Identification in a Multi-Output Linear Model.
  Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS 2025).

- **C. Kone**, M. Jourdan, E. Kaufmann.
  Pareto Set Identification with Posterior Sampling.
  Proceedings of the 28th International Conference on Artificial Intelligence and Statistics (AISTATS 2025).

- **C. Kone**, E. Kaufmann, L. Richert.
  Constrained Pareto Set Identification with Bandit Feedback.
  Proceedings of the Forty-Second International Conference on Machine Learning (ICML 2025).

Together, these publications form the core of the dissertation:

- Chapter 2–*Fixed-Budget PSI*–is based on our AISTATS 2024 paper introducing the Empirical Gap Elimination framework, the first theoretical treatment of PSI in this regime.

- Chapter 3–*Adaptive Algorithms for Fixed-Confidence PSI*–draws on our NeurIPS 2023 work proposing LUCB-type adaptive elimination strategies and relaxed objectives such as PSI-k.

- Chapter 4–*Multi-Output Linear Models for PSI*–follows our AISTATS 2025 paper extending PSI to structured linear models via optimal experimental design.

- Chapter 5–*Bayesian PSI*–is based on our AISTATS 2025 paper on posterior-sampling-based algorithms achieving asymptotic optimality.

- Chapter 6–*Constrained PSI*–builds on our ICML 2025 paper introducing feasibility-aware Pareto identification with linear constraints.

Additional Publications

- **C. Kone**, K. Jamieson.
  Optimal Posterior Sampling for Policy Identification in Tabular Markov Decision Processes.
  Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS 2026).

  Although not directly focused on PSI, this work builds on the posterior sampling principles developed in Chapter 5 to design computationally efficient algorithms for policy identification in reinforcement learning.

Chapter 2

Pareto Set Identification: The Fixed-Budget Setting

This chapter addresses the problem of *Pareto Set Identification* in the *fixed-budget* setting. Our motivation stems from applications such as clinical trials, where the number of participants is pre-specified and the exploration budget is thus limited. The goal is then to identify the Pareto set with minimal error.

We propose *Empirical Gap Elimination* (EGE), a family of algorithms that combine (i) a principled estimation of the difficulty of classifying each arm as Pareto-optimal or not, with (ii) elimination-style strategies tailored to the fixed-budget regime. We provide a theoretical analysis showing that two concrete instances, EGE-SR and EGE-SH, achieve an error probability that decreases exponentially with the budget, with an exponent supported by an information-theoretic lower bound.

This chapter is based on joint work with Émilie Kaufmann and Laura Richert, published in the proceedings of *AISTATS 2024*.

2.1 Introduction

As discussed in Chapter 1, the multi-armed bandit problem has been predominantly studied as a single-objective stochastic optimization problem, mainly for reward maximization, (Lattimore & Szepesvari 2020) or for identifying the arm with the largest expected value (best arm identification, (Even-Dar et al. 2002; Audibert & Bubeck 2010)). However, many practical problems involve multiple, possibly conflicting objectives. This motivates the study of multi-objective pure exploration, where the goal is to identify the set of arms whose mean performance vectors are Pareto optimal.

Previous work has mainly focused on the *fixed-confidence* formulation of Pareto Set Identification (PSI) (Zuluaga, Sergent, et al. 2013; Auer et al. 2016; Zuluaga, Krause, et al. 2016; Ararat & Tekin 2023): given a risk $\delta \in (0, 1)$, the agent stops at a random time τ and recommends a set $\widehat{S}_\tau$ such that $\mathbb{P}(\widehat{S}_\tau \neq \mathcal{S}^\star) \leqslant \delta$, while minimizing its stopping time τ.

In contrast, in the dual *fixed-budget* formulation, the total number of samples T is fixed in advance and the goal is to minimize the misidentification probability. This unexplored setting is particularly relevant in applications such as adaptive clinical trials, where the number of participants is known in advance due to ethical and/or financial constraints. In such contexts, the challenge is to allocate samples efficiently across competing treatment arms to maximize the probability of correctly identifying the most promising vaccine or drug combinations. Beyond clinical trials, similar trade-offs arise in multi-objective design problems such as optimizing hardware configurations for both latency and energy efficiency (Zuluaga, Sergent, et al. 2013), or in A/B/n testing for recommender systems involving multiple engagement metrics (Mehrotra et al. 2020).

In this chapter, we introduce the first algorithms for Pareto Set Identification in the fixed-budget setting. Our approach extends the principles of Successive Rejects (Audibert & Bubeck 2010) and Sequential Halving (Karnin et al. 2013); two canonical methods for best-arm identification, to the multi-objective case. We propose the *Empirical Gap Elimination* (EGE) framework, which relies on a novel arm-classification rule based on *empirical gaps*.

To handle practical situations where only a subset of the Pareto set needs to be identified, we further introduce the *k-relaxation* of PSI, denoted PSI-k. This relaxation requires the learner to return any subset of k Pareto-optimal arms, thereby interpolating between identifying a single good arm ($k = 1$) and recovering the full Pareto set ($k = K$). Such a formulation is particularly meaningful in clinical studies, where one may only need to shortlist a few good-performing strategies for further testing.

We show that the proposed algorithms, EGE-SR and EGE-SH, achieve instance-dependent exponential convergence rates matching known lower bounds up to logarithmic factors (Section 2.2).

Related work. Building on the context introduced in Chapter 1, where we reviewed the theoretical and practical motivations for pure exploration in multi-objective bandits, this chapter focuses on the fixed-budget formulation of Pareto Set Identification (PSI). While several works have addressed PSI in the fixed-confidence setting (Zuluaga, Sergent, et al.

2013; Auer et al. 2016; Zuluaga, Krause, et al. 2016; Ararat & Tekin 2023), no prior work provides theoretical guarantees for PSI when the sampling budget is fixed, despite its practical relevance in resource-limited domains such as clinical trials or recommender systems. In the single-objective case, classical algorithms like Successive Rejects and Sequential Halving achieve exponential decay of the misidentification probability with the budget (Audibert & Bubeck 2010; Karnin et al. 2013), yet no single strategy attains a universally optimal rate across all instances (Degenne 2023).

A common approach to extend fixed-confidence methods to the fixed-budget setting is to tune their risk parameter δ according to the available budget T and the instance hardness $H(\nu)$, as in the UGapEb algorithm (Gabillon et al. 2012).

Indeed, a fixed-confidence algorithm for PSI with stopping time τ typically takes a risk parameter $\delta \in (0,1)$ as input and satisfies guarantees of the form

$$\mathbb{P}\left(\widehat{S}_\tau = \mathcal{S}^\star \text{ and } \tau \leqslant C\, H(\nu) \log\left(d\frac{H(\nu)}{\delta}\right)\right) \geqslant 1-\delta,$$

where $\widehat{S}_\tau$ denotes the estimated Pareto set and $C > 0$ is a universal constant. Given a fixed sampling budget T, one may attempt to tune δ as a function of T and the instance hardness $H(\nu)$ by choosing δ_T such that $C\, H(\nu) \log\big(H(\nu)/\delta_T\big) \leqslant T$. Under this calibration, the corresponding error probability behaves approximately as

$$\delta_T \approx H(\nu)\, \exp\left(-\tfrac{T}{C\, H(\nu)}\right).$$

However, such oracle tuning presupposes knowledge of $H(\nu)$, which is rarely available in practice. In this chapter, we overcome this limitation by introducing the EGE framework for fixed-budget PSI. Our method adaptively estimates the classification difficulty of each arm and integrates these estimates into elimination-based allocation rules, leading to two practical algorithms: EGE-SR and EGE-SH.

Learning model. We consider a stochastic multi-armed bandit model with K arms, where each arm $a \in [K]$ is associated with an unknown probability distribution ν_a supported on $\mathbb{R}^d$, $d \geqslant 1$. The mean vector of arm a is denoted $\mu_a = \mathbb{E}_{X\sim\nu_a}[X] \in \mathbb{R}^d$, and we write $\mu = (\mu_1, \dots, \mu_K)$. Throughout, we assume subgaussian marginals: for every arm a and each coordinate $c \in [d]$, the random variable X_a^c (with $X_a \sim \nu_a$) is σ-subgaussian, that is,

$$\forall\, \lambda \in \mathbb{R}, \qquad \log \mathbb{E}[\exp(\lambda(X_a^c - \mathbb{E}[X_a^c]))] \leqslant \frac{\lambda^2\sigma^2}{2},$$

for some common variance proxy $\sigma^2 > 0$.

The learner interacts with this environment over R rounds, sequentially selecting arms from an active set $\mathcal{A}_r \subseteq [K]$ and observing independent samples $(Z_{r,a})_{a\in\mathcal{A}_r}$, where each $Z_{r,a} \sim \nu_a$ satisfies $\mathbb{E}[Z_{r,a} \,|\, a] = \mu_a$. Let $\mathbb{P}_\nu$ and $\mathbb{E}_\nu$ denote respectively the probability law and expectation under the joint process $\{Z_{r,a}\}_{r,a}$, and define the filtration $\mathcal{H}_r = \sigma(\mathcal{A}_1, Z_1, \dots, \mathcal{A}_r, Z_r)$, representing the information available up to round r. A sampling strategy is said to be adaptive if each selection A_t is $\mathcal{H}_{t-1}$-measurable. Given a

total sampling budget T, the learner outputs an estimator $\widehat{S}_T$ of the true Pareto-optimal set $\mathcal{S}^\star(\nu)$, based on the full observation history $\mathcal{H}_T$. The performance of a strategy is measured by its probability of misidentification

$$e_T(\nu) \coloneqq \mathbb{P}_\nu\Big(\widehat{S}_T \neq \mathcal{S}^\star(\nu)\Big) \ ,$$

which quantifies the probability of returning an incorrect Pareto set under a fixed sampling budget.

Definition 2.1.1. Given two arms $i, j \in [K]$, i is weakly (Pareto) dominated by j (denoted by $\mu_i \leqslant \mu_j$) if for any $c \in \{1, \dots, d\}$, $\mu_i^c \leqslant \mu_j^c$. The arm i is (Pareto) dominated by j ($\mu_i \preceq \mu_j$ or $i \preceq j$) if i is weakly dominated by j and there exists $c \in \{1, \dots, d\}$ such that $\mu_i^c < \mu_j^c$. The arm i is strictly (Pareto) dominated by j ($\mu_i \prec \mu_j$ or $i \prec j$) if for any $c \in \{1, \dots, d\}$, $\mu_i^c < \mu_j^c$.

The Pareto set $\mathcal{S}^\star(\nu)$ is defined as

$$\mathcal{S}^\star(\nu) \coloneqq \{i \in [K] \textit{s.t.} \nexists\, j \in [K] : \mu_i \prec \mu_j\} \, ,$$

and will be denoted by $\mathcal{S}^\star$ when ν is clear from the context. Any arm $a \in \mathcal{S}^\star$ will be called (Pareto) *optimal* and an arm $a \notin \mathcal{S}^\star$ is called *sub-optimal*.

To characterize the difficulty of the problem, we introduce below some quantities that allow us to measure how much an arm is dominated. For any arms i, j, we let

$$\begin{aligned} \mathrm{m}(i,j) &\coloneqq \min_{c \leqslant d} \left[\mu_j^c - \mu_i^c\right] , \\ \mathrm{M}(i,j) &\coloneqq \max_{c \leqslant d} \left[\mu_i^c - \mu_j^c\right] . \end{aligned}$$

If $\mu_i \not\prec \mu_j$, then $\mathrm{M}(i,j)$ is the smallest uniform increase of j that makes it dominate i. If $\mu_i \prec \mu_j$ then $\mathrm{m}(i,j)$ is the smallest increase of any component of i which makes it non-dominated by j.

We recall the gaps introduced in Chapter 1 and proven (Auer et al. 2016) to characterize fixed-confidence PSI. For a sub-optimal arm $i \notin \mathcal{S}^\star$,

$$\Delta_i \coloneqq \Delta_i^\star \coloneqq \max_{j \in \mathcal{S}^\star} \mathrm{m}(i,j) \, , \tag{2.1}$$

which is the smallest quantity that should be added component-wise to μ_i to make i appear Pareto optimal *w.r.t.* $\{\mu_k : k \in [K] \backslash \{i\}\}$. For an optimal arm $i \in \mathcal{S}^\star$,

$$\Delta_i \coloneqq \min\left\{\delta_i^+, \delta_i^-\right\} \tag{2.2}$$

where

$$\begin{aligned} \delta_i^+ &\coloneqq \min_{j \in \mathcal{S}^\star \backslash \{i\}} \min\left[\mathrm{M}(i,j), \mathrm{M}(j,i)\right] \ , \\ \delta_i^- &\coloneqq \min_{j \in [K] \backslash \mathcal{S}^\star} \left[(\mathrm{M}(j,i))_+ + \Delta_j\right] \ , \end{aligned}$$

with the convention $\min_{\emptyset} = +\infty$. δ_i^+ accounts for how much i is close to dominate (or to be dominated by) another optimal arm, while δ_i^- translates, in a sense, the smallest "margin" from an optimal arm i to the sub-optimal arms. The difficulty (*i.e.*, near-optimal sample complexity) of fixed-confidence PSI is characterized by the complexity term

$$H(\nu) \coloneqq \sum_{a=1}^{K} \frac{1}{\Delta_a^2} \,.$$

In this chapter, we show that it is also a relevant complexity measure for the fixed-budget setting. Our goal is to propose elimination algorithms that estimate the gaps of each arm online and sequentially discard and classify arms with larger gaps.

Toward this goal, we begin by rewriting and simplifying the gaps identified by Auer et al. 2016, which, in their current form, directly depend on the Pareto set, making them apparently difficult to estimate online. Recalling that $(x)_+ \coloneqq \max\{x, 0\}$, the following holds.

Lemma 2.1.2. *For any arm $k \in [K]$,*

$$\Delta_k = \begin{cases} \Delta_k^\star = \max_{j \in [K]} \mathrm{m}(k,j) & \text{if } k \notin \mathcal{S}^\star \\ \delta_k^\star & \text{if } k \in \mathcal{S}^\star \end{cases} \,,$$

where $\delta_k^\star \coloneqq \min_{j \neq k} \left[\mathrm{M}(k,j) \wedge (\mathrm{M}(j,k)_+ + (\Delta_j^\star)_+) \right]$. In particular, for any arm k, $\Delta_k = \max\{\Delta_k^\star, \delta_k^\star\}$.

The proof of this result is given in Section 2.5. This rewriting removes the explicit dependency on the Pareto Set. Below, we propose a generic elimination algorithm that relies on a simple empirical estimation of these gaps.

2.2 Algorithmic contributions

We introduce the family of Empirical Gap Elimination (EGE) algorithms. They are generic algorithms that take as input a number of rounds, an elimination scheme (dictating the fraction of arms discarded at the end of each round), and a sampling scheme, dictating the fraction of the budget allocated to each round.

2.2.1 Empirical Gap Elimination

An Empirical Gap Elimination algorithm uses a round-based structure. The algorithm is parameterized by a number of rounds $R > 0$, an arm schedule vector $\lambda = (\lambda_1, \ldots, \lambda_R, \lambda_{R+1}) \in [K]^{R+1}$ satisfying $\lambda_r > \lambda_{r+1}$ where λ_r indicates how many arms are active in round r, and an allocation vector $t = (t_1, \ldots, t_R) \in [T]^R$, where t_r is the number of samples gathered

from each active arm in round r. Given the maximal budget T, these vectors should further satisfy the following:

$$\lambda_1 = K, \qquad \lambda_{R+1} \in \{0,1\} \quad \text{and} \tag{2.3}$$

$$\sum_{r=1}^{R} \lambda_r t_r \leqslant T, \tag{2.4}$$

ideally with an equality.

In each round r, the algorithm maintains a set of active arms, denoted by $\mathcal{A}_r$, that is of size λ_r, and collects t_r new samples from each arm in $\mathcal{A}_r$. The total number of samples gathered from each arm in $\mathcal{A}_r$ in the first r rounds[1] is therefore

$$n_r = \sum_{s=1}^{r} t_s.$$

For each $a \in \mathcal{A}_r$, the empirical estimate of μ_a at the end of round r is denoted by $\hat{\mu}_{r,a} \coloneqq \frac{1}{n_r}\sum_{s=1}^{n_r} Z_{a,s}$ where $Z_{a,s}$ denote the s-th observation drawn *i.i.d.* from distribution ν_a. These estimates are used to carefully decide which arm to explore in the next round, based on an appropriate notion of *empirical gap*. We introduce the empirical quantities

$$\begin{aligned} \mathrm{m}(i,j;r) &\coloneqq \min_{c \leqslant d} \left[\hat{\mu}^c_{r,j} - \hat{\mu}^c_{r,i}\right], \\ \mathrm{M}(i,j;r) &\coloneqq \max_{c \leqslant d} \left[\hat{\mu}^c_{r,i} - \hat{\mu}^c_{r,j}\right] \end{aligned}$$

and the empirical Pareto set at the end of round r

$$\begin{aligned} S_r &\coloneqq \left\{ i \in \mathcal{A}_r : \nexists\, j \in \mathcal{A}_r \text{ such that } \hat{\mu}_{r,i} \prec \hat{\mu}_{r,j} \right\}, \\ &= \left\{ i \in \mathcal{A}_r : \forall\, j \in \mathcal{A}_r \backslash \{i\},\ \mathrm{M}(i,j;r) > 0 \right\}. \end{aligned}$$

Finally, we define for any arm $i \in \mathcal{A}_r$ the empirical gaps

$$\begin{aligned} \widehat{\Delta}^{\star}_{i,r} &\coloneqq \max_{j \in \mathcal{A}_r \backslash \{i\}} \mathrm{m}(i,j;r), \\ \widehat{\delta}^{\star}_{i,r} &\coloneqq \min_{j \in \mathcal{A}_r \backslash \{i\}} \left[\mathrm{M}(i,j;r) \ \wedge\ (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^{\star}_{i,r})_+) \right], \end{aligned}$$

and for any $i \in \mathcal{A}_r$

$$\widehat{\Delta}_{i,r} \coloneqq \begin{cases} \widehat{\Delta}^{\star}_{i,r} & \text{if } i \in \mathcal{A}_r \backslash S_r, \\ \widehat{\delta}^{\star}_{i,r} & \text{else.} \end{cases} \tag{2.5}$$

These gaps are empirical variants of the gaps stated in Lemma 2.1.2 evaluated on the active arms. In particular, it is simple to observe that

$$\widehat{\Delta}_{i,r} = \max\{\widehat{\delta}^{\star}_{i,r}, \widehat{\Delta}^{\star}_{i,r}\},$$

Algorithm 2.1: EGE: Empirical Gap Elimination

Require: number of rounds R; budget allocation t; and elimination scheme λ

Initialize: Set of active arms $\mathcal{A}_1 \leftarrow \{1, \dots, K\}$; set of arms (so far) identified as Pareto-optimal $\mathcal{B}_1 \leftarrow \emptyset$; set of discarded arms $\mathcal{D}_1 \leftarrow \emptyset$;

foreach $r = 1, 2, \dots, R$ **do**

  Collect t_r samples from each arm $a \in \mathcal{A}_r$;

  Compute S_r the empirical Pareto set ;

  Let $\mathcal{A}_{r+1}$ be the set of λ_{r+1} arms in $\mathcal{A}_r$ with the smallest empirical gaps $\widehat{\Delta}_{a,r}$;

  // ties broken in favor of arms in S_r

  $\mathcal{B}_{r+1} \leftarrow \mathcal{B}_r \cup \{S_r \cap (\mathcal{A}_r \backslash \mathcal{A}_{r+1})\}$;

  $\mathcal{D}_{r+1} \leftarrow \mathcal{D}_r \cup \{(\mathcal{A}_r \backslash \mathcal{A}_{r+1}) \backslash S_r\}$;

return $\widehat{S}_T = \mathcal{B}_{R+1} \cup \mathcal{A}_{R+1}$

which follows from the fact that when $i \in S_r$, an empirically Pareto-optimal arm $\widehat{\Delta}^\star_{i,r} < 0$ and $\widehat{\delta}^\star_{i,r} \geqslant 0$. On the other side, when $i \in \mathcal{A}_r \backslash S_r$, an empirical sub-optimal arm, we have $\widehat{\delta}^\star_{i,r} < 0$ while $\widehat{\Delta}^\star_{i,r} \geqslant 0$, as it follows from their definitions.

At the end of round r, the algorithm sorts the arms by increasing order of their empirical gaps: $\widehat{\Delta}_{(1),r} \leqslant \cdots \leqslant \widehat{\Delta}_{(\lambda_r),r}$ and we define $\mathcal{A}_{r+1} = \{(1), \dots, (\lambda_{r+1})\}$. In case of ties, *i.e.*, if $\widehat{\Delta}_{(\lambda_{r+1}),r} = \widehat{\Delta}_{(\lambda_{r+1}+m),r}$ for some m, we first add arms in S_r to $\mathcal{A}_{r+1}$. We emphasize that this tie-breaking rule is crucial in our analysis. Arms in $\mathcal{A}_{r+1} \backslash \mathcal{A}_r$ are further classified as optimal (and added to the set $\mathcal{B}_{r+1}$) or sub-optimal (and added to the set $\mathcal{D}_{r+1}$) based on whether or not they belong to S_r. The output of the algorithm is the set $\mathcal{B}_{R+1} \cup \mathcal{A}_{R+1}$.

2.2.2 Particular instances

Arm elimination algorithms have been proposed for different unidimensional fixed-budget identification tasks (Audibert & Bubeck 2010; Bubeck, T. Wang, et al. 2013; Karnin et al. 2013) with different elimination rules, that could also be rewritten featuring some (simpler) gaps. In these works, two different types of arm schedules and sampling allocations have been mainly investigated:

- In Successive Rejects (SR), one arm is deactivated in each round, *i.e.*, $R = K - 1$ and $\lambda^{\text{SR}}_r = K - r + 1$ for all $r \leqslant K$. The sampling allocation proposed by Audibert & Bubeck 2010 satisfies $t^{\text{SR}}_r = n^{\text{SR}}_r - n^{\text{SR}}_{r-1}$ with $n_r = \left\lceil \frac{1}{\overline{\log}(K)} \frac{T-K}{K+1-r} \right\rceil$ where $\overline{\log}(K) := 2^{-1} + \sum_{i=2}^{K} i^{-1}$ and $n_0 = 0$.

[1] We emphasize that EGE algorithms do not discard samples between rounds

- In Sequential Halving (SH), one half of the active arms is de-activated at the end of each round, that is $R = \lceil \log_2(K) \rceil$ and for all $r \in \{1, \dots, \lceil \log_2(K) \rceil\}$, $\lambda_{r+1}^{\mathrm{SH}} := \lceil \lambda_r^{\mathrm{SH}}/2 \rceil$ (we easily verify that $\lambda_{R+1}^{\mathrm{SH}} = 1$). The sampling allocation proposed by Karnin et al. 2013 is uniformly spread across rounds, that is $t_r^{\mathrm{SH}} := \left\lfloor \frac{T}{|\mathcal{A}_r| \lceil \log_2(K) \rceil} \right\rfloor$.

We refer to EGE-SR (resp. EGE-SH) as the instances of EGE using the same allocation as SR (resp. SH). Another instantiation of EGE may use the geometric allocation proposed by Karpov & Zhang 2022, generalizing the SH-type of allocation.

For $d = 1$, the PSI problem coincides with BAI, and EGE-SR (resp. EGE-SH) coincides with SR (resp. SH[2]). Indeed, in that case $S_r = \{\hat{a}_r\}$ reduces to the empirical best arm, $\widehat{\Delta}_{i,r} = \hat{\mu}_{\hat{a}_r,r} - \hat{\mu}_{i,r}$ for $i \neq \hat{a}_r$, $\widehat{\Delta}_{\hat{a}_r,r} = \min_{i \in \mathcal{A}_r \setminus \{\hat{a}_r\}} \widehat{\Delta}_{i,r}$. Then, our tie-breaking rule ensures that no arm is accepted before the last round, and at each round r, $\mathcal{A}_{r+1}$ is defined as the λ_{r+1} arms in $\mathcal{A}_r$ with the largest empirical means, and the final survival arm is recommended as optimal.

2.2.3 Beyond exact PSI

We propose an extension of EGE-SR to the PSI-k problem. In PSI-k, the learner aims to return a subset $\widehat{\mathcal{S}}_T \subseteq \mathcal{S}^\star$ of size k, or the full Pareto set if $|\mathcal{S}^\star| < k$. In the fixed-budget setting, the performance of an algorithm is measured through the *k-relaxed expected loss*:

$$e_{T,k}(\nu) := \mathbb{E}_\nu[\mathcal{L}(\widehat{\mathcal{S}}_T, k)],$$

where

$$\mathcal{L}(\widehat{\mathcal{S}}, k) := \begin{cases} \mathbb{1}\big(\widehat{\mathcal{S}} \subset \mathcal{S}^\star\big), & \text{if } |\widehat{\mathcal{S}}| = k, \\ \mathbb{1}\big(\widehat{\mathcal{S}} = \mathcal{S}^\star\big), & \text{otherwise.} \end{cases}$$

This loss penalizes the inclusion of any sub-optimal arm while tolerating partial identification of the Pareto set up to size k.

The EGE-SR algorithm can be naturally extended to PSI-k by adjusting its stopping condition to

$$\tau := \inf\{r : |\mathcal{B}_{r+1}| = k\} \wedge (K-1),$$

where $\mathcal{B}_r$ denotes the set of arms declared optimal after round r. Let N_τ be the total number of samples drawn at termination. In Section 2.5.2, we show that when the budget is large,

$$\mathbb{E}_\nu[\tau] \lesssim q \quad \text{and} \quad \mathbb{E}_\nu[N_\tau] \lesssim N_q,$$

where $q := K - |\mathcal{S}^\star| + k$. Intuitively, this means that in the worst case, the $(K - |\mathcal{S}^\star|)$ sub-optimal arms are eliminated before k optimal ones are confirmed, as some optimal arms may be needed to correctly classify borderline sub-optimal arms. Notably, q can be significantly smaller than $(K-1)$; for instance, when all arms are Pareto-optimal ($[K] = \mathcal{S}^\star$), we have $q = k$, yielding a substantial reduction in sample complexity.

[2]Besides the fact that the original SH algorithm for BAI discards samples collected in previous rounds to compute the empirical means.

Algorithm 2.2: EGE-SR-k: Any k-sized subset of the Pareto Set

Require: Parameter $k \in [K]$;number of rounds R; budget allocation t; and elimination scheme λ

Initialize: Set of active arms $\mathcal{A}_1 \leftarrow \{1, \dots, K\}$; set of arms (so far) identified as Pareto-optimal $\mathcal{B}_1 \leftarrow \emptyset$; set of discarded arms $\mathcal{D}_1 \leftarrow \emptyset$;

foreach $r = 1, 2, \dots, K-1$ **do**

  Collect t_r samples from each arm $a \in \mathcal{A}_r$;

  Compute S_r the empirical Pareto set

  Choose $i_r \in \operatorname{argmax}_{i \in \mathcal{A}_r} \widehat{\Delta}_{i,r}$;

  `// ties broken in favor of arms in` S_r

  $\mathcal{A}_{r+1} \leftarrow \mathcal{A}_r \backslash \{i_r\}$;

  $\mathcal{B}_{r+1} \leftarrow \mathcal{B}_r \cup \{S_r \cap \{i_r\}\}$;

  $\mathcal{D}_{r+1} \leftarrow \mathcal{D}_r \cup \{\{i_r\} \cap (\mathcal{A}_r \backslash S_r)\}$;

  if $|\mathcal{B}_{r+1}| = k$ **then**

    break and return $\widehat{S}_T \leftarrow \mathcal{B}_{r+1}$;

`// the identified Pareto set contains` k `or fewer arms`

return $\widehat{S}_T = \mathcal{B}_K \cup \mathcal{A}_K$

2.3 Main theoretical results

We present the main guarantees of EGE algorithms.

2.3.1 Upper bound on error probability

Exact PSI. We first propose an analysis of EGE for a generic number of rounds R, arm schedule λ, and sampling allocation t satisfying (2.3) and (2.4). It features the quantity

$$\widetilde{T}^{R,t,\lambda}(\nu) \coloneqq \min_{r \in [R]} \left[\sum_{s=1}^{r} t_s \right] \cdot \Delta^2_{(\lambda_{r+1}+1)},$$

in which the dependency in ν is captured in the gaps.

Theorem 2.3.1. *Let a budget $T \geqslant K$ be given and ν be a bandit with marginally σ-subgaussian arms. Then Empirical Gap Elimination with number of rounds R, budget allocation t, and elimination scheme λ satisfies*

$$e_T^{EGE}(\nu) \leqslant 2(K-1)\left|\mathcal{S}^\star\right| R d \cdot e^{-\frac{\widetilde{T}^{R,t,\lambda}(\nu)}{144\sigma^2}} .$$

This result shows that the probability of failure of EGE decreases exponentially fast with $\widetilde{T}^{R,t,\lambda}(\nu)$.

Introducing

$$H_2(\nu) \coloneqq \max_{k\in[K]} k\Delta_{(k)}^{-2},$$

where $k \mapsto (k)$ is a permutation such that $\Delta_{(1)} \leqslant \cdots \leqslant \Delta_{(K)}$, we obtain the following.

Corollary 2.3.2. *Let $T \geqslant K$ and ν be a bandit with σ-subgaussian marginals. Then EGE-SR satisfies*

$$e_T^{SR}(\nu) \leqslant 2(K-1)^2|\mathcal{S}^\star|d \cdot e^{-\frac{T-K}{144\sigma^2 H_2(\nu)\overline{\log}(K)}},$$

and for EGE-SH, $e_T^{SH}(\nu)$ is upper-bounded by

$$2(K-1)\lceil\log_2(K)\rceil\,|\,\mathcal{S}^\star\,|\,d \cdot e^{-\frac{T}{288\sigma^2 H_2(\nu)\lceil\log_2(K)\rceil}}.$$

Proof. For EGE-SR, we have $R = K-1$, $\lambda_r^{\mathrm{SR}} = K+1-r$ and $t_r^{\mathrm{SR}} = n_r^{\mathrm{SR}} - n_{r-1}^{\mathrm{SR}}$ where $n_r^{\mathrm{SR}} = \left\lceil \frac{1}{\overline{\log}(K)} \frac{T-K}{K+1-r} \right\rceil$, which yields

$$\begin{aligned}
\widetilde{T}^{R,\boldsymbol{t}^{\mathrm{SR}},\boldsymbol{\lambda}^{\mathrm{SR}}}(\nu) &\coloneqq \min_{r\in[K-1]} n_r^{\mathrm{SR}}\Delta^2_{(\lambda_{r+1}^{\mathrm{SR}}+1)},\\
&\geqslant \min_{r\in[K-1]} \frac{\Delta^2_{(K+1-r)}}{\overline{\log}(K)} \frac{T-K}{K+1-r},\\
&= \frac{T-K}{\overline{\log}(K)} \frac{1}{\max_{r\in\{2,\dots K\}} r\Delta_{(r)}^{-2}}\\
&= \frac{T-K}{\overline{\log}(K)} \frac{1}{H_2(\nu)}.
\end{aligned}$$

The analysis of EGE-SH is proven similarly in Section 2.5. □

The complexity measure $H_2(\nu)$ featured in our error exponent satisfies $H_2(\nu) \leqslant H(\nu) \leqslant H_2(\nu)\cdot\log(2K)$, following similar steps as in Audibert & Bubeck 2010. For BAI ($d=1$), we essentially recover the existing guarantees for SR and SH, whose error bounds also feature $H_2(\nu)$, up to constant factors inside the exponential and an extra multiplicative K factor for Sequential Halving. Still, to our knowledge, this is the first analysis of the variant of SH that does not discard samples between rounds, which often performs (much) better in practice.

In the general case ($d \geqslant 1$), we note that the bounds obtained for EGE-SR and EGE-SH are hard to compare: the latter has an improved polynomial dependence ($K\log_2(K)$ instead of K^2) but a worse constant inside the exponential. As we shall see in the experiments, both algorithms have actually pretty close performance (like in BAI, see Karnin et al. 2013). Moreover, they both outperform a simple baseline using Uniform Allocation (UA) and recommending the Pareto set of the empirical means. We note that Theorem 2.3.1 yields an upper bound on the error probability of this strategy (by choosing $R=1$, $t_1 = T/K$ and $\lambda_2 = 0$, for which $\widetilde{T}^{1,t,\lambda}(\nu) = n_1\Delta^2_{(\lambda_2+1)} = T/(K\Delta_{(1)}^{-2})$), which we add to our summary in Table 2.1. This bound can be much worse than that for EGE-SR/SH when the gaps are distinct.

Table 2.1: Upper bounds on $e_T(\nu)$ for different algorithms (up to constants). APE-FB is an "oracle" algorithm introduced in Kone, Kaufmann, et al. 2024 by converting a fixed-confidence PSI algorithm. APE-FB requires the parameter $H(\nu)$ to run.

Algorithm	Error Probability
EGE-SR	$K^2 \left\|\mathcal{S}^\star\right\| d \cdot e^{-T/(H_2(\nu)\log K)}$
EGE-SH	$K\log(K) \left\|\mathcal{S}^\star\right\| d \cdot e^{-T/(2H_2(\nu)\log K)}$
APE-FB*	$K\log(T) d \cdot e^{-T/H(\nu)}$
UA	$K \left\|\mathcal{S}^\star\right\| d \cdot e^{-T/(K\Delta_{(1)}^{-2})}$

PSI-k relaxation. To introduce the upper-bound on the expected loss $e_{T,k}(\nu)$, we define $\omega_{(k)}$ to be the k-th largest gap among the optimal arms: $\omega_{(k)} := \max_{i\in\mathcal{S}^\star}^k \Delta_i$ with $\omega_{(k)} = 0$ if $|\mathcal{S}^\star| < k$. Our bound features the complexity measure $H_2^{(k)}(\nu) := \max_{i\in[K]} i(\Delta_{(i)}^{(k)})^{-2}$ with the k-relaxed gaps

$$\Delta_i^{(k)} := \begin{cases} \max\left\{\Delta_i, \omega_{(k)}\right\} & \text{if} \quad i \in \mathcal{S}^\star \\ \Delta_i & \text{else.} \end{cases} \tag{2.6}$$

Theorem 2.3.3. *Let $k \in [K]$. EGE-SR-k satisfies*

$$e_{T,k}(\nu) \leqslant 2(K-1)^2 \left|\mathcal{S}^\star\right| d \cdot e^{-\frac{T-K}{144\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}}.$$

This result is particularly insightful when there are many optimal arms and some of them are easy to identify as such (large gaps). Indeed when $|\mathcal{S}^\star| \approx K$ and $k \ll |\mathcal{S}^\star|$, $H_2^{(k)}(\nu)$ can be an order of magnitude smaller than $H_2(\nu)$. We also note that when $k > |\mathcal{S}^\star|$ (and PSI-k reduces to PSI), $H_2^{(k)}(\nu) = H_2(\nu)$ and we recover the result of Theorem 2.3.2.

The theorem below bounds the expected stopping time and the number of samples used at stopping.

Theorem 2.3.4. *Fix $k < |\mathcal{S}^\star|$ and let $q := K - |\mathcal{S}^\star| + k$. Then*

$$\begin{aligned} \mathbb{E}[\tau] &\leqslant q + 2(K-1)|\mathcal{S}^\star|(K-q-1)qde^{-\frac{T-K}{144\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}} \quad \textit{and} \\ \mathbb{E}[N_\tau] &\leqslant N_q + 2(K-1)|\mathcal{S}^\star|(K-q-1)qdTe^{-\frac{T-K}{144\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}}. \end{aligned}$$

This result suggests that for this relaxed problem, the algorithm might not need to use the whole budget, in particular when T is large. For example, in the case where $[K] = \mathcal{S}^\star$, so $q = k$ and we roughly use N_k samples which can be way smaller than N_{K-1}, still, the probability of error improves upon the exact PSI setting.

2.3.2 Lower bound on the error probability

We present a lower bound for a class of instances. We define $\mathcal{I}^K$ to be the set of means $\mu := (\mu_1, \ldots, \mu_K)$ such that each sub-optimal arm i is only dominated by a single arm,

denoted by $i^\star$ (that has to belong to $\mathcal{S}^\star$) and that for each optimal arm j there exists a unique sub-optimal arm which is dominated by j, denoted by $\underline{j}$. We further assume that optimal arms are not too close to arms that they do not dominate: for any sub-optimal arm i and optimal arm j such that $\mu_i \not\prec \mu_j$,

$$\mathrm{M}(i,j) \geqslant 3\max(\Delta_i, \Delta_{\underline{j}}).$$

Let $\nu := (\nu_1, \dots, \nu_K)$ be an instance whose means $\mu \in \mathcal{I}^K$ and such that $\nu_i \sim \mathcal{N}(\mu_i, \sigma^2 I)$. For every $i \in [K]$ we define the alternative instance $\nu^{(i)} := (\nu_1, \dots \nu_i^{(i)}, \dots, \nu_K)$ in which only the mean of arm i is modified to:

$$\mu_i^{(i)} := \begin{cases} \mu_i - 2\Delta_i e_{c_{\underline{i}}} & \text{if } i \in \mathcal{S}^\star(\nu), \\ \mu_i + 2\Delta_i e_{c_i} & \text{else,} \end{cases} \tag{2.7}$$

where $e_1, \dots, e_d$ denotes the canonical basis of $\mathbb{R}^d$ and $c_i := \operatorname{argmin}_c[\mu_{i^\star}^c - \mu_i^c]$, with $\nu^{(0)} := \nu$. The following result is proven in Kone, Kaufmann, et al. 2024.

Theorem 2.3.5 (Theorem 2 of Kone, Kaufmann, et al. 2024)**.** *Let $\mu := (\mu_1, \dots, \mu_K) \in \mathcal{I}^K$ and $\nu := (\nu_1, \dots, \nu_K)$ where $\nu_i \sim \mathcal{N}(\mu_i, \sigma^2 I)$. For any algorithm* alg*, there exists $i \in \{0, \dots, K\}$ such that*

$$e_T^{\mathrm{alg}}(\nu^{(i)}) \geqslant \frac{1}{4} e^{-\frac{2T}{\sigma^2 H(\nu^{(i)})}}.$$

In particular, there exists some instances $\widetilde{\nu} \in \mathcal{M}^K$ such that $e_T^{\mathrm{alg}}(\tilde{\nu}) \geqslant \frac{1}{4}\exp(-\frac{2T}{\sigma^2 H(\widetilde{\nu})})$. In such instances, the decay rate of EGE-SR and EGE-SH is optimal up to constants and $\log(K)$ factors, and that of APE-FB is optimal up to constant factors, when the complexity is known.

2.3.3 Sketch of proof of Theorem 2.3.1

We define for any arms i, j and round r the events

$$\begin{aligned} \xi_{i,j,r} &:= \left\{ \left\| (\hat{\mu}_{i,n_r} - \hat{\mu}_{j,n_r}) - (\mu_i - \mu_j) \right\|_\infty \leqslant \gamma \Delta_{(\lambda_{r+1}+1)} \right\} \\ \mathcal{E}_\gamma^1 &:= \bigcap_{r\in[R]} \bigcap_{i\in\mathcal{S}^\star} \bigcap_{j\in[K]} \xi_{i,j,r}, \text{ for any } \gamma > 0. \end{aligned}$$

We shall prove that there exists some $\gamma > 0$ such that EGE does not make any error on the event $\mathcal{E}_\gamma^1$. That is, no sub-optimal arm is added to $(\mathcal{B}_r)_{1\leqslant r\leqslant R}$ and no optimal arm is added to $(\mathcal{D}_r)_{1\leqslant r\leqslant R}$, in any round r, and the possibly remaining arm in $\mathcal{A}_{R+1}$ is a Pareto-optimal arm.

To do so, an important step is to justify that any sub-optimal arm should be deactivated before the optimal arm that dominates it the most. More formally, for any sub-optimal arm i, we let

$$i^\star \in \operatorname*{argmax}_{j\in\mathcal{S}^\star} \mathrm{m}(i,j),$$

which by definition is such that $\Delta_i = \mathrm{m}(i, i^\star)$. For a sub-optimal arm i, we know that $i^\star \in \mathcal{S}^\star$ always exists. More importantly, $i^\star$ could be the only arm dominating i. Therefore,

it is crucial to ensure that i is no longer active before discarding $i^\star$, otherwise i could appear as optimal *w.r.t.* the remaining active arms.

Defining

$$\mathcal{P}_r \coloneqq \left\{\forall\, i \notin \mathcal{S}_\star, i \in \mathcal{A}_r \Rightarrow i^\star \in \mathcal{A}_r\right\}, \tag{2.8}$$

we state the following concentration result. The first one controls the deviation of the empirical $\mathrm{M}(i,j;r)$ and $\mathrm{m}(i,j;r)$ from their actual values.

Lemma 2.3.6. *On the event $\mathcal{E}_\gamma^1$, for all $r \in [R]$ and $i,j \in \mathcal{A}_r$, if $i \in \mathcal{S}^\star$ or $j \in \mathcal{S}^\star$ then,*

$$\begin{aligned} |\,\mathrm{M}(i,j;r) - \mathrm{M}(i,j)| &\leqslant \gamma\Delta_{(\lambda_{r+1}+1)} \text{ and} \\ |\,\mathrm{m}(i,j;r) - \mathrm{m}(i,j)| &\leqslant \gamma\Delta_{(\lambda_{r+1}+1)}\,. \end{aligned}$$

The second one builds on it to prove that the empirical gaps cannot be too far from their true gaps, if $\mathcal{P}_r$ holds. It is complementary to Lemma 2.5.2 stated in the main paper.

Lemma 2.3.7. *Let $\gamma > 0$ and assume $\mathcal{E}_\gamma^1$ holds. At round $r \in [R]$, for any sub-optimal arm $i \in \mathcal{A}_r$, if $i^\star \in \mathcal{A}_r$ and $i^\star$ does not empirically dominate i then $\Delta_i^\star \leqslant \gamma\Delta_{(\lambda_{r+1}+1)}$.*

Lemma 2.3.6, Lemma 2.3.7, and Lemma 2.3.8 are proven at the end of the chapter.

Lemma 2.3.8. *Assume that $\mathcal{E}_\gamma^1$ holds and let $r \in [R]$ such that $\mathcal{P}_r$ holds. Then for any $i \in \mathcal{A}_r$*

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -2\gamma\Delta_{(\lambda_{r+1}+1)} & \text{if } i \in \mathcal{S}^\star, \\ -\gamma\Delta_{(\lambda_{r+1}+1)} & \text{else.} \end{cases}$$

This result then permits us to prove by induction that $\mathcal{P}_r$ holds in any round r, when γ is small enough.

Lemma 2.3.9. *Let $\gamma < 1/6$. On the event $\mathcal{E}_\gamma^1$, for any $r \in [R+1]$, $\mathcal{P}_r$ holds. In particular, for any sub-optimal arm i, $i^\star$ cannot be deactivated before i.*

Proof. We assume that the event $\mathcal{E}_\gamma^1$ holds and we prove the result by induction on r. $\mathcal{P}_r$ trivially holds for $r = 1$ as all arms are active. Let $r \geqslant 1$ such that hat $\mathcal{P}_r$ holds. We shall prove that for any $i \in \mathcal{A}_{r+1} \cap (\mathcal{S}^\star)^\mathsf{c}$, $i^\star$ cannot be deactivated at the end of round r. A first observation is that the empirical gap $\widehat{\Delta}_{i,r}$ of each arm i satisfies

$$\widehat{\Delta}_{i,r} = \max\left\{\widehat{\Delta}_{i,r}^\star, \widehat{\delta}_{i,r}^\star\right\} \tag{2.9}$$

which follows from the fact that when $i \in S_r$, $\widehat{\Delta}_{i,r}^\star < 0 \leqslant \widehat{\delta}_{i,r}^\star$ and when $i \in \mathcal{A}_r \backslash S_r$, $\widehat{\delta}_{i,r}^\star < 0 \leqslant \widehat{\Delta}_{i,r}^\star$. Using Lemma 2.3.8 and the inductive hypothesis enables us to prove that

$$\forall i \in \mathcal{A}_r,\ \ \widehat{\Delta}_{i,r} \geqslant \Delta_i - 2\gamma\Delta_{(\lambda_{r+1}+1)}. \tag{2.10}$$

To prove that $\mathcal{P}_{r+1}$ holds, we proceed by contradiction and assume that there exists a sub-optimal arm $i \in \mathcal{A}_r$ (and therefore $i^\star \in \mathcal{A}_r$ by $\mathcal{P}_r$) such that $i \in \mathcal{A}_{r+1}$ and $i^\star \in \mathcal{A}_r \backslash \mathcal{A}_{r+1}$.

A first observation is that as there are λ_{r+1} arms in $\mathcal{A}_{r+1}$ and $i^\star$ is deactivated at the end of round r, there exists an arm $a_r \in \mathcal{A}_{r+1} \cup \{i^\star\}$ such that $\Delta_{a_r} \geqslant \Delta_{(\lambda_{r+1}+1)}$ and $\widehat{\Delta}_{i^\star} \geqslant \widehat{\Delta}_{a_r}$. We now consider two cases depending on whether $i^\star$ is empirically optimal or empirically sub-optimal.

Case 1: arm $i^\star \notin S_r$ *i.e.*, $i^\star$ is empirically sub-optimal. We have

$$\max_{j \in \mathcal{A}_r \backslash \{i^\star\}} \mathrm{m}(i^\star, j; r) \coloneqq \widehat{\Delta}^\star_{i^\star,r} = \widehat{\Delta}_{i^\star,r} \geqslant \widehat{\Delta}_{a_r,r}$$

Next, using Lemma 2.3.6 (on the LHS) and Equation (2.10) (on the RHS) yields

$$\begin{aligned} \max_{j \in \mathcal{A}_r \backslash \{i^\star\}} \mathrm{m}(i^\star, j) &\geqslant \Delta_{a_r} - 3\gamma\Delta_{(\lambda_{r+1}+1)}, \\ &\geqslant (1 - 3\gamma)\Delta_{(\lambda_{r+1}+1)} \ , \end{aligned}$$

where the last inequality follows since $\Delta_{a_r} \geqslant \Delta_{(\lambda_{r+1}+1)}$. As $i^\star$ is optimal, the LHS of the previous inequality is negative. So it follows that $0 \geqslant (1 - 3\gamma)\Delta_{(\lambda_{r+1}+1)}$ which yields a contradiction if $3\gamma < 1$.

Case 2: arm $i^\star \in S_r$ *i.e.*, $i^\star$ is empirically optimal.

We first prove that $i^\star$ does not empirically dominate i. Indeed, if i were dominated by $i^\star$, we would have $i \notin S_r$, so $\widehat{\Delta}_{i,r} = \widehat{\Delta}^\star_{i,r} > 0$ and since $i^\star \in S_r$,

$$\begin{aligned} \widehat{\Delta}_{i^\star,r} = \widehat{\delta}^\star_{i^\star,r} &\leqslant (\mathrm{M}(i, i^\star; r))_+ + (\widehat{\Delta}^\star_{i,r})_+, &(2.11)\\ &= 0 + \Delta_{i,r}. &(2.12) \end{aligned}$$

Recalling that $i \in \mathcal{A}_{r+1}$ we also have $\widehat{\Delta}_{i^\star,r} \geqslant \widehat{\Delta}_{i,r}$, which combined with Equation (2.12) yields

$$\widehat{\Delta}_{i^\star,r} = \widehat{\Delta}_{i,r}. \tag{2.13}$$

However, the tie-breaking rule of EGE ensures that Equation (2.13) is not possible since $i^\star \in S_r$ and it is deactivated while $i \in \mathcal{A}_{r+1} \cap (S_r)^\mathsf{c}$ (in case of an equality in the gaps, empirically sub-optimal arms are removed). Therefore i is not empirically dominated by $i^\star$. Hence, by Lemma 2.3.7

$$\Delta^\star_i \leqslant \gamma\Delta_{(\lambda_{r+1}+1)}. \tag{2.14}$$

Moreover, since $a_r \in \mathcal{A}_{r+1} \cup \{i^\star\}$, using the definition of $\widehat{\Delta}_{i^\star,r}$ yields

$$\mathrm{M}(i, i^\star; r)_+ + (\widehat{\Delta}^\star_{i,r})_+ \geqslant \widehat{\Delta}_{a_r,r}$$

Using Lemma 2.3.6, Lemma 2.3.8 applied to i and Equation (2.10) applied to a_r, we obtain

$$\begin{aligned} \mathrm{M}(i, i^\star)_+ + \Delta^\star_i &\geqslant \Delta_{a_r} - 5\gamma\Delta_{(\lambda_{r+1}+1)} \\ &\geqslant (1 - 5\gamma)\Delta_{(\lambda_{r+1}+1)} \end{aligned}$$

so

$$\Delta_i^\star \geqslant (1-5\gamma)\Delta_{(\lambda_{r+1}+1)}. \tag{2.15}$$

Combining this inequality with Equation (2.14) yields a contradiction if $6\gamma < 1$. □

As an immediate consequence of Lemma 2.3.9, if $\mathcal{A}_{R+1}$ contains one arm (*i.e.*, $\lambda_{R+1} = 1$) and $\mathcal{E}_\gamma^1$ holds, then it is an optimal arm.

Note that in BAI ($d = 1$), for any sub-optimal arm i, $i^\star$ is the unique best arm, and Lemma 2.3.9 is sufficient to ensure the correctness of EGE on the good event. However, in the general case, we must ensure that no optimal arm is rejected and no suboptimal arm is accepted. This is proven in the lemma below.

Lemma 2.3.10. *Let $\gamma < 1/6$. Then, on the event $\mathcal{E}_\gamma^1$, $\widehat{S}_T = \mathcal{S}^\star$.*

Proof. We show that on $\mathcal{E}_\gamma^1$, every arm is well classified, so the recommended set is the true Pareto optimal set. First by Lemma 2.3.9 we know that on $\mathcal{E}_\gamma^1$, the property

$$(i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^\mathsf{c} \Rightarrow i^\star \in \mathcal{A}_r)$$

holds for every $r \in [R]$. As in the proof of Lemma 2.3.9 (see Equation (2.10)), this enables us to prove in Lemma 2.3.8

$$\forall\, r \in [R], \forall\, i \in \mathcal{A}_r,\ \ \widehat{\Delta}_{i,r} \geqslant \Delta_i - 2\gamma\Delta_{(\lambda_{r+1}+1)} \tag{2.16}$$

We now establish that at the end of every round $r \in [R]$, no mis-classification can occur. That is for every arm $i \in \mathcal{A}_r \backslash \mathcal{A}_{r+1}$:

a) if $i \notin S_r$ (that is, i is added to $\mathcal{D}_r$) then $i \notin \mathcal{S}\star$,

b) if $i \in S_r$ (that is, i is added to $\mathcal{B}_r$) then $i \in \mathcal{S}^\star$.

Let $i \in \mathcal{A}_r \backslash \mathcal{A}_{r+1}$. Since $\lambda_{r+1} = |\mathcal{A}_{r+1}|$ should remain active at the end of round r, if i is removed at the end of the round, there exists $a_r \in \mathcal{A}_{r+1} \cup \{i\}$ such that $\Delta_{a_r} \geqslant \Delta_{(\lambda_{r+1}+1)}$ and a_r has a smaller empirical gap than i.

Case 1: $i \notin S_r$ *i.e.*, i is empirically sub-optimal. As $a_r \in \mathcal{A}_{r+1} \cup \{i\}$ has a smaller empirical gap than i, we have

$$\widehat{\Delta}_{i,r}^\star \ \geqslant\ \widehat{\Delta}_{a_r,r},$$

Assume by contradiction $i \in \mathcal{S}^\star$. Using Lemma 2.3.6 and Equation (2.16) yields

$$\max_{j \in \mathcal{A}_r \backslash \{i\}} \mathrm{m}(i,j) \ \ \geqslant\ \ \Delta_{a_r} - 3\gamma\Delta_{(\lambda_{r+1}+1)} \geqslant (1-3\gamma)\Delta_{(\lambda_{r+1}+1)}, \tag{2.17}$$

When $3\gamma < 1$, the RHS of Equation 2.17 is positive, so this inequality yields

$$\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j) > 0,$$

that is there exists $j \in \mathcal{A}_r$ such that

$$\mathrm{m}(i,j) > 0$$

so there exists j such that $\mu_i \prec \mu_j$, which contradicts the assumption $i \in \mathcal{S}^\star$. Therefore, $i \notin \mathcal{S}^\star$: i is a sub-optimal arm.

Case 2: $i \in S_r$ ***i.e.,*** i **is empirically optimal.** Since a_r has a larger empirical gap than i and $i \in S_r$, we have

$$\min_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{M}(i,j;r) \geqslant \widehat{\delta}_i^\star \geqslant \widehat{\Delta}_{a_r,r} \tag{2.18}$$

Assume by contradiction that i is sub-optimal. By Lemma 2.3.9 (for $\gamma < 1/6$), $i^\star \in \mathcal{A}_r$. Combining with (2.18) yields

$$\mathrm{M}(i,i^\star;r) \geqslant \widehat{\Delta}_{a_r,r}\,.$$

Further using Lemma 2.3.6 and Equation (2.16) yields

$$\mathrm{M}(i,i^\star) \geqslant (1-3\gamma)\Delta_{(\lambda_{r+1}+1)}.$$

Then by taking $6\gamma < 1$, the RHS of the inequality is positive so

$$\mathrm{M}(i,i^\star) > 0, \tag{2.19}$$

which is not possible as $\mu_i \prec \mu_{i^\star}$. Therefore, $i \in \mathcal{S}^\star$: it is an optimal arm.

This proves that on the event $\mathcal{E}_\gamma^1$, with $\gamma < 1/6$ we have $\mathcal{B}_R \subseteq \mathcal{S}^\star$ and $\mathcal{D}_R \subseteq (\mathcal{S}^\star)^\mathsf{c}$. Moreover, if there is a remaining active arm in $\mathcal{A}_{R+1}$ (which happens if and only if $\lambda_{R+1} = 1$), this arm has to be optimal by Lemma 2.3.9. Therefore, on $\mathcal{E}_\gamma^1$, the set recommended by EGE is $\widehat{S}_T = \mathcal{B}_R \cup \mathcal{A}_{R+1} = \mathcal{S}^\star$. □

Thanks to Hoeffding's inequality, we have

$$\begin{aligned}
\mathbb{P}((\mathcal{E}_\gamma^1)^\mathsf{c}) &\leqslant \sum_{r=1}^{R} \sum_{i \in \mathcal{S}^\star} \sum_{j \neq i} \sum_{c=1}^{d} \mathbb{P}\left(|(\hat{\mu}_{i,n_r}^c - \hat{\mu}_{j,n_r}^c) - (\mu_i^c - \mu_j^c)| > \gamma \Delta_{\lambda_{r+1}+1}\right), \\
&\leqslant 2|\mathcal{S}^\star|(K-1)d \sum_r \exp\left(-\frac{\gamma^2 n_r \Delta_{\lambda_{r+1}+1}^2}{4\sigma^2}\right) \quad \text{(Hoedding's inequality)}, \\
&\leqslant 2|\mathcal{S}^\star|(K-1)dR \exp\left(-\frac{\gamma^2 \widetilde{T}^{R,t,\lambda}(\nu)}{4\sigma^2}\right).
\end{aligned}$$

Applying the above to $\gamma_x = 1/6 - x$ for $x \geqslant 0$, and letting $x \to 0$ yields

$$e_T^{\mathrm{EGE}}(\nu) \leqslant 2|\mathcal{S}^\star|(K-1)dR \exp\left(-\frac{\widetilde{T}^{R,t,\lambda}}{144\sigma^2}\right).$$

The analysis of EGE-SR-k is given at the end of the chapter.

2.4 Numerical study and discussion

Experimental setup. We evaluate our algorithms on both synthetic and real-world–inspired instances, comparing them to Uniform Allocation (UA) and the fixed-budget APE-FB algorithm (Kone, Kaufmann, et al. 2023). APE-FB depends on a parameter $a_\alpha = \alpha \frac{25}{36} \frac{T-K}{H(\nu)}$ with $\alpha \in 1/10, 1, 10$, for which guarantees hold when $\alpha \leqslant 1$ and are optimal at $\alpha = 1$. We also include a heuristic variant, APE-FB-ADAPT, which estimates $H(\nu)$ online. Each experiment is averaged over $4{,}000$ random runs, reporting the $\log_{10}$ of the mean misidentification rate. Following Audibert & Bubeck 2010; Karnin et al. 2013, we set the sampling budget to $T = H(\nu)$. The R-round implementation of EGE has time complexity $\mathcal{O}(RK^2d)$ and memory complexity $\mathcal{O}(K^2)$.

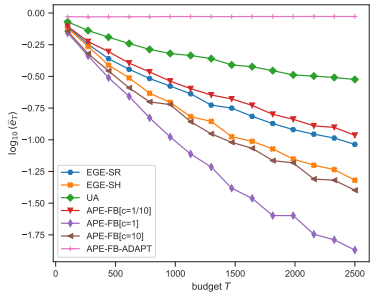


Figure 2.1: Application 1: COV-BOOST trial

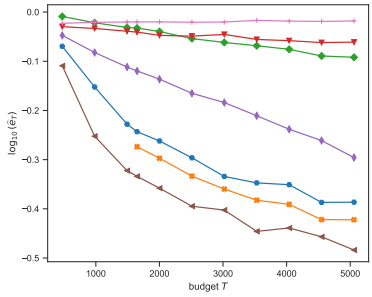


Figure 2.2: Application 2: Sorting Networks dataset.

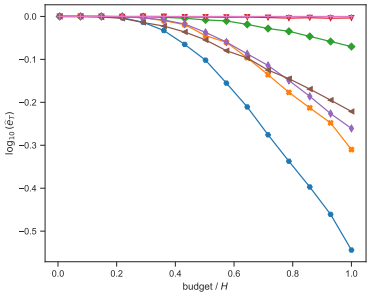


Figure 2.3: Arms on a convex Pareto set.

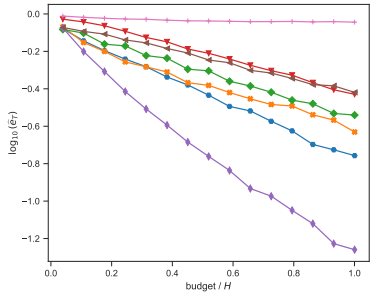


Figure 2.4: Each sub-optimal i is only dominated by $i^\star$.

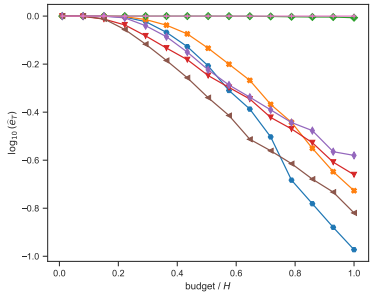


Figure 2.5: $K = 200$ arms on the unit circle.

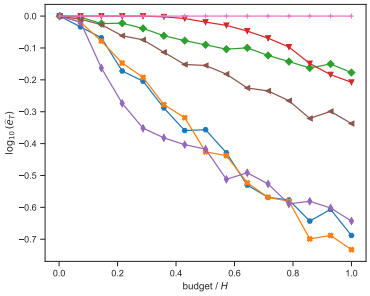


Figure 2.6: High dimension ($d = 10$) with 2 group of arms.

Datasets. Experiments are conducted on two real-world–inspired datasets. The first, COV-BOOST (Munro et al. 2021), is a phase II clinical trial on 2,883 participants evaluating the immunogenicity of $K = 20$ COVID-19 vaccination strategies, each defined by a triplet of vaccines used as first, second, and third doses. Using the three reported immunogenicity indicators—two antibody-based and one cellular—we construct a $d = 3$ Gaussian bandit under the log-normal assumption of Munro et al. 2021. The resulting model, reused in

subsequent chapters, simulates a realistic multi-objective pure-exploration task; full preprocessing details and statistics are given in Appendix 7. The second dataset, SNW (Zuluaga, Milder, et al. 2012), contains 206 sorting network architectures evaluated by two conflicting objectives: FPGA area (resource usage) and throughput (processing rate). Measurements are stochastic due to circuit randomness, and the data are modeled as a $d = 2$, $K = 206$ Gaussian bandit representing the trade-off between hardware cost and performance, with a fixed sampling budget of $T = 5{,}000$.

Synthetic instances. We test all algorithms on four synthetic instances covering various Pareto-set geometries, dimensions, and noise structures. The first instance (*Convex front*) has $K = 60$ arms in $d = 2$, with the first 20 arms placed on a smooth convex frontier $\mu_i = (x_i^2, 1/(4x_i^2))^\top$ for $x_i \in [0.55, 0.95]$, and the remaining 40 sampled from $\{(x, y) \in [0.1, 0.8]^2 : xy \leqslant 1/5\}$, forming a dense dominated region. The second instance (*Paired dominance*) includes $K = 10$ arms with $|\mathcal{S}^\star| = 2$, where each sub-optimal arm is dominated by a unique optimal one. We set $\mu_1 = (0.4, 0.75)^\top$, $\mu_2 = (0.75, 0.4)^\top$, and define $\mu_{2i+1} = (0.45 + 0.2^i, 0.35 - 0.2^i)^\top$, $\mu_{2i+2} = (0.10 + 0.20^i, 0.70 - 0.20^i)^\top$ for $i = 1, \dots, 4$. The third instance (*Circular set*) has $K = 200$ arms in $d = 2$, uniformly distributed on a unit circle with Gaussian noise $\sigma = 1/4$, where $\mu_i = (\cos \beta_i, \sin \beta_i)^\top$ for β_i uniformly sampled in two disjoint arcs. Finally, the fourth instance (*High-dimensional clusters*) explores $d = 10$ with $K = 50$ arms, the first 30 drawn uniformly from $[0.2, 0.45]^{10}$ and the remaining 20 from $[0.55, 0.75]^{10}$, forming two well-separated clusters.

Results. In all experiments, the uniform allocation (UA) baseline is largely outperformed by both EGE-SR and EGE-SH (with no clear ordering between the two algorithms). This is particularly the case when there are many arms and for complex Pareto sets. By estimating the hardness to classify each arm, EGE eventually allocates more samples to arms that are difficult to classify, leading to a smaller error probability. APE-FB is a good competitor to our EGE algorithms; however, it is not robust to the hyper-parameter a, which requires the knowledge of $H(\nu)$ to be properly tuned. Our proposed heuristic that estimates the complexity online fails dramatically.

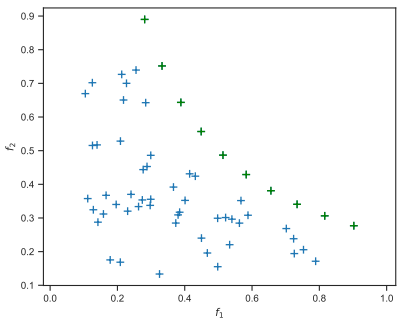


Figure 2.7: Synthetic Experiment 1: Group of arms on a convex Pareto set.

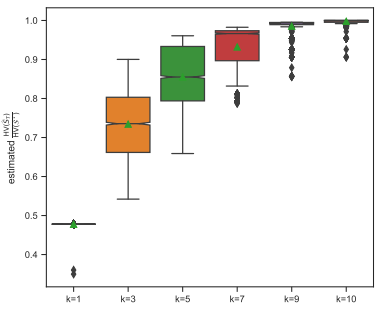


Figure 2.8: Hyper-volume fraction of the returned set on Exp.1 (convex Pareto set).

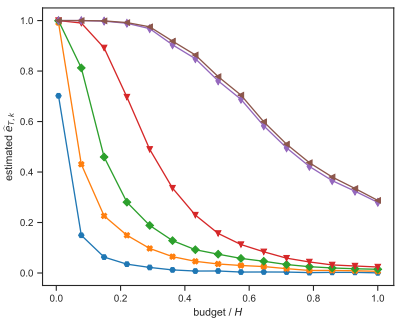


Figure 2.9: Estimated PSI-k loss for different values of k on Exp.1 (convex Pareto set).

Experiments on PSI-k. We further evaluate EGE-SR-k the convex Pareto set instance for different values of k, reporting the average $\log_{10}$ loss over 4,000 independent trials. The PSI-k loss is defined as

$$\mathcal{L}(\widehat{S}_T, k) = \begin{cases} \mathbb{1}\big(\widehat{S}_T \subset \mathcal{S}^\star\big), & \text{if } |\widehat{S}_T| = k, \\ \mathbb{1}\big(\widehat{S}_T = \mathcal{S}^\star\big), & \text{otherwise.} \end{cases}$$

The empirical results (Fig. 2.9) confirm the theoretical prediction that the misidentification probability decreases exponentially fast with the sampling budget T. Smaller values of k yield smaller losses, consistent with the monotonic increase of $H_2(\nu)^{(k)}$ with k. Moreover, for $k = |\mathcal{S}^\star|$, the loss can be smaller than for $k > |\mathcal{S}^\star|$. This reflects the structural difference between PSI-k, which stops once k optimal arms are identified, and PSI, which continues until all arms are classified, potentially introducing additional errors in later stages.

Hyper-volume analysis. To assess the representativeness of the returned sets, we use the hyper-volume (HV) metric (Daulton et al. 2020), which measures the region dominated by a set of points. Given a reference point $\boldsymbol{r}$ and a set $\mathcal{X} \subset \mathbb{R}^d$ the hyper-volume of S is

$$\mathrm{HV}(\mathcal{X}) = \boldsymbol{\lambda}\left(\bigcup_{\boldsymbol{x}\in S} [\boldsymbol{r}, \boldsymbol{x}]\right),$$

where $\boldsymbol{\lambda}$ denotes the Lebesgue measure on $\mathbb{R}^d$ and $[\boldsymbol{r}, \boldsymbol{x}]$ is the line segment connecting $\boldsymbol{r}$ and $\boldsymbol{x}$. To simplify notation we let for a discrete set $S \subset [K]$, $\mathrm{HV}(S) \coloneqq \mathrm{HV}(\{\mu_i : i \in S\})$. Since for any set S, $\mathrm{HV}(S)$ equals the hyper-volume of its Pareto front, we evaluate the normalized coverage

$$\alpha_{\mathrm{HV}} = \frac{\mathrm{HV}(\widehat{S}_T)}{\mathrm{HV}(\mathcal{S}^\star)}.$$

Higher values of α_{HV} indicate broader coverage of the Pareto-optimal region. In Fig. 2.8, even for $k = 1$, the identified arm already covers nearly half of the area dominated by the true Pareto set. This suggests that the most influential arms, those contributing most to the hyper-volume, are typically detected first by EGE-SR-k , confirming the efficiency of its allocation and discarding strategy.

Discussion. This chapter introduced the first algorithms for Pareto Set Identification in the fixed-budget setting. The proposed EGE framework, combined with sequential rejection or halving strategies, achieves an exponential decay of the misidentification probability with the budget, and its convergence exponent is provably tight for specific bandit instances. Empirical evaluations confirm that EGE-SR and EGE-SH not only outperform uniform allocation but also remain competitive with oracle algorithms requiring prior knowledge of the instance complexity $H(\nu)$.

2.5 Additional proofs

We present some proof elements for this chapter; we may refer to Kone, Kaufmann, et al. 2024 for complementary proofs.

2.5.1 Analysis of EGE

We recall for any arms i, j and round r the events

$$\begin{aligned}\xi_{i,j,r} &\coloneqq \left\{ \left\| (\hat{\mu}_{i,n_r} - \hat{\mu}_{j,n_r}) - (\mu_i - \mu_j) \right\|_\infty \leqslant \gamma \Delta_{(\lambda_{r+1}+1)} \right\} \\ \mathcal{E}_\gamma^1 &\coloneqq \bigcap_{r\in[R]} \bigcap_{i\in\mathcal{S}^\star} \bigcap_{j\in[K]} \xi_{i,j,r}, \text{ for any } \gamma > 0.\end{aligned}$$

For each sub-optimal arm i, we define $i^\star$ to be an arbitrary element in $\operatorname{argmax}_{k\in\mathcal{S}^\star} \mathrm{m}(i,k)$, which by definition yields $\Delta_i^\star = \mathrm{m}(i, i^\star)$. Introducing the property

$$\mathcal{P}_r \coloneqq \{\forall\, i \notin \mathcal{S}^\star,\ i \in \mathcal{A}_r \ \Rightarrow\ i^\star \in \mathcal{A}_r\},$$

the first step of the proof consists in proving that $\mathcal{P}_r$ holds for all r.

Corollary 2.3.2. *Let $T \geqslant K$ and ν be a bandit with σ-subgaussian marginals. Then EGE-SR satisfies*

$$e_T^{SR}(\nu) \leqslant 2(K-1)^2 |\mathcal{S}^\star| d \cdot e^{-\frac{T-K}{144\sigma^2 H_2(\nu)\overline{\log}(K)}},$$

and for EGE-SH, $e_T^{SH}(\nu)$ is upper-bounded by

$$2(K-1)\lceil\log_2(K)\rceil\,|\,\mathcal{S}^\star\,|\,d \cdot e^{-\frac{T}{288\sigma^2 H_2(\nu)\lceil\log_2(K)\rceil}}.$$

Proof. The case of EGE-SR has been proven in the main. For EGE-SH we have $R = \lceil\log_2(K)\rceil$, $n_r^{\mathrm{SH}} \coloneqq \sum_{s=1}^r t_s^{\mathrm{SH}} \geqslant \frac{T}{\lambda_r^{\mathrm{SH}}\lceil\log_2(K)\rceil}$ and $\lambda_r^{\mathrm{SH}} \leqslant 2(\lambda_{r+1}^{\mathrm{SH}} + 1)$. Then

$$\begin{aligned}\widetilde{T}^{R,t^{\mathrm{SH}},\lambda^{\mathrm{SH}}}(\nu) &\coloneqq \min_{r\in\{1,\dots,\lceil\log_2(K)\rceil\}} n_r^{\mathrm{SH}} \Delta^2_{(\lambda_{r+1}^{\mathrm{SH}}+1)}, \\ &\geqslant \min_{r\in\{1,\dots,\lceil\log_2(K)\rceil\}} \frac{\Delta^2_{(\lambda_{r+1}^{\mathrm{SH}}+1)}}{\lambda_{r+1}^{\mathrm{SH}}+1} \times \frac{T}{2\lceil\log_2(K)\rceil}, \\ &= \frac{T/(2\lceil\log_2(K)\rceil)}{\max_{r\in|\lceil\log_2(K)\rceil|}(\lambda_{r+1}^{\mathrm{SH}}+1)\Delta^{-2}_{(\lambda_{r+1}^{\mathrm{SH}}+1)}} \\ &\geqslant \frac{T}{2H_2(\nu)\lceil\log_2(K)\rceil}.\end{aligned}$$

□

Lemma 4.5.1. *On the event $\mathcal{E}_\gamma^1$, for all $r \in [R]$ and $i, j \in \mathcal{A}_r$, if $i \in \mathcal{S}^\star$ or $j \in \mathcal{S}^\star$ then,*

$$\begin{aligned}|\,\mathrm{M}(i,j;r) - \mathrm{M}(i,j)| &\leqslant \gamma\Delta_{(\lambda_{r+1}+1)} \text{ and} \\ |\,\mathrm{m}(i,j;r) - \mathrm{m}(i,j)| &\leqslant \gamma\Delta_{(\lambda_{r+1}+1)}\,.\end{aligned}$$

Proof. Assume $i \in \mathcal{S}^\star$ or $j \in \mathcal{S}^\star$. We have

$$\begin{aligned} |\,\mathrm{M}(i,j;r) - \mathrm{M}(i,j)| &= \left|\max_c \left[\hat{\mu}^c_{i,n_r} - \hat{\mu}^c_{j,n_r}\right] - \max_c \left[\mu^c_i - \mu^c_j\right]\right|, \\ &\overset{(a)}{\leqslant} \max_c \left|(\hat{\mu}^c_{i,n_r} - \hat{\mu}^c_{j,n_r}) - (\mu^c_i - \mu^c_j)\right|, \\ &= \left\|(\hat{\mu}_{i,n_r} - \hat{\mu}_{j,n_r}) - (\mu_i - \mu_j)\right\|_\infty, \\ &\overset{(b)}{\leqslant} \gamma\Delta_{(\lambda_{r+1}+1)}. \end{aligned}$$

where (a) follows by reverse triangle inequality and (b) holds on the event $\mathcal{E}^1_\gamma$. The second part of the lemma follows from

$$|\,\mathrm{m}(i,j;r) - \mathrm{m}(i,j)| \;=\; |-\mathrm{M}(i,j;r) + \mathrm{M}(i,j)|,$$

as $\mathrm{M}(i,j) = -\,\mathrm{m}(i,j)$ and $\mathrm{M}(i,j;r) = -\,\mathrm{m}(i,j;r)$. □

Lemma 2.3.7. *Let $\gamma > 0$ and assume $\mathcal{E}^1_\gamma$ holds. At round $r \in [R]$, for any sub-optimal arm $i \in \mathcal{A}_r$, if $i^\star \in \mathcal{A}_r$ and $i^\star$ does not empirically dominate i then $\Delta^\star_i \leqslant \gamma\Delta_{(\lambda_{r+1}+1)}$.*

Proof. We have by definition

$$\begin{aligned} \hat{\mu}_{i,n_r} \not\prec \hat{\mu}_{i^\star,n_r} &\implies \exists\, c \;:\hat{\mu}^c_{i,n_r} > \hat{\mu}^c_{i^\star,n_r} \\ &\implies \exists\, c \;:(\hat{\mu}^c_{i,n_r} - \mu^c_i) - (\hat{\mu}^c_{i^\star,n_r} - \mu^c_{i^\star}) > \mu^c_{i^\star} - \mu^c_i \geqslant \mathrm{m}(i,i^\star), \end{aligned}$$

so

$$\left\|(\hat{\mu}_{i,n_r} - \hat{\mu}_{i^\star,n_r}) - (\mu_i - \mu_{i^\star})\right\|_\infty \geqslant \Delta^\star_i,$$

which, on the event $\mathcal{E}^1_\gamma$ is only possible if

$$\Delta^\star_i \leqslant \gamma\Delta_{(\lambda_{r+1}+1)}\,.$$

□

We recall the definition of the property

$$\mathcal{P}_r = \{\forall\, i \notin \mathcal{S}^\star, i \in \mathcal{A}_r \Rightarrow i^\star \in \mathcal{A}_r\}\,.$$

and start by proving a first intermediate result towards proving Lemma 2.3.8.

Lemma 2.5.1. *Assume that $\mathcal{E}^c_1$ holds and let $r \in [R]$ such that $\mathcal{P}_r$ holds. Then for any $i \in \mathcal{A}_r$,*

$$(\widehat{\Delta}^\star_{i,r})_+ - (\Delta^\star_i)_+ \leqslant 2\gamma\Delta_{(\lambda_{r+1}+1)} \quad \textit{and} \quad (\widehat{\Delta}^\star_{i,r})_+ - (\Delta^\star_i)_+ \geqslant -\gamma\Delta_{(\lambda_{r+1}+1)}.$$

Proof. We define the $\max$ of an empty set to be $-\infty$. We first analyze the case $i \in \mathcal{S}^\star$. When $i \in \mathcal{S}^\star$, we have

$$(\Delta^\star_i)_+ = 0 = \left(\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j)\right)_+,$$

which yields

$$\begin{aligned}
|(\widehat{\Delta}^{\star}_{i,r})_+ - (\Delta^{\star}_i)_+| &= \left| \left(\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j;r) \right)_+ - \left(\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j) \right)_+ \right|, \\
&\leqslant \left| \left(\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j;r) \right) - \left(\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j) \right) \right| \quad \text{(since } |x_+ - y_+| \leqslant |x-y|\text{)}, \\
&\leqslant \max_{j \in \mathcal{A}_r \setminus \{i\}} \left| \mathrm{m}(i,j;r) - \mathrm{m}(i,j) \right|, \\
&\leqslant \gamma \Delta_{(\lambda_{r+1}+1)},
\end{aligned}$$

where the last inequality follows from Lemma 2.3.6. We now assume that i is a sub-optimal arm. Since $\mathcal{P}_r$ holds at round r, $i^\star \in \mathcal{A}_r$ and

$$\Delta^{\star}_i = \max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j).$$

Let $\hat{i} \in \operatorname{argmax}_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j;r)$ then

$$\begin{aligned}
(\widehat{\Delta}^{\star}_{i,r})_+ - (\Delta^{\star}_i)_+ &= (\mathrm{m}(i,\hat{i};r))_+ - (\mathrm{m}(i,i^\star))_+, &(2.20)\\
&\geqslant (\mathrm{m}(i,i^\star;r))_+ - (\mathrm{m}(i,i^\star))_+ \quad \text{(since } i^\star \in \mathcal{A}_r\text{)}. &(2.21)
\end{aligned}$$

We further note that

$$\begin{aligned}
|(\mathrm{m}(i,i^\star;r))_+ - (\mathrm{m}(i,i^\star))_+| &\leqslant |\mathrm{m}(i,i^\star;r) - \mathrm{m}(i,i^\star)|, \\
&\leqslant \gamma \Delta_{(\lambda_{r+1}+1)},
\end{aligned}$$

which follows from Lemma 2.3.6. Combining the latter inequality with (2.21) yields

$$(\widehat{\Delta}^{\star}_{i,r})_+ - (\Delta^{\star}_i)_+ \geqslant -\gamma \Delta_{(\lambda_{r+1}+1)}. \tag{2.22}$$

We also have

$$\begin{aligned}
(\widehat{\Delta}^{\star}_{i,r})_+ - (\Delta^{\star}_i)_+ &\leqslant (\mathrm{m}(i,\hat{i};r))_+ - (\mathrm{m}(i,\hat{i}))_+, \\
&\leqslant \left| \mathrm{m}(i,\hat{i};r) - \mathrm{m}(i,\hat{i}) \right|, \\
&= \left| \mathrm{M}(i,\hat{i}) - \mathrm{M}(i,\hat{i};r) \right|, \\
&\leqslant \left\| (\mu_i - \mu_{\hat{i}}) - (\hat{\mu}_{i,n_r} - \hat{\mu}_{\hat{i},n_r}) \right\|_\infty, \\
&= \left\| ((\mu_i - \mu_{i^\star}) - (\hat{\mu}_{i,n_r} - \hat{\mu}_{i^\star,n_r})) + \left((\mu_{i^\star} - \mu_{\hat{i}}) - \left(\hat{\mu}_{i^\star,n_r} - \hat{\mu}_{\hat{i},n_r} \right) \right) \right\|_\infty, \\
&\leqslant 2\gamma \Delta_{(\lambda_{r+1}+1)},
\end{aligned}$$

where the last inequality follows from the triangle inequality and Lemma 2.3.6. This proves the lemma. □

We then prove Lemma 2.5.2, which can be viewed as a "symmetric" version of Lemma 2.3.8.

Lemma 2.5.2. *Assume that $\mathcal{E}_\gamma^1$ holds. Let $r \in [R]$ such that $\mathcal{P}_r$ holds. Then, for any sub-optimal arm $i \in \mathcal{A}_r$,*

$$\left|\widehat{\Delta}_{i,r}^\star - \Delta_i^\star\right| \leqslant 2\gamma\Delta_{(\lambda_{r+1}+1)}$$

and for any optimal arm $i \in \mathcal{A}_r$,

$$\widehat{\delta}_{i,r}^\star \geqslant \Delta_i - 2\gamma\Delta_{(\lambda_{r+1}+1)}.$$

Proof. Let $i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^\mathsf{c}$. By assumption, $i^\star \in \mathcal{A}_r$, so

$$\Delta_i \coloneqq \Delta_i^\star = \max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j),$$

then we have

$$\begin{aligned}
|\widehat{\Delta}_{i,r}^\star - \Delta_i| &= \left|\left(\max_{j\in\mathcal{A}_r\setminus\{i\}} \mathrm{m}(i,j;r)\right) - \left(\max_{j\in\mathcal{A}_r\setminus\{i\}} \mathrm{m}(i,j)\right)\right|, \\
&\overset{(a)}{\leqslant} \max_{j\in\mathcal{A}_r\setminus\{i\}} |\,\mathrm{m}(i,j;r) - \mathrm{m}(i,j)|, \\
&\leqslant \max_{j\in\mathcal{A}_r\setminus\{i\}} |\,\mathrm{M}(i,j) - \mathrm{M}(i,j;r)| \\
&, \\
&\leqslant \max_{j\in\mathcal{A}_r\setminus\{i\}} \|(\mu_i - \mu_j) - (\hat{\mu}_{i,n_r} - \hat{\mu}_{j,n_r})\|_\infty \\
&= \max_{j\in\mathcal{A}_r\setminus\{i\}} \|((\mu_i - \mu_{i^\star}) - (\hat{\mu}_{i,n_r} - \hat{\mu}_{i^\star,n_r})) + ((\mu_{i^\star} - \mu_j) - (\hat{\mu}_{i^\star,n_r} - \hat{\mu}_{j,n_r}))\|_\infty, \\
&\overset{(b)}{\leqslant} 2\gamma\Delta_{(\lambda_{r+1}+1)},
\end{aligned}$$

where (a) follows by reverse triangle inequality and (b) follows by triangle inequality and Lemma 2.3.6. So we have proved the first statement of the lemma. Before proving the second statement, recall that

$$\widehat{\delta}_{i,r}^\star = \min_{j\in\mathcal{A}_r\setminus\{i\}} [\mathrm{M}(i,j;r) \ \wedge \ (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}_{j,r}^\star)_+)].$$

Let $i \in \mathcal{A}_r \cap \mathcal{S}^\star$. For any $j \in \mathcal{A}_r$ we have

$$\begin{aligned}
|\,\mathrm{M}(j,i;r)_+ - \mathrm{M}(j,i)_+| &\leqslant |\,\mathrm{M}(j,i;r) - \mathrm{M}(j,i)|, \\
&\leqslant \gamma\Delta_{(\lambda_{r+1}+1)} \quad \text{(Lemma 2.3.6)},
\end{aligned}$$

and similarly

$$|\,\mathrm{M}(i,j;r) - \mathrm{M}(i,j)| \leqslant \gamma\Delta_{(\lambda_{r+1}+1)}$$

which, combined, yields

$$\mathrm{M}(j,i;r)_+ \geqslant \mathrm{M}(j,i)_+ - \gamma\Delta_{(\lambda_{r+1}+1)} \quad \text{and} \tag{2.23}$$

$$\mathrm{M}(i,j;r) \geqslant \mathrm{M}(i,j) - \gamma\Delta_{(\lambda_{r+1}+1)}. \tag{2.24}$$

Letting $j \in \mathcal{A}_r$, we have by Lemma 2.5.1

$$(\widehat{\Delta}_{j,r}^\star)_+ - (\Delta_j^\star)_+ \geqslant -\gamma\Delta_{(\lambda_{r+1}+1)} \tag{2.25}$$

Combined, (2.25) and (2.23) yields

$$\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+ \;\geqslant\; \mathrm{M}(j,i)_+ + (\Delta^\star_j)_+ - 2\gamma\Delta_{(\lambda_{r+1}+1)}$$

for any $j \in \mathcal{A}_r$. Which combined with (2.24) yields

$$[\mathrm{M}(i,j;r) \wedge (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+)] \geqslant [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)] - 2\gamma\Delta_{(\lambda_{r+1}+1)}$$

for any $j \in \mathcal{A}_r$. Therefore,

$$\begin{aligned}
\widehat{\delta}^\star_{i,r} &\geqslant \left(\min_{j\in\mathcal{A}_r\setminus\{i\}} [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)]\right) - 2\gamma\Delta_{(\lambda_{r+1}+1)}, \\
&\geqslant \left(\min_{j\in[K]\setminus\{i\}} [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)]\right) - 2\gamma\Delta_{(\lambda_{r+1}+1)}, \\
&= \delta^\star_i - 2\gamma\Delta_{(\lambda_{r+1}+1)}.
\end{aligned}$$

□

Finally, we prove Lemma 2.3.8.

Lemma 2.3.8. *Assume that $\mathcal{E}^1_\gamma$ holds and let $r \in [R]$ such that $\mathcal{P}_r$ holds. Then for any $i \in \mathcal{A}_r$*

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -2\gamma\Delta_{(\lambda_{r+1}+1)} & \text{if } i \in \mathcal{S}^\star, \\ -\gamma\Delta_{(\lambda_{r+1}+1)} & \text{else.} \end{cases}$$

Proof of Lemma 2.3.8. Let $i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^\mathsf{c}$. Since $i \in (\mathcal{S}^\star)^\mathsf{c}$, $\Delta_i = \Delta^\star_i$. If $i \notin S_r$ then $\widehat{\Delta}_{i,r} = \widehat{\Delta}^\star_{i,r}$. If $i \in S_r$ then $\widehat{\Delta}_{i,r} = \widehat{\delta}^\star_{i,r}$ and

$$\widehat{\Delta}^\star_{i,r} \leqslant 0 \leqslant \widehat{\delta}^\star_{i,r}$$

which follows since $i \in S_r$. Therefore in both cases

$$\begin{aligned}
\widehat{\Delta}_{i,r} - \Delta_i &\geqslant \widehat{\Delta}^\star_{i,r} - \Delta^\star_i \\
&= \max_{j\in\mathcal{A}_r} \mathrm{m}(i,j;r) - \mathrm{m}(i,i_\star) \\
&\geqslant \mathrm{m}(i,i_\star;r) - \mathrm{m}(i,i_\star) \\
&\geqslant -\gamma\Delta_{(\lambda_{r+1}+1)},
\end{aligned}$$

where the last inequality uses Lemma 2.3.6.

Similarly, let $i \in \mathcal{A}_r \cap \mathcal{S}^\star$. We have $\Delta_i = \delta^\star_i$ (Lemma 2.1.2) and $\widehat{\Delta}_{i,r} = \widehat{\delta}^\star_{i,r}$ if $i \in S_r$. If $i \in \mathcal{A}_r \setminus S_r$ (empirically sub-optimal) then

$$\widehat{\Delta}_{i,r} = \widehat{\Delta}^\star_{i,r} \geqslant 0$$

and there exists $j \in \mathcal{A}_r$ such that $\hat{\mu}_{i,n_r} \prec \hat{\mu}_{j,n_r}$ so for this arm j, $\mathrm{M}(i,j;r) < 0$. Therefore $\widehat{\delta}^\star_{i,r} < 0$ and

$$\widehat{\delta}^\star_{i,r} < 0 \leqslant \widehat{\Delta}_{i,r} = \widehat{\Delta}^\star_{i,r}.$$

In all cases, we have

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \widehat{\delta}^\star_{i,r} - \Delta_i \geqslant -2\gamma\Delta_{(\lambda_{r+1}+1)},$$

where the last inequality follows from Lemma 2.5.2. □

2.5.2 Analysis of EGE-SR-k

In this section, we analyze EGE-SR-k and bound its stopping time. We define for any $\gamma > 0$ and $t \in [K-1]$:

$$\mathcal{E}_\gamma^2(t) \coloneqq \bigcap_{r\in[t]} \bigcap_{i\in\mathcal{S}^\star} \bigcap_{j\in[K]} \left\{ \left\| (\hat{\mu}_{i,n_r} - \hat{\mu}_{j,n_r}) - (\mu_i - \mu_j) \right\|_\infty \leqslant \gamma \Delta^{(k)}_{(K+1-r)} \right\}$$

and in particular

$$\mathcal{E}_\gamma^2 \coloneqq \mathcal{E}_\gamma^2(K-1).$$

We define also $\tilde{k} = \min\{k, |\mathcal{S}^\star|\}$.

For any round r, let $\alpha(r) = |\mathcal{B}_r|$ denote the number of arms so far identified as optimal at the beginning of round r. We denote these arms by $a_1, \dots, a_{\alpha(r)}$. We say that the algorithm has made an error before round r if $\mathcal{B}_r \cap (\mathcal{S}^\star)^\mathsf{c} \neq \emptyset$ or $\mathcal{D}_r \cap \mathcal{S}^\star \neq \emptyset$. We will show that on $\mathcal{E}_\gamma^2(t)$, the algorithm does not make any error until the end of round τ_k^t such that $\alpha(\tau_k^t + 1) = \tilde{k}$.

Lemma 2.5.3. *Let $\gamma < 1/6, t \in [R]$ and $k \leqslant |\mathcal{S}^\star|$. Let $\tau_k^t \coloneqq \min\{r \in [t] : \alpha(r+1) = \tilde{k}\} \wedge t$. On the event $\mathcal{E}_\gamma^2(t)$, the algorithm makes no error until the end round τ_k^t and for any $r \leqslant \tau_k^t + 1$, if $j \in \mathcal{A}_r$ is a sub-optimal arm then $j^\star \in \mathcal{A}_r$. In particular, on the event $\mathcal{E}_\gamma^2(t)$, $\{a_1, \dots, a_{\alpha(\tau_k^t+1)}\} \subset \mathcal{S}^\star$.*

Said otherwise, this lemma states that the first arms that will be declared as optimal by EGE-SR-k will be actually optimal if $\mathcal{E}_\gamma^2(t)$ holds for γ small enough.

Proof. In the sequel, we assume that $\mathcal{E}_\gamma^2(t)$ holds. We prove the correctness by induction on the round r. Let $r \in [t]$ and let $\mathcal{B}_r = \{a_1, \dots, a_{\alpha(r)}\}$ denote the arms so far identified as optimal. Let $\mathcal{P}_r$ be the property "for any sub-optimal arm $i \in \mathcal{A}_r$, $i^\star \in \mathcal{A}_r$ and no error occurred so far".

$\mathcal{P}_r$ trivially holds for $r = 1$ and $\alpha(1) = 0$. We now assume that it holds until the beginning of round r and that $\alpha(r) < \tilde{k}$. We will show that the arm i_r de-activated at the end of round r is well classified and that for any sub-optimal arm $j \in \mathcal{A}_{r+1}$, $j^\star \in \mathcal{A}_{r+1}$.

Lemma 2.5.4. *If $\mathcal{P}_r$ holds at round r and $\alpha(r) < \tilde{k}$, there exists $a \in \mathcal{A}_r$ such that $\Delta_a^{(k)} = \Delta_a$ and $\Delta_a^{(k)} \geqslant \Delta^{(k)}_{(K+1-r)}$.*

Proof. At the beginning of round r, it remains $K - r + 1$ active arms so there exists $a \in \mathcal{A}_r$ such that $\Delta_a^{(k)} \geqslant \Delta^{(k)}_{(K+1-r)}$. If a is sub-optimal then $\Delta_a^{(k)} = \Delta_a$. Otherwise, if a is an optimal arm, since $\alpha(r) < \tilde{k}$ and no error has occurred so for (by assumption) then there exists one of the optimal arms $a' \in \mathcal{A}_r$ (one of those with the $\tilde{k}$ largest gaps) such that $\Delta_{a'}^{(k)} = \Delta_{a'}$ and $\Delta^{(k)}_{(K+1-r)} \leqslant \Delta_a^{(k)} \leqslant \Delta_{a'}^{(k)}$.

□

Lemma 2.3.6 still holds for the event $\mathcal{E}_\gamma^2(t)$ with the modified gaps introduced earlier. We state the following lemma, which is similar to Lemma 2.5.2.

Lemma 2.5.5. *Assume that the event $\mathcal{E}_\gamma^2(t)$ holds. Let $r \in [t]$ and assume that for* any *sub-optimal arm $j \in \mathcal{A}_r$, $j^\star \in \mathcal{A}_r$. Then, for any sub-optimal arm $i \in \mathcal{A}_r$,*

$$|\widehat{\Delta}_{i,r}^\star - \Delta_i^\star| \leqslant 2\gamma\Delta_{(K+1-r)}^{(k)}$$

and for any optimal arm $i \in \mathcal{A}_r$,

$$\widehat{\delta}_{i,r}^\star \geqslant \Delta_i - 2\gamma\Delta_{(K+1-r)}^{(k)} \,.$$

As already noted in the proof of Lemma 2.3.9 and Lemma 2.3.8, it is simple to see that the empirical gap $\widehat{\Delta}_{i,r}$ of each arm i satisfies

$$\widehat{\Delta}_{i,r} = \max\left\{\widehat{\Delta}_{i,r}^\star, \widehat{\delta}_{i,r}^\star\right\}. \tag{2.26}$$

It follows from the fact that when $i \in S_r$, $\widehat{\Delta}_{i,r}^\star < 0 \leqslant \widehat{\delta}_{i,r}^\star$ and when $i \in \mathcal{A}_r \backslash S_r$, $\widehat{\delta}_{i,r}^\star < 0 \leqslant \widehat{\Delta}_{i,r}^\star$.

Using Lemma 2.5.5 and the inductive hypothesis $\mathcal{P}_r$ enables us to prove that

$$\forall i \in \mathcal{A}_r, \;\; \widehat{\Delta}_{i,r} \geqslant \Delta_i - 2\gamma\Delta_{(K+1-k)}^{(k)} \tag{2.27}$$

We split the proof in three steps to prove that $\mathcal{P}_{r+1}$ holds.

Step 1: i_r is well classified. Let i_r be empirically sub-optimal ($i_r \notin S_r$) and assume $i_r \in \mathcal{S}^\star$.

Since i_r is removed at the end of round r, by the inductive assumption and using Equation 2.27 and Lemma 2.3.6 (adapted) we have

$$\max_{j \in \mathcal{A}_r \backslash \{i_r\}} \mathrm{m}(i_r, j) \geqslant \Delta_i - 3\gamma\Delta_{(K+1-r)}^{(k)}, \;\; \forall i \in \mathcal{A}_r. \tag{2.28}$$

By Lemma 2.5.4, there exists $a \in \mathcal{A}_r$ such that $\Delta_a = \Delta_a^{(k)} \geqslant \Delta_{(K+1-r)}^{(k)}$. Applying (2.28) to this arm a yields

$$\max_{j \in \mathcal{A}_r \backslash \{i_r\}} \mathrm{m}(i_r, j) \geqslant (1 - 3\gamma)\Delta_{(K+1-r)}^{(k)}. \tag{2.29}$$

Recalling that $\gamma < 1/6$ we see that the LHS of (2.29) is positive, so there exists j such that

$$\mu_{i_r} \prec \mu_j$$

which contradicts the assumption $i_r \in \mathcal{S}^\star$.

Now, if i_r is empirically optimal, *i.e.*, $i_r \in S_r$ assume i_r is a sub-optimal arm that is $i_r \notin \mathcal{S}^\star$. By the hypothesis $\mathcal{P}_r$, $i_r^\star \in \mathcal{A}_r$ and by definition of $\widehat{\Delta}_{i_r,r}$ and since i_r is removed, we have (Lemma 2.3.6 and Equation 2.27)

$$\mathrm{M}(i_r, i_r^\star) \geqslant \Delta_i - 3\gamma\Delta_{(K+1-r)}^{(k)} \quad \forall i \in \mathcal{A}_r. \tag{2.30}$$

Using Lemma 2.5.4 as before, there exists $a \in \mathcal{A}_r$ such that $\Delta^{(k)}_{(K+1-r)} \leqslant \Delta^{(k)}_a = \Delta_a$. So the LHS of (2.30) is positive for $\gamma < 1/6$. That is

$$\mathrm{M}(i_r, i_r^\star) > 0,$$

which is impossible as $\mu_{i_r} \prec \mu_{i_r^\star}$. This concludes the proof of this part: i_r is well classified.

Step 2: If $i_r \in \mathcal{S}^\star$ is an optimal arm, then no active arm is "optimally" dominated by i_r:

$$\forall i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^\mathsf{c}, i^\star \neq i_r$$

To prove this, note that we can deduce from "Step 1" that $i_r \in S_r$. By contradiction, let $i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^\mathsf{c}$ such that $i^\star = i_r$. We claim that i is not dominated by i_r. Indeed, if i were dominated by i_r, we would have $i \notin S_r$ (*i.e.*, empirically sub-optimal) so

$$\widehat{\Delta}_{i,r} = \widehat{\Delta}^\star_{i,r} > 0\,,$$

and since $i_r \in S_r$, by definition of $\widehat{\delta}^\star_{i_r,r}$ and noting that $\widehat{\Delta}_{i_r,r} \geqslant \widehat{\Delta}_{i,r}$, we would have

$$\begin{aligned}\widehat{\Delta}_{i_r,r} &\leqslant (\mathrm{M}(i, i_r; r))_+ + (\widehat{\Delta}^\star_i) \\ &= 0 + \widehat{\Delta}_{i,r},\end{aligned}$$

that is $\widehat{\Delta}_{i_r,r} \leqslant \widehat{\Delta}_{i,r}$. However, our tie-breaking ensures that this inequality is impossible: since $i_r \in S_r, i \notin S_r$ and i_r is deactivated, we have $\widehat{\Delta}_{i_r,r} > \widehat{\Delta}_{i,r}$. Therefore, i is not empirically dominated by i_r. Moreover, since $i_r = i^\star$ (by assumption), Lemma 2.3.7 (it trivially holds with the modified gaps) yields

$$\Delta^\star_i \leqslant \gamma \Delta^{(k)}_{(K+1-r)}. \tag{2.31}$$

Recalling that i_r is deactivated, we have for any arm $j \in \mathcal{A}_r$,

$$(\mathrm{M}(i, i_r; r))_+ + (\widehat{\Delta}^\star_{i,r})_+ \geqslant \widehat{\Delta}_{j,r} \tag{2.32}$$

On the other side, there exists $a \in \mathcal{A}_r$ such that $\Delta^{(k)}_{(K+1-r)} \leqslant \Delta^{(k)}_a = \Delta_a$ (Lemma 2.5.4). Applying Equation (2.27) and Lemma 2.3.6 and taking $j = a$ in Equation 2.32 yields

$$(\mathrm{M}(i, i_r))_+ + (\Delta^\star_i)_+ \geqslant (1 - 5\gamma)\Delta^{(k)}_{(K+1-r)}$$

that is

$$\Delta^\star_i \geqslant (1 - 5\gamma)\Delta^{(k)}_{(K+1-r)}. \tag{2.33}$$

Both (2.31) and (2.33) cannot hold when $\gamma < 1/6$. So for any $i \in \mathcal{A}_r$

$$i^\star \neq i_r.$$

Which completes the proof of "Step 2". Combining, "Step 1&2" proves that $\mathcal{P}_{r+1}$ holds.

Step 3: Conclusion

We have proved that if $\mathcal{P}_r$ holds and $\alpha(r) < \tilde{k}$ then $\mathcal{P}_{r+1}$ holds. Note that if $i_r \in S_r$ then $\alpha(r+1) = \alpha(r) + 1$ otherwise $\alpha(r+1) = \alpha(r)$. Therefore, no error occurs until the end of round $\min\{r', t\}$ where r' is such that $\alpha(r'+1) = \tilde{k}$. As a consequence $\mathcal{B}_{\tau_k^t+1} = \{a_1, \dots, a_{\alpha(\tau_k^t+1)}\} \subset \mathcal{S}^\star$. And by the proof of "Step 2", if $j \in \mathcal{A}_{\tau_p^t}$ is a sub-optimal arm, then $j^\star \in \mathcal{A}_{\tau_p^t+1}$.

□

Theorem 2.3.3. *Let $k \in [K]$. EGE-SR-k satisfies*

$$e_{T,k}(\nu) \leqslant 2(K-1)^2 \left|\mathcal{S}^\star\right| d \cdot e^{-\frac{T-K}{144\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}}.$$

Proof. The proof is a direct consequence of Lemma 2.5.3 and Hoeffding's inequality. Indeed, Lemma 2.5.3 yields that EGE-SR-k is correct on the event $\mathcal{E}_\gamma^2$ for $\gamma < 1/6$. Therefore, for any $0 < \gamma < 1/6$

$$e_{T,k}(\nu) \leqslant \mathbb{P}((\mathcal{E}_\gamma^2)^\mathsf{c}).$$

Letting $0 < \gamma < 1/6$ fixed, by union bound and Hoeffding's inequality, it follows that

$$\mathbb{P}((\mathcal{E}_\gamma^2)^\mathsf{c}) \leqslant 2(K-1)^2|\mathcal{S}^\star|d\exp\left(-\gamma^2\frac{T-K}{4\sigma^2 H_2^{(k)}\overline{\log}(K)(\nu)}\right)$$

therefore,

$$e_{T,k}(\nu) \leqslant 2(K-1)^2|\mathcal{S}^\star|d\exp\left(-\gamma^2\frac{T-K}{4\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right)$$

and taking the limit to $1/6$ (as it holds for any $0 < \gamma < 1/6$) yields the expected bound. □

The theorem below bounds the expected stopping time and the number of samples used at stopping.

Theorem 2.5.6. *Fix $k < |\mathcal{S}^\star|$ and let $q \coloneqq K - |\mathcal{S}^\star| + k$. Then*

$$\begin{aligned}
\mathbb{E}[\tau] &\leqslant q + 2(K-1)|\mathcal{S}^\star|(K-q-1)qd\exp\left(-\frac{T-K}{144\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right) \text{ and} \\
\mathbb{E}[N_\tau] &\leqslant N_q + 2(K-1)|\mathcal{S}^\star|(K-q-1)qdT\exp\left(-\frac{T-K}{144\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right).
\end{aligned}$$

This result suggests that for this relaxed problem, the algorithm might not need to use the whole budget, in particular when T is large. For example, consider a setting $[K] = \mathcal{S}^\star$ then $q = k$ and we roughly use N_k samples which can be way smaller than N_{K-1}.

Proof. The idea is to show that if the algorithm has not stopped after round q then some high probability event must not hold. Let $\gamma < 1/6$ be fixed. We have

$$\begin{aligned}\mathbb{E}[\tau] &\leqslant q + \mathbb{E}[\tau \mathbb{1}(\tau > q)], \\ &\leqslant q + \sum_{s=q+1}^{K-1} \mathbb{P}(\tau \geqslant s).\end{aligned}$$

We claim that for any $s > q$,

$$\{\tau \geqslant s\} \subset \left(\mathcal{E}_\gamma^2(q)\right)^{\mathsf{c}}.$$

Indeed,

$$\begin{aligned}\tau \geqslant s &\implies \tau > q \\ &\implies \alpha(q+1) < k,\end{aligned}$$

However, by Lemma 2.5.3, if $\mathcal{E}_\gamma^2(q)$ holds and $\alpha(q+1) < k$ then no error has occurred until the end of round q, therefore $\mathcal{D}_q = (\mathcal{S}^\star)^{\mathsf{c}}$ and $|\mathcal{B}_q| = k$, which is not possible as $\alpha(q+1) < k$. So $\tau \geqslant s \implies (\mathcal{E}_\gamma^2(q)$ does not hold). Then

$$\begin{aligned}\mathbb{E}[\tau] &\leqslant q + \sum_{s=q+1}^{K-1} \mathbb{P}((\mathcal{E}_\gamma^2(q))^{\mathsf{c}}), &(2.34)\\ &\leqslant q + \sum_{s=q+1}^{K-1} 2(K-1)|\mathcal{S}^\star| qd \exp\left(-\gamma^2 \frac{T-K}{4\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right), &(2.35)\\ &\leqslant q + 2(K-1)|\mathcal{S}^\star|(K-q-1)qd \exp\left(-\gamma^2 \frac{T-K}{4\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right). &(2.36)\end{aligned}$$

Similarly, we have

$$\mathbb{E}[N_\tau] \leqslant N_q + \mathbb{E}[(N_\tau \mathbb{1}(\tau > q))].$$

Next,

$$\begin{aligned}\mathbb{E}[N_\tau] &\leqslant N_q + \sum_{s=q+1}^{K-1} N_s \mathbb{P}(\tau \geqslant s) \\ &\leqslant N_q + \sum_{s=q+1}^{K-1} N_s \mathbb{P}((\mathcal{E}_\gamma^2(q))^{\mathsf{c}}) \\ &\leqslant N_q + 2(K-1)|\mathcal{S}^\star| qd \exp\left(-\gamma^2 \frac{T-K}{4\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right) \sum_{s=q+1}^{K-1} N_s.\end{aligned}$$

Simple algebra yields for any $r \in [K-1]$,

$$\begin{aligned} N_r &= (K-r)n_r + \sum_{s=1}^{r} n_s, \\ &\leqslant \frac{T-K}{\overline{\log}(K)} + (K-r) + r + \frac{T-K}{\overline{\log}(K)} \sum_{s=1}^{r} \frac{1}{K+1-s} \\ &= \frac{T-K}{\overline{\log}(K)} \left(1 + \sum_{s=K+1-r}^{K} s^{-1}\right) + K \\ &\leqslant T. \end{aligned}$$

Therefore,

$$\mathbb{E}[N_\tau] \leqslant N_q + 2(K-1)|\mathcal{S}^\star|(K-q-1)qdT \exp\left(-\gamma^2 \frac{T-K}{4\sigma^2 H_2^{(k)}(\nu)\overline{\log}(K)}\right) \quad (2.37)$$

As (2.36) and (2.37) hold for any $\gamma < 1/6$, taking the limit for a sequence $\gamma \to 1/6$ yields the expected constants in the exponent, which completes the proof. □

2.5.3 Simplifying the sub-optimality gaps

In this section, we prove the lemma that allows us to remove the explicit dependency on $\mathcal{S}^\star$ in the expression of the sub-optimality gaps. Towards proving Lemma 2.1.2, we prove the result below.

Lemma 2.5.7. *For any sub-optimal arm a, there exists a Pareto optimal arm $a^\star$ such that $\mu_a \prec \mu_{a^\star}$ and $\Delta_a = \mathrm{m}(a, a^\star) > 0$. Moreover, for any $i \in [K] \setminus \mathcal{S}^\star$, $j \in \mathcal{S}^\star$*

i) $\max_{k \in \mathcal{S}^\star} \mathrm{m}(i,k) = \max_{k \in [K]} \mathrm{m}(i,k)$,

ii) If $i \in \operatorname{argmin}_{k \in [K]\setminus\{j\}} \mathrm{M}(j,k)$ then j is the unique arm such that $\mu_i \prec \mu_j$.

Proof. Assume there are $p < n$ dominated arms. Without loss of generality, we may assume they are $\mu_1, \dots, \mu_p$. Let $i_1 \leqslant p$. Suppose that no Pareto-optimal arm dominates μ_{i_1}. Since μ_{i_1} is not optimal, by the latter assumption, there exists $i_2 \leqslant p$ such that $\mu_{i_1} \prec \mu_{i_2}$. If μ_{i_2} is dominated by a Pareto optimal arm, this arm also dominates μ_{i_1} (strict dominance is transitive) which contradicts the initial assumption. If not, there exits $i_3 \leqslant p$ such that $\mu_{i_1} \prec \mu_{i_2} \prec \mu_{i_3}$. Again we can use the same reasoning as before for i_3. In any case we should stop in at most p steps, otherwise we would have $\mu_{i_1} \prec \mu_{i_2} \prec \cdots \prec \mu_{i_p}$ and μ_{i_p} should be dominated by a Pareto-optimal arm, otherwise it would be itself Pareto-optimal, which is not the case. Therefore, for any $a \in [K] \setminus \mathcal{S}^\star$, there exists $a^\star \in \mathcal{S}^\star$ such that $a^\star \prec a$ and $\Delta_a = \mathrm{m}(a, a^\star) > 0$.

Letting i be a sub-optimal arm, since for any $a \in [K] \setminus \mathcal{S}^\star$, there exists $a^\star \in \mathcal{S}^\star$ such that $a \prec a^\star$, it follows that

$$\forall c \in [d],\ \mu_a^c - \mu_i^c < \mu_{a^\star}^c - \mu_i^c,$$

which leads to $\mathrm{m}(i,a) \leqslant \mathrm{m}(i,a^\star)$, so

$$\max_{j \in [K]} \mathrm{m}(i,j) = \max_{j \in \mathcal{S}^\star} \mathrm{m}(i,j) > 0,$$

which completes the proof of the first point *i)*. For the second point, let $q \in [K] \setminus \mathcal{S}^\star$ and q' such that $q \prec q'$ and

$$q \in \operatorname*{argmin}_{a \in [K] \setminus \{j\}} \mathrm{M}(j,a).$$

By direct algebra, since $q \prec q'$, we have

$$\mathrm{M}(j,q') < \mathrm{M}(j,q),$$

which is impossible if $q' \neq j$ (because q belongs to the argmin). Therefore, if

$$q \in \operatorname*{argmin}_{a \in [K] \setminus \{j\}} \mathrm{M}(j,a)$$

is a sub-optimal arm, then j is the only arm such that $q \prec j$ (*i.e.*, $\mu_q \prec \mu_j$). □

We now prove the following result, which follows from the lemma above.

Lemma 2.5.8. *For any Pareto optimal arm i, $\Delta_i \leqslant \min_{j \neq i} \mathrm{M}(i,j)$.*

Proof of Lemma 2.5.8. If $\operatorname{argmin}_{j \neq i} \mathrm{M}(i,j) \subset \mathcal{S}^\star$, then the lemma follows from the definition of the gap of an optimal arm recalled in Section 4. If $\min_{j \neq i} \mathrm{M}(i,j) = \mathrm{M}(i,a)$, $a \notin \mathcal{S}^\star$, then, from Lemma 2.5.7, i is the unique arm which dominates a so $\Delta_a = \mathrm{m}(a,i)$ and using the definition of the gap of an optimal arm,

$$\begin{aligned} \Delta_i &\leqslant \mathrm{M}(a,i)_+ + \Delta_a, \\ &= 0 + \mathrm{m}(a,i) \leqslant \mathrm{M}(i,a), \end{aligned}$$

where we have used the fact that $\mathrm{m}(p,q) \leqslant \mathrm{M}(q,p)$ for any pair of arms p, q (which follows from the definition). Therefore, for an optimal arm i, we always have

$$\Delta_i \leqslant \min_{j \neq i} \mathrm{M}(i,j).$$

□

We can now finish by proving Lemma 2.1.2.

Proof of Lemma 2.1.2. For sub-optimal arms, the result follows from Lemma 2.5.7. It remains to prove the equality for sub-optimal arms. By definition, for an optimal arm i, we have

$$\Delta_i = \min\left\{\delta_i^+, \delta_i^-\right\},$$

where

$$\delta_i^+ := \min_{j\in\mathcal{S}^\star\setminus\{i\}} \min\left\{\mathrm{M}(i,j), \mathrm{M}(j,i)\right\} \text{ and } \delta_i^- := \min_{j\in[K]\setminus\mathcal{S}^\star} \left[(\mathrm{M}(j,i))_+ + \Delta_j^\star\right].$$

For any optimal arm i, $\Delta_i^\star \leqslant 0$ (by direct calculation), so introducing

$$\delta_i^{-'} := \min\left\{\delta_i^-, \min_{j\in\mathcal{S}^\star\setminus\{i\}} \mathrm{M}(j,i)\right\},$$

we have

$$\delta_i^{-'} = \min_{j\neq i}\left[\mathrm{M}(j,i)_+ + (\Delta_j^\star)_+\right]$$

Then, if

$$\min_{j\in\mathcal{S}^\star\setminus\{i\}} \mathrm{M}(i,j) = \min_{j\neq i} \mathrm{M}(i,j), \tag{2.38}$$

holds, the result simply follows as for any optimal arm i,

$$\begin{aligned}
\Delta_i = \min\left\{\delta_i^+, \delta_i^-\right\} &= \min\left\{\min_{j\in\mathcal{S}^\star\setminus\{i\}} \mathrm{M}(i,j), \delta_i^{-'}\right\},\\
&= \min(\min_{j\neq i} \mathrm{M}(i,j), \delta_i^{-'}),\\
&= \min_{j\neq i}\left[\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta_j^\star)_+)\right],\\
&= \delta_i^\star.
\end{aligned}$$

In the sequel, assume (2.38) does not hold, that is assume

$$\min_{j\in\mathcal{S}^\star\setminus\{i\}} \mathrm{M}(i,j) > \min_{j\neq i} \mathrm{M}(i,j).$$

From Lemma 2.5.7 we know that in this case, there exists a sub-optimal arm k such that i is the unique arm dominating k and

$$\Delta_k^\star = \mathrm{m}(k,i) \quad \text{and} \quad \min_{j\neq i} \mathrm{M}(i,j) = \mathrm{M}(i,k)\,.$$

Thus, we have

$$\begin{aligned}
\min_{j\neq i} \mathrm{M}(i,j) &= \mathrm{M}(i,k), &(2.39)\\
&\geqslant \mathrm{m}(k,i) = \Delta_k^\star, &(2.40)\\
&\geqslant \delta_i^{-'} \quad \text{(since } i \text{ dominates } k, \mathrm{M}(k,i)_+ = 0). &(2.41)
\end{aligned}$$

On the other side, as $\mu_k \prec \mu_i$, using the definition of Δ_i in particular δ_i^- directly yields

$$\begin{aligned}
\Delta_i &\leqslant \Delta_k^\star = \mathrm{m}(k,i) &(2.42)\\
&\leqslant \mathrm{M}(i,k) = \min_{j\neq i} \mathrm{M}(i,j) &(2.43)\\
&< \min_{j\in\mathcal{S}^\star\setminus\{i\}} \mathrm{M}(i,j). &(2.44)
\end{aligned}$$

We recall that

$$\Delta_i = \min\left\{\delta_i^{-'}, \min_{j\in\mathcal{S}^\star\setminus\{i\}} \mathrm{M}(i,j)\right\},$$

which combined with (2.44) yields

$$\Delta_i = \delta_i^{-'}$$

and further combining with (2.41) yields

$$\begin{aligned}\Delta_i = \delta_i^{-'} &= \min\left\{\min_{j\neq i} \mathrm{M}(i,j), \delta_i^{-'}\right\} \\ &= \delta_i^\star,\end{aligned}$$

which concludes the proof. □

Chapter 3

Adaptive Algorithms for Fixed-Confidence Pareto Set Identification

In this chapter, we study the problem of fixed-confidence identification of the Pareto set in multi-objective multi-armed bandits. Since the sample complexity required to exactly identify the Pareto set can be prohibitive, previous work has proposed relaxations that tolerate the inclusion of a few near-optimal arms. Here, we explore alternative and complementary relaxations that instead allow the learner to identify a relevant *subset* of the Pareto set.

We introduce a unified adaptive sampling strategy, called Adaptive Pareto Exploration (APE), which can be combined with different stopping rules to address multiple relaxations of the Pareto set identification problem within a single framework. We provide a theoretical analysis of the resulting algorithms, establishing instance-dependent sample complexity guarantees and quantifying the gains achieved when identifying at most k Pareto-optimal arms.

This chapter is based on joint work with Émilie Kaufmann and Laura Richert, published in the Proceedings of *NeurIPS 2023*.

3.1 Introduction

As discussed in Section 1.2 of the introductory chapter, fixed-confidence pure exploration problems aim to make statistically reliable decisions while explicitly controlling the probability of error through a prescribed risk δ. In this chapter, we revisit this setting for *Pareto Set Identification* (PSI) in multi-objective bandits, focusing on the design and analysis of *fully adaptive* algorithms that allocate samples where they are most informative.

Unlike in fixed-budget exploration—studied in Chapter 2—the total number of samples in this setting is not fixed in advance: sampling continues until the confidence in the identified set reaches the desired level $1-\delta$. This formulation directly aligns with regulatory and ethical requirements in clinical research, where one must control the probability of an incorrect recommendation while minimizing the expected number of enrolled participants.

Beyond clinical trials, active Pareto set identification is also relevant in domains such as hardware and software design (Zuluaga, Sergent, et al. 2013), or multi-objective recommender systems and A/B/n testing, where one jointly optimizes engagement, diversity, and platform objectives (Mehrotra et al. 2020).

Relaxed identification tasks. Before introducing our relaxations, we recall the notion of Pareto dominance. Given two mean vectors $\mu_i, \mu_j \in \mathbb{R}^d$, we say that arm i is (strictly) dominated by arm j, denoted $\mu_i \prec \mu_j$, if $\mu_i^c < \mu_j^c$ for all coordinates $c \in \{1, \dots, d\}$. Intuitively, j outperforms i on every objective. For any $\varepsilon \geqslant 0$, we define the ε-Pareto set as

$$\mathcal{S}_\varepsilon^\star := \left\{ i \in [K] : \nexists\, j \in [K] \text{ such that } \mu_i + \varepsilon \mathbf{1} \prec \mu_j \right\},$$

that is, the set of arms that remain non-dominated when all objectives are relaxed by an additive margin ε. In particular, $\mathcal{S}_0^\star$ coincides with the exact Pareto set $\mathcal{S}^\star$.

Exact identification of $\mathcal{S}^\star$ can be prohibitively costly when the Pareto frontier is large or when many arms are nearly optimal. We therefore consider several relaxed formulations that capture realistic decision-making constraints:

- ε_1**-PSI** (Auer et al. 2016): the returned set $\widehat{S}$ must contain all Pareto-optimal arms and may include arms that become Pareto-optimal after an additive ε_1 relaxation, *i.e.*, $\mathcal{S}^\star \subseteq \widehat{S} \subseteq \mathcal{S}_{\varepsilon_1}^\star$;
- $(\varepsilon_1, \varepsilon_2)$**-PSI**: generalizing the ε-cover of Zuluaga, Krause, et al. 2016, this criterion allows the learner to ignore arms that are nearly equivalent (within ε_2) to others already selected, which is particularly meaningful when several treatments exhibit indistinguishable efficacy;
- ε_1**-PSI-**k: introduced in Chapter 2 for the fixed-budget case, this version targets the identification of at most k promising candidates—such as a limited number of vaccine strategies that can proceed to later phases—balancing statistical confidence with practical feasibility.

These relaxations define distinct correctness criteria but share the same fixed-confidence goal: ensuring δ-correct identification while minimizing the expected number of samples.

Related work. The most closely related study is that of Auer et al. 2016, who introduced the ε_1-PSI relaxation and provided instance-dependent sample complexity bounds under subgaussian noise. In parallel, Zuluaga, Sergent, et al. 2013; Zuluaga, Krause, et al. 2016 proposed structured variants of PSI in which the mean vectors are smooth functions of arm descriptors modeled via Gaussian processes, leading to worst-case guarantees. Their notion of an ε-cover of the Pareto set is conceptually related to our $(\varepsilon_1, \varepsilon_2)$-PSI formulation. Both Auer's and Zuluaga's algorithms rely on *uniform exploration* over an active set combined with accept/reject mechanisms: an approach that can be sample-inefficient when the number of arms is large, as already observed in single-objective Best Arm Identification (Kaufmann & Kalyanakrishnan 2013; Kaufmann, Cappé, et al. 2016).

Our objective is to develop a fully adaptive exploration strategy that dynamically reallocates samples to the most ambiguous arms and objectives: those whose dominance relations remain uncertain.

Contributions. We introduce *Adaptive Pareto Exploration* (APE), a unified, adaptive framework for fixed-confidence PSI encompassing the three relaxations above. APE extends LUCB- and UGap-style principles from single-objective bandits (Gabillon et al. 2012; Kalyanakrishnan et al. 2012) to the multi-objective setting by maintaining confidence intervals on pairwise dominance gaps between arms. Depending on the chosen stopping and recommendation rules, APE can address ε_1-PSI, $(\varepsilon_1, \varepsilon_2)$-PSI, or ε_1-PSI-k tasks, with instance-dependent guarantees (Section 3.2). Our analysis further quantifies how relaxing the identification goal—through ε_1, ε_2, or k—provably reduces the required sample complexity. Empirical studies (Section 3.4) on synthetic and vaccine-inspired data corroborate these theoretical findings, showing that adaptive exploration yields substantial efficiency gains over uniform baselines.

Learning model. We formalize the *Pareto Set Identification* (PSI) problem and its relaxed variants. Let $K, d \in \mathbb{N}^\star$. Each arm $a \in [K] \coloneqq \{1, \dots, K\}$ is associated with an unknown distribution ν_a over $\mathbb{R}^d$ with mean vector $\mu_a = (\mu_a^1, \dots, \mu_a^d)^\top$. We denote the bandit instance by $\nu \coloneqq (\nu_1, \dots, \nu_K)$ and its mean matrix by $\mu \coloneqq (\mu_1, \dots, \mu_K)$. At each round $t = 1, 2, \dots$, the learner selects an arm $A_t \in [K]$ based on past observations and observes an independent draw $Z_t \sim \nu_{A_t}$ with $\mathbb{E}[Z_t | A_t] = \mu_{A_t}$. Each coordinate of Z_t is assumed to be *1-subgaussian* conditionally on A_t. We recall that a random variable X is σ-subgaussian if and only if

$$\forall\, \lambda \in \mathbb{R},\ \log \mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \leqslant \frac{\lambda^2 \sigma^2}{2}.$$

This standard assumption ensures concentration of empirical means around their expectations and covers both Gaussian and bounded observation models.

We denote by $\mathbb{P}_\nu$ the law of the stochastic process $(Z_t)_{t \geqslant 1}$ and by $\mathbb{E}_\nu$ the associated expectation. Let $\mathcal{H}_t \coloneqq \sigma(A_1, Z_1, \dots, A_t, Z_t)$ be the σ-algebra encoding the history up to round t. The algorithm stops sampling at some random time τ, and finally outputs a recommendation $\widehat{S}_\tau$. The goal is to make a correct guess with high probability while using as few samples as possible.

Depending on the application, different notions of correctness can be considered, controlled by parameters $\varepsilon_1, \varepsilon_2 \geqslant 0$ and $k \in [K]$:

Definition 3.1.1. A set $\widehat{S} \subseteq [K]$ is correct for ε_1-PSI if

$$\mathcal{S}^\star \subseteq \widehat{S} \subseteq \mathcal{S}^\star_{\varepsilon_1} ,$$

where $\mathcal{S}^\star_{\varepsilon_1}$ denotes the set of ε_1-Pareto-optimal arms.

Definition 3.1.2. A set $\widehat{S} \subseteq [K]$ is an $(\varepsilon_1, \varepsilon_2)$-cover of the Pareto set if $\widehat{S} \subseteq \mathcal{S}^\star_{\varepsilon_1}$ and, for any $i \notin \widehat{S}$, either $i \notin \mathcal{S}^\star$ or there exists $j \in \widehat{S}$ such that $\mu_i \prec \mu_j + \mathbf{1}\varepsilon_2$. This generalizes the ε-accurate set of Zuluaga, Krause, et al. 2016 (recovered for $\varepsilon_1 = \varepsilon_2 = \varepsilon$) and relaxes the requirement of returning all near-optimal arms.

Definition 3.1.3. A set $\widehat{S} \subseteq [K]$ is correct for ε_1-PSI-k if (i) $|\widehat{S}| = k$ and $\widehat{S} \subseteq \mathcal{S}^\star_{\varepsilon_1}$, or (ii) $|\widehat{S}| < k$ and $\mathcal{S}^\star \subseteq \widehat{S} \subseteq \mathcal{S}^\star_{\varepsilon_1}$.

Given an objective (one of the above tasks) and a risk parameter $\delta \in (0, 1)$, an algorithm is said to be δ-correct if, with probability at least $1 - \delta$, its recommendation $\widehat{S}_\tau$ satisfies the corresponding correctness criterion. The number of samples τ required to reach this guarantee defines the algorithm's sample complexity.

We now introduce two quantities that play a central role in characterizing the (Pareto) optimality and sub-optimality relations between arms. For any arm pair $i, j \in [K]$, define

$$\mathrm{m}(i, j) \coloneqq \min_{1 \leqslant c \leqslant d} \left[\mu^c_j - \mu^c_i\right], \quad \text{and} \quad \mathrm{M}(i, j) \coloneqq \max_{1 \leqslant c \leqslant d} \left[\mu^c_i - \mu^c_j\right] .$$

As explained in Chapter 1, these quantities capture the smallest and largest coordinate-wise performance gaps between the two arms.

3.2 Adaptive Pareto Exploration

We describe in this section our sampling rule, Adaptive Pareto Exploration and present three stopping and recommendation rules with which it can be combined to solve each of the proposed relaxations.

Let $N_{t,k} \coloneqq \sum_{s=1}^{t-1} \mathbb{1}(A_s = k)$ be the number of times arm k has been pulled up to round t and $\hat{\mu}_{t,k} \coloneqq N_{t,k}^{-1} \sum_{s=1}^{N_{t,k}} Z_{k,s}$ be the empirical mean of this arm at time t, where $Z_{k,s}$ denotes the s-th observation drawn *i.i.d.* from ν_k. For any arm pair $i, j \in [K]$, we let

$$\mathrm{m}(i, j; t) \coloneqq \min_{c \leqslant d} \left[\hat{\mu}^c_{t,j} - \hat{\mu}^c_{t,i}\right] \quad \text{and} \quad \mathrm{M}(i, j; t) \coloneqq \max_{c \leqslant d} \left[\hat{\mu}^c_{t,i} - \hat{\mu}^c_{t,j}\right] .$$

The empirical Pareto set is defined as

$$\begin{aligned} S(t) &\coloneqq \{i \in [K] : \nexists j \in [K] : \hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}\}, \\ &= \{i \in [K] : \forall j \in [K] \backslash \{i\}, \mathrm{M}(i, j; t) > 0\} . \end{aligned}$$

3.2.1 Generic algorithm(s)

Adaptive Pareto Exploration relies on a *lower/upper confidence bound* approach, similar to single-objective BAI algorithms like UGap (Gabillon et al. 2012), LUCB (Kalyanakrishnan et al. 2012), and LUCB++ (Simchowitz et al. 2017). These three algorithms identify, in each round, two contentious arms. A leader arm b_t, a current guess for the optimal arm (defined as the empirical best arm or the arm with the smallest upper bound on its sub-optimality gap), and c_t: a contender of b_t; the arm that is the most likely to outperform b_t (in all three algorithms, it is the arm with the largest upper confidence bound in $[K]\backslash\{b_t\}$). Then either both arms are pulled (LUCB, LUCB++) or the least explored among b_t and c_t is pulled (UGap).

The originality of our sampling rule lies in how to appropriately define b_t and c_t for the multi-objective setting. To define those, we suppose that there exists confidence intervals $[L^c_{i,j}(t,\delta), U^c_{i,j}(t,\delta)]$ on the difference of expected values for each pair of arms (i,j) and each objective $c \in \{1,2,\dots,d\}$, such that introducing

$$\mathcal{E}_t \coloneqq \bigcap_{i=1}^{K}\bigcap_{j\neq i}\bigcap_{c=1}^{d}\left\{L^c_{i,j}(t,\delta) \leqslant \mu^c_i - \mu^c_j \leqslant U^c_{i,j}(t,\delta)\right\} \text{ and } \mathcal{E} = \bigcap_{t=1}^{\infty}\mathcal{E}_t, \tag{3.1}$$

we have $\mathbb{P}(\mathcal{E}) \geqslant 1-\delta$. Concrete choices of these confidence intervals will be discussed in Section 3.2.2.

We further define

$$\mathrm{M}^-(i,j;t) \coloneqq \max_{c\leqslant d} L^c_{i,j}(t,\delta) \quad \text{and} \quad \mathrm{M}^+(i,j;t) \coloneqq \max_{c\leqslant d} U^c_{i,j}(t,\delta) \tag{3.2}$$

$$\mathrm{m}^-(i,j;t) \coloneqq -\mathrm{M}^+(i,j;t) \quad \text{and} \quad \mathrm{m}^+(i,j;t) \coloneqq -\mathrm{M}^-(i,j;t). \tag{3.3}$$

In the sequel, we simplify notation by omitting the dependence on δ and write $L^c_{i,j}(t)$ (resp. $U^c_{i,j}(t)$) in place of $L^c_{i,j}(t,\delta)$ (resp. $U^c_{i,j}(t,\delta)$).

Lemma 3.2.1. *For any round $t \geqslant 1$, if $\mathcal{E}_t$ holds, then for any $i,j \in [K]$, $\mathrm{M}^-(i,j;t) \leqslant \mathrm{M}(i,j) \leqslant \mathrm{M}^+(i,j;t)$ and $\mathrm{m}^-(i,j;t) \leqslant \mathrm{m}(i,j) \leqslant \mathrm{m}^+(i,j;t)$.*

Noting that $\mathcal{S}^\star_{\varepsilon_1} = \{i \in [K] : \forall j \neq i, \mathrm{M}(i,j) + \varepsilon_1 > 0\}$, we define the following set of arms that are likely to be ε_1-Pareto optimal:

$$\mathrm{OPT}^{\varepsilon_1}(t) \coloneqq \left\{i \in [K] : \forall\, j \in [K]\backslash\{i\}, \mathrm{M}^-(i,j;t) + \varepsilon_1 > 0\right\}.$$

Sampling rule. In round t, Adaptive Pareto Exploration samples A_t, the least pulled arm among two candidate arms b_t and c_t given by

$$\begin{aligned} b_t &\coloneqq \operatorname*{argmax}_{i\in[K]\backslash\,\mathrm{OPT}^{\varepsilon_1}(t)} \min_{j\neq i} \mathrm{M}^+(i,j;t), \\ c_t &\coloneqq \operatorname*{argmin}_{j\neq b_t} \mathrm{M}^-(b_t,j;t) \end{aligned}$$

The intuition for their definition is the following. Letting i be a fixed arm, note that $\mathrm{M}(i,j) > 0$ for some j, if and only if there exists an objective c such that $\mu^c_i > \mu^c_j$, *i.e.*, i is

not dominated by j. Moreover, the larger $\mathrm{M}(i,j)$, the more i is non-dominated by j in the sense that there exists c such that $\mu_i^c \gg \mu_j^c$. Therefore, i is strictly optimal if and only if for all $j \neq i$, $\mathrm{M}(i,j) > 0$ *i.e.*, $\alpha_i \coloneqq \min_{j\neq i} \mathrm{M}(i,j) > 0$.

The larger α_i, the more optimal i looks, in the sense that for each arm $j \neq i$, there exists a component d_j for which i is way better than j. As the α_i are unknown, we define b_t as the maximizer of an optimistic estimate of the α_i's. We further restrict the maximization to arms for which we are not already convinced that they are optimal (by Lemma 3.2.1, the arms in $\mathrm{OPT}^{\varepsilon_1}(t)$ are (nearly) Pareto optimal on the event $\mathcal{E}$). Then, we note that for a fixed arm i, $\mathrm{M}(i,j) < 0$ if and only if i is strictly dominated by j. And the smaller $\mathrm{M}(i,j)$, the closer j is to dominating i (or largely dominates it): for any component c, $\mu_i^c - \mu_j^c$ is small (or negative). Thus, for a fixed arm i, $\operatorname{argmin}_{j\neq i} \mathrm{M}(i,j)$ can be seen as the arm that is the closest to dominating i (or the one that dominates it by the largest margin). By minimizing a lower confidence bound on the unknown quantity $\mathrm{M}(b_t,j)$, our contender c_t can be interpreted as the arm that is the most likely to be (close to) dominating b_t. Gathering information on both b_t and c_t can be useful to check whether b_t can indeed be optimal.

Stopping and recommendation rule(s). Depending on the objective, Adaptive Pareto Exploration can be plugged in with different stopping rules, which are summarized in Table 3.1 with their associated recommendations. To define those, we introduce for all $i \in [K]$, $\varepsilon_1, \varepsilon_2 \geqslant 0$,

$$g_i^{\varepsilon_2}(t) \coloneqq \max_{j\neq i} \left[\mathrm{m}^-(i,j;t) + \varepsilon_2 \mathbb{1}\left(j \in \mathrm{OPT}^{\varepsilon_1}(t)\right)\right] \quad \text{and} \quad h_i^{\varepsilon_1}(t) \coloneqq \min_{j\neq i} \mathrm{M}^-(i,j;t) + \varepsilon_1.$$

and let $g_i(t) \coloneqq g_i^0(t)$. Introducing

$$Z_1^{\varepsilon_1}(t) \coloneqq \min_{i\in S(t)} h_i^{\varepsilon_1}(t), \text{ and } Z_2^{\varepsilon_1}(t) \coloneqq \min_{i\in S(t)^{\mathsf{c}}} \max\{g_i(t), h_i^{\varepsilon_1}(t)\},$$

for ε_1-PSI, our stopping rule is $\tau_{\varepsilon_1} \coloneqq \inf\{t \geqslant K : Z_1^{\varepsilon_1}(t) > 0 \wedge Z_2^{\varepsilon_1}(t) > 0\}$ and the associated recommendation is $\mathcal{O}(\tau_{\varepsilon_1})$ where

$$\mathcal{O}(t) \coloneqq S(t) \cup \left\{i \in S(t)^{\mathsf{c}} : \nexists j \neq i : \mathrm{m}^-(i,j;t) > 0\right\}$$

consists of the current empirical Pareto set plus some additional arms that have not yet been formally identified as sub-optimal. Those arms should be (ε_1)-Pareto optimal.

For $(\varepsilon_1,\varepsilon_2)$-PSI we define a similar stopping rule $\tau_{\varepsilon_1,\varepsilon_2}$ where the stopping statistics are respectively replaced by

$$Z_1^{\varepsilon_1,\varepsilon_2}(t) \coloneqq \min_{i\in S(t)} \max\{g_i^{\varepsilon_2}(t), h_i^{\varepsilon_1}(t)\} \text{ and } Z_2^{\varepsilon_1,\varepsilon_2}(t) \coloneqq \min_{i\in S(t)^{\mathsf{c}}} \max\{g_i^{\varepsilon_2}(t), h_i^{\varepsilon_1}(t)\}$$

with the convention $\min_\emptyset = +\infty$, and the recommendation is $\mathrm{OPT}^{\varepsilon_1}(\tau_{\varepsilon_1,\varepsilon_2})$.

To tackle the ε_1-PSI-k relaxation, we propose to couple τ_{ε_1} with an additional stopping condition checking whether $\mathrm{OPT}^{\varepsilon_1}(t)$ already contains k arms. That is, we stop at $\tau_{\varepsilon_1}^k \coloneqq \min\left\{\tau_{\varepsilon_1}, \tau^k\right\}$ where $\tau^k \coloneqq \inf\{t \geqslant K : |\mathrm{OPT}^{\varepsilon_1}(t)| \geqslant k\}$ with associated recommendation $\mathrm{OPT}^{\varepsilon_1}(\tau^k)$. Depending on the reason for stopping (τ_{ε_1} or τ^k), we follow the corresponding recommendation.

Table 3.1: Stopping conditions and associated recommendation

	Stopping Condition	Recommendation	Objective
τ_{ε_1}	$Z_1^{\varepsilon_1}(t) > 0 \ \wedge Z_2^{\varepsilon_1}(t) > 0$	$\mathcal{O}(\tau_{\varepsilon_1})$ or $\mathrm{OPT}^{\varepsilon_1}(\tau_{\varepsilon_1})$	ε_1-PSI
$\tau_{\varepsilon_1,\varepsilon_2}$	$Z_1^{\varepsilon_1,\varepsilon_2}(t) > 0 \ \wedge Z_2^{\varepsilon_1,\varepsilon_2}(t) > 0$	$\mathrm{OPT}^{\varepsilon_1}(\tau_{\varepsilon_1,\varepsilon_2})$	$(\varepsilon_1,\varepsilon_2)$-PSI
τ^k	$\lvert\mathrm{OPT}^{\varepsilon_1}(t)\rvert \geqslant k$	$\mathrm{OPT}^{\varepsilon_1}(\tau^k)$	ε_1-PSI-k

Lemma 3.2.2. *Assume $\mathcal{E}$ holds. For ε_1-PSI (resp. $(\varepsilon_1,\varepsilon_2)$-PSI, ε_1-PSI-k), Adaptive Pareto Exploration combined with the stopping rule τ_{ε_1} (resp. $\tau_{\varepsilon_1,\varepsilon_2}$, resp. $\tau^k_{\varepsilon_1}$) outputs a correct subset.*

Remark 3.1. We decoupled the presentation of the sampling rule from that of the "sequential testing" aspect (stopping and recommendation). We could even go further and observe that multiple tests could actually be run in parallel, for free. If we collect samples with APE (which only depends on ε_1), whenever one of the three stopping conditions given in Table 3.1 triggers, for any value of ε_2 or k, we can decide to stop and make the corresponding recommendation or continue and wait for another "more interesting" stopping condition to be satisfied. If $\mathcal{E}$ holds, a recommendation made at any such time will be correct for the objective associated to the stopping criterion (third column in Table 3.1).

Proof of Lemma 3.2.2.

Step 1: correctness for ε_1-PSI-k. We show the correctness of ε_1-PSI-k (for any k), which includes ε_1-PSI as a special case: ε_1-PSI-K. Assuming $\mathcal{E}$ holds, let $t = \tau^k_{\varepsilon_1}$ and $i \in \mathrm{OPT}^{\varepsilon_1}(t)$. Since $i \in \mathrm{OPT}^{\varepsilon_1}(t)$, for any $j \neq i$,

$$\mathrm{M}(i,j) + \varepsilon_1 \overset{\mathcal{E}}{\geqslant} \mathrm{M}^-(i,j;t) + \varepsilon_1 > 0,$$

that is $i \in \mathcal{S}^\star_{\varepsilon_1}$. Therefore, on the event $\mathcal{E}$, $\mathrm{OPT}^{\varepsilon_1}(t) \subseteq \mathcal{S}^\star_{\varepsilon_1}$. Thus, if the stopping has occurred because $|\,\mathrm{OPT}^{\varepsilon_1}(t)| \geqslant k$, since in this case $\mathrm{OPT}^{\varepsilon_1}(t) \subseteq \mathcal{S}^\star_{\varepsilon_1}$, all the recommended arms will be ε_1-Pareto optimal. On the contrary, if $|\,\mathrm{OPT}^{\varepsilon_1}(t)| < k$, then from the definition of $\tau^k_{\varepsilon_1}$ it holds that

$$Z_1^{\varepsilon_1}(t) > 0 \quad \text{and} \quad Z_2^{\varepsilon_1}(t) > 0.$$

We will show that $\mathcal{O}(t)$ recalled below:

$$\mathcal{O}(t) \ := \ S(t) \cup \big\{ i \in S(t)^{\mathsf{c}} : \nexists\, j \ \neq \ i : \mathrm{m}^-(i,j;t) > 0 \big\},$$

is an ε_1-Pareto set (at $t = \tau_{\varepsilon_1}$) and it is a subset of $\mathrm{OPT}^{\varepsilon_1}(t)$. We note that, for any $i \in \mathcal{O}(t)^{\mathsf{c}}$, by the definition, and since $Z_2(t) > 0$,

$$\exists j \in [K] \ \text{ such that } \ \mathrm{m}(i,j) \overset{\mathcal{E}}{\geqslant} \mathrm{m}^-(i,j;t) > 0,$$

so i is a sub-optimal arm. Therefore,

$$\mathcal{S}^\star \subseteq \mathcal{O}(t)\,.$$

Moreover, for any $i \in \mathcal{O}(t) \cap S(t)$, since $Z_1^{\varepsilon_1}(t) > 0$ we have $h_i^{\varepsilon_1}(t) > 0$, so $i \in \text{OPT}^{\varepsilon_1}(t)$, and, as above

$$\min_{j \in [K]\setminus\{i\}} \mathrm{M}(i,j) + \varepsilon_1 \overset{\mathcal{E}}{\geqslant} \min_{j \in [K]\setminus\{i\}} \mathrm{M}^-(i,j;t) + \varepsilon_1 > 0. \tag{3.4}$$

If $i \in \mathcal{O}(t) \cap S(t)^\mathsf{c}$, by definition of $\mathcal{O}(t)$, we have $g_i(t) < 0$. However, since $Z_2^{\varepsilon_1}(t) > 0$, $\max\{g_i(t), h_i^{\varepsilon_1}(t)\} > 0$ so we also have $h_i^{\varepsilon_1}(t) > 0$ and (3.4) applies. Thus, for any $i \in \mathcal{O}(t)$, again,

$$\min_{j \in [K]\setminus\{i\}} \mathrm{M}(i,j) + \varepsilon_1 > 0,$$

that is $i \in \mathcal{S}^\star_{\varepsilon_1}$, so $\mathcal{S}^\star \subseteq \mathcal{O}(t) \subseteq \mathcal{S}^\star_{\varepsilon_1}$ and $\mathcal{O}(t) \subseteq \text{OPT}^{\varepsilon_1}(t) \subseteq \mathcal{S}^\star_{\varepsilon_1}$, which implies that $\text{OPT}^{\varepsilon_1}(t)$ is itself an (ε_1)-Pareto set.

Finally the correctness for ε_1-PSI-k and ε_1-PSI follows on $\mathcal{E}$.

Step 2: correctness for $(\varepsilon_1, \varepsilon_2)$-PSI. Let $t = \tau_{\varepsilon_1,\varepsilon_2}$ and $i \in \text{OPT}^{\varepsilon_1}(t)$. Since $i \in \text{OPT}^{\varepsilon_1}(t)$, for any $j \neq i$, $\mathrm{M}(i,j) + \varepsilon_1 \overset{\mathcal{E}}{\geqslant} \mathrm{M}^-(i,j;t) + \varepsilon_1 > 0$ that is $i \in \mathcal{S}^\star_{\varepsilon_1}$. Therefore, on the event $\mathcal{E}$, $\text{OPT}^{\varepsilon_1}(t) \subseteq \mathcal{S}^\star_{\varepsilon_1}$. When the stopping time $\tau_{\varepsilon_1,\varepsilon_2}$ is reached, $Z_1^{\varepsilon_1,\varepsilon_2}(t) > 0$ and $Z_2^{\varepsilon_1,\varepsilon_2}(t) > 0$. Under this condition,

$$\text{OPT}^{\varepsilon_1}(t) \neq \emptyset.$$

Indeed, since $Z_1^{\varepsilon_1,\varepsilon_2}(t) > 0$ and $Z_2^{\varepsilon_1,\varepsilon_2}(t) > 0$, if $\text{OPT}^{\varepsilon_1}(t) = \emptyset$ then, by the stopping rule and since $\text{OPT}^{\varepsilon_1}(t) = \emptyset$, for any arm i, we would have $h_i^{\varepsilon_1}(t) < 0$ and $g_i^{\varepsilon_2}(t) > 0$. That is, for any arm $i \in [K]$,

$$\exists j \neq i \text{ such that } \mathrm{m}(i,j) \overset{\mathcal{E}}{>} \mathrm{m}^-(i,j,t) > 0,$$

so every arm would be strictly dominated, which is impossible since the Pareto set cannot be empty. Then, $\text{OPT}^{\varepsilon_1}(t) \neq \emptyset$ and for any $i \in \mathcal{O}(t)^\mathsf{c} = \text{OPT}^{\varepsilon_1}(t)^\mathsf{c}$, by the stopping rule it holds that $\max\{g_i^{\varepsilon_2}(t), h_i^{\varepsilon_1}(t)\} > 0$. Further, noting that for such arm $i \in \text{OPT}^{\varepsilon_1}(t)^\mathsf{c}$, $h_i^{\varepsilon_1}(t) < 0$, we thus have $g_i^{\varepsilon_2}(t) > 0$, which, recalling that

$$g_i^{\varepsilon_2}(t) \coloneqq \max_{j \neq i} \big[\mathrm{m}^-(i,j;t) + \varepsilon_2 \mathbb{1}\left(j \in \text{OPT}^{\varepsilon_1}(t)\right)\big],$$

implies

$$\exists j \neq i : \ \mathrm{m}^-(i,j;t) + \varepsilon_2 \mathbb{1}\left(j \in \text{OPT}^{\varepsilon_1}(t)\right) > 0,$$

thus, on the event $\mathcal{E}$ we have

$$\mathrm{m}(i,j) + \varepsilon_2 \mathbb{1}\left(j \in \text{OPT}^{\varepsilon_1}(t)\right) > 0.$$

Therefore, for such arm i, either

a) $\exists j \in [K]$ such that $\mathrm{m}(i,j) > 0$ that is $\mu_i \prec \mu_j$ or

b) $\exists j \in \text{OPT}^{\varepsilon_1}(t)$ such that $\mathrm{m}(i,j) + \varepsilon_2 > 0$ that is $\mu_i \prec \mu_j + \varepsilon_2 \mathbf{1}$.

Combining these, $\text{OPT}^{\varepsilon_1}(t) \subseteq \mathcal{S}^\star_{\varepsilon_1}$ and, for any $i \notin \text{OPT}^{\varepsilon_1}(t)$, either $i \notin \mathcal{S}^\star_{\varepsilon_1}$ (i is ε_1-sub-optimal) or there exists $j \in \text{OPT}^{\varepsilon_1}(t)$ such that $\mu_i \prec \mu_j + \varepsilon_2 \mathbf{1}$. Thus, $\text{OPT}^{\varepsilon_1}(t)$ is an $(\varepsilon_1, \varepsilon_2)$-cover of the Pareto set and APEis correct for $(\varepsilon_1, \varepsilon_2)$-PSI. □

3.2.2 Our instantiation

We propose to instantiate the algorithms with confidence intervals on the difference between pair of arms. For any pair $i, j \in [K]$, we define a function $\beta_{i,j}$ such that for any $c \in [d]$, $U^c_{i,j}(t) = \hat{\mu}^c_{t,i} - \hat{\mu}^c_{t,j} + \beta_{i,j}(t,\delta)$ and $L^c_{i,j}(t) = \hat{\mu}^c_{t,i} - \hat{\mu}^c_{t,j} - \beta_{i,j}(t,\delta)$. We use the following confidence bonus from Kaufmann & W.-M. Koolen 2021 for time-uniform concentration:

$$\beta_{i,j}(t,\delta) \coloneqq 2\sqrt{\left(C^g\left(\frac{\log\left(\frac{K_1}{\delta}\right)}{2}\right) + \sum_{a\in\{i,j\}} \log(4+\log(N_{t,a}))\right)\left(\sum_{a\in\{i,j\}} \frac{1}{N_{t,a}}\right)}, \quad (3.5)$$

where $K_1 \coloneqq K(K-1)d/2$ and $C^g \approx x + \log(x)$ is a calibration function. As a result, we have the simple expressions $\mathrm{M}^{\pm}(i,j;t) = \mathrm{M}(i,j;t) \pm \beta_{i,j}(t,\delta)$ and $\mathrm{m}^{\pm}(i,j;t) = \mathrm{m}(i,j;t) \pm \beta_{i,j}(t,\delta)$. As an example, we state in Algorithm 3.1 the pseudo-code of APE combined with the stopping rule suited for the k-relaxation of ε_1-PSI, which we refer to as the ε_1-APE-k algorithm.

Algorithm 3.1: APE: Adaptive Pareto Exploration

Require : (optional) ε_1-PSI relaxation parameter $\varepsilon_1 \geqslant 0$; (optional) parameter $\varepsilon_2 \geqslant 0$; (optional) k-relaxation parameter $k \in [K]$

Initialize: sample each arm once, set $t = K$, $N_{K,i} = 1$ for any $i \in [K]$

foreach $t = K+1, \ldots,$ **do**

  // Empirical non-dominated set at time t

  $S(t) \leftarrow \{\, i \in [K] : \forall j \in [K]\backslash\{i\},\ \mathrm{M}(i,j;t) > 0 \,\}$

  // Relaxed ε_1-Pareto candidates

  $\mathrm{OPT}^{\varepsilon_1}(t) \leftarrow \{\, i \in [K] : \forall j \in [K]\backslash\{i\},\ \mathrm{M}(i,j;t) - \beta_{i,j}(t,\delta) + \varepsilon_1 > 0 \,\}$

  if $\left|OPT^{\varepsilon_1}(t)\right| \geqslant k$ **then**

    break and output $\mathrm{OPT}^{\varepsilon_1}(t)$

  if $Z_1^{\varepsilon_1}(t) > 0 \ \wedge Z_2^{\varepsilon_1}(t) > 0$ **then**

    break and output

    $\mathcal{O}(t) = S(t) \bigcup \{i \in S(t)^{\mathsf{c}} : \nexists j \neq i : \mathrm{m}(i,j;t) - \beta_{i,j}(t,\delta) > 0\}$

  // Leader / Most promising arm outside $\mathrm{OPT}^{\varepsilon_1}(t)$

  $b_t \leftarrow \underset{i\in[K]\backslash \mathrm{OPT}^{\varepsilon_1}(t)}{\operatorname{argmax}} \ \min_{j\neq i}\left[\mathrm{M}(i,j;t) + \beta_{i,j}(t,\delta)\right]$

  // Challenger / Most critical opponent against b_t

  $c_t \leftarrow \underset{j\neq b_t}{\operatorname{argmin}}\left[\mathrm{M}(b_t,j;t) - \beta_{b_t,j}(t,\delta)\right]$

  // Sample the less explored of $\{b_t, c_t\}$

  Sample $A_t \leftarrow \underset{i\in\{b_t,c_t\}}{\operatorname{argmin}} N_{t,i}$; $t \leftarrow t+1$; update statistics

Other instantiations are possible, such as the ones based on individual confidence bounds of the form $U_{i,j}(t) = U_i(t) - L_j(t)$ where $[L_i(t), U_i(t)]$ is a confidence interval on μ_i (Kalyanakrishnan et al. 2012; Auer et al. 2016) or refined KL confidence sequences (Kaufmann & Kalyanakrishnan 2013).

Special case: Best Arm Identification. When $d = 1$, APEreduces to sampling at each round t, the least sampled among

$$b_t \coloneqq \operatorname*{argmax}_i \big[\min_{j\neq i} U_{i,j}(t)\big] \,, \quad \text{and} \quad c_t \coloneqq \operatorname*{argmin}_{j\neq b_t} L_{b_t,j}(t) \,. \tag{3.6}$$

To compare against LUCB (Kalyanakrishnan et al. 2012) and UGap (Gabillon et al. 2012) which use confidence intervals on single arms, we would have $\beta_{i,j}(t,\delta) \coloneqq \beta_i(t,\delta)+\beta_j(t,\delta)$, and $L_i(t) \coloneqq \hat{\mu}_{t,i} - \beta_i(t,\delta)$ and $U_i(t) \coloneqq \hat{\mu}_{t,i} + \beta_i(t,\delta)$: lower and upper confidence bounds on μ_i. Then (3.6) rewrites as

$$b_t \coloneqq \operatorname*{argmax}_i \big[U_i(t) - \max_{j\neq i} L_j(t)\big] \,, \quad \text{and} \quad c_t \coloneqq \operatorname*{argmax}_{j\neq b_t} U_j(t) \,.$$

This resembles the sampling rule of UGap, which defines

$$b_t^{\text{UGap}} \coloneqq \operatorname*{argmax}_i \big[L_i(t) - \max_{j\neq i} U_j(t)\big], \quad \text{and} \quad c_t^{\text{UGap}} \coloneqq \operatorname*{argmax}_{j\neq b_t} U_j(t),$$

and also pulls the least sampled so far. We further note that when $\varepsilon_1 = 0$, for any $i \in S(t)^{\mathsf{c}}, g_i(t) > h_i^0(t)$. Indeed, by definition,

$$h_i^0(t) = \min_{j\neq i}\big[\mathrm{M}(i,j;t) - \beta_{i,j}(t,\delta)\big] = \min_{j\neq i}\big[-\mathrm{m}(i,j;t) - \beta_{i,j}(t,\delta)\big]$$

and since $i \in S(t)^{\mathsf{c}}$, there exists $i^\star$ such that $\mathrm{m}(i,i^\star;t) > 0$ (*i.e.*, $\hat{\mu}_{t,i} \prec \hat{\mu}_{t,i^\star}$) and so

$$-\mathrm{m}(i,i^\star;t) - \beta_{i,i^\star}(t,\delta) < \mathrm{m}(i,i^\star;t) - \beta_{i,i^\star}(t,\delta).$$

Therefore,

$$\min_{j\neq i}\big[-\mathrm{m}(i,j;t) - \beta_{i,j}(t,\delta)\big] \coloneqq h_i^0(t) < \max_{j\neq i}\big[\mathrm{m}(i,j;t) - \beta_{i,j}(t,\delta)\big] \coloneqq g_i(t).$$

Thus for $\varepsilon_1 = 0$, $Z_2^0(t) = \min_{i\in S(t)^{\mathsf{c}}} g_i(t)$. In particular, when $d = 1, \varepsilon_1 = 0$, the stopping time τ_0 can be simplified to $\tau_0 = \inf\{t \geqslant K : Z_1^0(t) > 0\}$. Letting $\hat{a}_t$ denote the empirical best arm after t rounds, the stopping rule of APE (with the instantiation based on confidence intervals on pairs of arms) reduces to

$$\tau_0 = \inf\left\{ t \geqslant K : \forall i \neq \hat{a}_t, \frac{(\hat{\mu}_{\hat{a}_t,t} - \hat{\mu}_{t,i})^2}{2\left(\frac{1}{N_{t,\hat{a}_t}} + \frac{1}{N_{t,i}}\right)} \geqslant 2C^g\left(\frac{\log(K_1/\delta)}{2}\right) + 2\sum_{a\in\{\hat{a}_t,i\}} \log(4+\log(N_{t,a}))\right\}$$

which is very close to a Generalized Likelihood Ratio (GLR) stopping rule assuming Gaussian rewards with variance 1 (which is also known to be correct for subgaussian rewards) Garivier & Kaufmann 2016; Kaufmann & W.-M. Koolen 2021.

3.3 Main theoretical results

We state our main theorem on the sample complexity of our algorithms and give a sketch of its proof. First, let us introduce some quantities that are needed to state the theorem.

As presented in Chapter 1, the sample complexity of ε_1-Pareto set identification scales as a sum over the arms i of $1/(\Delta_i \vee \varepsilon_1)^2$ where Δ_i is called the sub-optimality gap of arm i and is defined as follows. For a sub-optimal arm $i \notin \mathcal{S}^\star(\mu)$,

$$\Delta_i \coloneqq \max_{j \in \mathcal{S}^\star(\mu)} \mathrm{m}(i,j),$$

which is the smallest quantity that should be added component-wise to μ_i to make i appear Pareto optimal *w.r.t.* $\{\mu_i : i \in [K]\}$. For a Pareto optimal arm $i \in \mathcal{S}^\star(\mu)$, the definition is more involved:

$$\Delta_i \coloneqq \begin{cases} \min_{j \in [K]\setminus\{i\}} \Delta_j & \text{if } \mathcal{S}^\star(\mu) \coloneqq \{i\} \\ \min\{\delta_i^+, \delta_i^-\} & \text{else,} \end{cases}$$

where

$$\delta_i^+ \coloneqq \min_{j \in \mathcal{S}^\star(\mu)\setminus\{i\}} \min\{\mathrm{M}(i,j), \mathrm{M}(j,i)\} \text{ and } \delta_i^- \coloneqq \min_{j \in [K]\setminus\mathcal{S}^\star(\mu)} \left[(\mathrm{M}(j,i))_+ + \Delta_j\right],$$

with $(x)_+ \coloneqq \max\{x, 0\},\ \forall x \in \mathbb{R}$. For $d = 1$, these gaps match the classical gaps in Best Arm Identification (Audibert & Bubeck 2010; Kaufmann, Cappé, et al. 2016).

We also introduce some additional notation required to express the contribution of the k-relaxation. Let $1 \leqslant k \leqslant K$. For any arm i, let $\omega_i = \min_{j \neq i} \mathrm{M}(i,j)$ and define

$$\omega^k \coloneqq \max_{i \in [K]}^{k} \omega_i, \qquad \mathcal{S}^{\star,k} \coloneqq \operatorname*{argmax}_{i \in [K]}^{1\ldots k} \omega_i,$$

with the k-th max and the first k argmax operators. Observe that $\omega^k > 0$ if and only if $\left|\mathcal{S}^\star(\mu)\right| \geqslant k$.

Theorem 3.3.1. *Let $\delta \in (0,1)$ be a risk parameter, $\varepsilon_1, \varepsilon_2 \geqslant 0$, $K \geqslant 2$, and $k \leqslant K$. Consider a bandit instance ν whose arm marginals are 1-subgaussian. Then, with probability at least $1-\delta$, the APE algorithm returns a correct recommendation for either the ε_1-PSI-k or the $(\varepsilon_1, \varepsilon_2)$-PSI task (with the corresponding stopping rule), and it stops after at most*

$$\sum_{a \in [K]} \frac{88}{\widetilde{\Delta}_a^2} \log\left(\frac{2K(K-1)d}{\delta} \log\left(\frac{12e}{\widetilde{\Delta}_a}\right)\right)$$

samples, where for each $a \in [K]$,

$$\widetilde{\Delta}_a \coloneqq \begin{cases} \max\{\Delta_a, \varepsilon_1, \omega^k\}, & \textit{for the } \varepsilon_1\textit{-PSI-}k \textit{ task,} \\ \max\{\Delta_a, \varepsilon_1\}, & \textit{for the } (\varepsilon_1, \varepsilon_2)\textit{-PSI task.} \end{cases}$$

Our theoretical results extend and refine existing bounds for approximate and relaxed Pareto set identification.

When $k = K$, the proposed algorithm provides a δ-correct solution for ε_1-PSI, which corresponds to the ε_1-PSI-K task. It improves upon Auer et al. 2016 by constant factors and replaces their $\log \Delta^{-2}$ dependence with a milder $\log \log \Delta^{-1}$ term, while nearly matching their lower bound (Theorem 1.3.1).

Our strongest result concerns the ε_1-PSI-k relaxation, which introduces a new notion of sub-optimality gap, quantifying how the relaxation parameter k reduces the effective sample complexity. In particular, for any arm $i \in \mathcal{S}^\star \setminus \mathcal{S}^{\star,k}$, we have $\max\{\Delta_i, \omega^k\} = \omega^k$, indicating that one should not pay more than the cost of identifying the k-th optimal arm. This observation parallels the findings of Roy Chaudhuri & Kalyanakrishnan 2019 for the "top-k among the best-m" problem. However, the authors have shown the relaxation only for the best m arms, while our result shows that even the sub-optimal arms should be sampled less.

For the general $(\varepsilon_1, \varepsilon_2)$-PSI formulation, our upper bound does not make the dependence on ε_2 explicit, yet nearly matching lower bounds suggest that this dependence may vanish in some cases. In particular, when $d = 1$, $\varepsilon_1 = 0$, and $\varepsilon_2 > 0$ with a unique optimal arm $a_\star$, $(0, \varepsilon_2)$-PSI reduces to exact Best Arm Identification. Indeed, in this case, any $\widehat{S} \subseteq \mathcal{S}^\star = \{a_\star\}$ trivially satisfies the covering condition. The corresponding lower bounds (Garivier & Kaufmann 2016; Kaufmann, Cappé, et al. 2016; Simchowitz et al. 2017) are thus independent of ε_2. This argument extends to the multi-objective setting when the Pareto set is a singleton, confirming that ε_2 has no impact on the intrinsic difficulty of such instances.

3.3.1 Sketch of proof of Theorem 3.3.1

Proof. Using Proposition 24 of Kaufmann & W.-M. Koolen 2021 it is simple to prove that the choice of $\beta_{i,j}$ in (3.5) yields $\mathbb{P}(\mathcal{E}) \geqslant 1 - \delta$ for the good event $\mathcal{E}$ defined in (3.1). Combining this result with Lemma 3.2.2 yields that APE is correct for each task with probability at least $1 - \delta$.

The idea of the remaining proof is to show that under the event $\mathcal{E}$, if APEhas not stopped at the end of round t, then the selected arm A_t has not been explored enough. The first lemma showing this is specific to the stopping rule $\tau^k_{\varepsilon_1}$ used for ε_1-PSI-k.

Lemma 3.3.2. *Let $\varepsilon_1 \geqslant 0$ and $k \in [K]$. If $\mathcal{E}_t$ holds and $t < \tau^k_{\varepsilon_1}$ then $\omega^k \leqslant 2\beta_{A_t,A_t}(t,\delta)$.*

The following two lemmas are more general as they apply to different stopping rules.

Lemma 3.3.3. *Let $\varepsilon_1 \geqslant 0$. Let $\tau = \tau^k_{\varepsilon_1}$ for some $k \in [K]$ or $\tau = \tau_{\varepsilon_1,\varepsilon_2}$ for some $\varepsilon_2 \geqslant 0$. If $\mathcal{E}_t$ holds and $t < \tau$ then $\Delta_{A_t} \leqslant 2\beta_{A_t,A_t}(t,\delta)$.*

Lemma 3.3.4. *Let $\varepsilon_1 \geqslant 0$ and τ be as in Lemma 3.3.3. If $\mathcal{E}_t$ holds and $t < \tau$ then $\varepsilon_1 \leqslant 2\beta_{A_t,A_t}(t,\delta)$.*

The proofs of these three lemmas are given at the end of the chapter. They heavily rely on the specific definition of b_t and c_t. In particular, to prove Lemma 3.3.3 and 3.3.4, we first establish that when $t < \tau$, any arm $j \in [K]$ satisfies $\mathrm{m}(b_t, j; t) \leqslant \beta_{b_t,j}(t, \delta)$.

The upper bound on the sample complexity follows, since using the lemmas above, we have that, if ε_1-APE-k has not stopped during round t, *i.e.*, $t < \tau^k_{\varepsilon_1}$ and the event $\mathcal{E}_t$ holds, then

- $\omega^k \leqslant 2\beta_{A_t,A_t}(t, \delta)$ (Lemma 3.3.2),
- $\Delta_{A_t} \leqslant 2\beta_{A_t,A_t}(t, \delta)$ (Lemma 3.3.3),
- $\varepsilon_1 \leqslant 2\beta_{A_t,A_t}(t, \delta)$ (Lemma 3.3.4)

hold simultaneously. Introducing $\widetilde{\Delta}_a := \max\{\omega^k, \varepsilon_1, \Delta_a\}$,

$$\begin{aligned}
\tau^k_{\varepsilon_1} \mathbb{1}(\mathcal{E}) - 1 &\leqslant \sum_{t=1}^{\infty} \mathbb{1}(\mathcal{E}) \mathbb{1}\left(\tau^k_{\varepsilon_1} > t\right), \\
&\leqslant \sum_{t=1}^{\infty} \mathbb{1}\big(\max\{\omega^k, \varepsilon_1, \Delta_{A_t}\} \leqslant 2\beta_{A_t,A_t}(t, \delta)\big) \\
&\leqslant \sum_{a=1}^{K} \sum_{t=1}^{\infty} \mathbb{1}\Big(\{A_t = a\} \wedge \{\widetilde{\Delta}_a \leqslant 2\beta_{a,a}(t, \delta)\}\Big) \\
&\leqslant \sum_{a=1}^{K} \inf\{n \geqslant 2 : \widetilde{\Delta}_a > 2\beta^n\},
\end{aligned}$$

where β^n is the expression of $\beta_{i,j}(t, \delta)$ when $N_{t,i} = N_{t,j} = n$, that is

$$\beta^n = 2\sqrt{\left(C^g\left(\frac{\log\left(\frac{K_1}{\delta}\right)}{2}\right) + 2\log(4 + \log(n))\right)\frac{2}{n}}.$$

Then, an inversion result (see Lemma 19 in Kone, Kaufmann, et al. 2023) yields

$$\inf\{s \geqslant 2 : 2\beta^s < \widetilde{\Delta}_a\} \leqslant \frac{88}{\widetilde{\Delta}_a^2} \log\left(\frac{2K(K-1)d}{\delta} \log\left(\frac{12e}{\widetilde{\Delta}_a}\right)\right).$$

Therefore,

$$\tau^k_{\varepsilon_1} \mathbb{1}(\mathcal{E}) \leqslant \sum_{a \in [K]} \frac{88}{\widetilde{\Delta}_a^2} \log\left(\frac{2K(K-1)d}{\delta} \log\left(\frac{12e}{\widetilde{\Delta}_a}\right)\right).$$

The same reasoning applies to $\tau_{\varepsilon_1,\varepsilon_2}$ as $\tau_{\varepsilon_1,\varepsilon_2} \leqslant \tau^K_{\varepsilon_1}$ holds with probability 1. □

LUCB-like instantiation. Theorem 3.3.1 bounds the sample complexity of APE on $\mathcal{E}$ but as, for many algorithms in pure-exploration, we do not control what happens on $\mathcal{E}^{\mathsf{c}}$. However, when APEis run with confidence bonuses similar to LUCB1 (Kalyanakrishnan et al. 2012), in the form

$$\beta_i(t,\delta) := \sqrt{\frac{2}{N_{t,i}}\log\left(\frac{5Kdt^4}{2\delta}\right)}, \quad \text{and} \quad \forall\,(i,j)\in[K], \quad \beta_{i,j}(t,\delta)=\beta_i(t,\delta)+\beta_j(t,\delta), \tag{3.7}$$

we are able to tightly upper bound the expected stopping time. The following result is proven in Kone, Kaufmann, et al. 2023 following the lines of the proof of Theorem 3.3.1.

Theorem 3.3.5 (Expected sample complexity of APE (Kone, Kaufmann, et al. 2023, Theorem 5)). *Let $\varepsilon_1,\varepsilon_2\geqslant 0$, $k\leqslant K$, and ν be a bandit instance with 1-subgaussian marginals. When APE is run with the parameters $(\beta_i)_i$ defined in (3.7), it is δ-correct for either ε_1-PSI-k or $(\varepsilon_1,\varepsilon_2)$-PSI. Moreover, its expected sample complexity satisfies*

$$\mathbb{E}_\nu[\tau]\leqslant \mathcal{O}\left(H(\nu)\,\log\frac{d\,H(\nu)}{\delta}\right),$$

where

$$H(\nu) := \sum_{a=1}^{K}\widetilde{\Delta}_a^{-2}, \qquad \text{with} \quad \widetilde{\Delta}_a := \begin{cases}\max\{\Delta_a,\varepsilon_1,\omega^k\}, & \text{for the } \varepsilon_1\text{-PSI-}k \text{ task},\\ \max\{\Delta_a,\varepsilon_1\}, & \text{for the } (\varepsilon_1,\varepsilon_2)\text{-PSI task}.\end{cases}$$

Kone, Kaufmann, et al. 2023 prove the following lower bound showing that in some scenarios, APE is optimal for ε_1-PSI-k, up to $d\log(K)$ and constant multiplicative terms. We note that for ε_1-PSI a lower bound featuring the gaps Δ_a and ε_1 was already derived by Auer et al. 2016.

Theorem 3.3.6 (Theorem 2 of Kone, Kaufmann, et al. 2023). *There exists a bandit instance ν with $|\mathcal{S}^\star|=p\geqslant 3$ such that for $k\in\{p,p-1,p-2\}$, any δ-correct algorithm for 0-PSI-k verifies*

$$\mathbb{E}_\nu[\tau_\delta]\geqslant \frac{1}{d}\log\frac{1}{\delta}\sum_{a=1}^{K}\frac{1}{(\Delta_a^k)^2},$$

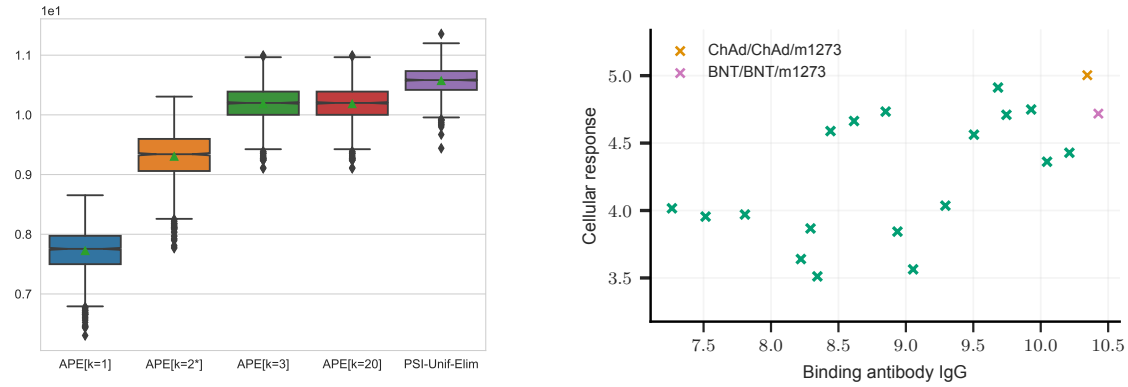
where $\Delta_a^k := \Delta_a+\omega^k$ and τ_δ is the stopping time of the algorithm.

3.4 Numerical study and discussion

Experimental setup. We compare *Adaptive Pareto Exploration* (APE) to the algorithm of Auer et al. 2016 (denoted PSI-Unif-Elim). For a fair comparison, both use our pairwise confidence bonuses $\beta_{i,j}(t,\delta)$.[1] Since anytime confidence bounds are known to be conservative, we set $K_1=1$ in (3.5) instead of its theoretical union-bound value. Unless stated otherwise, $\delta=0.1$. Across all runs, the empirical error is well below δ.

[1] Auer et al. 2016 already suggested this heuristic; we instantiate it for both methods.

Real-world-inspired evaluation (COV-BOOST). We reuse the COV-BOOST instance introduced in Chapter 2 and fully documented in Appendix 7. Briefly, the bandit has $K = 20$ arms (vaccine strategies) and $d = 3$ immunogenicity indicators derived from the processed trial data; we simulate a multivariate Gaussian model using the means and variances reported there. All preprocessing details, including the log-normal assumption and pooled variance estimates, are deferred to Appendix 7. We set $\varepsilon_1 = 0$ and compare PSI-Unif-Elim to APE-k for several values of k.



(a) Log empirical sample complexity (all arms). **(b)** Two immunogenicity indicators (illustration).

Figure 3.1: COV-BOOST: APE-k vs. PSI-Unif-Elim at $\delta = 0.1$ (2000 runs).

Since $|\mathcal{S}^\star| = 2$, APE (with $k \geqslant 3$) outperforms PSI-Unif-Elim on exact PSI. The k-relaxation substantially reduces sample complexity: APE-2 stops as soon as two optimal arms are certified, whereas for $k = 3$ remaining arms must be ruled out as suboptimal, explaining the clear gap between $k = 2$ and $k = 3$.

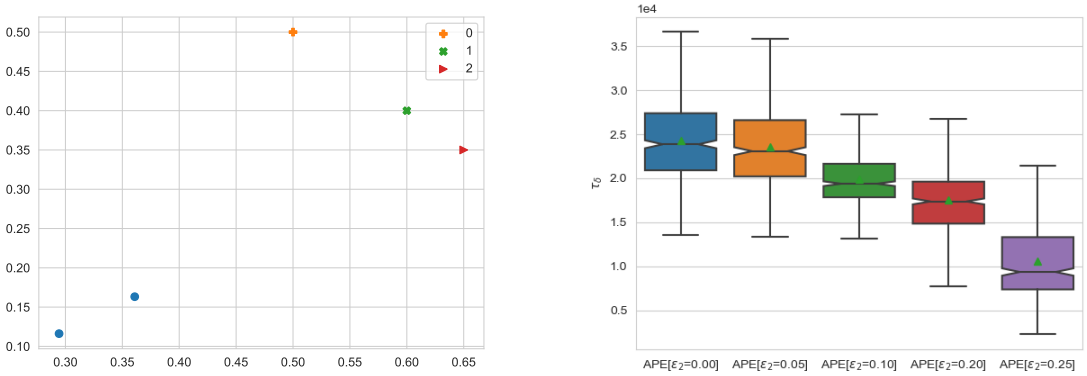
Random Bernoulli instances. We generate 2000 multivariate Bernoulli bandits with $K = 5$ and $d \in \{2, 4, 8\}$ (independent marginals), set $\delta = 0.1$ and $\varepsilon_1 = 0.005$ (for runtime), and average the sample complexity over runs.

Table 3.2: Average sample complexity over 2000 random instances ($K = 5$). Average $|\mathcal{S}^\star|$ is $(2.30,\ 4.06,\ 4.93)$ for $d = 2, 4, 8$.

	ε_1-APE-1	ε_1-APE-2	ε_1-APE-3	ε_1-APE-4	ε_1-APE-5	ε_1-PSI-Unif-Elim
$d = 2$	811	39,530	109,020	145,777	150,844	190,625
$d = 4$	214	6,410	19,908	68,061	124,001	157,584
$d = 8$	119	204	405	1,448	20,336	27,270

APE (with $k = K$) uses 20–25% fewer samples than PSI-Unif-Elim, with larger gains as d increases–and consequently $|\mathcal{S}^\star|$ as well. The k-relaxation drastically lowers budgets relative to exact PSI.

APE on $(\varepsilon_1, \varepsilon_2)$-PSI. We now examine the empirical behavior of $(\varepsilon_1, \varepsilon_2)$-APE for identifying an $(\varepsilon_1, \varepsilon_2)$-cover. We set $\varepsilon_1 = 0$, $\delta = 0.01$, and vary $\varepsilon_2 \in \{0, 0.05, 0.1, 0.2, 0.25\}$. Results are averaged over 2000 independent runs with different random seeds on the same instance (described in Figure 3.2a). The environment is a multivariate Bernoulli bandit with $K = 5$ and $d = 2$ (independent marginals). In this toy example, $(\varepsilon_1, \varepsilon_2)$-PSI is particularly meaningful since several arms are nearly equivalent: three Pareto-optimal vectors are manually chosen, while the remaining two are drawn uniformly at random.



(a) Instance.

(b) Average sample complexity.

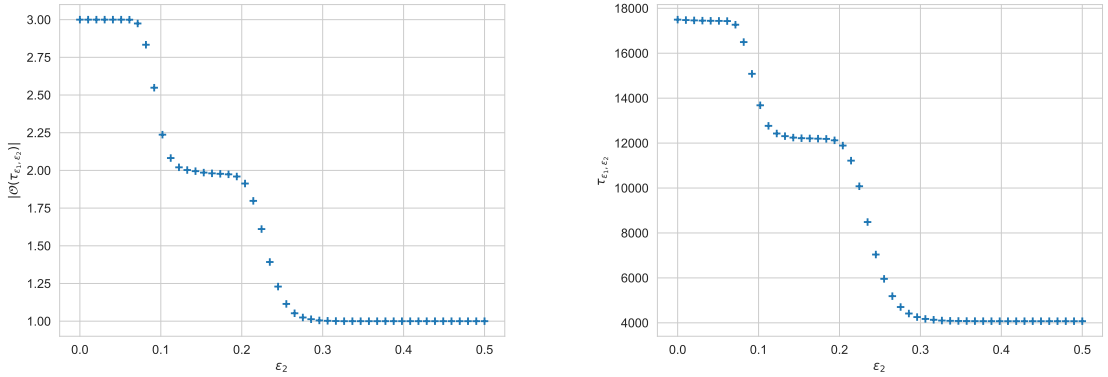
Figure 3.2: Toy instance with $\mathcal{S}^\star = \{0, 1, 2\}$. The difference on the x- and y-axes is 0.1 between arms 0 and 1, and 0.05 between arms 1 and 2. Right: average sample complexity over 2000 runs.

As shown in Figure 3.2b, the sample complexity decreases as ε_2 increases. This trend is confirmed in Figure 3.3, which reports both the average sample complexity and the mean size of the returned cover for 50 evenly spaced values of $\varepsilon_2 \in [0, 0.5]$. Each major drop in Figure 3.3b corresponds to a value of ε_2 at which APE removes an (otherwise optimal) arm from the cover to save samples (see Figure 3.3a). Thus, a decrease in sample complexity roughly coincides with a reduction in the number of arms included in the final recommendation.

For large values of ε_2 (*e.g.*, > 0.3), the sample complexity plateaus (Figure 3.3b). This behavior stems from the fact that, even when ε_2 is large, the algorithm must still certify at least one truly optimal arm (as reflected by the single-element covers in Figure 3.3a). Empirically, this limiting complexity is close to that of 0-APE-1 on the same instance (≈ 4073 samples).

Finally, Figure 3.4 shows the frequency with which each arm appears in the recommended set for three representative values of ε_2. For $\varepsilon_2 = 0.15$, arm 0 is always recommended, while the others appear in about half of the runs. For $\varepsilon_2 = 0.4$, the algorithm almost always returns only arm 0—the one with the largest ω_i term, *i.e.*, the easiest to identify as optimal.

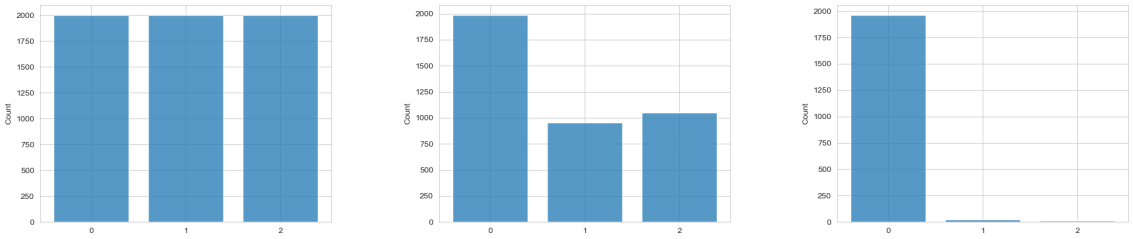
Sampling allocations of APE vs. PSI-Unif-Elim. We now examine how the sampling allocations of APE and PSI-Unif-Elim differ on specific instances. In some configurations,



(a) Average $|\mathcal{O}(\tau_{\varepsilon_1,\varepsilon_2})|$ vs. ε_2.

(b) Sample complexity vs. ε_2.

Figure 3.3: Average size of the returned cover (left) and corresponding sample complexity (right), both averaged over 2000 runs. The empirical error probability was negligible in all cases.



(a) $\varepsilon_2 = 0.05$.

(b) $\varepsilon_2 = 0.15$.

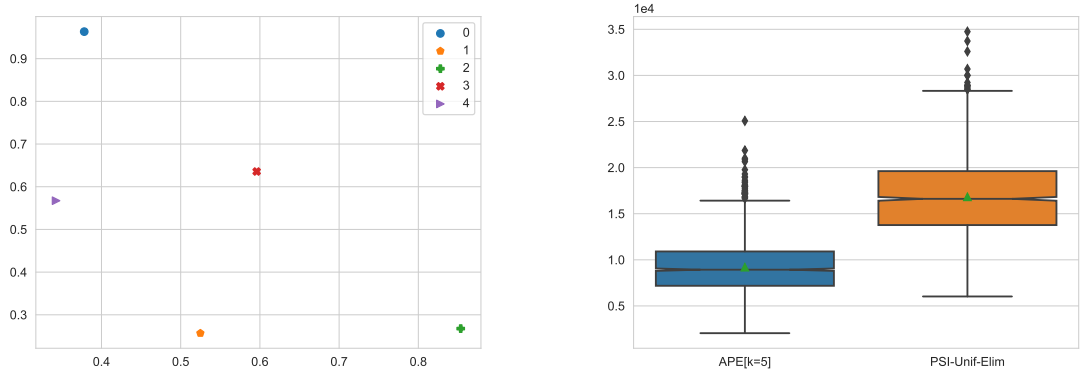
(c) $\varepsilon_2 = 0.4$.

Figure 3.4: Frequency of each arm in the recommended set across 2000 runs for three representative values of ε_2.

APE requires up to three times fewer samples than PSI-Unif-Elim. This improvement stems from PSI-Unif-Elim's conservative strategy: it continues sampling arms already identified as (nearly) optimal until these arms are proven not to dominate any remaining active ones. For instance, in Figure 3.5a, arm 2 is easily recognized as optimal, yet it slightly dominates the suboptimal arm 1. Consequently, PSI-Unif-Elim keeps sampling arm 2 until arm 1 is eventually removed from the active set—likely when it is shown to be dominated by arm 3. Our adaptive sampling rule avoids this inefficiency by dynamically reallocating samples to the most uncertain comparisons instead of repeatedly pulling confirmed optimal arms.

As shown in Figure 3.5b, APE achieves roughly half the sample complexity of PSI-Unif-Elim on this instance. Table 3.3 details the average number of pulls taken by PSI-Unif-Elim divided by that of 0-APE-K for each arm. The primary source of inefficiency is evident: arm 2 is sampled nearly six times more often by PSI-Unif-Elim, highlighting its tendency to oversample dominant arms that are already known to be optimal.

Discussion. Across real-world-inspired and synthetic settings, APE consistently improves over PSI-Unif-Elim on exact PSI and delivers large additional savings under the k- and



(a) Instance.

(b) Average sample complexity.

Figure 3.5: Illustrative instance with $\mathcal{S}^\star = \{0, 2, 3\}$. Right: average sample complexity across repeated trials. APE avoids oversampling already-certified optimal arms.

Table 3.3: Average number of pulls under PSI-Unif-Elim divided by that under 0-APE-K for each arm. The largest gap arises from arm 2, which PSI-Unif-Elim oversamples to maintain dominance guarantees.

Arm	0	1	2	3	4
Average pull ratio (PSI-Unif-Elim /APE)	3.08	1.36	5.68	1.28	1.40

$(\varepsilon_1, \varepsilon_2)$-relaxations. On average (Table 3.2), APE achieves about a 20% reduction in sample complexity for exact PSI, with substantially larger gains on specific instances where multiple Pareto-optimal arms dominate a suboptimal arm. As emphasized in Chapter 1 and Chapter 2, these relaxations align with practical decision constraints (*e.g.*, limited advancement slots or near-duplicate treatments), making APE both statistically and operationally attractive.

3.5 Additional proofs

In this section, we gather results used in the correctness and sample-complexity analyses of APE, and provide their proofs. We recall the definition of the events

$$\mathcal{E}_t = \bigcap_{i=1}^{K} \bigcap_{j \neq i} \bigcap_{c=1}^{d} \left\{ L_{i,j}^c(t) \leqslant \mu_i^c - \mu_j^c \leqslant U_{i,j}^c(t) \right\} \quad \text{and} \quad \mathcal{E} = \bigcap_{t=1}^{\infty} \mathcal{E}_t \ .$$

3.5.1 Probability of the good event

Proof of Lemma 3.2.1. This result follows from the definition of $\mathcal{E}_t$. Since

$$\mathcal{E}_t \ := \bigcap_{i=1}^{K} \bigcap_{j \neq i} \bigcap_{c=1}^{d} \left\{ L_{i,j}^c(t) \leqslant \mu_i^c - \mu_j^c \leqslant U_{i,j}^c(t) \right\} ,$$

if $\mathcal{E}_t$ holds, then for any i, j

$$\mathrm{M}^-(i,j;t) \coloneqq \max_c L^c_{i,j}(t) \leqslant \mathrm{M}(i,j) \coloneqq \max_c \big[\mu^c_i - \mu^c_j\big] \leqslant \max_c U^c_{i,j}(t) \coloneqq \mathrm{M}^+(i,j;t),$$

and the second point follows by noting that $\mathrm{m}(i,j) = -\mathrm{M}(i,j)$ and

$$\mathrm{m}^+(i,j;t) \coloneqq -\mathrm{M}^-(i,j;t); \mathrm{m}^-(i,j;t) \coloneqq -\mathrm{M}^+(i,j;t)$$

for any pair of arms. □

We note that when the algorithm uses a confidence bonus of the form $\big(\hat{\mu}^c_{t,i} - \hat{\mu}^c_{t,j} \pm \beta_{i,j}(t,\delta)\big)$,

$$\begin{aligned}
\mathrm{M}^+(i,j;t) &\coloneqq \max_c U^c_{i,j}(t) = \max_c \big[\hat{\mu}^c_{t,i} - \hat{\mu}^c_{t,j}\big] + \beta_{i,j}(t,\delta) = \mathrm{M}(i,j;t) + \beta_{i,j}(t,\delta),\\
\mathrm{M}^-(i,j;t) &\coloneqq \max_c L^c_{i,j}(t) = \max_c \big[\hat{\mu}^c_{t,i} - \hat{\mu}^c_{t,j}\big] - \beta_{i,j}(t,\delta) = \mathrm{M}(i,j;t) - \beta_{i,j}(t,\delta),
\end{aligned}$$

and the previous lemma implies that on $\mathcal{E}_t$,

$$\big|\,\mathrm{M}(i,j) - \mathrm{M}(i,j,t)\big| \;\leqslant\; \beta_{i,j}(t,\delta) \quad \text{and} \quad \big|\,\mathrm{m}(i,j) - \mathrm{m}(i,j;t)\big| \;\leqslant\; \beta_{i,j}(t,\delta),$$

which is extensively used in our sample-complexity analysis.

3.5.2 Sample complexity

Lemma 3.3.2. *Let $\varepsilon_1 \geqslant 0$ and $k \in [K]$. If $\mathcal{E}_t$ holds and $t < \tau^k_{\varepsilon_1}$ then $\omega^k \leqslant 2\beta_{A_t,A_t}(t,\delta)$.*

Proof. First, note that if $k > |\mathcal{S}^\star|$, then the lemma holds trivially since $\omega^k < 0$. In the sequel, we assume $\mathcal{E}_t$ holds and $k \leqslant |\mathcal{S}^\star|$. If $t < \tau^k_{\varepsilon_1}$ then it holds that $|\,\mathrm{OPT}^{\varepsilon_1}(t)| < k$. So $\mathcal{S}^{\star,k} \cap \mathrm{OPT}^{\varepsilon_1}(t)^{\mathsf{c}} \neq \emptyset$. Let $i \in \mathcal{S}^{\star,k} \cap \mathrm{OPT}^{\varepsilon_1}(t)^{\mathsf{c}}$, we have

$$\begin{aligned}
\omega^k &\leqslant \omega_i = \min_{j \in [K]\setminus\{i\}} \mathrm{M}(i,j),\\
&\overset{(a)}{\leqslant} \min_{j \in [K]\setminus\{i\}} \mathrm{M}(i,j;t) + \beta_{i,j}(t,\delta),\\
&\overset{(b)}{\leqslant} \min_{j \in [K]\setminus\{b_t\}} \mathrm{M}(b_t,j;t) + \beta_{b_t,j}(t,\delta),\\
&\leqslant \mathrm{M}(b_t,c_t;t) + \beta_{b_t,c_t}(t,\delta),\\
&\overset{(c)}{\leqslant} 2\beta_{b_t,c_t}(t,\delta),\\
&\leqslant 2\beta_{A_t,A_t}(t,\delta),
\end{aligned}$$

where (a) uses that $\mathcal{E}_t$ holds and Lemma 3.2.1, (b) uses the definition of b_t and (c) follows from the definition of c_t and the fact that $b_t \notin \mathrm{OPT}^{\varepsilon_1}(t)$, which yields $\mathrm{M}(b_t,c_t;t) \leqslant \beta_{b_t,c_t}(t,\delta)$. The last inequality follows since A_t is the least sampled among b_t, c_t, and β is decreasing. □

Lemma 3.3.3. *Let $\varepsilon_1 \geqslant 0$. Let $\tau = \tau^k_{\varepsilon_1}$ for some $k \in [K]$ or $\tau = \tau_{\varepsilon_1,\varepsilon_2}$ for some $\varepsilon_2 \geqslant 0$. If $\mathcal{E}_t$ holds and $t < \tau$ then $\Delta_{A_t} \leqslant 2\beta_{A_t,A_t}(t,\delta)$.*

Before proving Lemma 3.3.3, we recall the following lemma, proven in Chapter 2, which is used to derive an upper bound on the gap of an optimal arm.

Lemma 2.5.8. *For any Pareto optimal arm i, $\Delta_i \leqslant \min_{j\neq i} \mathrm{M}(i,j)$.*

Proof of Lemma 3.3.3. Assume that $\mathcal{E}_t$ holds. We consider four different cases depending on whether b_t and c_t are optimal or sub-optimal.

Case 1.1: b_t is a Pareto optimal arm. From the definition of the gap of an optimal arm and using Lemma 2.5.8 it follows $\Delta_{b_t} \leqslant \mathrm{M}(b_t,c_t)$ which on $\mathcal{E}_t$ and using Lemma 3.2.1 yields

$$\Delta_{b_t} + \varepsilon_1 \leqslant \mathrm{M}(b_t,c_t;t) + \beta_{b_t,c_t}(t,\delta) + \varepsilon_1 \tag{3.8}$$

then, noting that there exists $j \in [K]\setminus\{b_t\}$ such that $\mathrm{M}(b_t,j,t) + \varepsilon_1 \leqslant \beta_{b_t,j}(t,\delta)$, by definition of c_t, we have

$$\mathrm{M}(b_t,c_t;t) + \varepsilon_1 \leqslant \beta_{b_t,c_t}(t,\delta), \tag{3.9}$$

therefore,

$$\Delta_{b_t} + \varepsilon_1 \;\leqslant\; 2\beta_{b_t,c_t}(t,\delta).$$

Case 1.2 b_t is a sub-optimal arm. By definition of c_t and using $\mathrm{M} = -\mathrm{m}$, we have

$$c_t \in \operatorname*{argmax}_{j\in[K]\setminus\{b_t\}} \mathrm{m}(b_t,j;t) + \beta_{b_t,j}(t,\delta), \tag{3.10}$$

then, from the definition of the gap of a sub-optimal arm and since $\mathcal{E}_t$ holds, we know that there exists an arm $b^\star_t$ such that

$$\begin{aligned}
\Delta_{b_t} = \mathrm{m}(b_t,b^\star_t) &\leqslant \mathrm{m}(b_t,b^\star_t,t) + \beta_{b_t,b^\star_t}(t,\delta),\\
&\overset{(a)}{\leqslant} \mathrm{m}(b_t,c_t;t) + \beta_{b_t,c_t}(t,\delta),\\
&\overset{(b)}{\leqslant} 2\beta_{b_t,c_t}(t,\delta).
\end{aligned}$$

where (a) uses the definition of c_t and (b) uses Lemma 3.5.1.

Case 2.1 c_t is a Pareto optimal arm. If b_t is also an optimal arm, it follows that $\Delta_{c_t} \leqslant \mathrm{M}(b_t,c_t)$ which on $\mathcal{E}_t$ yields $\Delta_{c_t} \leqslant \mathrm{M}(b_t,c_t;t) + \beta_{b_t,c_t}(t,\delta)$, then, similarly to case 1.1, we have $\mathrm{M}(b_t,j;t) + \varepsilon_1 \leqslant \beta_{b_t,j}(t,\delta)$ so

$$\Delta_{c_t} + \varepsilon_1 \;\leqslant\; 2\beta_{b_t,c_t}(t,\delta)\,.$$

Now, assume b_t is a sub-optimal arm. Then, by definition, $\Delta_{c_t} \leqslant \mathrm{M}(b_t, c_t)_+ + \Delta_{b_t}$. Using a similar reasoning to case 1.2, it holds that $\Delta_{b_t} \leqslant \mathrm{m}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta)$, so

$$\begin{aligned}
\Delta_{c_t} &\leqslant \mathrm{M}(b_t, c_t)_+ + \Delta_{b_t}, \\
&\leqslant (\mathrm{M}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta))_+ + \mathrm{m}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta), \\
&= (-\mathrm{m}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta))_+ + \mathrm{m}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta), \\
&\overset{(a)}{\leqslant} \max(2\beta_{b_t,c_t}(t,\delta), \mathrm{m}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta)) \\
&\overset{(b)}{\leqslant} 2\beta_{b_t,c_t}(t,\delta).
\end{aligned}$$

where (a) follows from $(x-y)_+ + (x+y) \leqslant \max(x+y, 2x)$ and (b) follows from $\mathrm{m}(b_t, c_t; t) \leqslant \beta_{b_t,c_t}(t,\delta)$ (Lemma 3.5.1).

Case 2.2 c_t is a sub-optimal arm. We know that there exists an arm $c_t^\star$ such that $\Delta_{c_t} = \mathrm{m}(c_t, c_t^\star)$. If $c_t^\star = b_t$ then, since $\mathrm{m}(j,i) \leqslant \mathrm{M}(i,j)$ (follows from the definition), we have

$$\begin{aligned}
\Delta_{c_t} = \mathrm{m}(c_t, c_t^\star) &= \mathrm{m}(c_t, b_t), \\
&\leqslant \mathrm{M}(b_t, c_t), \\
&\overset{(a)}{\leqslant} \mathrm{M}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta), \\
&\overset{(b)}{\leqslant} 2\beta_{b_t,c_t}(t,\delta),
\end{aligned}$$

where (a) follows from $\mathcal{E}_t$ and (b) has been already justified in the case 1.1. If $b_t \neq c_t^\star$, then by definition of c_t, we have

$$\mathrm{m}(b_t, c_t; t) + \beta_{b_t,c_t}(t,\delta) \geqslant \mathrm{m}(b_t, c_t^\star; t) + \beta_{b_t,c_t^\star}(t,\delta),$$

which implies that there exists $c \in [d]$ such that

$$\hat{\mu}^c_{t,c_t} - \hat{\mu}^c_{t,b_t} + \beta_{b_t,c_t}(t,\delta) \geqslant \hat{\mu}^c_{t,c_t^\star} - \hat{\mu}^c_{t,b_t} + \beta_{b_t,c_t^\star}(t,\delta) \overset{\mathcal{E}_t}{\geqslant} \mu^c_{c_t^\star} - \mu^c_{b_t},$$

then recalling that $\beta_{i,j} = \beta_{j,i}$,

$$\mu^c_{c_t} - \mu^c_{b_t} + 2\beta_{b_t,c_t}(t,\delta) \overset{\mathcal{E}_t}{\geqslant} (\hat{\mu}^c_{t,c_t} - \hat{\mu}^c_{t,b_t} - \beta_{b_t,c_t}(t,\delta)) + 2\beta_{b_t,c_t}(t,\delta) \geqslant \mu^c_{c_t^\star} - \mu^c_{b_t}.$$

Together, these imply that there exists $c \in [d]$ such that

$$\mu^c_{c_t^\star} - \mu^c_{c_t} \leqslant 2\beta_{b_t,c_t}(t,\delta),$$

so

$$\Delta_{c_t} = \min_c \left[\mu^c_{c_t^\star} - \mu^c_{c_t}\right] \leqslant 2\beta_{b_t,c_t}(t,\delta).$$

Combining all cases, we have proved that if $t < \min(\tau^k_{\varepsilon_1}, \tau_{\varepsilon_1,\varepsilon_2})$ then both

$$\Delta_{b_t} \leqslant 2\beta_{b_t,c_t}(t,\delta) \quad \text{and} \quad \Delta_{c_t} \leqslant 2\beta_{b_t,c_t}(t,\delta) \tag{3.11}$$

holds. Further noting that A_t is the least sampled among b_t, c_t and β is non-increasing, $\beta_{b_t,c_t}(t,\delta) \leqslant \beta_{A_t,A_t}(t,\delta)$, (3.11) yields

$$\Delta_{A_t} \leqslant 2\beta_{A_t,A_t}(t,\delta),$$

which completes the proof of Lemma 3.3.3. □

The following lemma holds for each of the stopping times $\tau_{\varepsilon_1}, \tau_{\varepsilon_1,\varepsilon_2}$ and $\tau^k_{\varepsilon_1}$.

Lemma 3.3.4. *Let $\varepsilon_1 \geqslant 0$ and τ be as in Lemma 3.3.3. If $\mathcal{E}_t$ holds and $t < \tau$ then $\varepsilon_1 \leqslant 2\beta_{A_t,A_t}(t,\delta)$.*

Proof. By Lemma 3.5.1, we have $\mathrm{m}(b_t, c_t; t) \leqslant \beta_{b_t,c_t}(t,\delta)$ or equivalently

$$\mathrm{M}(b_t, c_t; t) \geqslant -\beta_{b_t,c_t}(t,\delta)\ . \tag{3.12}$$

Then, knowing that $b_t \notin \mathrm{OPT}^{\varepsilon_1}(t)$, there exists an arm j such that $\varepsilon_1 + \mathrm{M}(b_t, j; t) \leqslant \beta_{b_t,j}(t,\delta)$. Using further the definition of c_t, it follows that $\varepsilon_1 + \mathrm{M}(b_t, c_t; t) \leqslant \beta_{b_t,c_t}(t,\delta)$. Combining this with inequality (3.12) and noting that A_t is the least sampled among b_t, c_t yields

$$\beta_{A_t,A_t}(t,\delta) \geqslant \beta_{b_t,c_t}(t,\delta) \geqslant \varepsilon_1/2.$$

□

Lemma 3.5.1. *Let $\varepsilon_1 \geqslant 0, \varepsilon_2 \geqslant 0$ and $k \in [K]$. If $\tau = \tau^k_{\varepsilon_1}$ or $\tau = \tau_{\varepsilon_1,\varepsilon_2}$, the following holds. If $t < \tau$ then for any $j \in [K]$, $\mathrm{m}(b_t, j; t) \leqslant \beta_{b_t,j}(t,\delta)$.*

Proof. The proof is split into two steps.

Step 1: If $t < \tau^k_{\varepsilon_1}$ then for any $j \in [K]$, $\mathrm{m}(b_t, j; t) \leqslant \beta_{b_t,j}(t,\delta)$.

First, note that $t < \min(\tau^k_{\varepsilon_1}, \tau_{\varepsilon_1})$ implies that $Z^{\varepsilon_1}_1(t) \leqslant 0$ or $Z^{\varepsilon_1}_2(t) \leqslant 0$. By definition of b_t and noting that $\mathrm{M}(i,j;t) = -\mathrm{m}(i,j;t)$, we have

$$b_t \in \operatorname*{argmin}_{i \in \mathrm{OPT}^{\varepsilon_1}(t)^{\mathsf{c}}} \max_{j \neq i} \mathrm{m}(i,j;t) - \beta_{i,j}(t,\delta). \tag{3.13}$$

so that if there exists j such that $\mathrm{m}(b_t, j; t) > \beta_{b_t,j}(t,\delta)$, then

$$\max_{j \neq b_t} \mathrm{m}(b_t, j; t) - \beta_{b_t,j}(t,\delta) > 0,$$

therefore,

$$\forall i \in \mathrm{OPT}^{\varepsilon_1}(t)^{\mathsf{c}}\ , \max_{j \neq i} \mathrm{m}(i,j;t) - \beta_{i,j}(t,\delta) > 0 \quad \textit{i.e.}, \quad g_i(t) > 0. \tag{3.14}$$

Furthermore, for any $i \in \mathrm{OPT}^{\varepsilon_1}(t)$, $h^{\varepsilon_1}_i(t) > 0$. Combining these, if there exists j such that $\mathrm{m}(b_t, j; t) > \beta_{b_t,j}(t,\delta)$ then, $Z^{\varepsilon_1}_1(t) > 0$ and $Z^{\varepsilon_1}_2(t) > 0$.

Step 2: If $t < \tau_{\varepsilon_1,\varepsilon_2}$ then for any $j \in [K]$, $\mathrm{m}(b_t, j; t) \leqslant \beta_{b_t,j}(t, \delta)$.

Recall that by definition $t < \tau_{\varepsilon_1,\varepsilon_2}$ implies that $Z_1^{\varepsilon_1,\varepsilon_2}(t) \leqslant 0$ or $Z_2^{\varepsilon_1,\varepsilon_2}(t) \leqslant 0$. Using (3.13), if there exists j such that $\mathrm{m}(b_t, j; t) > \beta_{b_t,j}(t, \delta)$, then

$$\max_{j \neq b_t} \mathrm{m}(b_t, j; t) - \beta_{b_t,j}(t, \delta) > 0.$$

Combining this with

$$g_i^{\varepsilon_2}(t) \coloneqq \max_{j \in [K]\setminus\{i\}} \left[\mathrm{m}(i, j; t) - \beta_{i,j}(t, \delta) + \varepsilon_2 \mathbb{1}\left(j \in \mathrm{OPT}^{\varepsilon_1}(t)\right)\right],$$

yields

$$\forall i \in \mathrm{OPT}^{\varepsilon_1}(t)^{\mathsf{c}}, 0 < \max_{j \neq i} \left[\mathrm{m}(i, j; t) - \beta_{i,j}(t, \delta)\right] \leqslant g_i^{\varepsilon_2}(t). \tag{3.15}$$

Furthermore, since we have

$$\forall i \in \mathrm{OPT}^{\varepsilon_1}(t), h_i^{\varepsilon_1}(t) > 0, \tag{3.16}$$

the initial assumption would yield that for any arm i, $\max\{h_i^{\varepsilon_1}(t), g_i^{\varepsilon_2}(t)\} > 0$, so $Z_1^{\varepsilon_1,\varepsilon_2}(t) > 0$ and $Z_2^{\varepsilon_1,\varepsilon_2}(t) > 0$. We conclude that if $\tau = \tau_{\varepsilon_1}^k$ or $\tau_{\varepsilon_1,\varepsilon_2}$, $t < \tau$ implies that for any $j \in [K], \mathrm{m}(b_t, j; t) \leqslant \beta_{b_t,j}(t, \delta)$. □

Chapter 4

Linear Model for Pareto Set Identification

This chapter extends the study of *Pareto Set Identification* (PSI) to a structured *multi-output linear bandit* model, where each arm is represented by a feature vector in $\mathbb{R}^h$ and its mean outcome in $\mathbb{R}^d$ depends linearly on this feature via a shared, unknown parameter matrix $\theta \in \mathbb{R}^{h\times d}$.

Building upon the foundations established in Chapters 2 and 3, we introduce, to the best of our knowledge, the first *optimal-design-based algorithms* for structured PSI and provide a unified analysis in both the fixed-budget and fixed-confidence settings. Our results show that the statistical complexity of linear PSI depends only on the h arms with the smallest suboptimality gaps, leading to substantial reductions in sample complexity when the number of arms is large.

Thi chapter is based on joint work with Émilie Kaufmann and Laura Richert, published in the proceedings of *AISTATS 2025*.

4.1 Introduction

In previous chapters, we studied *Pareto Set Identification* (PSI) in the unstructured setting (fixed-budget and fixed-confidence), where each arm is a distribution over $\mathbb{R}^d$ and no structure is assumed between the arms. In this chapter, we incorporate additional structure through a *multi-output linear model*, assuming that each arm k is associated with a known feature vector $x_k \in \mathbb{R}^h$ and that its mean reward vector satisfies

$$\mu_k = \theta^\top x_k, \qquad \theta \in \mathbb{R}^{h \times d}.$$

This setting, which we refer to as *(multi-output) linear PSI*, generalizes both classical linear bandits ($d = 1$) and multi-objective bandits (with one-hot features, $h = K$). The goal remains to identify the set of arms whose mean vectors are not uniformly dominated by any other, while minimizing the number of samples required to achieve a given confidence level. We bridge the settings of the two previous chapters by extending gap-based PSI algorithms to the linear, multi-objective case.

In many applications, arms exhibit observable structure: for instance, candidate vaccines may be characterized by design features such as antigen dose, adjuvant type, or administration schedule, and their immunogenicity responses can be measured along several endpoints (antibody titers, neutralization, T-cell activity, Munro et al. 2021). Such features often induce correlations across arms that can be exploited to accelerate learning. Incorporating this structure allows the learner to infer the expected responses of untested candidates and substantially reduces the sample complexity required for identifying promising treatments. The same principle applies to engineering problems–such as algorithm configuration or hardware design–where arms correspond to structured configurations with measurable descriptors (Almer et al. 2011a).

Related work. The multi-output linear model was first explored by S. Lu et al. 2019 in the context of *Pareto regret minimization*. It can also be seen as a special case of the multi-output kernel regression model of Zuluaga, Krause, et al. 2016 under a linear kernel, for which the authors proposed the ε-PAL algorithm. Parameterized correctly for linear fixed-confidence PSI, ε-PAL has a sample complexity of order $(h^2/\Delta_{\min}^2)\log(1/\delta)$.

Kim et al. 2023 extended the algorithm of Auer et al. 2016 using a robust estimator to jointly minimize Pareto regret and identify the Pareto set, achieving a bound of order $(h/\Delta_{\min}^2)\log(1/\delta)$.

When $d = 1$, linear PSI reduces to the classical *linear bandit* pure exploration problem, whose sample complexity and optimal design connections have been extensively studied (Soare et al. 2014; Tao et al. 2018; Degenne, Ménard, et al. 2020). Soare et al. 2014 first explored the link with optimal experimental design (Pukelsheim 2006), showing that the minimal sample complexity can be expressed as an optimal (XY) design. They proposed the first algorithms with sample complexity scaling in $(h/\Delta_{\min}^2)\log(1/\delta)$ where $\Delta_{\min}$ is the smallest gap in the model. Tao et al. 2018 proposed a novel estimator of the regression parameter and a G-optimal-design-based algorithm, with a sample complexity in $\sum_{i=1}^{h} \Delta_{(i)}^{-2}\log(1/\delta)$

where $\Delta_{(1)} \leqslant \cdots \leqslant \Delta_{(h)}$ are the h smallest gaps, which improves upon the complexity of the unstructured setting when $K \gg h$. Some algorithms even match the minimal sample complexity either in the asymptotic regime $\delta \to 0$ (Degenne, Ménard, et al. 2020; Jedra & Proutiere 2020b) or within multiplicative factors (Fiez et al. 2019). Some adaptive algorithms, such as LinGapE (Xu et al. 2018), are also very effective in practice, but without provably improving over unstructured algorithms in all instances.

The fixed-budget setting has been studied by Azizi et al. 2022; Yang & Tan 2022, who propose algorithms based on Sequential Halving (Karnin et al. 2013) where in each round, the active arms are sampled according to a G-optimal design. The best guarantees are those obtained by Yang & Tan 2022 who show that a budget T of order $\log_2(h) \sum_{i=1}^{h} \Delta_{(i)}^{-2} \log(1/\delta)$ is sufficient to ensure an error probability at most δ.

Katz-Samuels, Jain, et al. 2020 propose an elimination algorithm for both fixed-confidence and fixed-budget settings, using optimal design to minimize the *Gaussian width*–a complexity measure that may better capture non-asymptotic error rates. Extending such notions to linear PSI is challenging due to the complex structure of the set of alternative models with a different Pareto set.

Contributions. We formalize the *linear Pareto Set Identification* (linear PSI) problem as an extension of PSI to structured arms with features in $\mathbb{R}^h$, allowing the learner to exploit shared information across arms through their linear structure. We propose GEGE (G-optimal Empirical Gap Estimation), the first PSI algorithm based on G-optimal experimental design, applicable to both fixed-confidence and fixed-budget regimes. At each round, arms are sampled according to a G-optimal design computed over the feature space, ensuring maximal information gain for estimating the relevant gaps between arms.

Our unified analysis establishes instance-dependent sample complexity bounds showing that the hardness of linear PSI scales with the h smallest suboptimality gaps: replacing the unstructured dependence on all K arms with $\sum_{i=1}^{h} \Delta_{(i)}^{-2}$ instead of $K/\Delta_{\min}^2$. This yields a substantial reduction in sample complexity compared to the unstructured algorithms of Chapters 2 and 3. In the fixed-confidence setting, GEGE achieves identification within $\mathcal{O}(\log(1/\Delta_{(1)}))$ adaptive rounds, while in the fixed-budget setting, it is, to our knowledge, the first algorithm for multi-output linear bandits with near-optimal performance guarantees.

Learning model. We formalize the linear PSI problem. Let $d, h \in \mathbb{N}^\star$ with $h \leqslant K$. Each arm $k \in [K]$ is associated with a known feature vector $x_k \in \mathbb{R}^h$ and an unknown mean reward vector $\mu_k \in \mathbb{R}^d$ satisfying

$$\mu_k = \theta^\top x_k, \qquad \theta \in \mathbb{R}^{h \times d},$$

where θ is an unknown parameter matrix. Let $\mathcal{X} \coloneqq (x_1, \ldots, x_K)^\top$ and $[K] \coloneqq \{1, \ldots, K\}$.

Definition 4.1.1 (Pareto dominance)**.** For any two arms $i, j \in [K]$, arm i is *weakly dominated* by j if $\mu_i^c \leqslant \mu_j^c$ for all $c \in [d]$. It is *dominated* ($i \preceq j$) if it is weakly dominated and strictly smaller in at least one coordinate, and *strictly dominated* ($i \prec j$) if all inequalities are strict.

The (strict) *Pareto set* is then defined as

$$\mathcal{S}^\star \;=\; \big\{i \in [K] : \nexists j \in [K] \setminus \{i\},\ \mu_i \prec \mu_j\big\}.$$

At each round t, the learner selects an arm $A_t \in [K]$ and observes

$$y_t = \theta^\top x_{A_t} + \eta_t,$$

where η_t is centered and, conditionally on A_t, has σ-subgaussian marginals[1]. The goal is to identify $\mathcal{S}^\star$ as efficiently as possible.

In the *fixed-confidence* setting, given $\delta \in (0,1)$, the learner collects samples up to a stopping time τ and outputs $\widehat{S}_\tau$ such that $\mathbb{P}(\widehat{S}_\tau \neq \mathcal{S}^\star) \leqslant \delta$ while minimizing τ. In the *fixed-budget* setting, sampling stops after T rounds, and the objective is to minimize the error probability $e_T \coloneqq \mathbb{P}(\widehat{S}_T \neq \mathcal{S}^\star)$.

Throughout the chapter, we use the following notation. $\triangle_n$ denotes the probability simplex in $\mathbb{R}^n$. For a positive semidefinite matrix $A \in \mathbb{R}^{n\times n}$ and a vector $x \in \mathbb{R}^n$, we write $\|x\|_A^2 \coloneqq x^\top A x$, and $x(i)$ denotes the i-th coordinate of x. For scalars $a, b \in \mathbb{R}$, $a \wedge b \coloneqq \min\{a,b\}$ and $(a)_+ \coloneqq \max\{a, 0\}$.

4.1.1 Complexity measures for Pareto Set Identification

As discussed in Chapters 2 and 3, the difficulty of Pareto Set Identification (PSI) in the unstructured setting can be quantified through suitable notions of *suboptimality gaps*, which measure how confidently one can classify each arm as dominated or non-dominated. These gaps, first introduced by Auer et al. 2016 and further refined in Chapter 2, rely on the quantities

$$\mathrm{m}(i,j) \coloneqq \min_{c\in[d]}[\,\mu_j^c - \mu_i^c\,], \qquad \mathrm{M}(i,j) \coloneqq -\mathrm{m}(i,j),$$

which respectively capture how much arm j dominates arm i, or how much j must be shifted component-wise to dominate i. In particular, $\mathrm{m}(i,j) > 0$ if and only if $i \prec j$, while $\mathrm{M}(i,j) > 0$ if and only if $i \not\preceq j$.

For each arm i, the *suboptimality gap* Δ_i (see Lemma 1 in Kone, Kaufmann, et al. 2024) is defined as

$$\Delta_i \coloneqq \begin{cases} \Delta_i^\star, & \text{if } i \notin \mathcal{S}^\star, \\ \min\limits_{j\neq i}\big[\mathrm{M}(i,j) \wedge \big(\mathrm{M}(j,i)_+ + (\Delta_j^\star)_+\big)\big], & \text{otherwise,} \end{cases} \tag{4.1}$$

with

$$\Delta_i^\star \coloneqq \max_{j\in[K]} \mathrm{m}(i,j)\,.$$

Intuitively, for a dominated arm i, Δ_i represents the minimal shift required to make it non-dominated, while for an optimal arm i, it measures its closest separation from both other optimal and suboptimal arms.

[1] A centered random variable X is σ-subgaussian if $\log \mathbb{E}[\exp(\lambda X)] \leqslant \lambda^2\sigma^2/2$ for all $\lambda \in \mathbb{R}$.

In the sequel, we assume without loss of generality that $\Delta_1 \leqslant \Delta_2 \leqslant \ldots \leqslant \Delta_K$. Following Chapters 2 and 3, we recall the global complexity measures

$$H_1(\nu) \coloneqq \sum_{i=1}^{K} \Delta_i^{-2} \quad \text{and} \quad H_2(\nu) \coloneqq \max_{i\in[K]} \frac{i}{\Delta_i^2},$$

which govern the sample complexity in the fixed-confidence and fixed-budget settings, respectively.

Structured complexity under linearity. In the linear setting, the learner estimates the low-dimensional parameter $\theta \in \mathbb{R}^{h\times d}$ rather than the K individual means, leading to a substantial reduction in complexity. We therefore define the analogous quantities

$$H_{1,\text{lin}}(\nu) \coloneqq \sum_{i=1}^{h} \frac{1}{\Delta_i^2}, \qquad H_{2,\text{lin}}(\nu) \coloneqq \max_{i\in[h]} \frac{i}{\Delta_i^2}, \tag{4.2}$$

which characterize the hardness of the linear PSI problem in the fixed-confidence and fixed-budget settings, respectively. Since these depend only on the h smallest gaps, $H_{1,\text{lin}}(\nu) \leqslant H_1(\nu)$ and $H_{2,\text{lin}}(\nu) \leqslant H_2(\nu)$, illustrating the gain achieved by leveraging linear structure.

4.1.2 Least-squares estimation and optimal designs

Given n arm selections $a_1, \ldots, a_n$, define $X_n \coloneqq (x_{a_1} \ldots x_{a_n})^\intercal \in \mathbb{R}^{n\times h}$ and $Y_n \coloneqq (y_1 \ldots y_n)^\intercal \in \mathbb{R}^{n\times d}$. We define the information matrix as $V_n \coloneqq X_n^\intercal X_n = \sum_{i=1}^{K} N_n(i) x_i x_i^\intercal \in \mathbb{R}^{h\times h}$ where $N_{n,i}$ denotes the number of observations from arm i among the n samples. More generally, given $w \in \mathbb{R}_+^K$, we define $V(w) \coloneqq \sum_{i=1}^{K} w(i) x_i x_i^\intercal$.

The multi-output regression model can be written in matrix form as $Y_n = X_n\theta + H_n$ where $H_n \coloneqq (\eta_1 \ldots \eta_n)^\intercal$ is the noise matrix. The least-squares estimate $\hat{\theta}_n$ of the matrix θ is defined as the matrix minimizing the least-squares error $\text{Err}_n(A) \coloneqq \|X_n A - Y_n\|_{\text{F}}^2$. Computing the gradient of the loss yields $V_n\hat{\theta}_n = X_n^\top Y_n$. If the matrix V_n is nonsingular, the least-squares estimator can be written

$$\hat{\theta}_n = V_n^{-1} X_n^\intercal Y_n.$$

In the course of our elimination algorithm, we will compute least-squares estimates based on observations from a restricted number of arms, and we will face the case in which V_n is singular. In this case, different choices have been made in prior work on linear bandits: Alieva et al. 2021 define a custom "pseudo-inverse" while Yang & Tan 2022 define new contexts $\widetilde{x}_i$ that are projections of the x_i onto a subspace of dimension $\text{rank}(\mathcal{X}_S)$ where $\mathcal{X}_S \coloneqq (x_i : i \in S)^\intercal$ and S is the set of arms that are active. We adopt an approach close to the latter, which is described below. Let the singular value decomposition of $(\mathcal{X}_S)^\intercal$ be $USV^\intercal$ where U, V are orthogonal matrices and $B \coloneqq (u_1, \ldots, u_m)$ is formed with the first m columns of U where $m = \text{rank}(\mathcal{X}_S)$. We then define

$$V_n^\dagger \coloneqq B(B^\intercal V_n B)^{-1} B^\intercal \quad \text{and} \quad \hat{\theta}_n = V_n^\dagger X_n^\intercal Y_n. \tag{4.3}$$

The following result addresses the statistical uncertainty of this estimator.

Lemma 4.1.2. *If $\Sigma \in \mathbb{R}^{d\times d}$ is the covariance matrix of the noise η_t and $a_1, \dots, a_n$ are deterministically chosen, then for any $x_i \in \{x_{a_1}, \dots, x_{a_n}\}$, $\mathrm{Cov}(\hat{\theta}_n^\intercal x_i) = \|x_i\|^2_{V_n^\dagger}\Sigma$.*

Hence, achieving uniform estimation accuracy across arms amounts to selecting $\{a_1, \dots, a_n\}$ so as to minimize $\max_{i\in S} \|x_i\|^2_{V_n^\dagger}$. Relaxing this combinatorial problem leads to the *G-optimal design*:

$$w_S^\star \in \operatorname*{argmin}_{w\in\triangle_{|S|}} \max_{i\in S} \|\widetilde{x}_i\|^2_{\widetilde{V}(w)^{-1}}, \tag{4.4}$$

where $\widetilde{x}_i = B^\top x_i$ are the projected features and $\widetilde{V}(w) = \sum_{i\in S} w(i)\,\widetilde{x}_i\widetilde{x}_i^\top$, with the convention that $w(i)$ indexes arms in S. Intuitively, $w_S^\star$ specifies a sampling distribution over the active arms that yields the most uniform estimation accuracy of their mean responses under the projected model (4.3). A procedure for computing $w_S^\star$ is provided in Appendix H of Kone, Kaufmann, et al. 2025a.

4.2 Algorithmic contribution

Our elimination algorithms operate in rounds. They progressively eliminate a portion of arms and classify them as optimal or suboptimal based on empirical estimation of their gaps. In each round, a sampling budget is allocated among the surviving arms based on a G-optimal design.

4.2.1 Optimal designs and gap estimation

At round r, we denote by $\mathcal{A}_r$ the set of arms that are still active. To estimate the means and, henceforth, the gaps, we first compute an estimate of the regression matrix denoted $\hat{\theta}_r$. This estimate is obtained by carefully sampling the arms using an integral rounding of a G-optimal design.

Algorithm 4.1 takes as input a set of arms S, a budget N and chooses some N arms to pull (with repetitions) based on an integer rounding of $w_S^\star$, a continuous G-optimal design over the set $\{\widetilde{x}_i, i \in S\}$ of (transformed) features associated with the arms (see Section 4.1.2). Several rounding procedures have been proposed in the literature, and we use that of Allen-Zhu et al. 2017, henceforth referred to as ROUND, which satisfies the following guarantees, as proven in Lemma 12 of Kone, Kaufmann, et al. 2025a.

Lemma 4.2.1 (Lemma 12 of Kone, Kaufmann, et al. 2025a)**.** *ROUND$(N, \widetilde{\mathcal{X}}_S, w_S^\star, \kappa)$ outputs a sequence of arms $(a_1, \dots, a_N) \in S^N$ such that*

$$\max_{i\in S} \|x_i\|^2_{V_N^\dagger} \leqslant (1 + 6\kappa)\,\frac{F_S(w_S^\star)}{N},$$

where $F_S(w_S^\star)$ is the optimal value of (4.4).

Algorithm 4.1: OptEstimator: Least-squares estimation from G-optimal design

Require: Subset $S \subset [K]$, sample size N, precision κ

// Feature transformation (Section 4.1.2)

Compute the transformed features $\widetilde{\mathcal{X}}_S = (B^\intercal x_i, i \in S)$ with B as defined in Section 4.1.2 // G-optimal design over transformed features and sampling

Compute a G-optimal design $w_S^\star$ over the set $\widetilde{\mathcal{X}}_S$

Pull $(a_1, \dots, a_N) \leftarrow \text{ROUND}(N, \widetilde{\mathcal{X}}_S, w_S^\star, \kappa)$ and collect responses $y_1, \dots, y_N$

Compute $V_N^\dagger$ as in Eq. (4.3) and compute the OLS estimator on the samples collected

$$\hat{\theta} \leftarrow V_N^\dagger \sum_{t=1}^{N} x_{a_t} y_t^\intercal$$

return: $\hat{\theta}$

By the Kiefer–Wolfowitz theorem (Kiefer & Wolfowitz 1960), $F_S(w_S^\star) = h_S$, the dimension of $\text{span}\{x_i : i \in S\}$. This observation is key to the following concentration result.

Lemma 4.2.2. *Let $S \subset [K]$, $\kappa \in (0, 1/3]$ and $N \geqslant 5h_S/\kappa^2$ where $h_S = \dim\big(\text{span}\{x_i : i \in S\}\big)$. The output $\hat{\theta}$ of OptEstimator(S, N, κ) satisfies for all $\varepsilon > 0$ and $i \in S$*

$$\mathbb{P}\left(\|(\theta - \hat{\theta})^\intercal x_i\|_\infty \geqslant \varepsilon\right) \leqslant 2d \exp\left(-\frac{N\varepsilon^2}{2(1+6\kappa)\sigma^2 h_S}\right).$$

Proof of Lemma 4.2.2. By assumption, the noise vector has σ-subgaussian marginals. It is easy (see *e.g.*, proof of Lemma 11 in Kone, Kaufmann, et al. 2025a) to see that for any $i \in S$, the marginals of $(\theta - \hat{\theta})x_i$ are $\sigma\|X_N^\intercal V_N^\dagger x_i\|_2$-subgaussian. Then, direct algebra yields

$$\begin{aligned}
\|X_N^\intercal V_N^\dagger x_i\|_2^2 &= x_i^\intercal V_N^\dagger V_N V_N^\dagger x_i \\
&= x_i^\intercal \left(B_S(B_S^\intercal V_N B_S)^{-1} B_S^\intercal\right) V_N \left(B_S(B_S^\intercal V_N B_S)^{-1} B_S^\intercal\right) x_i \\
&= x_i^\intercal B_S(B_S^\intercal V_N B_S)^{-1} B_S^\intercal x_i \\
&= x_i^\intercal V_N^\dagger x_i = \|x_i\|_{V_N^\dagger}^2.
\end{aligned}$$

Next, by concentration of subgaussian variables (see *e.g.*, Lattimore & Szepesvari 2020)

$$\begin{aligned}
\mathbb{P}(\|(\theta - \hat{\theta})^\intercal x_i\|_\infty \geqslant \varepsilon) &\leqslant 2d \exp\left(-\frac{\varepsilon^2}{2\sigma^2 \|x_i\|_{V_N^\dagger}^2}\right) \\
&\leqslant 2d \exp\left(-\frac{\varepsilon^2}{2\sigma^2 \max_{k \in S} \|x_k\|_{V_N^\dagger}^2}\right),
\end{aligned}$$

and the result follows from Lemma 4.2.1. □

Empirical gaps. Once the parameter $\hat{\theta}_r$ has been obtained as an output of Algorithm 4.1 with $S = \mathcal{A}_r$ and an appropriate value of the budget N, we compute estimates of the mean vectors as $\hat{\mu}_{i,r} := \hat{\theta}_r^\intercal x_i$ and the empirical Pareto set of active arms,

$$S_r := \{i \in \mathcal{A}_r : \nexists j \in \mathcal{A}_r : \hat{\mu}_{i,r} \prec \hat{\mu}_{j,r}\}.$$

In both the fixed-confidence and fixed-budget settings, at round r, after collecting new samples from the surviving arms, GEGE discards a fraction of the arms based on the empirical estimation of their gaps. We first introduce the empirical quantities used to compute the gaps:

$$\begin{aligned} \mathrm{M}(i,j;r) &:= \max_{c\in[d]} [\hat{\mu}^c_{i,r} - \hat{\mu}^c_{j,r}], \quad \text{and} \\ \mathrm{m}(i,j;r) &:= \min_{c\in[d]} [\hat{\mu}^c_{j,r} - \hat{\mu}^c_{i,r}]. \end{aligned}$$

We define for any arm $i \in \mathcal{A}_r$,

$$\widehat{\Delta}^\star_{i,r} := \max_{j\in\mathcal{A}_r} \mathrm{m}(i,j;r)\,,$$

and the empirical estimates of the PSI gaps as:

$$\widehat{\Delta}_{i,r} := \begin{cases} \widehat{\Delta}^\star_{i,r} & \text{if } i \in \mathcal{A}_r \backslash S_r \\ \widehat{\delta}^\star_{i,r} & \text{if } i \in S_r \end{cases} \tag{4.5}$$

with $\widehat{\delta}^\star_{i,r} := \min_{j\in\mathcal{A}_r\backslash\{i\}} [\mathrm{M}(i,j;r) \wedge (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+)]$; the empirical estimates of the gaps introduced earlier in Section 4.1.1 and already used in Chapter 2.

Unlike BAI, which only requires rejecting suboptimal arms, PSI must explicitly classify each arm as either Pareto-optimal or dominated, since the cardinality of the Pareto set is unknown. These classification mechanisms are detailed in the following sections for the fixed-budget and fixed-confidence settings.

Final output. In both cases, letting $\mathcal{A}_r$ be the set of active arms and $\mathcal{B}_r$ be the set of arms already classified as optimal at the beginning of round r, GEGE outputs $\mathcal{B}_{\tau+1} \cup \mathcal{A}_{\tau+1}$ as the candidate Pareto-optimal set, where τ denotes the final round. Moreover, $\mathcal{A}_{\tau+1}$ contains at most one arm.

GEGE for ε-PSI. Algorithm 4.2 can be easily modified to identify an ε-Pareto set. As introduced in Auer et al. 2016 and studied in Chapter 3, an ε-Pareto set S_ε is such that $S^\star \subset S_\varepsilon$ and for any arm $i \in S_\varepsilon$, $\Delta^\star_i \leqslant \varepsilon$: it contains the Pareto set and possibly some suboptimal arms that are (ε)-close to be optimal. To identify an ε-Pareto set, we relax the stopping condition: instead of stopping when it remains only one active arm (*i.e.*, $|\mathcal{A}_r| \leqslant 1$), we stop when $(|\mathcal{A}_r| \leqslant 1 \;\text{ or }\; \varepsilon_r \leqslant \varepsilon/4)$ holds. After stopping, the same set is recommended, namely $\mathcal{A}_{\tau+1} \cup \mathcal{B}_{\tau+1}$. The guarantees of GEGE under this modification are discussed in Section E.5 of Kone, Kaufmann, et al. 2025a.

Algorithm 4.2: GEGE: G-optimal Empirical Gap Elimination (fixed-confidence)

Require: risk parameter $\delta \in (0,1)$; noise proxy σ^2

Initialize: set of active arms $\mathcal{A}_1 \leftarrow [K]$; set of arms (so far) optimal $\mathcal{B}_1 \leftarrow \emptyset$; set of discarded arms $\mathcal{D}_1 \leftarrow \emptyset$; $r \leftarrow 1$

while $|\mathcal{A}_r| > 1$ **do**

// Schedule, confidence split, and local dimension

$\varepsilon_r \leftarrow \dfrac{1}{2 \cdot 2^r}$; $\delta_r \leftarrow \dfrac{6\delta}{\pi^2 r^2}$; $h_r \leftarrow \dim\big(\mathrm{span}\{x_i : i \in \mathcal{A}_r\}\big)$

// Round budget via G-optimal design complexity

$t_r \leftarrow \left\lceil \dfrac{32(1+3\varepsilon_r)\,\sigma^2\, h_r}{\varepsilon_r^2} \log\Big(\dfrac{|\mathcal{A}_r|\, d}{2\,\delta_r}\Big) \right\rceil$

// Compute estimate from near G-optimal allocation over $\mathcal{A}_r$

$\hat{\theta}_r \leftarrow \textsc{OptEstimator}\big(\mathcal{A}_r,\ t_r,\ \varepsilon_r\big)$

Compute S_r and empirical gaps $\widehat{\Delta}_{i,r}$ via (4.5)

// Accept strong Pareto-optimal arms; discard strong suboptimal arms

$\mathcal{B}_{r+1} \leftarrow \mathcal{B}_r \cup \{\, i \in S_r : \widehat{\Delta}_{i,r} \geqslant \varepsilon_r \,\}$

$\mathcal{D}_{r+1} \leftarrow \mathcal{D}_r \cup \{\, i \in \mathcal{A}_r \setminus S_r : \widehat{\Delta}_{i,r} \geqslant \varepsilon_r/2 \,\}$

// Shrink active set and advance round

$\mathcal{A}_{r+1} \leftarrow \mathcal{A}_r \setminus \big(\mathcal{D}_{r+1} \cup \mathcal{B}_{r+1}\big)$

$r \leftarrow r+1$

return $\mathcal{B}_r \cup \mathcal{A}_r$

4.2.2 Fixed-budget algorithm

Algorithm 4.3, operates over $\lceil \log_2(h) \rceil$ rounds, with an equal budget of $T/\lceil \log_2(h) \rceil$ allocated per round. By construction $\left|\mathcal{A}_{\lceil \log_2(h) \rceil + 1}\right| = 1$. At the end of round r, the $\lceil h/2^r \rceil$ arms with the smallest empirical gaps are kept active while the remaining arms are discarded and classified as Pareto optimal (added to $\mathcal{B}_{r+1}$) if they are empirically optimal (belonging to set S_r) and deemed suboptimal otherwise. If a tie occurs, we break it in favor of empirically Pareto-optimal arms (*i.e.*, keep arms in S_r when possible). This is crucial to prove the guarantees on the algorithm, as sketched in Section 4.3.

Algorithm 4.3: GEGE: G-optimal Empirical Gap Elimination (fixed-budget)

Require : Total budget T; feature dimension h

Initialize: $R \leftarrow \lceil \log_2(h) \rceil$; $t_r \leftarrow \left\lceil \frac{T}{R} \right\rceil$: per-round allocation size (equal split across rounds) ; set of active arms $\mathcal{A}_1 \leftarrow [K]$; $\mathcal{B}_1 \leftarrow \emptyset$; $\mathcal{D}_1 \leftarrow \emptyset$;

for $r = 1$ **to** $\lceil \log_2(h) \rceil$ **do**

  // G-optimal design over current active set

  $\hat{\theta}_r \leftarrow \textsc{OptEstimator}(\mathcal{A}_r,\ t_r,\ 1/3)$;

  // Execute allocation, update estimates, form Pareto set and gaps

  Compute S_r (empirical Pareto set) and gaps $\widehat{\Delta}_{i,r}$ via (4.5);

  // Keep the $\lceil h/2^r \rceil$ smallest-gap arms for next round

  $m_r \leftarrow \left\lceil \frac{h}{2^r} \right\rceil$; $\mathcal{A}_{r+1} \leftarrow$ the m_r arms in $\mathcal{A}_r$ with smallest $\widehat{\Delta}_{i,r}$;

  // Tie-breaks favor arms in S_r

  break ties by keeping $i \in S_r$ when possible;

  $\mathcal{B}_{r+1} \leftarrow \mathcal{B}_r \cup \big(S_r \cap (\mathcal{A}_r \setminus \mathcal{A}_{r+1})\big)$;

  $\mathcal{D}_{r+1} \leftarrow \mathcal{D}_r \cup \big((\mathcal{A}_r \setminus \mathcal{A}_{r+1}) \setminus S_r\big)$;

$\mathcal{B}_{R+1} \cup \mathcal{A}_{R+1}$

4.3 Main theoretical results

We now turn to the theoretical analysis of the proposed algorithms. In this section, we establish finite-sample guarantees for both the fixed-budget and fixed-confidence versions of GEGE, beginning with the fixed-budget setting.

4.3.1 Fixed-budget

In the fixed-budget setting, Algorithm 4.3 satisfies the following guarantees.

Theorem 4.3.1. *The probability of error of Algorithm 4.3 run with budget $T \geqslant 45h \log_2 h$ is at most*

$$\exp\left(-\frac{T}{1200\sigma^2 H_{2,lin}\lceil \log_2 h\rceil} + \log C(h,d,K)\right),$$

where $C(h,d,K) = 2d\,(K + h + \lceil \log_2 h\rceil)$.

To the best of our knowledge, GEGE is the first algorithm with theoretical guarantees for fixed-budget linear PSI. Our result shows that in this setting, the probability of error scales only with the first h gaps. In Chapter 2, we introduced EGE-SH, an algorithm for fixed-budget PSI in the unstructured setting whose probability of error is essentially upper-bounded by

$$\exp\left(-\frac{T}{288\sigma^2 H_2 \log_2 K} + \log(2d(K-1)|\mathcal{S}^\star|\log_2 K)\right).$$

Therefore, GEGE largely improves upon EGE-SH when $K \gg h$.

When $K = h$ and $x_1, \dots, x_K$ are the canonical $\mathbb{R}^h$-basis, both algorithms coincide[2], thus, GEGE can be seen as a generalization of EGE-SH.

The following lower bound for linear PSI in the fixed-budget setting shows that GEGE is optimal in the worst case, up to constants and a $\log_2(h)$ factor.

Theorem 4.3.2. *Let $\mathbb{W}_H$ be the set of instances with complexity $H_{2,lin}$ smaller than H. For any budget T, letting $\widehat{S}_T^{\text{alg}}$ be the output of an algorithm* alg*, it holds that*

$$\inf_{\text{alg}} \sup_{\nu \in \mathbb{W}_H} \mathbb{P}_\nu(\widehat{S}_T^{\text{alg}} \neq \mathcal{S}^\star(\nu)) \geqslant \frac{1}{4}\exp\left(-\frac{2T}{H\sigma^2}\right).$$

4.3.2 Fixed-confidence

We now turn to the fixed-confidence regime, where the learner adaptively collects samples until it can identify the Pareto set with confidence $1-\delta$.

Theorem 4.3.3. *With probability at least $1-\delta$, Algorithm 4.2 identifies the Pareto set using at most*

$$\log_2(2/\Delta_1) + \mathcal{O}\left(\sum_{i=2}^{h} \frac{\sigma^2}{\Delta_i^2} \log\left(\frac{Kd}{\delta}\log_2\left(\frac{2}{\Delta_i}\right)\right)\right)$$

samples and $\lceil \log_2(1/\Delta_1)\rceil$ rounds.

This result shows that the complexity of Algorithm 4.2 scales only with the first h gaps. In particular, when $K \gg h$ using our algorithm substantially reduces the sample complexity of PSI. In Table 4.1, we compare the sample complexity of GEGE to that of existing fixed-confidence PSI algorithms, showing that GEGE enjoys stronger guarantees than its competitors. We emphasize that both Kim et al. 2023 and Zuluaga, Krause, et al.

Table 4.1: Sample complexity up to constant multiplicative terms of different algorithms for PSI in the fixed-confidence setting.

Algorithm	Upper-Bound on τ_δ	Linear PSI
Zuluaga, Krause, et al. 2016	$\left(\frac{h^2}{\Delta^2_{\min}}\right) \log^3\left(\frac{Kd}{\delta}\right)$	✔
Kone, Kaufmann, et al. 2023	$\sum_{i=1}^{K} \frac{1}{\Delta_i^2} \log(\frac{Kd}{\delta} \log(\frac{1}{\Delta_i}))$	✘
Kim et al. 2023	$\frac{h}{\Delta^2_{\min}} \log(\frac{d(h \vee K)}{\delta \Delta^2_{\min}})$	✔
GEGE (Ours)	$\sum_{i=1}^{h} \frac{1}{\Delta_i^2} \log(\frac{Kd}{\delta} \log(\frac{1}{\Delta_i}))$	✔

2016 use uniform sampling and do not exploit an optimal design, which prevents them from reaching the guarantees given in Theorem 4.3.3.

The result below upper bounds the sample complexity by leveraging properties of the optimal design.

Theorem 4.3.4. *With probability at least $1-\delta$, the sample complexity of Algorithm 4.2 is upper bounded by*

$$\mathcal{O}\left(\log\frac{1}{\Delta_1} \inf_{w\in\triangle_K} \max_{k\in[K]} \frac{2\sigma^2\|x_k\|^2_{V(w)^{-1}}}{\Delta_k^2} \log\frac{Kd}{\Delta_1\delta}\right).$$

This provides an alternative characterization of the sample complexity, though it is not directly comparable to Theorem 4.3.3 in general.

We state a lower bound, showing that our algorithm is essentially minimax optimal for linear PSI.

Theorem 4.3.5. *For any $K, d, h \in \mathbb{N}$, there exists a set $\mathcal{M}(K,d,h)$ of linear PSI instances such that for $\nu \in \mathcal{M}(K,d,h)$ and for any δ-correct algorithm for linear PSI, with probability at least $1-\delta$,*

$$\tau_\delta^{\mathcal{A}} \geqslant \Omega\left(H_{1,lin}(\nu)\log\frac{1}{\delta}\right).$$

Remark 4.1. When $K = h$ and $x_1, \dots, x_K$ form the canonical $\mathbb{R}^h$ basis, we recover the classical PSI problem. We note that, unlike its fixed-budget version, GEGE does not coincide with an existing PSI identification algorithm. Instead, it matches the optimal guarantees of Kone, Kaufmann, et al. 2023 while needing only $\lceil\log(1/\Delta_1)\rceil$ rounds of adaptivity, which is the first fixed-confidence PSI algorithm having this property. Such a batched algorithm may be desirable in some applications, *e.g.*, in clinical trials where measuring different biological indicators of efficacy can take time.

[2]Up to the fact that EGE-SH does not discard samples between rounds

4.3.3 Sketch of proofs

Before sketching our proof strategy, we highlight a key property of PSI that makes the analysis different from classical BAI settings. Let a be a (Pareto) suboptimal arm. From (4.1), there exists $a^\star \in \mathcal{S}^\star$ such that $\Delta_a = \mathrm{m}(a, a^\star)$ and importantly, $a^\star$ could be the unique arm dominating a. Therefore, discarding $a^\star$ before a may result in the latter appearing as optimal in the remaining rounds, thus leading to the misidentification of the Pareto set.

To avoid this, an elimination algorithm for PSI should guarantee that if a suboptimal arm a is active, then $a^\star$ is also active. We introduce the following event

$$\mathcal{P}_r \coloneqq \{\forall\, s \leqslant r : \forall i \in (\mathcal{S}^\star)^\mathsf{c},\ i \in \mathcal{A}_s \Rightarrow i^\star \in \mathcal{A}_s\}.$$

An important aspect of our proofs is to control the occurrence of $\mathcal{P}_\infty$ (by convention, if $\mathcal{P}_t$ holds and $\mathcal{A}_s = \emptyset$ for any $s \geqslant t$, then $\mathcal{P}_\infty$ holds). The first step of the proof is to show that when $\mathcal{P}_r$ holds, we can control the deviations of the empirical gaps, which is essential to guarantee the correctness of GEGE and to control its sample complexity in the fixed-confidence setting. We now define for $\eta > 0$, the good event

$$\mathcal{E}^r(\eta) \coloneqq \left\{\forall\, i, j \in \mathcal{A}_r : \left\|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\right\|_\infty \leqslant \eta\right\}. \tag{4.6}$$

Letting $n_r = |\mathcal{A}_r|$ and λ a constant to be specified, we introduce

$$\mathcal{E}^\lambda_{\mathrm{fb}} \coloneqq \cap_{r=1}^{\lceil \log_2(h)\rceil} \mathcal{E}^r(\lambda \Delta_{n_{r+1}+1})\,, \quad \text{and} \quad \mathcal{E}_{\mathrm{fc}} \coloneqq \cap_{r=1}^{\infty} \mathcal{E}^r(\varepsilon_r/2)\,.$$

We then prove by concentration and induction the following key result.

Proposition 4.3.6. *Let $\lambda \in (0, 1/5)$ and assume $\mathcal{E}_{\mathrm{fc}}$ (resp. $\mathcal{E}^\lambda_{\mathrm{fb}}$ in fixed-budget) holds. Then at any round r, $\mathcal{P}_r$ holds and for all arm $i \in \mathcal{A}_r$,*

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -\eta_r & \text{if } i \in \mathcal{S}^\star \\ -\eta_r/2 & \text{else,} \end{cases} \quad \text{where} \quad \eta_r \coloneqq \begin{cases} 2\lambda\Delta_{n_{r+1}+1} & \text{(fixed-budget)} \\ \varepsilon_r & \text{(fixed-confidence).} \end{cases}$$

Building on this result, we show that the recommendation of Algorithm 4.3 is correct on $\mathcal{E}^\lambda_{\mathrm{fb}}$, so its probability of error is upper-bounded by $\inf_{\lambda\in(0,1/5)} \mathbb{P}(\mathcal{E}^\lambda_{\mathrm{fb}})$. We conclude the proof of Theorem 4.3.1 by upper-bounding this probability (see Section 4.5).

Similarly, using Proposition 4.3.6 we prove the correctness of Algorithm 4.2 on $\mathcal{E}_{\mathrm{fc}}$: at any round r, $\mathcal{B}_r \subset \mathcal{S}^\star$ and $\mathcal{D}_r \subset (\mathcal{S}^\star)^\mathsf{c}$. To further upper bound its sample complexity, we need an additional result to control the size of $\mathcal{A}_r$.

Lemma 4.3.7. *The following statement holds for Algorithm 4.2 on the event $\mathcal{E}_{\mathrm{fc}}$: for all $p \in [K]$, after $\lceil \log(1/\Delta_p)\rceil$ rounds it remains less than p active arms. In particular, GEGE stops after at most $\lceil \log(1/\Delta_1)\rceil$ rounds.*

The proof of this lemma is given in section 4.5.3. To get the sample complexity bound of Theorem 4.3.3, some extra arguments are needed. We sketch some elements below (the

full proof is given in section 4.5). Assume $\mathcal{E}_{\text{fc}}$ holds and let τ_δ be the sample complexity of Algorithm 4.2. Lemma 4.3.7 yields $\tau_\delta \leqslant \sum_{r=1}^{\lceil \log(1/\Delta_1)\rceil} \mathcal{O}(h_r/\varepsilon_r^2)$ with $h_r \leqslant |\mathcal{A}_r|$.

Using Lemma 4.3.7, we introduce "checkpoints rounds" between which we control $|\mathcal{A}_r|$ and thus h_r. Let the sequence $(\alpha_s)_{s\geqslant 0}$ defined as $\alpha_0 = 0$ and $\alpha_s = \lceil \log_2(1/\Delta_{\lfloor h/2^s \rfloor})\rceil$, for $s \geqslant 1$. Simple calculation yields $\alpha_{\lfloor \log_2(h)\rfloor} = \lceil \log_2(1/\Delta_1)\rceil$ and $\{1, \dots, \lceil \log_2(1/\Delta_1)\rceil\} = \cup_{s=1}^{\lfloor \log_2(h)\rfloor} [\![1+\alpha_{s-1}, \alpha_s]\!]$. Therefore

$$\tau_\delta \leqslant \sum_{s=1}^{\lfloor \log_2(h)\rfloor} \sum_{r=\alpha_{s-1}+1}^{\alpha_s} \mathcal{O}(|\mathcal{A}_r|/\varepsilon_r^2)\,.$$

Now by Lemma 4.3.7, for $r > \alpha_s$, $|\mathcal{A}_r| \leqslant \lfloor h/2^s \rfloor$, so essentially $\tau_\delta \leqslant \sum_{s=1}^{\lfloor \log_2(h)\rfloor} \mathcal{O}(4^{\alpha_s}\lfloor h/2^s \rfloor)$. Carefully re-indexing this sum and addressing a few more technicalities, we obtain the result in Theorem 4.3.3. Showing that $\mathbb{P}(\mathcal{E}_{\text{fc}}) \geqslant 1-\delta$, from Lemma 4.2.2 completes the proof of Theorem 4.3.3 (see Section 4.5)

On the other side, combining the Kiefer–Wolfowitz theorem

$$\begin{aligned}
\tau_\delta &\leqslant \sum_{r=1}^{\lceil \log(1/\Delta_1)\rceil} \mathcal{O}\left(\frac{\sigma^2 h_r}{\varepsilon_r^2}\log\frac{Kd2^r}{\delta}\right) \\
&\leqslant \sum_{r=1}^{\lceil \log(1/\Delta_1)\rceil} \mathcal{O}\left(\inf_{w\in\triangle_K}\max_{k\in\mathcal{A}_r}\|x_k\|^2_{V(w)^{-1}}\frac{\sigma^2}{\varepsilon_r^2}\log\frac{Kd2^r}{\delta}\right) \quad \text{(Kiefer–Wolfowitz)}.
\end{aligned}$$

Next, note that if $k \in \mathcal{A}_r$ then $\Delta_k \leqslant 2\varepsilon_{r-1}$ (otherwise it would have been discarded at the end of round $r-1$). Therefore, $1/\varepsilon_r^2 \leqslant 4/\Delta_k^2$ for any arm $k \in \mathcal{A}_r$. Putting these displays together yields

$$\begin{aligned}
\tau_\delta &\leqslant \sum_{r=1}^{\lceil \log(1/\Delta_1)\rceil} \mathcal{O}\left(\sigma^2 \inf_{w\in\triangle_K}\max_{k\in\mathcal{A}_r}\frac{\|x_k\|^2_{V(w)^{-1}}}{\Delta_k^2}\log\frac{Kd2^r}{\delta}\right) \\
&\leqslant \mathcal{O}\left(\log\frac{1}{\Delta_1}\inf_{w\in\triangle_K}\max_{k\in[K]}\frac{2\sigma^2\|x_k\|^2_{V(w)^{-1}}}{\Delta_k^2}\log\frac{Kd}{\Delta_1\delta}\right).
\end{aligned}$$

4.4 Numerical study and discussion

We now empirically evaluate the proposed algorithms on both synthetic and real-world instances. Our experiments aim to validate the theoretical guarantees derived in Section 4.3, and to illustrate the statistical and computational advantages of exploiting the linear structure in Pareto set identification.

Experimental setup. In the fixed-budget setting, we compare GEGE-FB with two unstructured PSI algorithms from Chapter 2–EGE-SH and EGE-SR–as well as a uniform sampling

baseline. In the fixed-confidence setting, we benchmark GEGE-FC against APE, the fully adaptive algorithm introduced in Chapter 3, and PAL (Zuluaga, Sergent, et al. 2013), a Gaussian-process–based method instantiated here with a linear kernel. For the synthetic instances, we fix feature vectors $x_1, \dots, x_h$ and add additional arms $x_{h+1}, \dots, x_K$. A common regression matrix θ is shared across all instances. Given $K \geqslant h$, we construct a linear PSI instance ν_K by augmenting the base feature set with the additional arms $x_{h+1}, \dots, x_K$, chosen such that the first h arms have identical minimal gaps. Consequently, the complexity terms $H_{1,\text{lin}}$ and $H_{2,\text{lin}}$ remain constant as K grows. Unless stated otherwise, we set $h = 8$ and $d = 2$.

For the real-world dataset, we evaluate the algorithms on the NoC dataset (Almer et al. 2011b), a bi-objective hardware design benchmark involving $d = 2$ performance metrics–energy consumption and runtime–across $K = 259$ implementations, each described by $h = 4$ features for hardware implementation (chip area, bandwidth, etc.)

For each algorithm, we report (i) the empirical probability of misidentification in the fixed-budget case, and (ii) the empirical distribution of the sample complexity in the fixed-confidence case, both averaged over 500 independent runs. We fix $\delta = 0.01$ for fixed-confidence experiments and $T = \lceil H_{2,\text{lin}} \rceil$ for fixed-budget ones.

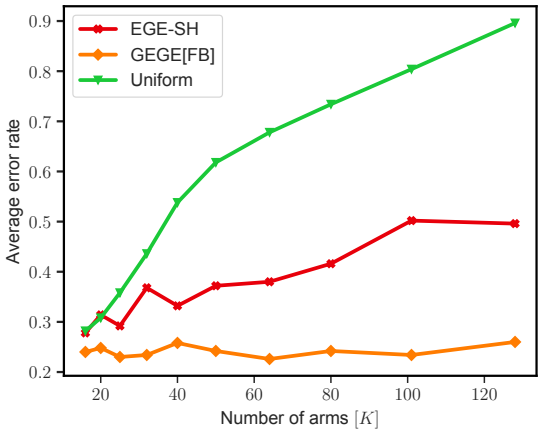


Figure 4.1: Average misidentification rate vs. K (synthetic data).

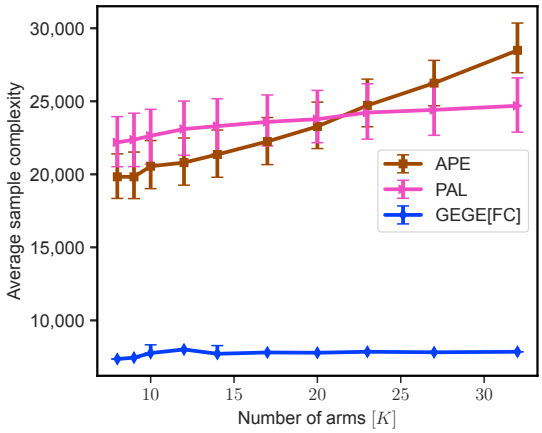


Figure 4.2: Average sample complexity vs. K (synthetic data).

Summary. Theoretical results (Theorems 4.3.1 and 4.3.3) predict that the sample complexity of GEGE depends on the number of arms K only through logarithmic factors. This trend is confirmed empirically: in Figure 4.1, the misidentification rate of GEGE-FB remains almost constant as K increases, whereas that of EGE-SH/SR rises sharply. Similarly, in Figure 4.2, the sample complexity of GEGE-FC shows only mild dependence on K, unlike the unstructured baseline.

On the real-world NoC experiment, GEGE consistently outperforms all competitors. In the fixed-confidence regime (Figure 4.4), it requires substantially fewer samples to achieve δ-correct identification than both APE and PAL. In the fixed-budget regime (Figure 4.3),

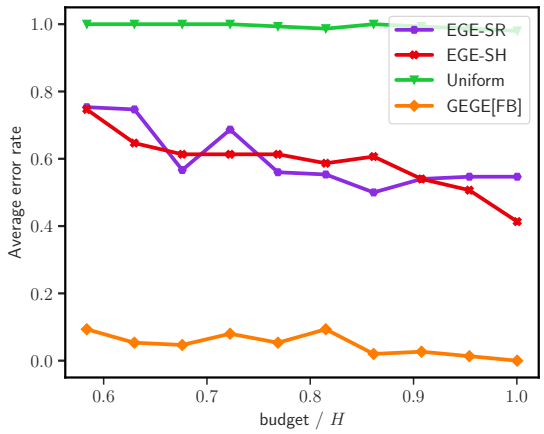


Figure 4.3: Average misidentification rate vs. T (NoC experiment).

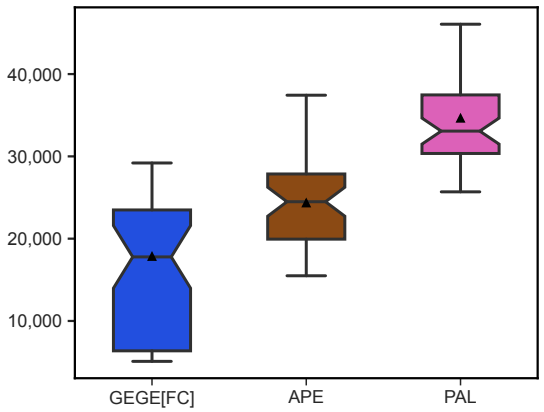


Figure 4.4: Empirical sample complexity (NoC experiment).

the probability of misidentification is reduced by up to 50% compared to EGE-SH. Notably, EGE-SH requires $T \geqslant K \log_2 K \approx 2000$ to operate on this instance, while GEGE-FB succeeds with $T \geqslant \lceil \log_2 h \rceil$. Each experiment runs in under 10 seconds for instances with up to $K = 500$ and $d = 8$.

Discussion. Overall, GEGE provides the first near-optimal algorithms for Pareto set identification in multi-output linear bandits. By leveraging optimal design principles, it efficiently estimates both mean responses and dominance gaps, yielding strong theoretical guarantees and consistent empirical performance across settings. In the fixed-budget regime, GEGE-FB achieves near-minimax guarantees; in the fixed-confidence regime, GEGE-FC remains the only algorithm with provably limited adaptivity.

While the complexity terms of GEGE depend only on the h smallest gaps, logarithmic dependencies in K remain due to standard union bounds. Following the empirical-process techniques of Katz-Samuels, Jain, et al. 2020, future work could explore whether these logarithmic factors can be eliminated. Finally, extending the optimal design approach to high-dimensional or kernelized feature spaces–along the lines of Camilleri, Jamieson, et al. 2021—would be a promising direction for PSI in rich function classes.

4.5 Additional proofs

In the subsequent sections, r will always denote a round of GEGE, which should be clear from the context. We then denote by $\mathcal{A}_r$ active arms at round r and by $\hat{\theta}_r$ the empirical estimate of θ at round r, computed by a call to Algorithm 4.1. By convention we let $\max_\emptyset = -\infty$. For any suboptimal arm i, there exists a Pareto-optimal arm $i^\star$ (not necessarily unique) such that $\Delta_i = \mathrm{m}(i, i^\star)$. More generally given a suboptimal i we denote by $i^\star$ any arm of $\operatorname{argmax}_{j \in \mathcal{S}^\star} \mathrm{m}(i, j)$. At a round r, we let

$$\mathcal{P}_r := \{\forall\, s \in \{1, \dots, r\},\ \forall\, i \in \mathcal{A}_s, i \in (\mathcal{S}^\star)^\mathsf{c} \Rightarrow i^\star \in \mathcal{A}_s\}\,, \tag{4.7}$$

with $\mathcal{P} = \mathcal{P}_\infty$.

In particular, for a suboptimal arm i, $i^\star \in \mathcal{A}_s$ should be understood as

$$\mathcal{A}_s \cap (\operatorname*{argmax}_{j \in \mathcal{S}^\star} \mathrm{m}(i,j)) \neq \emptyset$$

. If for some τ, $\mathcal{P}_\tau$ is true and $\mathcal{A}_{\tau+1} = \emptyset$ then we say that $\mathcal{P}$ holds.

4.5.1 Empirical gaps: Proof of Proposition 4.3.6

In both the fixed-budget and fixed-confidence setting, the proof proceeds by induction on the round r.

To establish this first result, we need the following intermediate lemmas, proved at the end of the section.

Lemma 4.5.1. *At any round r and for any arms $i, j \in \mathcal{A}_r$ it holds that*

$$\begin{aligned} \left|\, \mathrm{M}(i,j;r) - \mathrm{M}(i,j) \right| &\leqslant \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty \text{ and} \\ \left|\, \mathrm{m}(i,j;r) - \mathrm{m}(i,j) \right| &\leqslant \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty. \end{aligned}$$

Lemma 4.5.2. *At any round r, for any suboptimal arm $i \in \mathcal{A}_r$, if $i^\star \in \mathcal{A}_r$ and $i^\star$ does not empirically dominate i then $\Delta_i^\star < \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_{i^\star})\|_\infty$.*

Deviations of the gaps when $\mathcal{P}_r$ holds. In this part, we control the deviations of the empirical gaps when proposition $\mathcal{P}_r$ holds.

Lemma 4.5.3. *Assume that the proposition $\mathcal{P}_r$ holds at some round r. Then for any arm $i \in \mathcal{A}_r$ it holds that*

$$\left| (\widehat{\Delta}_{i,r}^\star)_+ - (\Delta_i^\star)_+ \right| \leqslant \left| \widehat{\Delta}_{i,r}^\star - \Delta_i^\star \right| \leqslant \gamma_{i,r}$$

where $\gamma_{i,r} := \max_{j \in \mathcal{A}_r} \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty$.

Proof. This inequality is a direct consequence of Lemma 4.5.1 and the relation $|x_+ - y_+| \leqslant |x - y|$) which holds for any $x, y \in \mathbb{R}$. Note that for a Pareto-optimal arm i we trivially have $(\Delta_i^\star)^+ = 0 = (\max_{j \in \mathcal{A}_r} \mathrm{m}(i,j))_+$. And for a suboptimal arm $i \in \mathcal{A}_r$, as $i^\star \in \mathcal{A}_r$ (from proposition $\mathcal{P}_r$) we have $\Delta_i^\star = \mathrm{m}(i,i^\star) = \max_{j \in \mathcal{A}_r} \mathrm{m}(i,j)$. Thus for any arm $i \in \mathcal{A}_r$ we have

$$\begin{aligned} \left| (\widehat{\Delta}_{i,r}^\star)_+ - (\Delta_i^\star)_+ \right| &= \left| (\max_{j \in \mathcal{A}_r} \mathrm{m}(i,j;r))_+ - (\max_{j \in \mathcal{A}_r} \mathrm{m}(i,j))_+ \right|, \\ &\leqslant \left| (\max_{j \in \mathcal{A}_r} \mathrm{m}(i,j;r)) - (\max_{j \in \mathcal{A}_r} \mathrm{m}(i,j)) \right|, \\ &\leqslant \max_{j \in \mathcal{A}_r} \left| \mathrm{m}(i,j;r) - \mathrm{m}(i,j) \right|, \\ &\leqslant \max_{j \in \mathcal{A}_r} \left\| (\hat{\theta}_r - \theta)^\intercal (x_i - x_j) \right\|_\infty = \gamma_{i,r}, \end{aligned}$$

where the last inequality follows from Lemma 4.5.1. □

Lemma 4.5.4. *If the proposition $\mathcal{P}_r$ holds at some round r then for any arm $i \in \mathcal{A}_r$,*

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -2\gamma_r & \text{if } i \in \mathcal{S}^\star, \\ -\gamma_{i,r} & \text{else,} \end{cases}$$

where $\gamma_{i,r} \coloneqq \max_{j\in\mathcal{A}_r} \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty$ *and* $\gamma_r \coloneqq \max_{i\in\mathcal{A}_r} \gamma_{i,r}$.

Proof. We first prove the result a suboptimal arm i. From the proposition $\mathcal{P}_r$, as $i \in \mathcal{A}_r$ we have $i^\star \in \mathcal{A}_r$ so $\Delta_i = \max_{j\in\mathcal{A}_r} \mathrm{m}(i,j)$ and we recall that

$$\widehat{\Delta}_{i,r} \coloneqq \max\left(\widehat{\Delta}^\star_{i,r}, \widehat{\delta}^\star_{i,r}\right). \tag{4.8}$$

Note that by reverse triangle we have for any arm $i \in \mathcal{A}_r$ (suboptimal or not)

$$\begin{aligned} \left|\left(\max_{j\in\mathcal{A}_r} \mathrm{m}(i,j;r)\right) - \left(\max_{j\in\mathcal{A}_r} \mathrm{m}(i,j)\right)\right| &\leqslant \max_{j\in\mathcal{A}_r} |\,\mathrm{m}(i,j;r) - \mathrm{m}(i,j)|, &(4.9)\\ &\leqslant \max_{j\in\mathcal{A}_r} \left\|(\hat{\theta}_r - \theta)^T (x_i - x_j)\right\|_\infty = \gamma_{i,r}. &(4.10)\end{aligned}$$

where the last inequality follows from Lemma 4.5.1. If i a suboptimal arm ($i \notin \mathcal{S}^\star$) then as $\Delta_i = \Delta^\star_i$, it follows

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \widehat{\Delta}^\star_{i,r} - \Delta^\star_i$$

therefore

$$\begin{aligned} \widehat{\Delta}_{i,r} - \Delta_i &\geqslant -|\widehat{\Delta}^\star_{i,r} - \Delta^\star_i| \\ &= -|(\max_{j\in\mathcal{A}_r} \mathrm{m}(i,j;r)) - (\max_{j\in\mathcal{A}_r} \mathrm{m}(i,j))| \\ &\geqslant -\gamma_{i,r} \quad \text{(see (4.10)).} \end{aligned}$$

Now we assume i is a Pareto-optimal arm ($i \in \mathcal{S}^\star$) so that

$$\Delta_i = \delta^\star_i.$$

Combining with Eq. (4.8) yields

$$\widehat{\Delta}_{i,r} - \Delta_{i,r} \geqslant \widehat{\delta}^\star_{i,r} - \delta^\star_i,$$

where we recall that

$$\widehat{\delta}^\star_{i,r} = \min_{j\in\mathcal{A}_r\setminus\{i\}} [\mathrm{M}(i,j;r) \ \wedge \ (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+)]$$

and

$$\delta^\star_i \coloneqq \min_{j\in[K]\setminus\{i\}} [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)].$$

As for any $x, y \in \mathbb{R}$ we have $|x^+ - y^+| \leqslant |x - y|$, the following holds for any $i, j \in \mathcal{A}_r$

$$\begin{aligned} |\,\mathrm{M}(j,i;r)^+ - \mathrm{M}(j,i)^+| &\leqslant |\,\mathrm{M}(j,i;r) - \mathrm{M}(j,i)| && (4.11)\\ &\leqslant \gamma_{j,r}. && (4.12)\end{aligned}$$

From Lemma 4.5.3 we have for any $j \in \mathcal{A}_r$

$$(\widehat{\Delta}^\star_{j,r})_+ - (\Delta^\star_j)_+ \geqslant -\gamma_{j,r}. \tag{4.13}$$

Combining (4.12) and (4.13) yields for any $j \in \mathcal{A}_r$

$$\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+ \geqslant \mathrm{M}(j,i)_+ + (\Delta^\star_j)_+ - 2\gamma_{j,r}, \tag{4.14}$$

which in addition to $\mathrm{M}(j,i;r) \geqslant \mathrm{M}(j,i) - \gamma_{j,r}$ yields

$$[\mathrm{M}(i,j;r) \wedge (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+)] \geqslant [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)] - 2\gamma_{j,r}\,,$$

for any arm $j \in \mathcal{A}_r$. Thus taking the $\min$ over $i \in \mathcal{A}_r$ yields

$$\begin{aligned} \widehat{\delta}^\star_{i,r} &= \min_{j \in \mathcal{A}_r \setminus \{i\}} [\mathrm{M}(i,j;r) \wedge (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+)] \\ &\geqslant \min_{j \in \mathcal{A}_r \setminus \{i\}} [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)] - 2\gamma_r, \\ &\geqslant \min_{j \in [K] \setminus \{i\}} [\mathrm{M}(i,j) \wedge (\mathrm{M}(j,i)_+ + (\Delta^\star_j)_+)] - 2\gamma_r, \\ &= \delta^\star_i - 2\gamma_r \end{aligned}$$

which concludes the proof of this lemma. □

Building on this result, we show that $\mathcal{P}_\infty$ holds in the fixed-confidence and fixed-budget settings.

Fixed-budget setting. We recall the definition of the good event for any $\lambda > 0$.

$$\mathcal{E}^{r,\lambda}_{\mathrm{fb}} = \left\{ \forall\, i, j \in \mathcal{A}_r : \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty \leqslant \lambda \Delta_{n_{r+1}+1} \right\}$$

and $\mathcal{E}^\lambda_{\mathrm{fb}} := \cap_{r=1}^{\lceil \log_2(h) \rceil} \mathcal{E}^{r,\lambda}_{\mathrm{fb}}$. We prove that proposition $\mathcal{P}_\infty$ holds on the event $\mathcal{E}^\lambda_{\mathrm{fb}}$ for some any $\lambda \in (0, 1/5)$.

Lemma 4.5.5. *The proposition holds $\mathcal{P}_\infty$ on the event $\mathcal{E}^\lambda_{\mathrm{fb}}$ for any $\lambda \in (0, 1/5)$: at any round $r \in \{1, \dots, \lceil \log_2 h \rceil + 1\}$ and for any arm $i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^{\mathsf{c}}$, $i^\star \in \mathcal{A}_r$.*

Proof. We prove $\mathcal{P}_\infty$ by induction on the round r. In the sequel we assume $\mathcal{E}^\lambda_{\mathrm{fb}}$ holds. We also assume $\mathcal{P}_r$ is true until some round r. As $\mathcal{E}^\lambda_{\mathrm{fb}}$ holds, we have by application of Lemma 4.5.4: for any arm $i \in \mathcal{A}_r$,

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -2\lambda \Delta_{n_{r+1}+1} & \text{if } i \in \mathcal{S}^\star \\ -\lambda \Delta_{n_{r+1}+1} & \text{else.} \end{cases} \tag{4.15}$$

We shall prove that if a Pareto-optimal arm i is discarded at the end of round r then there exists no arm suboptimal $j \in \mathcal{A}_{r+1}$ such that $j^\star = i$. Since i is removed and $|\mathcal{A}_{r+1}| = n_{r+1}$ there exists $k_r \in \mathcal{A}_{r+1} \cup \{i\}$ such that

$$\Delta_{k_r} \geqslant \Delta_{n_{r+1}+1}. \tag{4.16}$$

If i is empirically suboptimal then as it is discarded we have

$$\widehat{\Delta}_{i,r} = \widehat{\Delta}^\star_{i,r} \geqslant \widehat{\Delta}_{k,r}$$

for any arm $k \in \mathcal{A}_{r+1}$. So $\widehat{\Delta}^\star_{i,r} \geqslant \widehat{\Delta}_{k_r,r}$ thus using (4.15) and (4.16) it comes that

$$\begin{aligned} \max_{q \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,q) &\geqslant \Delta_{n_{r+1}+1} - 3\lambda\Delta_{n_{r+1}+1} \\ &= (1-3\lambda)\Delta_{n_{r+1}+1} \end{aligned}$$

and the latter inequality is not possible for $\lambda < 1/3$ as the LHS of the inequality is negative as i is a Pareto-optimal arm.

Next we assume that i is empirically optimal. We claim that j is not dominated by i. To see this, first note that as $j \in \mathcal{A}_{r+1}$ we have

$$\widehat{\Delta}_{i,r} \geqslant \widehat{\Delta}_{j,r} \tag{4.17}$$

so that as i is empirically optimal, if j was empirically dominated by i we would have

$$\widehat{\Delta}_{i,r} \leqslant \mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+ = \widehat{\Delta}_{j,r}. \tag{4.18}$$

Combining (4.17) and (4.18) yields $\widehat{\Delta}_{i,r} = \widehat{\Delta}_{j,r}$, i is empirically optimal and j is empirically suboptimal. However, our breaking rule ensures that this case cannot occur. Therefore, j is not dominated by i. But, by assumption, j is such that $j^\star = i$ and we have proved that i does not empirically dominate j so by Lemma 4.5.2

$$\Delta_j \leqslant \|(\hat{\theta}_r - \theta)^\intercal (x_j - x_i)\|_\infty$$

which on the event $\mathcal{E}_{\mathrm{fb}}$ yields

$$\Delta_j \leqslant \lambda\Delta_{n_{r+1}+1}. \tag{4.19}$$

On the other side, as i is discarded as an empirically optimal arm we have

$$\widehat{\Delta}_{i,r} = \widehat{\delta}^\star_{i,r} \geqslant \widehat{\Delta}_{k,r}$$

for any arm $k \in \mathcal{A}_{r+1}$. Since $k_r \in \mathcal{A}_{r+1} \cup \{i\}$ it comes $\widehat{\delta}^\star_{i,r} \geqslant \widehat{\Delta}_{k_r,r}$ thus using (4.15) and (4.16) yields

$$\mathrm{M}(j,i)_+ + \Delta_j \geqslant \Delta_{n_{r+1}+1} - 4\lambda\Delta_{n_{r+1}+1}$$

which further combined with (4.19) yields

$$\mathrm{M}(j,i)_+ \geqslant (1-5\lambda)\Delta_{n_{r+1}+1}.$$

However, as $j^\star = i$ we have $\mathrm{M}(j,i)_+ = 0$ so the latter inequality is not possible as long as $\lambda < 1/5$. Combining these, we have proved that if $\mathcal{P}_r$ holds, then for any Pareto-optimal arm i which is removed at the end of round r, there does not exist an arm $j \in \mathcal{A}_{r+1}$ such that $j^\star = i$. So $\mathcal{P}_{r+1}$ holds. Finally noting that $\mathcal{P}_r$ trivially holds for $r = 1$ we conclude that $\mathcal{P}_\infty$ holds on the event $\mathcal{E}_{\mathrm{fb}}^\lambda$ for any $\lambda < 1/5$. □

Combining this result with Lemma 4.5.4 and assuming $\mathcal{E}_{\mathrm{fb}}^\lambda$ holds then yields at any round $r \in \{1, \dots, \lceil \log_2 h \rceil\}$ and for any arm $i \in \mathcal{A}_r$:

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -2\lambda\Delta_{n_{r+1}+1} & \text{if } i \in \mathcal{S}^\star \\ -\lambda\Delta_{n_{r+1}+1} & \text{else,} \end{cases} \tag{4.20}$$

which proves Proposition 4.3.6 in the fixed-budget setting.

Fixed-confidence setting. We recall below the good events we study in the fixed-confidence setting:

$$\mathcal{E}_{\mathrm{fc}}^r = \left\{ \forall\, i,j \in \mathcal{A}_r : \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty \leqslant \varepsilon_r/2 \right\}$$

and $\mathcal{E}_{\mathrm{fc}} \coloneqq \cap_{r=1}^\infty \mathcal{E}_{\mathrm{fc}}^r$.

Lemma 4.5.6. *The proposition $\mathcal{P}_\infty$ holds on the event $\mathcal{E}_{fc}$: at any round r for any arm $i \in \mathcal{A}_r \cap (\mathcal{S}^\star)^\mathsf{c}$, $i^\star \in \mathcal{A}_r$.*

Proof. We prove the proposition by induction on the round r. Note that the proposition $\mathcal{P}_r$ trivially holds for $r = 1$. Assume the property holds until the beginning of some round r. Let $i \in \mathcal{S}^\star$ be an optimal arm and assume i is discarded at the end of round r. We will prove that there exists no suboptimal arm $j \in \mathcal{A}_{r+1}$ such that $j^\star = i$. Recall that when i is discarded, we have either $i \in S_r$ (empirically optimal) or $i \notin S_r$ (empirically suboptimal). We analyze both cases below. If $i \notin S_r$ then it holds that

$$\widehat{\Delta}_{i,r} \geqslant \varepsilon_r/2,$$

then, as $i \notin S_r$ it follows that $\widehat{\Delta}_{i,r} = \widehat{\Delta}_i^\star \coloneqq \max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j;r)$, so

$$\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j;r) \geqslant \varepsilon_r/2$$

which using Lemma 4.5.1 and assuming event $\mathcal{E}_{\mathrm{fc}}^r$ holds would yield

$$\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j) > 0.$$

The latter inequality is not possible as $i \in \mathcal{S}^\star$ is a Pareto-optimal arm. Therefore, on $\mathcal{E}_{\text{fc}}^r$, when $i \in \mathcal{S}^\star$ is discarded we have $i \in S_r$.

Next, we analyze the case $i \in S_r$: that is i is discarded and classified as optimal. In this case it follows from the definition of $\widehat{\Delta}_{i,r}$ that

$$\min_{j \in \mathcal{A}_r \setminus \{i\}} [\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+] \geqslant \varepsilon_r. \tag{4.21}$$

Let $j \in \mathcal{A}_{r+1} \cap (\mathcal{S}^\star)^c$ be such that $j^\star = i$. If j is empirically optimal then $(\widehat{\Delta}^\star_{j,r})_+ = 0$ thus $\mathrm{M}(j,i;r)_+ \geqslant \varepsilon_r$. On the contrary, if j is empirically suboptimal, then because it has not been removed at the end of round r it holds that

$$\widehat{\Delta}^\star_{j,r} < \varepsilon_r/2,$$

which combined with (4.21) yields $\mathrm{M}(j,i;r)_+ > \varepsilon_r/2$. Thus, in both cases we have $\mathrm{M}(j,i;r)_+ > \varepsilon_r/2$ which using Lemma 4.5.1 and assuming event $\mathcal{E}_{\text{fc}}^r$ would imply that

$$\mathrm{M}(j,i)_+ > 0,$$

which is impossible as, by assumption $j^\star = i$, so j is dominated by i.

Together, these imply that on $\mathcal{E}_{\text{fc}}$, if $\mathcal{P}_r$ holds then $\mathcal{P}_{r+1}$ holds. Since the property trivially holds for $r = 1$ we have proved that the property $\mathcal{P}_r$ holds at any round when $\mathcal{E}_{\text{fc}}$ holds. □

Combining this result with Lemma 4.5.4 proves that, on the event $\mathcal{E}_{\text{fc}}$, for any round r and for any arm $i \in \mathcal{A}_r$

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -\varepsilon_r & \text{if } i \in \mathcal{S}^\star \\ -\varepsilon_r/2 & \text{else,} \end{cases} \tag{4.22}$$

which proves Proposition 4.3.6 in the fixed-confidence setting.

4.5.2 Probability of error: Proof of Theorem 4.3.1

Proof of Theorem 4.3.1. We first prove the correctness of GEGE on the event $\mathcal{E}_{\text{fb}}^\lambda$ for some λ small enough. Let us assume $\mathcal{E}_{\text{fb}}^\lambda$ holds which by Proposition 4.3.6 implies that $\mathcal{P}_\infty$ holds and at round r, we have for any arm $i \in \mathcal{A}_r$

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -2\lambda\Delta_{n_{r+1}+1} & \text{if } i \in \mathcal{S}^\star \\ -\lambda\Delta_{n_{r+1}+1} & \text{else.} \end{cases} \tag{4.23}$$

We recall the definition of the good event for any $\lambda > 0$,

$$\mathcal{E}_{\text{fb}}^{r,\lambda} = \left\{ \forall\, i,j \in \mathcal{A}_r : \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty \leqslant \lambda\Delta_{n_{r+1}+1} \right\}$$

and $\mathcal{E}_{\text{fb}}^{\lambda} := \cap_{r=1}^{\lceil \log_2(h) \rceil} \mathcal{E}_{\text{fb}}^{r,\lambda}$. Applying Lemma 4.5.1 on this event then yields for all arms $i, j \in \mathcal{A}_r$,

$$|\,\mathrm{M}(i,j;r) - \mathrm{M}(i,j)| \leqslant \lambda \Delta_{n_{r+1}+1} \text{ and} \tag{4.24}$$

$$|\,\mathrm{m}(i,j;r) - \mathrm{m}(i,j)| \leqslant \lambda \Delta_{n_{r+1}+1}. \tag{4.25}$$

Let i be an arm discarded at the end of round r. Since i is discarded and $|\mathcal{A}_{r+1}| = n_{r+1}$ there exists $k_r \in \mathcal{A}_{r+1} \cup \{i\}$ such that

$$\Delta_{k_r} \geqslant \Delta_{n_{r+1}+1}. \tag{4.26}$$

If $i \notin S_r$ that is i is empirically suboptimal then

$$\widehat{\Delta}_{i,r} = \widehat{\Delta}_{i,r}^{\star} \geqslant \widehat{\Delta}_{k_r,r},$$

then, recalling that

$$\widehat{\Delta}_{i,r}^{\star} := \max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j;r)$$

and further applying (4.23) to k_r and using (4.25) yields

$$\max_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{m}(i,j) \geqslant (1 - 3\lambda)\Delta_{n_{r+1}+1}$$

which for $\lambda < 1/3$ implies that $\max_{j \in \mathcal{A}_r} \mathrm{m}(i,j) > 0$, that is there exists $j \in \mathcal{A}_r$ such that $\mu_i \prec \mu_j$ so i is a suboptimal arm. Next, assume $i \in S_r$ (*i.e.*, i is empirically Pareto-optimal). In this case we have $\widehat{\Delta}_{i,r} = \widehat{\delta}_{i,r}^{\star} \geqslant \widehat{\Delta}_{k_r,r}$.

We recall that

$$\widehat{\delta}_{i,r}^{\star} = \min_{j \in \mathcal{A}_r \setminus \{i\}} [\mathrm{M}(i,j;r) \wedge (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}_{i,r}^{\star})_+)].$$

Applying (4.23) to k_r and using (4.24), it follows that

$$\min_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{M}(i,j) \geqslant (1 - 3\lambda)\Delta_{n_{r+1}+1}.$$

Thus, for $\lambda < 1/3$, we have $\min_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{M}(i,j) > 0$. Therefore, no active arm at round r dominates i (based on their true means), which, together with proposition $\mathcal{P}_\infty$, yields that i is a Pareto-optimal arm (otherwise, we would have $i^\star \in \mathcal{A}_r$ that dominates i).

Combining these, we have proved that for any $\lambda < 1/5$ (we need $\lambda < 1/5$ for $\mathcal{P}_\infty$ to hold), Algorithm 4.3 does not make any error on the event $\mathcal{E}_{\text{fb}}^{\lambda}$. It then follows that the probability of error of GEGE is at most

$$\inf_{\lambda \in (0,1/5)} \mathbb{P}\left((\mathcal{E}_{\text{fb}}^{\lambda})^c\right) \tag{4.27}$$

Now we upper-bound Eq. (4.27), which will conclude the proof. Let $\lambda \in (0, 1/5)$ be fixed. We have by union bound

$$\begin{aligned}\mathbb{P}\left((\mathcal{E}_{\text{fb}}^{\lambda})^{c}\right) &\leqslant \sum_{r=1}^{\lceil \log_2 h \rceil} \mathbb{E}\left[\mathbb{P}\left((\mathcal{E}_{\text{fb}}^{r,\lambda})^{c} | \mathcal{A}_r\right)\right] \\ &\leqslant \sum_{r=1}^{\lceil \log_2 h \rceil} \mathbb{E}\left[\sum_{i \in \mathcal{A}_r} \mathbb{P}(\|(\widehat{\theta}_r - \theta)^{\intercal} x_i\|_{\infty} > \frac{1}{2}\lambda \Delta_{n_{r+1}+1} | \mathcal{A}_r)\right]\end{aligned}$$

Note that for i fixed, we can use Lemma 4.2.2 with $\kappa = 1/3$ and the conditions of this theorem are satisfied as the budget per phase is $T/\log_2(h) \geqslant 45h$ (recall from the theorem that GEGE is run with $T \geqslant 45h\log_2(h)$). Thus, applying this theorem yields

$$\begin{aligned}\mathbb{P}\left((\mathcal{E}_{\text{fb}}^{\lambda})^{\mathsf{c}}\right) &\leqslant 2d \sum_{r=1}^{\lceil \log_2 h \rceil} n_r \mathbb{E}\left[\exp\left(-\frac{\lambda^2 \Delta_{n_{r+1}+1}^2 T}{24\sigma^2 h_r \log_2 h\rceil}\right)\right] \\ &\leqslant 2d \sum_{r=1}^{\lceil \log_2 h \rceil} n_r \exp\left(-\frac{\lambda^2 T \Delta_{n_{r+1}+1}^2}{24\sigma^2 \min(h, n_r)\lceil \log_2 h\rceil}\right), \quad \text{as } h_r \leqslant \min(n_r, h).\end{aligned}$$

Then, note that

$$\begin{aligned}\frac{\Delta_{n_{r+1}+1}^2}{\min(h, n_r)} &= \frac{\Delta_{\lceil h/2^r \rceil + 1}^2}{\lceil h/2^{r-1} \rceil} \\ &= \frac{\Delta_{\lceil h/2^r \rceil + 1}^2}{\lceil h/2^r \rceil + 1} \frac{\lceil h/2^r \rceil + 1}{\lceil h/2^{r-1} \rceil} \\ &\geqslant \frac{\Delta_{\lceil h/2^r \rceil + 1}^2}{\lceil h/2^r \rceil + 1} \frac{h/2^r + 1}{h/2^{r-1} + 1} \\ &\geqslant \frac{\Delta_{\lceil h/2^r \rceil + 1}^2}{\lceil h/2^r \rceil + 1} \frac{1}{2},\end{aligned}$$

which follows as $(x+1)/(2x+1) = (1/2) + (1/2)/(2x+1) \geqslant \frac{1}{2}$ for $x \geqslant 0$. Therefore,

$$\begin{aligned}\frac{\Delta_{n_{r+1}+1}^2}{\min(h, n_r)} &\geqslant \frac{1}{2} \frac{\Delta_{\lceil h/2^r \rceil + 1}^2}{\lceil h/2^r \rceil + 1} \\ &\geqslant \frac{1}{2H_{2,\text{lin}}}.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}\left((\mathcal{E}_{\text{fb}}^{\lambda})^{c}\right) &\leqslant 2\exp\left(-\frac{\lambda^2 T}{48\sigma^2 H_{2,\text{lin}} \lceil \log_2 h \rceil} + \log(d)\right) \sum_{r=1}^{\lceil \log_2 h \rceil} n_r \\ &\leqslant 2\left(K + h + \lceil \log_2 h \rceil\right) \exp\left(-\frac{\lambda^2 T}{48\sigma^2 H_{2,\text{lin}} \lceil \log_2 h \rceil} + \log(d)\right)\end{aligned}$$

Finally, it follows that

$$\inf_{\lambda \in (0,1/5)} \mathbb{P}\left((\mathcal{E}_{\text{fb}}^{\lambda})^{c}\right) \leqslant 2\left(K + h + \lceil \log_2 h \rceil\right) \exp\left(-\frac{T}{1200\sigma^2 H_{2,\text{lin}} \lceil \log_2 h \rceil} + \log(d)\right),$$

which concludes the proof. □

4.5.3 Sample complexity: Proof of Theorem 4.3.3

We prove the theoretical guarantees in the fixed-confidence setting. We prove the correctness of Algorithm 4.2 and we prove the sample complexity bound of Theorem 4.3.3 and some key lemmas. We first prove the correctness of the fixed-confidence variant of GEGE.

Proof of the correctness. We need to prove that the final recommendation of Algorithm 4.2 is correct: that is we should show that: at any round r, $\mathcal{B}_r \subset \mathcal{S}^\star$ and $\mathcal{D}_r \subset (\mathcal{S}^\star)^\mathsf{c}$.

Lemma 4.5.7. *On the event $\mathcal{E}_{\text{fc}}$, Algorithm 4.2 identifies the correct Pareto set.*

Proof of Lemma 4.5.7. In this part let τ denotes the stopping time of Algorithm 4.2. We assume $\mathcal{E}_{\text{fc}}$ holds.

Using Proposition 4.3.6: for any round $r \leqslant \tau$ for any (Pareto) suboptimal $i \in \mathcal{A}_r$ we have $i^\star \in \mathcal{A}_r$. We then prove the correctness of the algorithm as follows. Let i be an arm that is removed at the end of some round r. Assume $i \in S_r$ then, as i is discarded and empirically optimal we have $\widehat{\Delta}_{i,r} = \widehat{\delta}_i^\star \geqslant \varepsilon_r$. In particular, it holds that

$$\min_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{M}(i, j; r) \geqslant \varepsilon_r$$

which using Lemma 4.5.1 on the event $\mathcal{E}_{\text{fc}}$ yields

$$\min_{j \in \mathcal{A}_r \setminus \{i\}} \mathrm{M}(i, j) > \varepsilon_r/2 > 0,$$

that is no active arm dominates i. Combined with proposition $\mathcal{P}_\infty$ (cf Lemma 4.5.6) the latter inequality yields $i \in \mathcal{S}^\star$. Now assume we have $i \notin S_r$: i is discarded and it is empirically suboptimal. Then

$$\widehat{\Delta}_{i,r} = \max_{j \in \mathcal{A}_r} \mathrm{m}(i, j; r) \geqslant \varepsilon_r/2,$$

so using Lemma 4.5.1 again on event $\mathcal{E}_{\text{fc}}$ it follows that there exists $j \in \mathcal{A}_r$ such that $\mathrm{m}(i, j) > 0$: that is $i \notin \mathcal{S}^\star$. In all cases, we have proved that if $\mathcal{E}_{\text{fc}}$ holds then for any arm i discarded at some round r,

$$i \in \mathcal{B}_{r+1} \iff i \in \mathcal{S}^\star.$$

Note that if $\mathcal{A}_\tau$ is non-empty, then it contains a single arm and because $\mathcal{P}_\infty$ holds, this arm is also Pareto optimal. □

Thus, Algorithm 4.2 is correct on $\mathcal{E}_{\text{fc}}$. Before proving Theorem 4.3.3 we need Lemma 4.3.7 to control the size of the active set $\mathcal{A}_r$ in the fixed-confidence setting.

Size of the active set. We prove the following result that controls the size of the active set.

Proof of Lemma 4.3.7. By Lemma 4.5.6 we have on the event $\mathcal{E}_{\text{fc}}$: for any round r and for any arm $i \in \mathcal{A}_r$,

$$\widehat{\Delta}_{i,r} - \Delta_i \geqslant \begin{cases} -\varepsilon_r & \text{if } i \in \mathcal{S}^\star \\ -\varepsilon_r/2 & \text{else.} \end{cases}$$

Then let $p \in [K]$ and let us assume an arm $i \in \{p, \dots, K\}$ is still active at round $r = \lceil \log_2(1/\Delta_p) \rceil$. We have $\widehat{\Delta}_{i,r} \geqslant \Delta_i - \varepsilon_r$ with $\varepsilon_r = 1/2^{r+1}$ and $\Delta_i \geqslant \Delta_p$ which combined with $\widehat{\Delta}_{i,r} \geqslant \Delta_i - \varepsilon_r$ yields

$$\widehat{\Delta}_{i,r} \geqslant \Delta_p - \varepsilon_r. \tag{4.28}$$

As $r = \lceil \log_2(1/\Delta_p) \rceil$, it holds that $2\varepsilon_r \leqslant \Delta_p$ so Eq. (4.28) yields $\widehat{\Delta}_{i,r} \geqslant \varepsilon_r$ thus i will be discarded at the end of round r that is any arm $i \in \{p, \dots, K\}$ will be discarded at the end of round $\lceil \log_2(1/\Delta_p) \rceil$. □

We now prove the main lemma on the sample complexity of GEGE in the fixed-confidence setting.

Bound on the sample complexity of Theorem 4.3.3. We provide an upper bound on the sample complexity of the algorithm.

Proof. We assume $\mathcal{E}_{\text{fc}}$ holds. The correctness of Algorithm 4.2 is then proven in Lemma 4.5.7 and Lemma 4.3.7 upper-bounds the number of rounds before termination. It remains to bound the sample complexity of the algorithm on $\mathcal{E}_{\text{fc}}$ and compute $\mathbb{P}(\mathcal{E}_{\text{fc}})$ to conclude.

By Lemma 4.3.7 an upper-bound on $|\mathcal{A}_r|$ for some specific rounds. Interestingly we can bound the sample complexity between consecutive "checkpoints rounds". In what follows, we rewrite the complexity as a sum of number of pulls between these intermediate "checkpoints rounds". Let us introduce the sequence $\{\alpha_s : s \geqslant 0\}$ defined as $\alpha_0 = 0$ and for any $s \geqslant 1$, $\alpha_s = \lceil \log_2(1/\Delta_{\lfloor h/2^s \rfloor}) \rceil$. We assume *w.l.o.g* that the sequence is non-decreasing and that the gaps are bounded in $(0, 1)$ (otherwise, we could start the sequence $(\alpha)_s$ from arms with gap smaller than 1). Simple calculation shows that $\alpha_{\lfloor \log_2(h) \rfloor} = \lceil \log_2(1/\Delta_1) \rceil$ and

$$\{1, \dots, \lceil \log_2(1/\Delta_1) \rceil\} = \bigcup_{s=1}^{\lfloor \log_2(h) \rfloor} [\![1 + \alpha_{s-1}, \alpha_s]\!]. \tag{4.29}$$

Introducing

$$T_r \coloneqq \frac{32(1+3\varepsilon_r)\sigma^2 h_r}{\varepsilon_r^2} \log\left(\frac{dn_r}{\delta_r}\right),$$

where $n_r = |\mathcal{A}_r|$, we have $t_r = \lceil T_r \rceil$, so $t_r \leqslant T_r + 1$. Using (4.29) then leads to

$$\begin{aligned} \sum_{r=1}^{\lceil \log_2(1/\Delta_1) \rceil} T_r &= \sum_{s=0}^{\lfloor \log_2(h) \rfloor - 1} \sum_{r=\alpha_s+1}^{\alpha_{s+1}} T_r \\ &=: \sum_{s=0}^{\lfloor \log_2(h) \rfloor - 1} N_s \end{aligned}$$

where $N_s = \sum_{r=\alpha_s+1}^{\alpha_{s+1}} T_r$ is "the number of arms pulls" between round $(\alpha_s + 1)$ and α_{s+1}.

Next we bound the term N_s for $s \in \{0, \dots, \lfloor \log_2(h) \rfloor - 1\}$. We recall that $h_r \leqslant \min(h, n_r)$ as, $n_r = |\mathcal{A}_r|$ is the number of active arms at round r and h_r is the dimension of the space spanned by the features of the active arms. Using Lemma 4.3.7 on $\mathcal{E}_{\text{fc}}$, it holds that for $r \geqslant \alpha_s + 1$

$$n_r \leqslant \begin{cases} K & \text{if } s = 0 \\ \lfloor h/2^s \rfloor - 1 & \text{if } s \geqslant 1 \end{cases} \tag{4.30}$$

Therefore for $s \in \{0, \dots, \lfloor \log_2(h) \rfloor - 1\}$ and for any $r \geqslant \alpha_s + 1$, we simply have $\min(h, n_r) \leqslant \lfloor h/2^s \rfloor$, so $h_r \leqslant \lfloor h/2^s \rfloor$ and even $h_r \leqslant \lfloor h/2^s \rfloor - 1$ if $s > 0$. In particular, it holds that

$$h_r \leqslant 2\lfloor h/2^{s+1} \rfloor \quad \text{for } r \geqslant \alpha_s + 1.$$

It then follows that

$$\begin{aligned}
\widetilde{N}_s = N_s/(32(1+3\varepsilon_1)\sigma^2) &= \sum_{r=\alpha_s+1}^{\alpha_{s+1}} T_r/(32(1+3\varepsilon_1)\sigma^2) && (4.31)\\
&\leqslant 2\lfloor h/2^{s+1} \rfloor \log\left(\frac{Kd}{\delta_{\alpha_{s+1}}}\right) \sum_{r=\alpha_s+1}^{\alpha_{s+1}} \frac{1}{\varepsilon_r^2} && (4.32)\\
&= 8\lfloor h/2^{s+1} \rfloor \log\left(\frac{Kd}{\delta_{\alpha_{s+1}}}\right) \sum_{r=\alpha_s+1}^{\alpha_{s+1}} 4^r && (4.33)\\
&\leqslant 8\lfloor h/2^{s+1} \rfloor \log\left(\frac{Kd}{\delta_{\alpha_s+1}}\right) \sum_{r=1}^{\alpha_{s+1}} 4^r && (4.34)\\
&= \frac{32\lfloor h/2^{s+1} \rfloor}{3} \log\left(\frac{Kd}{\delta_{\alpha_{s+1}}}\right) (4^{\alpha_{s+1}} - 1) && (4.35)
\end{aligned}$$

then further using that

$$\alpha_s \geqslant \begin{cases} \log_2(1/\Delta_{\lfloor h/2^s \rfloor}) & \text{if } s \geqslant 1 \\ 0 & \text{if } s = 0 \end{cases}$$

yields

$$4^{\alpha_{s+1}} \leqslant \frac{1}{\Delta^2_{\lfloor h/2^{s+1} \rfloor}}$$

which combined with (4.35) yields

$$\widetilde{N}_s \leqslant \frac{32\sigma^2\lfloor h/2^{s+1} \rfloor}{3\Delta^2_{\lfloor h/2^{s+1} \rfloor}} \log\left(\frac{Kd}{\delta_{\alpha_{s+1}}}\right). \tag{4.36}$$

We can now bound $N = \sum_s N_s$ in terms of the suboptimality gaps:

$$\widetilde{N} = \sum_{s=0}^{\lfloor \log_2 h \rfloor - 1} \widetilde{N}_s \tag{4.37}$$

$$\leqslant \frac{32\sigma^2}{3} \sum_{s=0}^{\lfloor \log_2 h \rfloor - 1} \frac{\lfloor h/2^{s+1} \rfloor}{\Delta^2_{\lfloor h/2^{s+1} \rfloor}} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_{\lfloor h/2^{s+1} \rfloor}) \rceil^2}{6\delta} \right), \tag{4.38}$$

$$= \frac{32\sigma^2}{3} \sum_{s=1}^{\lfloor \log_2 h \rfloor} \frac{\lfloor h/2^{s} \rfloor}{\Delta^2_{\lfloor h/2^{s} \rfloor}} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_{\lfloor h/2^{s} \rfloor}) \rceil^2}{6\delta} \right) \tag{4.39}$$

Then, recalling that by assumption $\Delta_1 \leqslant, \ldots, \leqslant \Delta_K$, one can observe that the mapping from $[K]$ to $(0, \infty)$,

$$u \mapsto \frac{1}{\Delta_u^2} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_u) \rceil^2}{6\delta} \right)$$

is non-increasing and it is easy to check that

$$\lfloor h/2^s \rfloor - \lceil \lfloor h/2^s \rfloor / 2 \rceil + 1 \geqslant \frac{1}{2} \lfloor h/2^s \rfloor$$

therefore

$$\frac{\lfloor h/2^s \rfloor}{\Delta^2_{\lfloor h/2^s \rfloor}} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_{\lfloor h/2^s \rfloor}) \rceil^2}{6\delta} \right) \leqslant 2 \sum_{u=\lceil \lfloor h/2^s \rfloor / 2 \rceil}^{\lfloor h/2^s \rfloor} \frac{1}{\Delta_u^2} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_u) \rceil^2}{6\delta} \right) \tag{4.40}$$

Combining (4.39) and (4.40) yields

$$N \leqslant \frac{64\sigma^2}{3} \sum_{s=1}^{\lfloor \log_2 h \rfloor} \sum_{u=\lceil \lfloor h/2^s \rfloor / 2 \rceil}^{\lfloor h/2^s \rfloor} \frac{1}{\Delta_u^2} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_u) \rceil^2}{6\delta} \right) \tag{4.41}$$

Now let us introduce for any s, the set of integers $\mathcal{I}_s = [\![\lceil \lfloor h/2^s \rfloor / 2 \rceil, \lfloor h/2^s \rfloor]\!]$. We have

$$\bigcup_{s=1}^{\lfloor \log_2 h \rfloor} \mathcal{I}_s \subset \{2, \ldots, h\}.$$

We show that for any $p, q \in \{1, \ldots, \lfloor \log_2(h) \rfloor\}$ if $|p - q| \geqslant 2$ then $\mathcal{I}_p \cap \mathcal{I}_q = \emptyset$. Assuming $p \leqslant q$ we claim that

$$\lfloor h/2^{p+2} \rfloor < \lceil \lfloor h/2^p \rfloor / 2 \rceil \tag{4.42}$$

Assume otherwise, then $\lfloor h/2^{p+2} \rfloor \geqslant \lceil \lfloor h/2^p \rfloor / 2 \rceil \geqslant \lfloor h/2^p \rfloor / 2$ so

$$h/2^{p+1} \geqslant \lfloor h/2^p \rfloor$$

which is impossible since for any $p \in \{0, \ldots, \lfloor \log_2(h) \rfloor - 1\}$, $h/2^p \geqslant 1$. Therefore we have proved (4.42) and for any $q \geqslant p + 2$ it holds that

$$\lfloor h/2^q \rfloor \leqslant \lfloor h/2^{p+2} \rfloor < \lceil \lfloor h/2^p \rfloor / 2 \rceil$$

thus $\mathcal{I}_q \cap \mathcal{I}_p = \emptyset$ and for any $i \in \{2, \dots, h\}$, i belongs to no more than 2 of the subsets $\mathcal{I}_1, \dots \mathcal{I}_{\lfloor \log_2 h \rfloor}$, thus we have

$$\begin{aligned}
\widetilde{N} &\leqslant \frac{64}{3}\sigma^2 \sum_{s=1}^{\lfloor \log_2 h \rfloor} \sum_{u=\lceil \lfloor h/2^s \rfloor /2 \rfloor}^{\lfloor h/2^s \rfloor} \frac{1}{\Delta_u^2} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_u) \rceil^2}{6\delta}\right) && (4.43)\\
&\leqslant \frac{128}{3}\sigma^2 \sum_{i=2}^{h} \frac{1}{\Delta_i^2} \log\left(\frac{\pi^2 K d \lceil \log_2(1/\Delta_i) \rceil^2}{6\delta}\right) && (4.44)\\
&\leqslant \frac{128}{3}\sigma^2 \sum_{i=2}^{h} \frac{1}{\Delta_i^2} \log\left(\frac{\pi^2 K d \log_2(2/\Delta_i)^2}{6\delta}\right) && (4.45)\\
&\leqslant \frac{256}{3}\sigma^2 \sum_{i=2}^{h} \frac{1}{\Delta_i^2} \log\left(\frac{Kd}{\delta} \log_2\left(\frac{2}{\Delta_i}\right)\right) && (4.46)
\end{aligned}$$

Then, from Lemma 4.5.6 it holds that with probability at least $1-\delta$ the sample complexity N^δ of GEGE is upper-bounded as

$$\log_2(2/\Delta_1) + \mathcal{O}\left(\sum_{i=2}^{h} \frac{\sigma^2}{\Delta_i^2} \log\left(\frac{Kd}{\delta} \log_2\left(\frac{1}{\Delta_i}\right)\right)\right),$$

□

where $\mathcal{O}(\cdot)$ hides universal multiplicative constant. Therefore, we have shown the sample complexity bound and the correctness on $\mathcal{E}_{\text{fc}}$. Thus, proving that $\mathbb{P}(\mathcal{E}_{\text{fc}}) \geqslant 1-\delta$ will conclude the proof.

Probability of the good event $\mathcal{E}_{\text{fc}}$.

Proof. At round r,

$$\mathbb{P}\left((\mathcal{E}_{\text{fc}}^r)^c \mid \mathcal{A}_r\right) \leqslant \sum_{i \in \mathcal{A}_r} \mathbb{P}\left(\|(\widehat{\theta}_r - \theta)^\intercal x_i\|_\infty > \varepsilon_r/4 | \mathcal{A}_r\right)$$

Then, note that at round r, Algorithm 4.2 calls OptEstimator with precision $\varepsilon_r/2$ and budget t_r and by design we have $t_r \geqslant 20 h_r/\varepsilon_r^2$, so using Lemma 4.2.2, it follows

$$\begin{aligned}
\mathbb{P}\left((\mathcal{E}_{\text{fc}}^r)^c \mid \mathcal{A}_r\right) &\leqslant 2d \exp\left(-\frac{t_r \varepsilon_r^2}{32(1+3\varepsilon_r)\sigma^2 h_r}\right)\\
&\leqslant \delta_r/|\mathcal{A}_r|
\end{aligned}$$

which follows by plugging in the value of t_r. Therefore, by union bound over $\mathcal{A}_r$ and r it holds that $\mathbb{P}\left(\mathcal{E}_{\text{fc}}\right) \geqslant 1 - \sum_{r \geqslant 1} \delta_r \geqslant 1-\delta$. □

This concludes the proof of Theorem 4.3.3.

4.5.4 Lower bounds

We now prove minimax lower bounds in both fixed-confidence and fixed-budget settings. We recall the lower bound below for unstructured PSI in the fixed confidence setting.

Theorem 4.5.8 (Theorem 17 of Auer et al. 2016). *For any set of operating points $\mu_i \in [1/4, 3/4]^d$, $i = 1, \dots, K$, there exist distributions $(\mathcal{D}_i)_{1 \leqslant i \leqslant K}$ such that with probability at least $1 - \delta$, any δ-correct algorithm for linear PSI requires at least*

$$\Omega\left(\sum_{i=1}^{K} \frac{1}{\widetilde{\Delta}_i^2} \log(\delta^{-1})\right)$$

samples to identify the Pareto set. Where for any suboptimal arm $\widetilde{\Delta}_i = \Delta_i$ and for an optimal arm $\widetilde{\Delta}_i = \delta_i^+$.

In particular, there exist instances where $\Delta_i = \delta_i^+$ for any Pareto-optimal arm i. Thus, this result shows that H_1 is a good proxy to measure the complexity of PSI in the fixed-confidence setting.

The proof of such results often relies on the celebrated change of distribution technique (see *e.g.*Kaufmann, Cappé, et al. 2016) which, given the instance $\nu := (\nu_1, \dots, \nu_K)$, shifts the mean of ν_i for an arm i while keeping the others fixed constant. However, in linear PSI, the arms' means are correlated through Θ. So, in general, Theorem 4.5.8 does not directly apply to linear PSI. We recall below our lower bound for linear PSI in the fixed-confidence setting.

Theorem 4.3.5. *For any $K, d, h \in \mathbb{N}$, there exists a set $\mathcal{M}(K, d, h)$ of linear PSI instances such that for $\nu \in \mathcal{M}(K, d, h)$ and for any δ-correct algorithm for linear PSI, with probability at least $1 - \delta$,*

$$\tau_\delta^{\mathcal{A}} \geqslant \Omega\left(H_{1,lin}(\nu) \log \frac{1}{\delta}\right).$$

Proof of Theorem 4.3.5. The idea of the proof is to transform an unstructured bandit instance into a linear PSI instance. Let ν be a bandit instance with $K \geqslant 2$ arms and dimension $d \geqslant 1$ and with means $\mu_1, \dots, \mu_K \in [0, 1]^d$. Let $e_1, \dots e_h$ denote the canonical basis of $\mathbb{R}^h$. We define a linear PSI instance ν_{lin} with features

$$x_i = \begin{cases} e_i & \text{if } i \leqslant h \\ \mathbf{0} & \text{else.} \end{cases}$$

We assume that the learner knows that $\mu_i \in [0, 1]^d$ for any arm i. We claim that with this information, an "efficient" algorithm for PSI should not pull arms from $\{h + 1, \dots, K\}$. To see this, first note that these arms will be suboptimal, so $\mathcal{S}^\star \subset [h]$. Moreover, even if an arm $i \in \{h + 1, \dots, K\}$ dominates another arm $j \in \{1, \dots, h\}$, as j is not Pareto-optimal there exists another arm $j^\star \in \mathcal{S}^\star \subset \{1, \dots, h\}$ which dominates j with a larger margin,

so is "cheaper" to pull. Therefore, the complexity of ν_{lin} reduces to the complexity of a linear bandit $\widetilde{\nu}_{\text{lin}}$ with only h arms. As the features in $x_1, \dots, x_h$ form the canonical $\mathbb{R}^h$ basis, $\widetilde{\nu}_{\text{lin}}$ reduces to an unstructured bandit instance with (uncorrelated) means $\widetilde{\mu}_i = \theta^\intercal x_i$, $i = 1, \dots, h$. Therefore, by choosing $\mu_1, \dots, \mu_h \in [1/4, 3/4]^d$, we can apply Theorem 4.5.8 to $\widetilde{\nu}_{\text{lin}}$. □

The result proven above holds for a class of instances $\mathcal{M}(K, d, h)$ with the covariates defined as above and with matrix coefficients in $[1/4, 3/4]^d$. For the fixed-budget setting, Kone, Kaufmann, et al. 2024 proved a lower bound for a class of instances. We recall their result below after introducing some notation.

Their lower bound applies to the class of instances $\mathcal{M}$ defined as follows. $\mathcal{M}$ contains the instances such that each suboptimal arm i is only dominated by a Pareto-optimal arm denoted by $i^\star$ and that for each optimal arm j there exists a unique suboptimal arm which is dominated by j, denoted by $\underline{j}$. Moreover, for any instance in $\mathcal{M}$ the authors require its Pareto-optimal arms not to be close to the suboptimal arms they don't dominate: for any suboptimal arm i and Pareto-optimal arm j such that $\mu_i \not\prec \mu_j$,

$$\mathrm{M}(i, j) \geqslant 3 \max(\Delta_i, \Delta_{\underline{j}}).$$

Let $\nu := (\nu_1, \dots, \nu_K)$ be an unstructured instance whose means belongs to $\mathcal{M}$ and with isotropic multi-variate normal arms $\nu_i \sim \mathcal{N}(\mu_i, \sigma^2 I)$. For every $i \in [K]$, define the alternative instance $\nu^{(i)} := (\nu_1, \dots \nu_i^{(i)}, \dots, \nu_K)$ in which *only* the mean of arm i is shifted:

$$\mu_i^{(i)} := \begin{cases} \mu_i - 2\Delta_i \widetilde{e}_{d_{\underline{i}}} & \text{if } i \in \mathcal{S}^\star(\nu), \\ \mu_i + 2\Delta_i \widetilde{e}_{d_i} & \text{else,} \end{cases} \tag{4.47}$$

where $\widetilde{e}_1, \dots, \widetilde{e}_d$ denotes the canonical basis of $\mathbb{R}^d$ and for any arm i, $d_i := \operatorname{argmin}_{c \in [d]} [\mu^c_{i^\star} - \mu^c_i]$. Defining $\nu^{(0)} := \nu$, the theorem below holds.

Theorem 4.5.9 (Theorem 5 of Kone, Kaufmann, et al. 2024)**.** *Let $\nu = (\nu_1, \dots, \nu_K)$ be an instance in $\mathcal{M}$ with means $\mu := (\mu_1 \dots \mu_K)^\intercal$ and $\nu_i \sim \mathcal{N}(\mu_i, \sigma^2 I)$. For any algorithm $\mathcal{A}$, there exists $i \in \{0, \dots, K\}$ such that $H(\nu^{(i)}) \leqslant H(\nu)$ and the probability of error $\mathcal{A}$ on $\nu^{(i)}$ is at least*

$$\frac{1}{4} \exp\left(-\frac{2T}{\sigma^2 H(\nu^{(i)})} \right).$$

As explained above for the fixed-confidence setting. The proof of this lower bound also uses the change of distribution lemma. In the instances $\nu^{(i)}$ introduced above, only the mean of arm i must change *w.r.t.* $\nu^{(0)}$. Therefore, Theorem 4.5.9 does not apply to general instances in linear PSI. We recall our lower bound for linear PSI in the fixed-budget setting.

Theorem 4.3.2. *Let $\mathbb{W}_H$ be the set of instances with complexity $H_{2,lin}$ smaller than H. For any budget T, letting $\widehat{S}_T^{\text{alg}}$ be the output of an algorithm* alg*, it holds that*

$$\inf_{\text{alg}} \sup_{\nu \in \mathbb{W}_H} \mathbb{P}_\nu(\widehat{S}_T^{\text{alg}} \neq \mathcal{S}^\star(\nu)) \geqslant \frac{1}{4} \exp\left(-\frac{2T}{H\sigma^2} \right).$$

Proof of Theorem 4.3.2. Let H be fixed and recall that $\mathbb{W}_H : \{\nu_{\text{lin}} : H_{2,\text{lin}}(\nu) \leqslant H\}$ is the set of linear PSI instances with complexity less than H. The proof of Theorem 4.3.2 follows similar lines to Theorem 4.3.5. Let ν be an un-structured bandit instance with $K \geqslant 2$ arms, dimension $d \geqslant 1$, with means $\mu_1, \dots, \mu_K \in [0,1]^d$ and such that $H_2(\nu) \leqslant H$. We construct a linear PSI instance ν_{lin} from an unstructured multi-dimensional instance ν by setting $x_i := e_i$ for any $i \leqslant h$ and for $i > h$, $x_i = \mathbf{0}$ where $e_1, \dots, e_h$ is the canonical $\mathbb{R}^h$-basis. We also assume that the agent knows that $\mu_i \in [0,1]^d$ for any arm i.

For ν_{lin}, the arms $\{h+1, \dots, K\}$ are necessarily suboptimal, so $\mathcal{S}^\star \subset [h]$. Thus, to identify the Pareto set, an efficient algorithm should reduce to pulling arms in $\{1, \dots, h\}$. Indeed, as explained in the proof of Theorem 4.3.5 even if an arm $i \in \{h+1, \dots, K\}$ dominates another arm $j \in \{1, \dots, h\}$, as j is not Pareto-optimal there exists another arm $j^\star \in \mathcal{S}^\star \subset \{1, \dots, h\}$ which is "cheaper" to pull as it dominates j with a larger margin. ν_{lin} reduces to a linear bandit $\widetilde{\nu}_{\text{lin}}$ with only h arms and since the features $x_1, \dots, x_h$ forms the canonical basis of $\mathbb{R}^h$, $\widetilde{\nu}_{\text{lin}}$ is an un-structured bandit instance with (un-correlated) means $\widetilde{\mu}_i = \theta^\intercal x_i$, $i = 1, \dots, h$. Therefore, by choosing $\widetilde{\nu} := (\nu_1, \dots, \nu_h)$ that belongs to $\mathcal{M}$, we can apply Theorem 4.5.9 which yields

$$\max_{i \in \{0,\dots,K\}} \mathbb{P}_{\widetilde{\nu}^{(i)}}(S_T^{\text{alg}} \neq \mathcal{S}^\star(\widetilde{\nu}^{(i)})) \geqslant \frac{1}{4} \exp\left(-\frac{2T}{H\sigma^2}\right)$$

where by construction $\widetilde{\nu}^{(i)}$ (see construction above) is also a linear PSI instance. Then, further noting that $H \geqslant H_2(\nu) \geqslant H_2(\widetilde{\nu})$ and by Theorem 4.5.9 for any $i \leqslant h$ $H_{2,\text{lin}}(\widetilde{\nu}) \geqslant H_2(\widetilde{\nu}^{(i)})$. Then recalling that ν_{lin} is equivalent to $\widetilde{\nu}$ it comes

$$\inf_{\text{alg}} \sup_{\nu \in \mathbb{W}_H} \mathbb{P}_\nu(S_T^{\text{alg}} \neq \mathcal{S}^\star(\nu)) \geqslant \frac{1}{4} \exp\left(-\frac{2T}{H\sigma^2}\right),$$

which is the claimed result. □

4.5.5 Concentration lemmas

Proof of Lemma 4.5.1. We have

$$\begin{aligned}
|\,\mathrm{M}(i,j;r) - \mathrm{M}(i,j)| &= \left|\max_c \left[\hat{\mu}_{i,r}^c - \hat{\mu}_{j,r}^c\right] - \max_c \left[\mu_i^c - \mu_j^c\right]\right|, \\
&\overset{(i)}{\leqslant} \max_c \left|(\hat{\mu}_{i,r}^c - \hat{\mu}_{j,r}^c) - (\mu_i^c - \mu_j^c)\right|, \\
&= \|(\hat{\mu}_{i,r} - \hat{\mu}_{j,r}) - (\mu_i - \mu_j)\|_\infty, \\
&= \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_j)\|_\infty.
\end{aligned}$$

where (i) follows from the reverse triangle inequality. The second part of the lemma is a direct consequence of the relation $\mathrm{M}(i,j) = -\,\mathrm{m}(i,j)$ as well as $\mathrm{M}(i,j;r) = -\,\mathrm{m}(i,j;r)$ that holds for any pair of arms i, j. □

Proof of Lemma 4.5.2. Since $i^\star$ does not empirically dominate i it holds that $\mathrm{M}(i, i^\star; r) > 0$ so $\mathrm{M}(i, i^\star; r) - \mathrm{M}(i, i^\star) > -\mathrm{M}(i, i^\star)$. Then noting that

$$-\mathrm{M}(i, i^\star) = \mathrm{m}(i, i^\star) = \Delta_i$$

yields $\mathrm{M}(i, i^\star; r) - \mathrm{M}(i, i^\star) > \Delta_i$. Therefore

$$\begin{aligned} \Delta_i = \Delta_i^\star &< \mathrm{M}(i, i^\star; r) - \mathrm{M}(i, i^\star) \\ &\leqslant \|(\hat{\theta}_r - \theta)^\intercal (x_i - x_{i^\star})\|_\infty, \end{aligned}$$

where the last inequality is a consequence of Lemma 4.5.1. □

Chapter 5

Posterior Sampling for Pareto Set Identification

In previous chapters, we studied Pareto Set Identification (PSI) through deterministic, confidence-sequence-based algorithms that offer strong finite-sample guarantees but suffer from three main limitations: the confidence level δ must be fixed *a priori*, asymptotic optimality is not achieved as $\delta \to 0$, and correlations between objectives are typically ignored.

This chapter introduces a posterior-sampling approach that overcomes these issues. Focusing on the unstructured multivariate Gaussian setting with possibly correlated objectives, we propose PSIPS (*Pareto Set Identification with Posterior Sampling*), an algorithm that uses posterior draws in both the sampling rule and the stopping rule. PSIPS removes the need to pre-specify δ, achieves asymptotic optimality from both Bayesian and frequentist perspectives, and naturally exploits correlations between objectives.

Beyond its theoretical guarantees, PSIPS remains computationally tractable by avoiding costly optimization oracles. This chapter is based on joint work with Marc Jourdan and Émilie Kaufmann, published in the proceedings of *AISTATS 2025*, where PSIPS is further analyzed in the more general linear PSI setting.

5.1 Introduction

In previous chapters, we investigated Pareto Set Identification (PSI) using deterministic confidence-sequence-based methods that provide rigorous finite-time guarantees. While effective in practice, these approaches exhibit three key limitations. First, they require fixing the confidence parameter δ before data collection, forcing practitioners to predefine an acceptable risk level. Second, although these methods achieve sample complexity of $\mathcal{O}(H(\nu)\log(H(\nu)d/\delta))$ where $H(\nu) := \sum_{k\in[K]} \Delta_k^{-2}$ depends on instance-specific gaps, they fail to achieve *asymptotic optimality*—their sample complexity does not match the information-theoretic lower bound as $\delta \to 0$. Evidence from best-arm identification suggests LUCB-style algorithms incur suboptimal constants in this regime (Kejriwal et al. 2025). Third, while not assuming independence between objectives, they do not explicitly exploit correlations common in applications such as efficacy-toxicity trade-offs in clinical trials.

This chapter addresses these limitations through a *posterior sampling* approach. We propose PSIPS (Posterior Sampling for Pareto Set Identification), an algorithm that uses random draws from a posterior distribution for both exploration and stopping decisions. This algorithmic design eliminates the need to pre-specify δ, naturally accounts for correlations across objectives through the covariance structure Σ, and achieves asymptotic optimality. Specifically, we establish that $\limsup_{\delta\to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)}$ matches the information-theoretic lower bound, answering a fundamental question left open by previous chapters.

We focus on the unstructured Gaussian setting where each arm $i \in [K]$ yields independent d-dimensional observations from $\mathcal{N}(\mu_i, \Sigma)$ with known covariance Σ. This formulation explicitly captures correlations between objectives, enabling the learner to transfer information between outcomes and reduce sample complexity. These properties make PSIPS particularly relevant for adaptive clinical trials, where multiple correlated outcomes must be analyzed simultaneously, for instance, in dose-finding trials where higher efficacy correlates with higher toxicity.

Related work. The algorithms of Chapters 3 and 4, along with Auer et al. 2016, rely on confidence intervals requiring δ to be specified in advance. While practical, these methods target finite-time guarantees rather than asymptotic optimality and do not exploit the covariance structure. Crepon et al. 2024 achieved asymptotic optimality through gradient-based optimization inspired by the information-theoretic lower bound, but their approach requires solving $\mathcal{O}(K^d)$ convex programs per iteration and assumes independent objectives ($\Sigma = \sigma^2 I_d$). More broadly, asymptotically optimal algorithms based on optimization or game theory (Garivier & Kaufmann 2016; Degenne, W. M. Koolen, et al. 2019; Degenne, Ménard, et al. 2020) provide strong guarantees but remain computationally expensive in multi-objective settings.

Posterior sampling approaches for pure exploration extend Thompson Sampling's success from regret minimization (Thompson 1933; Kaufmann, Korda, et al. 2012) to identification problems. Russo 2016 introduced Top-Two Thompson Sampling for best-arm identification with optimal posterior contraction but no stopping rule. S. Wang & Zhu 2022 proposed TS-Explore using posterior samples for exploration and stopping, though its gap-based

sampling remains asymptotically suboptimal. The PEPS algorithm (Z. Li et al. 2024) achieves optimal Bayesian convergence in linear bandits but lacks fixed-confidence guarantees. Most multi-objective algorithms assume independent objectives (Auer et al. 2016; Zuluaga, Krause, et al. 2016; Katz-Samuels & Scott 2018), limiting efficiency when outcomes are correlated.

In summary, existing PSI algorithms either rely on confidence-sequence-based methods that require a pre-specified δ, or they achieve asymptotic optimality at the cost of computational intractability and restrictive independence assumptions.

PSIPS integrates posterior sampling into both sampling and stopping, achieving asymptotic optimality while remaining computationally efficient and naturally exploiting objective correlations.

Learning model. We consider the standard multi-armed bandit framework in the multi-objective setting with dimension $d \geqslant 2$. The collection of means is denoted by $\mu = (\mu_i)_{i\in[K]} \in \mathbb{R}^{K\times d}$. Throughout this chapter, we focus on the *Gaussian case*:

$$\nu_i = \mathcal{N}(\mu_i, \Sigma), \quad i \in [K],$$

where the covariance matrix Σ is known and identical across arms. The off-diagonal entries of Σ encode dependence–for instance, between toxicity and efficacy in dose-finding trials–allowing the learner to transfer information across outcomes. Such correlations are common in clinical and engineering applications and play a key role in improving sample efficiency.

At each time t, the learner selects an arm $A_t \in [K]$ based on past observations and obtains a sample $Z_t \in \mathbb{R}^d$ with conditional distribution ν_{A_t} given A_t. An *algorithm* specifies both a *sampling rule* that determines A_t and a *recommendation rule* that outputs a candidate Pareto set $\widehat{S}_t \subseteq [K]$ before each arm pull. Formally, both A_t and $\widehat{S}_t$ are $\mathcal{H}_t$-measurable, where the σ-algebra $\mathcal{H}_t := \sigma(U_1, A_1, Z_1, \ldots, A_{t-1}, Z_{t-1}, U_t)$ represents the *history* before time t, and $U_t \sim \mathcal{U}([0,1])$ captures any exogenous randomness used by the algorithm.

For completeness, we recall that an arm i is said to be (Pareto) dominated by an arm j, written $\mu_i \prec \mu_j$, if $\mu_i^c \leqslant \mu_j^c$ for all $c \in [d]$ and the inequality is strict for at least one coordinate. The *Pareto set* is then

$$\mathcal{S}^\star(\mu) := \{\, i \in [K] : \nexists j \in [K] \setminus \{i\},\ \mu_i \prec \mu_j \,\},$$

that is, the collection of arms not strictly dominated by any other.

The empirical allocation over arms is $N_t := (N_{t,a})_{a\in[K]}$, where $N_{t,a} := \sum_{s\in[t-1]} \mathbb{1}(A_s = a)$, so that $\frac{N_t}{t-1} \in \triangle_K := \{w \in \mathbb{R}_+^K : \sum_{a\in[K]} w_a = 1\}$ belongs to the probability simplex. For vector $x \in \mathbb{R}^d$, $x(c)$ and x^{c} denote c-th coordinate. In the sequel, we let $\mathcal{I} := \mathbb{R}^d$.

Lower Bound. To formalize the notion of asymptotic optimality, we now characterize the minimal number of samples required by any δ-correct algorithm to reliably identify the Pareto set. A δ-correct algorithm must be able to statistically distinguish the true mean matrix μ from all *alternative instances* that would induce a different Pareto set. We define

this set of alternatives as

$$\mathrm{Alt}(\mathcal{S}^\star(\mu)) \coloneqq \{\lambda \in \mathbb{R}^{K\times d} : \mathcal{S}^\star(\lambda) \neq \mathcal{S}^\star(\mu)\}.$$

Intuitively, these are the configurations of mean vectors that would lead the learner to recommend an incorrect Pareto set. The requirement to separate μ from $\mathrm{Alt}(\mathcal{S}^\star)$ leads to a fundamental information-theoretic lower bound on the expected stopping time (Garivier & Kaufmann 2016; Garivier, Ménard, et al. 2019; Crepon et al. 2024).

Lemma 5.1.1. *Let $\mathcal{D}$ be the class of d-dimensional Gaussian distributions with known covariance Σ. For any δ-correct algorithm on $\mathcal{D}^K$, and any bandit instance $\nu \in \mathcal{D}^K$ we have*

$$\mathbb{E}_\nu[\tau_\delta] \;\geqslant\; T^\star(\nu)\,\log\!\Big(\tfrac{1}{2.4\delta}\Big),$$

where $T^\star(\nu)$, called the characteristic time*, is defined by*

$$T^\star(\nu)^{-1} \;\coloneqq\; \sup_{w\in\triangle_K} \inf_{\lambda\in Alt(\mathcal{S}^\star(\mu))} \left[\sum_{k=1}^{K} \frac{w_k}{2}\big\|\mu_k - \lambda_k\big\|^2_{\Sigma^{-1}}\right].$$

The set of maximizers of the outer optimization, denoted by $w^\star(\nu)$, corresponds to the asymptotically optimal allocation of samples across arms.

This result establishes that no δ-correct algorithm can, in expectation, identify the Pareto set faster than $T^\star(\nu)\log(1/(2.4\delta))$ samples. An algorithm is said to be *asymptotically optimal* if it matches *exactly* this lower bound asymptotically, *i.e.*, if

$$\forall\, \nu \in \mathcal{D}^K,\ \lim_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta]}{\log\frac{1}{\delta}} \;=\; T^\star(\nu)\,. \tag{5.1}$$

Chernoff stopping rules. In fixed-confidence pure exploration, stopping rules determine when to terminate sampling. The classical *Chernoff* or *generalized likelihood ratio* (GLR) stopping rule (Garivier & Kaufmann 2016) stops when

$$\tau_\delta^{\mathrm{GLR}} = \inf\{t \geqslant 1 | \mathrm{GLR}(t) > \beta(t,\delta)\}, \quad \text{with}$$

$$\mathrm{GLR}(t) = \inf_{\lambda\in\mathrm{Alt}(\widehat{\mathcal{S}}_t)} \sum_{i=1}^{K} \frac{N_{t,i}}{2}\|\hat{\mu}_{t,i} - \lambda_i\|^2_{\Sigma^{-1}}\,. \tag{5.2}$$

While the GLR rule ensures δ-correctness and can achieve asymptotic optimality with a proper sampling rule, its evaluation is computationally demanding.

To the best of our knowledge, computing (5.2) is only tractable for the setting with independent objectives, *i.e.*, diagonal Σ. Crepon et al. 2024 proposed the best-known procedure in that restrictive setting, which requires solving $\mathcal{O}(Kd^3|\widehat{S}_t|^d)$ convex problems at each round. Their computationally costly procedure is difficult to extend to correlated objectives.

Sampling rules. Given a stopping rule, the sampling rule should make it stop as soon as possible. While gap-based sampling rules are suboptimal in their asymptotic allocation, the asymptotic optimality of Track-and-Stop (Garivier & Kaufmann 2016) comes at the cost of computational intractability since it computes $w_t \in w^\star(\hat{\mu}_t)$ at time t. The game-based approach arises from viewing the lower bound as the value of a two-player zero-sum game (Degenne, W. M. Koolen, et al. 2019). Using two no-regret learning algorithms against each other yields a saddle-point algorithm that sequentially approximates $T^\star(\nu)^{-1}$. The popularized instance uses the best response as a min learner, *i.e.*,

$$\lambda^{\mathrm{BR}} : w \mapsto \operatorname*{argmin}_{\lambda \in \mathrm{Alt}(\widehat{S}_t)} \sum_{k=1}^{K} \frac{w_k}{2} \|\hat{\mu}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}} \,, \tag{5.3}$$

whose computational cost is the same as the GLR statistic (5.2). For the PS stopping rule, we rely on an inflated posterior sampling (PS) learner for the min learner. Therefore, our algorithm can tackle structured settings with correlated objectives and has a low computational cost.

Contributions. As the main contribution, we propose the PSIPS algorithm, which is the first computationally efficient, asymptotically optimal PSI algorithm with correlated objectives. PSIPS builds on posterior sampling in both the PS stopping rule and the min learner of the game-based sampling rule. By removing the oracle calls to (5.2) or (5.3), PSIPS deals with the PSI structure in a computationally efficient way.

From a frequentist perspective, it is shown to be asymptotically optimal for PSI with correlated objectives (Theorem 5.3.1). From a Bayesian perspective, the posterior probability that PSIPS misidentifies the Pareto set decays exponentially fast (as a function of t) with the optimal rate $1/T^\star(\nu)$ (Theorem 5.3.2). Our experiments on both synthetic and real-world instances showcase the superior performance of our algorithm in terms of sample complexity and computational cost.

5.2 PSI with Posterior Sampling

We propose the PSIPS (PSI with Posterior Sampling) algorithm for PSI with correlated objectives. PSIPS combines the PS stopping rule and the game-based sampling rule using a PS min learner.

5.2.1 The Posterior Sampling (PS) Stopping Rule

We introduce the posterior sampling (PS) stopping rule. To specify the PS stopping rule, a budget function $M : \mathbb{N} \times (0,1) \to \mathbb{R}_+$ and an inflation function $c : \mathbb{N} \times (0,1) \to \mathbb{R}_+$ are given. The PS stopping rule stops when $M(t-1,\delta)$ independent realizations from a posterior distribution inflated by $c(t-1,\delta)$ all yield the same Pareto set as the current empirical estimate $\widehat{S}_t$, suggesting that $S^\star(\mu) = \widehat{S}_t$. Conditioned on $\mathcal{H}_t$, let $(v_t^m)_{m \in [M(t,\delta)]}$ be *i.i.d.* draws

Algorithm 5.1: PSIPS: Pareto Set Identification with Posterior Sampling

Require: budget function M; inflation function c; exploration allocation $w_{\exp} > 0$; fored exploration rate $\alpha > 0$; posterior inflation rate function η; AdaHedge learner $\mathcal{L}^{[K]}$

Initialize: pull each arm once and compute $\hat{\mu}_1$

for $t = 1, 2, \ldots$ **do**

// Current empirical Pareto set

$\widehat{S}_t \leftarrow \mathcal{S}^\star(\hat{\mu}_t)$;

Initialize: $m \leftarrow 0$, $m_t \leftarrow +\infty$, and $m_{t,\delta} \leftarrow +\infty$;

// Posterior sampling and stopping rule

while $\max\{m_t, m_{t,\delta}\} = +\infty$ **do**

$m \leftarrow m + 1$; sample $v_t^m \sim \bigotimes_{k=1}^K \mathcal{N}(0_d, \Sigma/N_{t,k})$;

if $m_{t,\delta} = +\infty$ **then**

$\theta_{t,k}^m \leftarrow \hat{\mu}_{t,k} + \sqrt{c(t-1,\delta)}\, v_{t,k}^m \quad \forall\, k$;

if $\theta_t^m \in \mathit{Alt}(\widehat{S}_t)$ **then**

$m_{t,\delta} \leftarrow m$;

else if $m \geqslant M(t-1,\delta)$ **then**

return $\widehat{S}_t$;

if $m_t = +\infty$ **then**

$\lambda_{t,k}^m \leftarrow \hat{\mu}_{t,k} + \eta_t^{-1/2} v_{t,k}^m \quad \forall\, k$;

if $\lambda_t^m \in \mathit{Alt}(\widehat{S}_t) \cap \mathcal{I}_t^K$ **then**

$m_t \leftarrow m$;

// Learner update and mixture exploration

Get w_t from learner $\mathcal{L}^{[K]}$ and set $\lambda_t \leftarrow \lambda_t^{m_t}$;

$$\widetilde{w}_t \leftarrow (1-\gamma_t) w_t + \gamma_t w_{\exp}, \quad \gamma_t = t^{-\alpha}.$$

// Sampling and observation

Sample $A_t \sim \widetilde{w}_t$ and collect $Z_t \sim \nu_{A_t}$;

// Feed back to learner

Feed gain

$$g_t(w) = \sum_{k=1}^K \frac{w_k}{2} \left\| \lambda_{t,k} - \hat{\mu}_{t,k} \right\|_{\Sigma^{-1}}^2 \quad \text{to } \mathcal{L}^{[K]}$$

from the centered posterior distribution $\Pi_t := \bigotimes_{k=1}^K \mathcal{N}(0_d, \Sigma/N_{t,k})$. Each v_t^m is a collection of $\mathbb{R}^d$ vectors $(v_{t,1}^m, \dots, v_{t,K}^m)$. For all $m \in [M(t-1,\delta)]$, let $\theta_{t,k}^m := \hat{\mu}_{t,k} + \sqrt{c(t-1,\delta)} v_{t,k}^m$. Then,

$$\tau_\delta^{\mathrm{PS}} := \inf\{t \geqslant 1 | \forall m \in [M(t-1,\delta)],\ \mathcal{S}^\star(\theta_t^m) = \widehat{S}_t\}\,.$$

When $\tau_\delta^{\mathrm{PS}} < +\infty$, we recommend $\widehat{S}_{\tau_\delta^{\mathrm{PS}}} := \mathcal{S}^\star(\hat{\mu}_{\tau_\delta^{\mathrm{PS}}})$. When $t < \tau_\delta^{\mathrm{PS}}$, let $m_{t,\delta} := \inf\{m | \theta_t^m \in \mathrm{Alt}(\widehat{S}_t)\}$.

Computational cost. To reduce the computational cost of PSIPS, we draw $(v_t^m)_m$ sequentially as the PS stopping condition is often infringed before $M(t,\delta)$ realizations are drawn (see Figure 5.6).

Compared to the oracle call to (5.2), testing that $\theta_t^m \notin \mathrm{Alt}(\widehat{S}_t)$ has a lower computational cost. It scales as the sum of the costs of membership to $\mathrm{Alt}(\widehat{S}_t)^\mathsf{c}$, which is at most $\mathcal{O}(d|\widehat{S}_t|\max\{|\widehat{S}_t|, K - |\widehat{S}_t|\})$ since, thanks to Lemma 1 in Crepon et al. 2024

$$\mathrm{Alt}(\widehat{S}_t)^\mathsf{c} = \left(\bigcap_{(i,j)\in \widehat{S}_t^2, i\neq j} \bigcup_{c\in[d]} \{\lambda \mid \lambda_i^c \geqslant \lambda_j^c\} \right) \bigcap \left(\bigcap_{i\notin \widehat{S}_t} \bigcup_{j\in\widehat{S}_t} \bigcap_{c\in[d]} \{\lambda | \lambda_j^c \geqslant \lambda_i^c\} \right)\,. \tag{5.4}$$

To update the candidate answer, we check whether $\hat{\mu}_t \in \mathrm{Alt}(\widehat{S}_{t-1})^\mathsf{c}$, in which case $\widehat{S}_t = \widehat{S}_{t-1}$. Otherwise, we compute $\widehat{S}_t = \mathcal{S}^\star(\hat{\mu}_t)$. While a naive implementation has cost at most $\mathcal{O}(dK^2)$, Kung et al. 1975 proposed an algorithm having a cost scaling as $\mathcal{O}(K(\log K)^{\max\{1,d-2\}})$.

Correctness. Lemma 5.2.2 exhibits choices of budget $M(t,\delta)$ and inflation $c(t,\delta)$ ensuring δ-correctness in the setting with independent or correlated objectives. Letting $X \sim \mathcal{N}(0,1)$, we denote by $R(x) := \frac{\mathbb{P}(X>x)}{f_X(x)}$, the Mills ratio (Mills 1926) of X, with f_X, the density of X.

We introduce the concentration event $\mathcal{E}_\delta := \bigcap_{t\in\mathbb{N}} \mathcal{E}_{t,\delta}$ with

$$\mathcal{E}_{t,\delta} := \left\{ \sum_{k=1}^K N_{t,k} \left\|\mu_k - \hat{\mu}_{t,k}\right\|_{\Sigma^{-1}}^2 \leqslant 2\beta(t-1,\delta) \right\}. \tag{5.5}$$

The lemma below gives choices of $\beta(t,\delta)$ such that $\mathbb{P}_\nu(\mathcal{E}_\delta^\mathsf{c}) \leqslant \delta$. These thresholds satisfy

$$\lim_{\delta\to 0} \beta(\cdot,\delta)/\log(1/\delta) = 1\,,$$

and $\beta(t,\cdot) =_{+\infty} \mathcal{O}(\log\log t)$. This result is proven in Lemma 3 of Kone, Jourdan, et al. 2025.

Lemma 5.2.1 (Kone, Jourdan, et al. 2025). *Let $s > 1$, ζ be the Riemann ζ function and $\overline{W}_{-1}(x) := -W_{-1}(-e^{-x})$ for all $x \geqslant 1$, where W_{-1} is the negative branch of the Lambert W function. Let $\mathcal{E}_\delta$ be defined as in (5.5). Then, we have $\mathbb{P}_\nu(\mathcal{E}_\delta^\mathsf{c}) \leqslant \delta$ by taking*

$$\beta(t,\delta) = \frac{dK}{2}\overline{W}_{-1}\left(\frac{2}{dK}\log\frac{e^{Ks}\zeta(s)^K}{\delta} + \frac{2s}{d}\log\left(1 + \frac{d}{2s}\log\frac{t}{K}\right) + 1\right)\,.$$

We can now state the calibration of M and c to ensure δ-correctness of the PS stopping rule.

Lemma 5.2.2. *Let $\delta \in (0,1)$, $s > 1$, and let ζ denote the Riemann zeta function. Define*

$$r(\delta, n) := \left(\frac{1}{\sqrt{2\pi}} R\left(\sqrt{\frac{2}{n} \log(1/\delta)} \right) \right)^n,$$

where R is the Mills ratio of $\mathcal{N}(0,1)$. Regardless of the sampling rule, the PS stopping rule with parameters

$$c(t,\delta) = \frac{\beta(t,\delta/2)}{\log(1/\delta)}, \quad M(t,\delta) = \left\lceil \frac{\log(2t^s\zeta(s)/\delta)}{\delta \cdot q(t,\delta)} \right\rceil,$$

ensures δ-correctness for any $\delta \in (0,1)$, where $q(t,\delta)$ is defined as follows:

- *If Σ is diagonal:*
$$q(t,\delta) = \min\left\{ r(\delta, d),\ r(\delta, d + |\widehat{S}_{t+1}|) \right\}.$$
- *If Σ is non-diagonal:*
$$q(t,\delta) = \det(\Sigma\bar{\sigma})^{-1/2} \min\left\{ r\left(\delta^{\frac{d_\Sigma}{d}}, d\right),\ r\left(\delta^{\frac{d_\Sigma + |\widehat{S}_{t+1}|}{d + |\widehat{S}_{t+1}|}}, d + |\widehat{S}_{t+1}|\right) \right\},$$
where $\bar{\sigma} = |||\Sigma^{-1}|||$ and $d_\Sigma = \|1_d\|^2_{(\bar{\sigma}\Sigma)^{-1}}$.

Moreover, these choices satisfy

$$\limsup_{\delta \to 0} \frac{c(t,\delta) \log M(t,\delta)}{\log(1/\delta)} \leqslant 1.$$

5.2.2 Game-based sampling rule

We introduce a game-based sampling rule (Degenne & W. Koolen 2019) using a PS min learner. By combining a max learner playing w_t and a min learner playing λ_t, it yields a saddle-point algorithm which approximates $T^\star(\nu)^{-1}$.

Max learner. We opt for AdaHedge (De Rooij et al. 2014) as the max learner. We add forced exploration by mixing the played allocation w_t with an exploration allocation $w_{\text{exp}} \in \triangle_K$ with positive weights. Formally, we pull $A_t \sim \widetilde{w}_t$ with $\widetilde{w}_t := (1-\gamma_t)w_t + \gamma_t w_{\text{exp}}$ where $\gamma_t = 1/t^\alpha$ with $\alpha \in (0,1)$. Forced exploration ensures that $\widehat{S}_t$ converges to $S^\star(\mu)$ despite the initial fluctuation of the min learning space $\text{Alt}(\widehat{S}_t)$. The cost of forced-exploration is only logarithmic for α close to 1.

Min learner. As in Z. Li et al. 2024, we propose a min learner based on posterior samples. The PS min learner draws independent realizations from a posterior distribution inflated by some parameter $\eta_t > 0$ until one disagrees with $\widehat{S}_t$. The parameter η_t will act as the learning rate as the posterior distribution is interpreted as the sampling distribution of continuous exponential weights for the quadratic loss.

Formally, let $(v_t^m)_m$ be *i.i.d.* draws from Π_t. For all $m \geqslant 1$, let $\lambda_{t,k}^m := \hat{\mu}_{t,k} + \eta_t^{-1/2} v_{t,k}^m \; \forall \; k$. Then,

$$\lambda_t := \lambda_t^{m_t} \text{ with } m_t = \inf\{m | \; \lambda_t^m \in \mathcal{I}_t^K \cap \mathrm{Alt}(\widehat{S}_t)\} \; , \tag{5.6}$$

where $\mathcal{I}_t^K \subset \mathcal{I}^K$ is a bounded set and we recall that $\Pi_t := \bigotimes_{k=1}^K \mathcal{N}(0_d, \Sigma/N_{t,k})$. Intuitively, $\lambda_t^{m_t}$ is a randomized approximation of the best-response oracle $\lambda^{\mathrm{BR}}(N_t)$ as in (5.3) (projected onto $\mathcal{I}_t^K$).

We re-use the realizations $(v_t^m)_m$ in the PS min learner of our sampling rule with a different inflation parameter. Equivalently, we could draw λ_t from the distribution $\bigotimes_{k=1}^K \mathcal{N}(\hat{\mu}_{t,k}, \eta_t^{-1}\Sigma/N_{t,k})$ truncated to $\mathcal{I}_t^K \cap \mathrm{Alt}(\widehat{S}_t)$. The inflation η_t is chosen as a function of an upper bound on the magnitude of the stochastic loss, and $\mathcal{I}_t^K$ is defined to control the magnitude of this loss. We refer the reader to Kone, Jourdan, et al. 2025 for more details on how η_t and $\mathcal{I}_t^K$ are precisely defined sequentially.

Computational cost. At time $t < \tau_\delta^{\mathrm{PS}}$, we sequentially re-use the realizations $(v_t^m)_{m \in [m_{t,\delta}]}$ drawn by the PS stopping rule with a different inflation parameter. When $m_{t,\delta} < m_t$, we draw fresh realizations sequentially. Despite the different inflations, m_t is expected to be close to $m_{t,\delta}$, *i.e.*, $\mathcal{O}(M(t,\delta))$ (see Figure 5.6). The computational cost of (5.6) is lower than a best-response oracle to (5.2).

5.3 Main theoretical results

We show that under a generic condition on (c, M), which is satisfied in Lemma 5.2.2, the expected sample complexity of PSIPS is asymptotically optimal (Theorem 5.3.1). From a Bayesian perspective, the posterior probability of misidentifying the Pareto set decays at an asymptotically tight exponential rate (Theorem 5.3.2).

Sample complexity. Theorem 5.3.1 shows that PSIPS is asymptotically optimal when using suitable choices of budget and inflation. In comparison, known upper bounds on $\mathbb{E}_\nu[\tau_\delta]$ were either only worst-case optimal (see Chapter 3), or restricted to the setting with independent objectives (Crepon et al. 2024).

Theorem 5.3.1. *For any budget M and inflation c such that* $\limsup_{\delta \to 0} \frac{c(t,\delta) \log M(t,\delta)}{\log(1/\delta)} \leqslant 1$, *PSIPS satisfies that*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\nu[\tau_\delta^{PS}]}{\log \frac{1}{\delta}} \leqslant T^\star(\nu) \quad \textit{and} \quad \mathbb{P}\left(\limsup_{\delta \to 0} \frac{\tau_\delta^{PS}}{\log \frac{1}{\delta}} \leqslant T^\star(\nu)\right) = 1, \quad \forall \nu \in \mathcal{D}^K \; .$$

Lemma 5.2.2 gives choices of (c, M) satisfying the above condition and ensuring δ-correctness. Theorem 5.3.1 also holds on the set $\widetilde{\mathcal{D}}$ of multivariate Σ-subgaussian, where $\kappa \in \widetilde{\mathcal{D}}$ with mean μ implies that $\mathbb{E}_{X\sim\kappa}[e^{u^\intercal(X-\mu)}] \leqslant e^{\frac{1}{2}u^\intercal\Sigma u}$ for all $u \in \mathbb{R}^d$ (Degenne & Perchet 2016). This includes distributions with Bernoulli marginals and distributions with bounded support. Optimality is only achieved for multivariate Gaussian distributions.

Posterior probability of misidentification. Theorem 5.3.2 shows that the posterior probability that PSIPS misidentifies the Pareto set decays exponentially fast (as a function of t) with a rate $1/T^\star(\nu)$, which is shown to be asymptotically optimal by a lower bound.

In comparison, similar Bayesian guarantees in the literature were restricted to BAI (*e.g.*, TTTS in Russo 2016 and PEPS in Z. Li et al. 2024).

Theorem 5.3.2. *Let $\widetilde{\Pi}_t := \bigotimes_{k=1}^K \mathcal{N}(\hat{\mu}_{t,k}, \Sigma/N_{t,k})$ be the posterior distribution under a flat Gaussian prior (without inflation). For any bandit instance $\nu \in \mathcal{D}^K$, it holds with probability 1 that*

$$\limsup_{t\to+\infty} -\frac{1}{t}\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathit{Alt}(\mathcal{S}^\star)) \leqslant T^\star(\nu)^{-1},$$

for any algorithm, and PSIPS almost surely satisfies that

$$\liminf_{t\to+\infty} -\frac{1}{t}\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathit{Alt}(\mathcal{S}^\star)) \geqslant T^\star(\nu)^{-1}\,.$$

5.3.1 Sketch of proofs

Proof sketch of Lemma 5.2.2

To prove Lemma 5.2.2 we first establish the following result.

Lemma 5.3.3. *For all c, M, the PS stopping rule satisfies*

$$\begin{aligned}\mathbb{P}_\nu(\tau_\delta^{PS} < +\infty, \widehat{S}_{\tau_\delta^{PS}} \neq \mathcal{S}^\star) \leqslant \delta/2 + \mathbb{E}\Bigg[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right)\mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right)\cdot \\ \exp\left(-M(t-1,\delta)\mathbb{P}_\nu\left(\theta_t \in \mathit{Alt}(\widehat{S}_t)|\mathcal{H}_t\right)\right)\Bigg].\end{aligned} \tag{5.7}$$

Proof. Using the definition of the PS stopping rule, we obtain

$$\begin{aligned}
\mathbb{P}_\nu\left(\tau_\delta^{\text{PS}} < \infty, \widehat{S}_{\tau_\delta^{\text{PS}}} \neq \mathcal{S}^\star\right) &\leqslant \delta/2 + \mathbb{P}_\nu\left(\mathcal{E}_{\delta/2} \cap \{\tau_\delta^{\text{PS}} < \infty, \widehat{S}_{\tau_\delta^{\text{PS}}} \neq \mathcal{S}^\star\}\right) \\
&\leqslant \delta/2 + \sum_{t\geqslant 1} \mathbb{P}_\nu\left(\mathcal{E}_{t,\delta/2} \cap \{\tau_\delta^{\text{PS}} = t, \widehat{S}_{\tau_\delta^{\text{PS}}} \neq \mathcal{S}^\star\}\right) \\
&= \delta/2 + \mathbb{E}\left[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \mathbb{P}_\nu\left(\tau_\delta^{\text{PS}} = t | \mathcal{H}_t\right)\right] \\
&\leqslant \delta/2 + \mathbb{E}\Bigg[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \cdot \\
&\qquad \mathbb{P}_\nu\left(\forall m \leqslant M(t-1,\delta), \theta_t^m \notin \text{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)\Bigg] \\
&= \delta/2 + \mathbb{E}\Bigg[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \cdot \\
&\qquad \mathbb{P}_\nu\left(\theta_t \notin \text{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)^{M(t-1,\delta)}\Bigg] \\
&= \delta/2 + \mathbb{E}\Bigg[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \cdot \\
&\qquad \left(1 - \mathbb{P}_\nu\left(\theta_t \in \text{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)\right)^{M(t-1,\delta)}\Bigg]
\end{aligned}$$

which follows since the samples $\theta_t, \theta_t^1, \ldots, \theta_t^m$ are *i.i.d.* given $\mathcal{H}_t$. Further recalling that $1 - x \leqslant \exp(-x)$, it follows

$$\begin{aligned}
\mathbb{P}_\nu\left(\tau_\delta^{\text{PS}} < \infty, \widehat{S}_{\tau_\delta^{\text{PS}}} \neq \mathcal{S}^\star\right) \leqslant \delta/2 + \mathbb{E}\Bigg[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \cdot \\
\exp\left(-M(t-1,\delta)\mathbb{P}_\nu\left(\theta_t \in \text{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)\right)\Bigg].
\end{aligned} \tag{5.8}$$

□

Lemma 5.3.3 shows that choosing c, M such that

$$\mathbb{E}\left[\sum_{t\geqslant 1} \mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \exp\left(-M(t-1,\delta)\mathbb{P}_\nu\left(\theta_t \in \text{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)\right)\right] \leqslant \delta/2$$

will ensure the correctness of the PS stopping rule. However, it is not possible to pick $M(t-1,\delta) \propto \mathbb{P}_\nu(\theta_t \in \text{Alt}(\widehat{S}_t)|\mathcal{H}_t)^{-1}$ as the latter quantity is intractable due to multidimensional integration over a non-trivial domain. Instead, we will derive tractable lower bounds on $\mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \mathbb{P}_\nu(\theta_t \in \text{Alt}(\widehat{S}_t)|\mathcal{H}_t)$. This will ensure δ-correctness of the PS

stopping rule, regardless of the sampling rule. Let $\widehat{\Pi}_t \coloneqq \bigotimes_{k=1}^K \mathcal{N}(\hat{\mu}_{t,k}, c(t-1,\delta)\Sigma/N_{t,k})$. From the above display, it is sufficient to lower bound

$$\mathbb{1}\big(\mathcal{E}_{t,\delta/2} \cap \{\widehat{S}_t \neq \mathcal{S}^\star\}\big) \exp\Big(-M(t-1,\delta)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t))\Big)$$

by $\frac{\delta}{2\zeta(s)t^s}$ since their sum is smaller than $\delta/2$. Therefore, we should lower bound $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t))$ under the event $\mathcal{E}_{t,\delta/2} \cap \{\widehat{S}_t \neq \mathcal{S}^\star\}$ to conclude the proof.

Calibration of $M(t,\delta)$. Having $\widehat{S}_t \neq \mathcal{S}^\star$ implies that $\mu \in \mathrm{Alt}(\widehat{S}_t)$, see (5.4). Either (1) there exists $(i,j) \in \widehat{S}_t$ with $i \neq j$ such that $\mu_i(c) < \mu_j(c)$ for all $c \in [d]$, and we define $\mathcal{J}_t = \{j\}$. Or (2) there exists $i \notin \widehat{S}_t$ and $c \in [d]^{|\widehat{S}_t|}$ such that $\mu_j(c_j) < \mu_i(c_j)$ for all $j \in \widehat{S}_t$, and $\mathcal{J}_t = \widehat{S}_t$.

We can show that

$$\mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t)) \geqslant \prod_{k\in\{i\}\cup\mathcal{J}_t} \mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}\left(X > \sqrt{\frac{N_{t,k}}{c(t-1,\delta)}}(\mu_k - \hat{\mu}_{t,k})\right),$$

where $X > x$ denotes $X(c) > x(c)$ for all $c \in [d]$.

When Σ is diagonal, we have $\mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}(X > x) = \prod_{c\in[d]} \mathbb{P}_{X\sim\mathcal{N}(0,\Sigma_{c,c})}(X > x(c))$. When Σ is not diagonal, we derive a lower bound on $\mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}(X > x)$ for any vector $x \in \mathbb{R}^d$ (Lemma 5.5.3).

To further lower bound those quantities, we introduce the ratio of the tail distribution to the Gaussian density function, known as the Mills ratio (Mills 1926). Under the event $\mathcal{E}_{t,\delta/2}$, we have $\sum_{k=1}^K N_{t,k}\|\mu_k - \hat{\mu}_{t,k}\|^2_{\Sigma^{-1}} \leqslant 2\beta(t-1,\delta/2)$. Using the fact that the Mills ratio is non-increasing and log-convex (Baricz 2008), we obtain

$$\mathbb{1}\left(\mathcal{E}_{t,\delta/2} \cap \{\widehat{S}_t \neq \mathcal{S}^\star\}\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t)) \geqslant \delta q(t-1,\delta)$$

by taking $c(t,\delta) = \beta(t,\delta/2)/\log(1/\delta)$. This concludes the proof by considering $M(t,\delta) = \left\lceil \frac{\log(2t^s\zeta(s)/\delta)}{\delta q(t,\delta)} \right\rceil$.

In contrast, the δ-correctness of the GLR stopping rule is obtained quite simply by concentration. Indeed, we have $\mathrm{GLR}(t) \leqslant \sum_{k=1}^K \frac{N_{t,k}}{2}\|\hat{\mu}_{t,k} - \mu_k\|^2_{\Sigma^{-1}} \leqslant \beta(t-1,\delta)$ under $\mathcal{E}_{t,\delta} \cap \{\widehat{S}_t \neq \mathcal{S}^\star\}$, hence $\mathbb{P}_\nu(\mathcal{E}_\delta \cap \{\tau_\delta^{\mathrm{GLR}} < +\infty, \widehat{S}_{\tau_\delta^{\mathrm{GLR}}} \neq \mathcal{S}^\star\}) = 0$. Under a similar event, proving the δ-correctness of the PS stopping rule requires controlling the randomness of $\widehat{\Pi}_t|\mathcal{H}_t$ by deriving anti-concentration results on $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t))$.

These results are derived in Section 5.5.1 where Lemma 5.2.2 is proven.

Proof sketch of Theorem 5.3.1

Studying the expected sample complexity of an algorithm using the PS stopping rule requires controlling the randomness of $\widehat{\Pi}_t|\mathcal{H}_t$ by deriving concentration results on $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t))$

since a direct union bound yields that

$$\mathbb{P}_\nu(\tau_\delta^{\mathrm{PS}} > t) \leqslant M(t-1,\delta)\mathbb{E}_\nu[\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t))] \,.$$

Now, we invoke the following lemma to write $\mathrm{Alt}(\widehat{S}_t)$ as a finite union of convex sets.

Lemma 5.3.4 (Convex decomposition of alternative set; Crepon et al. 2024; Kone, Jourdan, et al. 2025). *For any parameter $\mu \in \mathcal{I}^K$ with $p \coloneqq |S^\star(\mu)|$, the alternative set Alt(μ) can be decomposed as the union of $n \coloneqq p(p-1) + (K-p)d^p$ convex subsets.*

We then combine this lemma with the result below on Gaussian concentration on convex sets.

Lemma 5.3.5 (Gaussian tail bound for convex sets; D. Lu & W. V. Li 2009). *Let $X \sim \mathcal{N}(\mu, A)$ with $\mu \in \mathbb{R}^n$ and covariance matrix A. For any convex set $C \subset \mathbb{R}^n$,*

$$\mathbb{P}(X \in C) \leqslant \frac{1}{2}\exp\left(-\frac{1}{2}\inf_{\lambda\in C}\|\lambda-\mu\|^2_{A^{-1}}\right) \,.$$

As $\widehat{\Pi}_t|\mathcal{H}_t$ is a Kd-sized Gaussian vector with block diagonal covariance, this result relates $\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t))$ to the GLR statistic defined in (5.2).

Lemma 5.3.4 proves that there exists n_t and convex sets $C_1, \dots, C_{n_t}$ such that $\mathrm{Alt}(\widehat{S}_t) = \cup_{i\in[n_t]} C_i$, and $\widehat{\Pi}_t|\mathcal{H}_t = \bigotimes_{k=1}^K \mathcal{N}(\hat{\mu}_{t,k}, c(t-1,\delta)\Sigma/N_{t,k})$, so, thanks to Lemma 5.3.5, and convexity of C_i, it follows that

$$\mathbb{P}(\theta_t \in C_i|\mathcal{H}_t) \leqslant \frac{1}{2}\exp\left(-\inf_{\theta\in C_i}\left[\sum_{k=1}^K N_{t,k}\frac{\|\hat{\mu}_{t,k}-\theta_k\|^2_{\Sigma^{-1}}}{2c(t-1,\delta)}\right]\right) \,.$$

Therefore, by union bound we have

$$\begin{aligned}
\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t)) &\coloneqq \mathbb{P}\left(\theta_t \in \cup_{i\in[n_t]}(C_i)|\mathcal{H}_t\right) \\
&\leqslant \frac{1}{2}\sum_{i\in[n_t]}\exp\left(-\inf_{\lambda\in C_i}\left[\sum_{k=1}^K N_{t,k}\frac{\|\hat{\mu}_{t,k}-\lambda_k\|^2_{\Sigma^{-1}}}{2c(t-1,\delta)}\right]\right) \\
&\leqslant \frac{n_t}{2}\exp\left(-\inf_{\lambda\in\mathrm{Alt}(\widehat{S}_t)}\left[\sum_{k=1}^K N_{t,k}\frac{\|\hat{\mu}_{t,k}-\lambda_k\|^2_{\Sigma^{-1}}}{2c(t-1,\delta)}\right]\right) \\
&= \frac{n_t}{2}\exp\left(-\frac{\mathrm{GLR}(t)}{c(t-1,\delta)}\right)
\end{aligned}$$

and from Lemma 5.3.4, $n_t = p_t(p_t-1) + (K-p_t)d^{p_t}$, with $p_t \coloneqq |\widehat{S}_t|$. From the above displays, we have

$$\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t)) \leqslant \alpha_0\exp(-\mathrm{GLR}(t)/c(t-1,\delta)) \,,$$

where $\alpha_0 \leqslant K(K+d^K)/2$. Now the event $\{\tau_\delta^{\text{PS}} > t\}$ is linked to the GLR statistic; it suffices to lower bound $\text{GLR}(t)$ to conclude the proof.

Studying the saddle-point convergence of our algorithm shows that there exist $(\Xi_t)_{t\geqslant 1}$ and $T_0 \in \mathbb{N}$ such that for all $t > T_0$, when Ξ_t holds, $\text{GLR}(t) \geqslant (t-1)/T^\star(\nu) - f(t)$ where $f(t) =_{+\infty} o(t)$ and $\sum_{t\in\mathbb{N}} \mathbb{P}(\mathcal{E}_t^{\mathsf{c}}) = C_0 < +\infty$. Then, we have

$$\mathbb{E}_\nu[\tau_\delta^{\text{PS}}] \leqslant T_0 + C_0 + \sum_{t>T_0} \alpha_0 \exp\left(-\frac{t-1}{T^\star(\nu)c(t-1,\delta)} + \frac{f(t)}{c(t-1,\delta)} + \log M(t-1,\delta)\right).$$

Using that $\limsup_{\delta\to 0} \frac{c(t,\delta)\log M(t,\delta)}{\log(1/\delta)} \leqslant 1$, a direct inversion of the above sum concludes the proof.

The complete proof is given at the end of the chapter.

5.4 Numerical study and discussion

Experimental setup. We evaluate the performance of PSIPS on real-world-inspired and synthetic instances. As benchmarks, we consider the following PSI algorithms: APE (Chapter 3), the gradient-based algorithm of Crepon et al. 2024 denoted as GAPSI, the *oracle* algorithm which samples arms according to the optimal weights, *i.e.*, $A_t \sim w^\star(\nu)$, and round-robin *uniform* sampling (RR).

APE relies on confidence bounds to pull an arm at each round and to define its stopping rule. While APE can be extended to the structured setting, it does not exploit the correlation between objectives as its confidence intervals only consider the marginals' variance. GAPSI computes a supergradient of the GLR (5.2) and uses the GLR stopping rule. Likewise, GAPSI has no efficient implementation or guarantees for correlated objectives.

Since GAPSI is computationally expensive, we exclude it from the experiment in the linear setting (due to the large number of arms), and we run it as heuristic in experiments with correlated objectives.

In our experiments, we use the heuristics $c(t,\delta) = 1 + \frac{\log\log t}{\log(1/\delta)}$ and $M(t,\delta) = \frac{1}{\delta}\log\frac{t}{\delta}$. For a fair comparison, APE uses $\beta(t,\delta) = \log(1/\delta) + \log\log t$ in its bounds, hence improving its performance. We report the averaged results over 500 independent runs with boxplots or shaded areas for standard deviation. The observed empirical error is an order of magnitude lower than δ.

CovBoost trial. As in prior PSI work, we use the dataset from Munro et al. 2021 to simulate a bandit instance for PSI. The dataset contains average responses for patient cohorts in a COVID-19 trial, including 20 vaccine strategies. Among the indicators recorded to measure the efficacy of each strategy, following Kone, Kaufmann, et al. 2023; Crepon et al. 2024, we keep three of them: titers of neutralizing antibodies and immunoglobulin G and the wild-type cellular response (see Table 7.1). The ideal vaccine strategy would maximize all three indicators, but the average response for the considered metrics reveals a Pareto set of two vaccine strategies in the trial.

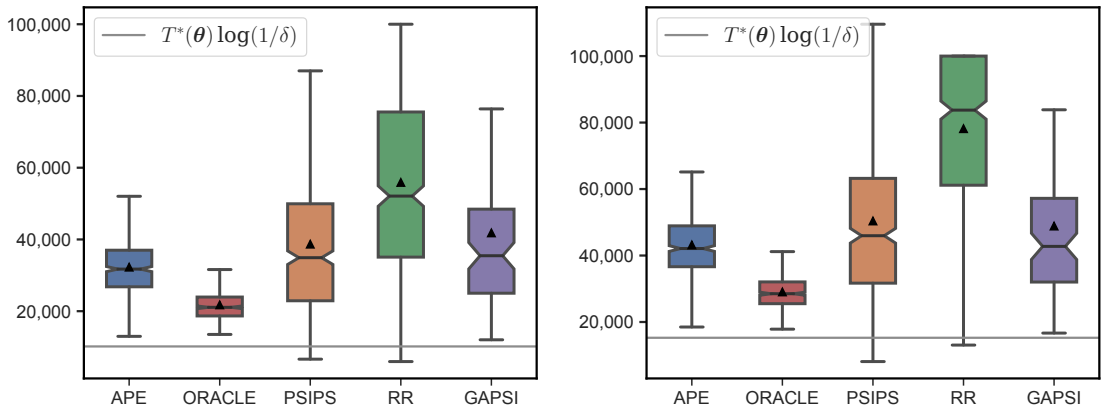


Figure 5.1: Empirical stopping times on the covid19 experiment with $\delta = 0.01$ (left) and $\delta = 0.001$ (right).

Figure 5.1 shows that PSIPS has a high variability of stopping time, which was expected due to the use of posterior sampling. While PSIPS outperforms uniform sampling, it performs on average slightly worse than APE and oracle on the CovBoost instance. For $\delta = 0.1$, we reported an average sample complexity of 20456 for PSIPS compared to 17909 reported for GAPSI by Crepon et al. 2024. However, due to the high computational cost of their algorithm, we averaged its performance over 100 runs.

Correlated Objectives. We evaluate the impact of correlated objectives on the performance of PSIPS. Consider a 5-arm Gaussian instance with means $\mu_1 = (0.73, 1.20)^\top$, $\mu_2 = (0.45, -0.63)^\top$, $\mu_3 = (0.63, 1.28)^\top$, $\mu_4 = (0.94, 2.31)^\top$, and $\mu_5 = (2.08, 1.48)^\top$. The noise covariance matrix Σ_ρ has unit diagonal and constant off-diagonal entries equal to $\rho \in (-1, 1)$, controlling the correlation between objectives ($\rho > 0$ for positive and $\rho < 0$ for negative correlation). Using an approximate convex solver and the exponentiated gradient algorithm, we estimate the theoretical complexity $T^\star_{\Sigma_\rho}(\nu)$ defined in Lemma 5.1.1. As shown in Figure 5.3, negative correlations reduce the theoretical complexity (characteristic time) compared to the uncorrelated case ($\rho = 0$), reflecting the additional information gained from anti-correlated objectives.

Figure 5.2 shows the sample complexity of PSIPS versus ρ, including GAPSI as heuristic. It reveals that PSIPS has a decreasing sample complexity for stronger (negative) correlation between objectives, reducing by up to a factor of 3 relative to the uncorrelated case ($\rho = 0$). Due to its asymptotic optimality, PSIPS inherits the properties of $T^\star(\nu)$, which appears to decrease with negative ρ on this specific instance. Both GAPSI and APE's performance are independent of ρ.

Robustness on random instances. To evaluate the robustness of PSIPS, we measure its performance on 250 randomly picked Bernoulli (marginals) and Gaussian instances with $K = 5, d = 2$ and $\Sigma = I_2/2$. Figure 5.4 demonstrates the robustness of PSIPS, which remains competitive on subgaussian instances.

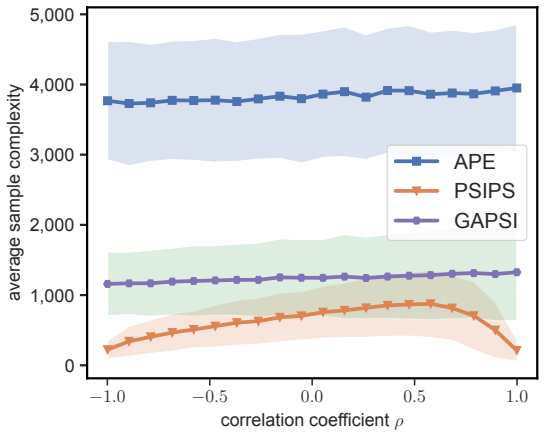


Figure 5.2: Empirical impact of the correlation ρ on the sample complexity.

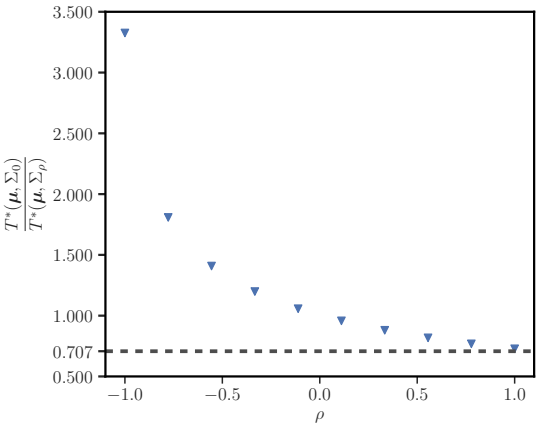


Figure 5.3: Theoretical impact of the correlation coefficient on the complexity $T^\star$.

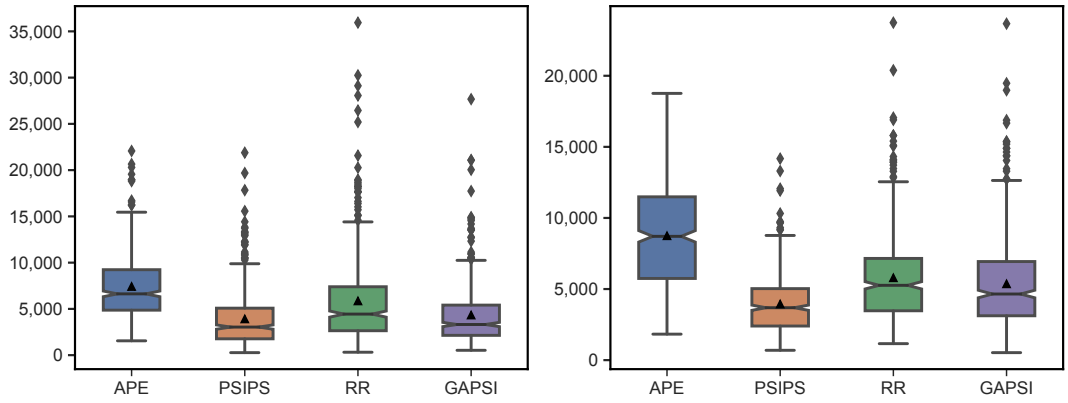


Figure 5.4: Empirical stopping time on random Gaussian (left) and Bernoulli (right) instances.

Computational cost. We evaluate the number of rejection samples used by PSIPS on a Gaussian instance with $\Sigma = I_2/2$ defined as $\mu_1 = (1, 1)^\intercal$ and $\mu_i = R_{\pi/5}\mu_{i-1}$ for $i \in \{2, 3, 4, 5\}$ where $R_{\pi/5}$ is the $\pi/5$ rotation matrix. Without actually stopping, we record both m_t and $m_{t,\delta}$ at each round, averaged over 1000 runs. We also report the cumulative runtime of PSIPS and APE on the CovBoost experiment.

Figure 5.5 shows that the computational cost of PSIPS is comparable to APE, which is orders of magnitudes smaller than GAPSI (see Figure 4 in Crepon et al. 2024). Figure 5.6 reveals that, while finding an alternative is initially faster, more rejections are needed before finding an alternative when the posterior concentrates. It is precisely at this time that the PS stopping rule triggers.

Discussion. We proposed the PSIPS algorithm for Pareto set identification with correlated objectives. By leveraging posterior sampling in both the stopping and the sampling rules, PSIPS is a computationally efficient algorithm that deals with structure and correlation, without the costly oracle calls required by existing algorithms. We show that PSIPS is asymptotically optimal both from a frequentist and Bayesian perspective. Moreover, PSIPS

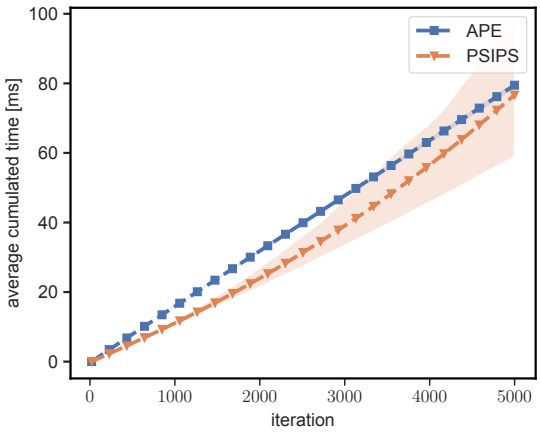


Figure 5.5: Average runtime for the first T iterations in the COVID-19 experiment.

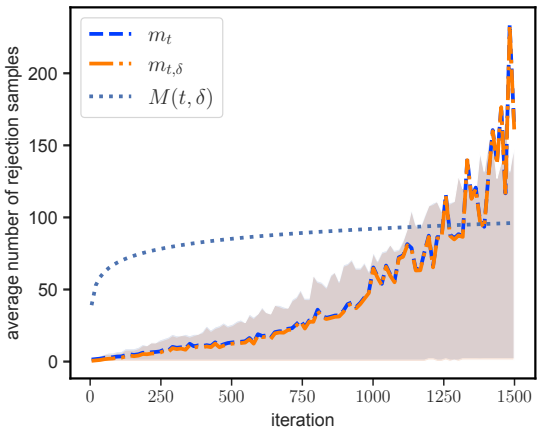


Figure 5.6: Average per-round number of rejection samples $m_t, m_{t,\delta}$.

has competitive empirical performance.

Despite being a theoretically convenient distributional assumption, using Gaussian distributions with known covariance matrix is too restrictive for many applications. Therefore, generalizing our results to other classes of distributions is another interesting direction for future work. For example, when the covariance matrices are unknown and different for each arm, the estimation of the Pareto set is intertwined with estimating the correlation structure.

5.5 Additional proofs

We prove detail additional results in this section. Complementary proofs can be found in Kone, Jourdan, et al. 2025.

5.5.1 Stopping rule

In this section, we prove the correctness of the PS (Posterior Sampling) stopping rule. We recall the definition of $\tau_\delta^{\mathrm{PS}}$:

$$\tau_\delta^{\mathrm{PS}} := \inf\left\{t | \forall m \in [M(t-1,\delta)],\ \lambda_t^m \notin \mathrm{Alt}(\widehat{S}_t)\right\},$$

where, conditioned on $\mathcal{H}_t, \theta_t, \theta_t^1, \dots, \theta_t^m$ are *i.i.d.* samples from $\mathcal{N}(\hat{\mu}_{t,k}, c(t-1,\delta)\Sigma/N_{t,k})$. To prove the correctness of this stopping rule, we will have to control the randomness in the posterior, which we do by introducing the concentration event $\mathcal{E}_\delta := \bigcap_{t\in\mathbb{N}} \mathcal{E}_{t,\delta}$ with

$$\mathcal{E}_{t,\delta} := \left\{\sum_{k=1}^{K} N_{t,k} \left\|\mu_k - \hat{\mu}_{t,k}\right\|^2_{\Sigma^{-1}} \leqslant 2\beta(t-1,\delta)\right\},$$

for which, correct values of the thresholds are given in the lemma below.

The following result shows that to prove the δ-correctness, it is sufficient to show some anti-concentration of the posterior under the event $\mathcal{E}$. $\mathbb{P}_\nu(\mathcal{E}_{\delta/2} \cap \{\tau_\delta^{\mathrm{PS}} < +\infty, \widehat{S}_{\tau_\delta^{\mathrm{PS}}} \neq \mathcal{S}^\star\}) \leqslant \delta/2$.

Next, we calibrate c, M to ensure the correctness of the PS stopping rule. The quantity to control is

$$\mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right) \mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right) \mathbb{P}_\nu(\theta_t \in \mathrm{Alt}(\widehat{S}_t)|\mathcal{H}_\infty) \text{ with } (\theta_t - \widehat{\theta}_t)|\mathcal{H}_t \sim \bigotimes_{k=1}^{K} \mathcal{N}(0_d, c(t-1,\delta)\Sigma/N_{t,k}),$$

and

$$\mathrm{Alt}(\widehat{S}_t) = \left(\bigcup_{i \neq j \in \widehat{S}_t^2} \left\{\lambda \coloneqq (\lambda_1, \dots, \lambda_K) \in \mathcal{I}^K | \lambda_i \prec \lambda_j\right\} \right) \bigcup \left(\bigcup_{i \notin \widehat{S}_t} \left\{\lambda \coloneqq (\lambda_1, \dots, \lambda_K) \in \mathcal{I}^K | \forall j \in \widehat{S}_t, \lambda_i \not\prec \lambda_j\right\} \right).$$

Indeed, recalling that $\mathrm{Alt}(S)$ is the set of parameters for which the Pareto set differs from S. To change the Pareto set, we either add an arm or remove one from it. For a parameter instance $\mu \coloneqq (\mu_1 \dots \mu_K)^\intercal$, if $i \in S$ and there exists $j \in S$ such that $\mu_i \prec \mu_j$, then as i is a dominated arm in the instance μ, we will have $S^\star(\mu) \neq S$. Similarly, if for some $i \notin S$ it holds that for all $j \in S$, $\mu_i \not\prec \mu_j$, then we cannot have $S^\star(\mu) = S$, otherwise, as $i \notin \mathcal{S}^\star(\mu)$, an arm from $S^\star(\mu) = S$ would have dominated i. This is formally shown in Lemma 5.3.4, where it is further proven that

$$\mathrm{Alt}(S) = \left(\bigcup_{i \in S^c} \bigcup_{\bar{d}^i \in [d]^S} \left\{\lambda | \forall j \in S, \lambda_i(d^i(\sigma(j))) \geqslant \lambda_j(d^i(\sigma(j)))\right\} \right) \bigcup \left(\bigcup_{(i,j) \in S^2, i \neq j} \left\{\lambda \in \mathcal{I}^K | \lambda_i \prec \lambda_j\right) \right)$$

where σ is any permutation that maps S onto $\{1, \dots, |S|\}$. In the sequel, we let

$$\begin{aligned} \mathrm{Alt}^-(S) &\coloneqq \bigcup_{(i,j) \in S^2, i \neq j} \left\{\lambda | \lambda_i \prec \lambda_j\right\} \quad \text{and} \\ \mathrm{Alt}^+(S) &\coloneqq \bigcup_{i \in S^c} \bigcup_{\bar{d}^i \in [d]^S} \left\{\lambda | \forall j \in S : \lambda_i(d^i(\sigma(j))) \geqslant \lambda_j(d^i(\sigma(j)))\right\}, \end{aligned}$$

so that $\mathrm{Alt}(S) = \mathrm{Alt}^+(S) \cup \mathrm{Alt}^-(S)$. We prove the following lemma.

Lemma 5.5.1. *Letting $Y \sim\sim \mathcal{N}(0_d, \Sigma)$, it holds that*

$$\mathbb{1}\big(\widehat{S}_t \neq \mathcal{S}^\star\big)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathit{Alt}(\widehat{S}_t)) \geqslant \max\Bigg\{ \max_{i\notin\widehat{S}_t, \bar{d}\in[d]^{|\widehat{S}_t|}} \mathbb{1}\Bigg(\forall j \in \widehat{S}_t, \mu_i(\bar{d}_j^\sigma) \geqslant \mu_j(\bar{d}_j^\sigma)\Bigg) \cdot\mathbb{P}_Y\Bigg(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \hat{\mu}_{t,i})\Bigg)\cdot \prod_{j\in\widehat{S}_t} \mathbb{P}_{X\sim\mathcal{N}(0,1)}\Bigg(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \hat{\mu}_{t,j}\|_{\Sigma^{-1}}\Bigg),$$
$$\max_{i\neq j\in\widehat{S}_t^2,} \mathbb{1}\left(\mu_i \prec \mu_j\right)\mathbb{P}_Y\Bigg(Y \geqslant \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}}\left((\hat{\mu}_{t,i} - \hat{\mu}_{t,j}) - (\mu_i - \mu_j)\right)\Bigg)\Bigg\}.$$

Proof.

$$\begin{aligned}
\mathbb{1}\Big(\widehat{S}_t \neq \mathcal{S}^\star\Big)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}^-(\widehat{S}_t)) &\geqslant \mathbb{1}\Big(\widehat{S}_t \neq \mathcal{S}^\star\Big)\max_{(i,j)\in\widehat{S}_t, i\neq j}\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\{\mu\in\mathcal{I}^K : \mu_i \prec \mu_j\}) \\
&\geqslant \max_{(i,j)\in\widehat{S}_t, i\neq j}\mathbb{1}\Big(\mu_i \prec \mu_j\Big)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\{\mu\in\mathcal{I}^K : \mu_i \prec \mu_j\}) \\
&= \max_{(i,j)\in\widehat{S}_t, i\neq j}\mathbb{1}\left(\mu_i \prec \mu_j\right)\mathbb{P}(\theta_{t,i} \prec \theta_{t,j}|\mathcal{H}_t),
\end{aligned}$$

where we have $\theta_{t,i}|\mathcal{H}_t \sim \mathcal{N}(\hat{\mu}_{t,i}, c(t-1,\delta)\Sigma/N_{t,i})$. Therefore, we have

$$\begin{aligned}
\mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}^-(\widehat{S}_t)) &\geqslant \max_{(i,j)\in\widehat{S}_t, i\neq j}\mathbb{1}\left(\mu_i \prec \mu_j\right)\mathbb{P}((\theta_{t,i} - \theta_{t,j}) - (\hat{\mu}_{t,i} - \hat{\mu}_{t,j}) \\
&\quad \prec -(\hat{\mu}_{t,i} - \hat{\mu}_{t,j})|\mathcal{H}_t) \\
&\geqslant \max_{(i,j)\in\widehat{S}_t, i\neq j}\mathbb{1}\left(\mu_i \prec \mu_j\right)\mathbb{P}((\theta_{t,i} - \theta_{t,j}) - (\hat{\mu}_{t,i} - \hat{\mu}_{t,j}) \prec \\
&\quad (\mu_i - \mu_j) - (\hat{\mu}_{t,i} - \hat{\mu}_{t,j})|\mathcal{H}_t)
\end{aligned}$$

then, observe that $(\theta_{t,i} - \theta_{t,j})|\mathcal{H}_t \sim \mathcal{N}\left(\hat{\mu}_{t,i} - \hat{\mu}_{t,j}, (N_{t,i}^{-1} + N_{t,j}^{-1})c(t-1,\delta)\Sigma\right)$, so that letting $Y \sim \mathcal{N}(0_d, \Sigma)$, we have

$$\begin{aligned}
&\mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}^-(\widehat{S}_t)) \\
&\geqslant \max_{i\neq j\in\widehat{S}_t^2} \mathbb{1}\left(\mu_i \prec \mu_j\right)\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\Bigg(Y \prec \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}}\left((\mu_i - \mu_j) - (\hat{\mu}_{t,i} - \hat{\mu}_{t,j})\right)\Bigg) \\
&\geqslant \max_{i\neq j\in\widehat{S}_t^2,} \mathbb{1}\left(\mu_i \prec \mu_j\right)\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\Bigg(Y \geqslant \sqrt{\left(\frac{1}{N_{t,i}} + \frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}}\left((\hat{\mu}_{t,i} - \hat{\mu}_{t,j}) - (\mu_i - \mu_j)\right)\Bigg).
\end{aligned}$$

We now prove a similar result on $\text{Alt}^+(\widehat{S}_t)$. We have

$$
\begin{aligned}
&\mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\text{Alt}^+(\widehat{S}_t)) \\
&\geqslant \max_{i\notin\widehat{S}_t,\bar{d}\in[d]^{|S_t|}} \mathbb{1}\left(\forall j\in\widehat{S}_t, \mu_i(\bar{d}_j^\sigma)\geqslant\mu_j(\bar{d}_j^\sigma)\right)\mathbb{P}(\forall j\in\widehat{S}_t, \theta_{t,i}(\bar{d}_j^\sigma)\geqslant\theta_{t,j}(\bar{d}_j^\sigma)|\mathcal{H}_t) \\
&\geqslant \max_{i\notin\widehat{S}_t,\bar{d}\in[d]^{|\widehat{S}_t|}} \mathbb{1}\left(\forall j\in\widehat{S}_t, \mu_i(\bar{d}_j^\sigma)\geqslant\mu_j(\bar{d}_j^\sigma)\right)\mathbb{P}(\forall j\in\widehat{S}_t, (\theta_{t,i}-\mu_i)(\bar{d}_j^\sigma)\geqslant(\theta_{t,j}-\mu_j)(\bar{d}_j^\sigma)|\mathcal{H}_t) \\
&\geqslant \max_{i\notin\widehat{S}_t,\bar{d}\in[d]^{|\widehat{S}_t|}} \mathbb{1}\left(\forall j\in\widehat{S}_t, \mu_i(\bar{d}_j^\sigma)\geqslant\mu_j(\bar{d}_j^\sigma)\right)\mathbb{P}((\theta_{t,i}-\mu_i)>0_d|\mathcal{H}_t)\prod_{j\in\widehat{S}_t}\mathbb{P}_\nu((\theta_{t,j}-\mu_j)(\bar{d}_j^\sigma)>0|\mathcal{H}_t)
\end{aligned}
$$

then, noting again that $(\theta_{t,i}-\hat{\mu}_{t,i})|\mathcal{H}_t\sim\mathcal{N}(0, c(t-1,\delta)\Sigma/N_{t,i})$, it follows that

$$
\begin{aligned}
\mathbb{1}\left(\widehat{S}_t\neq\mathcal{S}^\star\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\text{Alt}^+(\widehat{S}_t)) \geqslant \max_{i\notin\widehat{S}_t,\bar{d}\in[d]^{|\widehat{S}_t|}} \mathbb{1}\left(\forall j\in\widehat{S}_t, \mu_i(\bar{d}_j^\sigma)\geqslant\mu_j(\bar{d}_j^\sigma)\right)\cdot \\
\mathbb{P}_Y\left(Y>\sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i-\hat{\mu}_{t,i})\right)\prod_{j\in\widehat{S}_t}\mathbb{P}_\nu((\theta_{t,j}-\mu_j)(\bar{d}_j^\sigma)>0|\mathcal{H}_t).
\end{aligned}
\tag{5.9}
$$

Further noting that by Cauchy-Schwarz inequality,

$$
(\hat{\mu}_{t,j}-\mu_j)(\bar{d}_j^\sigma) \;\leqslant\; \|\widetilde{e}_{\bar{d}_j^\sigma}\|_\Sigma\|\mu_j-\hat{\mu}_{t,j}\|_{\Sigma^{-1}},
$$

it follows for any $j\in\widehat{S}_t$ that

$$
\begin{aligned}
\mathbb{P}((\theta_{t,j}-\mu_j)(\bar{d}_j^\sigma)>0|\mathcal{H}_t) &= \mathbb{P}((\theta_{t,j}-\hat{\mu}_{t,j})(\bar{d}_j^\sigma)>(\mu_j-\hat{\mu}_{t,j})(\bar{d}_j^\sigma)|\mathcal{H}_t) \\
&\geqslant \mathbb{P}((\theta_{t,j}-\widehat{\mu}_{t,j})(\bar{d}_j^\sigma)>\|\widetilde{e}_{\bar{d}_j^\sigma}\|_\Sigma\|\mu_j-\hat{\mu}_{t,j}\|_{\Sigma^{-1}}|\mathcal{H}_t) \\
&= \mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X>\sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j-\hat{\mu}_{t,j}\|_{\Sigma^{-1}}\right),
\end{aligned}
$$

which follows from $(\theta_{t,j}-\hat{\mu}_{t,j})(\bar{d}_j^\sigma)|\mathcal{H}_t\sim\mathcal{N}(0, c(t-1,\delta)\|\widetilde{e}_{\bar{d}_j^\sigma}\|/N_{t,\delta})$. Therefore, combining the last display with (5.9) yields

$$
\begin{aligned}
\mathbb{1}\big(\widehat{S}_t\neq\mathcal{S}^\star\big)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\text{Alt}^+(\widehat{S}_t)) \geqslant \max_{i\notin\widehat{S}_t,\bar{d}\in[d]^{|\widehat{S}_t|}} \mathbb{1}\left(\forall j\in\widehat{S}_t, \mu_i(\bar{d}_j^\sigma)\geqslant\mu_j(\bar{d}_j^\sigma)\right) \\
\cdot\,\mathbb{P}_Y\left(Y>\sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i-\hat{\mu}_{t,i})\right)\cdot\prod_{j\in\widehat{S}_t}\mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X>\sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j-\hat{\mu}_{t,j}\|_{\Sigma^{-1}}\right),
\end{aligned}
$$

which completes the proof. □

Special Case: PSI with Independent Marginals In this section, we specialize Lemma 5.5.1 to the case where Σ is diagonal. In this case, $\Sigma = \mathrm{diag}(\{\sigma_c^2\}_{c\in[d]})$ and

$$(*) = \mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y \geqslant \sqrt{\left(\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}}\left((\hat{\mu}_{t,i}-\hat{\mu}_{t,j})-(\mu_i-\mu_j)\right)\right) =$$

$$\prod_{c\in[d]}\mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X > \underbrace{\sqrt{\left(\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}\right)^{-1}\frac{1}{\sigma_c^2 c(t-1,\delta)}}\left((\hat{\mu}_{t,i}-\hat{\mu}_{t,j})-(\mu_i-\mu_j)\right)(c)}_{Z_{i,j}(c)}\right),$$

which we may rewrite with Mills ratio as $R(x) \coloneqq \frac{\mathbb{P}(X>x)}{f_X(x)}$ so

$$\mathbb{P}(X>x) = R(x)\exp(-\frac{1}{2}x^2)\frac{1}{\sqrt{2\pi}} = \widetilde{R}(x)\exp(-\frac{1}{2}x^2),$$

and

$$\begin{aligned}(*) &= \exp\left(-\frac{1}{2}\sum_{c\in[d]}Z_{i,j}(c)\right)\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c))\\ &= \exp\left(-\left(\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}\sum_{c\in[d]}\frac{\left((\hat{\mu}_{t,i}-\hat{\mu}_{t,j})-(\mu_i-\mu_j)\right)(c)^2}{2\sigma_c^2}\right)\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c))\end{aligned}$$

and by Cauchy-Schwarz inequality,

$$\left|\left((\hat{\mu}_{t,i}-\hat{\mu}_{t,j})-(\mu_i-\mu_j)\right)(c)\right| \leqslant \sqrt{\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}}\sqrt{N_{t,i}(\hat{\mu}_{t,i}-\mu_i)(c)^2 + N_{t,j}(\mu_j-\hat{\mu}_{t,j})(c)^2}.$$

Therefore

$$(*) \geqslant \exp\left(-\frac{1}{c(t-1,\delta)}\sum_{k\in\{i,j\}}\frac{N_{t,k}}{2}\|\hat{\mu}_{t,k}-\mu_k\|^2_{\Sigma^{-1}}\right)\prod_{c\in[d]}\widetilde{R}(Z_{i,j}(c)), \tag{5.10}$$

and using the following lemma,

Lemma 5.5.2. *The Mills ratio R is decreasing, log-convex and satisfies for all $(x_1,\dots,x_p)\in\mathbb{R}^p$,*

$$R\left(\frac{1}{p}\sum_{i=1}^p x_i\right)^p \leqslant \prod_{i=1}^p R(x_i).$$

we obtain

$$\begin{aligned}\prod_{c\in[d]} \widetilde{R}(Z_{i,j}(c)) &\geqslant \widetilde{R}\left(\frac{1}{d}\sum_{c\in[d]} Z_{i,j}(c)\right)^d \\ &\overset{(a)}{\geqslant} \widetilde{R}\left(\frac{1}{d}\sum_{c\in[d]} \sqrt{\frac{1}{\sigma_c^2 c(t-1,\delta)}}\sqrt{N_{t,i}(\hat{\mu}_{t,i}-\mu_i)(c)^2 + N_{t,j}(\mu_j-\widehat{\mu}_{t,j})(c)^2}\right)^d \\ &\overset{(b)}{\geqslant} \widetilde{R}\left(\frac{\sqrt{2}}{\sqrt{c(t-1,\delta)d}}\sqrt{\sum_{k\in\{i,j\}}\frac{N_{t,k}}{2}\|\hat{\mu}_{t,k}-\mu_k\|_{\Sigma^{-1}}^2}\right)^d,\end{aligned}$$

where (a) follows since $\widetilde{R}$ is decreasing (Lemma 5.5.2) and (b) follows from this monotonicity and Cauchy-Schwarz. Combining the last display with (5.10), we prove that on the event $\mathcal{E}_{t,\delta/2}$, we have

$$(*) \geqslant \exp\left(-\frac{\beta(t-1,\delta/2)}{c(t-1,\delta)}\right)\widetilde{R}\left(\sqrt{\frac{2\beta(t-1,\delta/2)}{c(t-1,\delta)d}}\right)^d$$

Choosing

$$c(t,\delta) \coloneqq \frac{\beta(t,\delta/2)}{\log(1/\delta)}$$

yields,

$$(*) \geqslant \delta\widetilde{R}\left(\sqrt{\frac{2\log(1/\delta)}{d}}\right)^d = r(\delta,d) \quad \text{with} \quad r(\delta,n) = \widetilde{R}\left(\sqrt{\frac{2\log(1/\delta)}{n}}\right)^n$$

and proceeding identically, we prove that under the event $\mathcal{E}_{t,\delta/2}$,

$$\begin{aligned}&\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i-\hat{\mu}_{t,i})\right)\prod_{j\in\widehat{S}_t}\mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j-\hat{\mu}_{t,j}\|_{\Sigma^{-1}}\right) \\ &\geqslant \delta\widetilde{R}\left(\sqrt{\frac{2\log(1/\delta)}{d+|\widehat{S}_t|}}\right)^{d+|\widehat{S}_t|}.\end{aligned}$$

All these combined, we proved that

$$\begin{aligned}\mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right)\mathbb{1}\left(\widehat{S}_t \neq \mathcal{S}^\star\right)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t)) &\geqslant \delta\max\Big\{\mathbb{1}\left(\mu\in\mathrm{Alt}^-(\widehat{S}_t)\right)r(\delta,d), \\ &\quad \mathbb{1}\left(\mu\in\mathrm{Alt}^+(\widehat{S}_t)\right)r(\delta,d+|\widehat{S}_t|)\Big\} \\ &\geqslant \delta\min\{r(\delta,d), r(\delta,d+|\widehat{S}_t|)\}\end{aligned}$$

which follows since $\widehat{S}_t \neq \mathcal{S}^\star \Longrightarrow \mu \in \mathrm{Alt}(\widehat{S}_t) = \mathrm{Alt}^+(\widehat{S}_t) \cup \mathrm{Alt}^-(\widehat{S}_t)$.

The conclusion is immediate by taking $M(t,\delta) = \left\lceil \frac{\log(2t^s\zeta(s)/\delta)}{\delta q(t,\delta)} \right\rceil$ where

$$q(t,\delta) = \min\left\{ r(\delta, d), r(\delta, d + |\widehat{S}_t|) \right\}.$$

Moreover, it is known (Birnbaum 1942) that for $x \geqslant 0$,

$$R(x) \geqslant \frac{2}{x + \sqrt{x^2+4}}$$

and $R(x) \sim \frac{1}{x}, x \to +\infty$. The Mills ratio of the standard normal is implemented in common scientific libraries. To compute $M(t,\delta)$, we just need to compute at most the values $(r(\delta, d+k))_{k\in[K]\cup\{0\}}$ at the initialisation of the algorithm. Using the lower bound on the Mills ratio, one can easily check that

$$\frac{\log(M(t,\delta))c(t,\delta)}{\log(1/\delta)} \underset{\delta\to 0}{\leqslant} 1.$$

Proof of Lemma 5.5.2. The monotonicity is well-known and simple to prove by noting that

$$\begin{aligned} R(x) &:= \exp\left(x^2/2\right) \int_x^\infty \exp\left(-t^2/2\right) dt \\ &= \exp\left(x^2/2\right) \int_0^\infty \exp\left(-(t+x)^2/2\right) dt \\ &= \int_0^\infty \exp\left(-tx - t^2/2\right) dt, \end{aligned}$$

from which we have $R(x+\alpha^2) = \int_0^\infty \exp(-t\alpha^2)\exp\left(-tx - t^2/2\right) dt < R(x)$. The log-convexity is proven in Theorem 2.5 of Baricz 2008. From the log-convexity and using Jensen inequality, we have

$$\begin{aligned} \log R\left(\frac{1}{p}\sum_{i=1}^p x_i\right) &\leqslant \frac{1}{p}\sum_{i=1}^p \log R\left(x_i\right) \\ &= \frac{1}{p}\log\left(\prod_{i=1}^p R(x_i)\right) \end{aligned}$$

which by monotonicity of the $\log$ proves the claimed statement. □

PSI with a full covariance matrix. We specialize Lemma 5.5.1 to the case where the covariance Σ is non-diagonal. To derive this result, we will use the following lemma.

Lemma 5.5.3. *Let Σ a covariance matrix, V diagonal matrix such that $(V - \Sigma^{-1})$ is psd, and define $d_\Sigma := \|1_d\|^2_{(V^{-1/2}\Sigma^{-1}V^{-1/2})}$. Then, for all $x \in \mathbb{R}^d$, it holds that*

$$\mathbb{P}_{X\sim\mathcal{N}(0_d,\Sigma)}\left(X \geqslant x\right) \geqslant (2\pi)^{-d/2}\det(V\Sigma)^{-1/2}\exp\left(-\frac{1}{2}x^\intercal\Sigma^{-1}x\right)\prod_{c\in[d]} R(e_c^\intercal V^{-1/2}\Sigma^{-1}x),$$

and in particular, $\prod_{c\in[d]} R(e_c^\intercal V^{-1/2}\Sigma^{-1}x) \geqslant R\left(\|x\|_{\Sigma^{-1}}\sqrt{d_\Sigma}/d\right)^d$.

Applying this result with $V = \bar{\sigma} I_d$ where $\bar{\sigma} = \||\Sigma^{-1}|\|$, we have

$$\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \hat{\mu}_{t,i})\right) \geqslant \frac{1}{\sqrt{(2\pi)^d \det(\bar{\sigma}\Sigma)}} \exp\left(-\frac{1}{2}\|\mu_i - \hat{\mu}_{t,i}\|^2_{\Sigma^{-1}}\right) \cdot \prod_{c\in[d]} R(e_c^\intercal V^{-1/2}\Sigma^{-1}u_t),$$

with $u_t \coloneqq \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \hat{\mu}_{t,i})$, therefore, letting

$$(**) \coloneqq \mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y > \sqrt{\frac{N_{t,i}}{c(t-1,\delta)}}(\mu_i - \hat{\mu}_{t,i})\right) \prod_{j\in\widehat{S}_t} \mathbb{P}_{X\sim\mathcal{N}(0,1)}\left(X > \sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \hat{\mu}_{t,j}\|_{\Sigma^{-1}}\right)$$

we have

$$(**) \geqslant \det(\bar{\sigma}\Sigma)^{-1/2} \exp\left(-\sum_{k\in\{i\}\cup\widehat{S}_t} \frac{1}{2}\|\mu_k - \hat{\mu}_{t,k}\|^2_{\Sigma^{-1}}\right) \cdot \prod_{j\in\widehat{S}_t} \widetilde{R}\left(\sqrt{\frac{N_{t,j}\|\mu_j - \hat{\mu}_{t,j}\|^2_{\Sigma^{-1}}}{c(t-1,\delta)}}\right) \prod_{c\in[d]} \widetilde{R}(e_c^\intercal V^{-1/2}\Sigma^{-1}u_t),$$

then note that by Lemma 5.5.2, we have

$$\prod_{j\in\widehat{S}_t} \widetilde{R}\left(\sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \hat{\mu}_{t,j}\|_{\Sigma^{-1}}\right) \prod_{c\in[d]} \widetilde{R}(e_c^\intercal V^{-1/2}\Sigma^{-1}u_t/d) \geqslant \widetilde{R}\left(\frac{1_d^\intercal V^{-1/2}\Sigma^{-1}u_t + 1_{|\widehat{S}_t|}^\intercal h_t}{d + |\widehat{S}_t|}\right)^{d+|\widehat{S}_t|}$$

with $h_t \coloneqq \left(\sqrt{\frac{N_{t,j}}{c(t-1,\delta)}}\|\mu_j - \hat{\mu}_{t,j}\|_{\Sigma^{-1}}\right)_{j\in|\widehat{S}_t|}$ and by Cauchy-Swcharz inequality, replacing $V = \bar{\sigma} I_d$,

$$\begin{aligned}
1_d^\intercal V^{-1/2}\Sigma^{-1}u_t + 1_{|\widehat{S}_t|}^\intercal h_t &\leqslant \sqrt{\|1_d\|^2_{(\bar{\sigma}\Sigma)^{-1}} + \|1_{|\widehat{S}_t|}\|^2}\sqrt{\|u_t\|^2_{\Sigma^{-1}} + \|h_t\|^2} \\
&= \sqrt{d_\Sigma + |\widehat{S}_t|}\left(\sqrt{\frac{1}{c(t-1,\delta)}\sum_{k\in\{i\}\cup\widehat{S}_t} N_{k,t}\|\mu_k - \hat{\mu}_{t,k}\|^2_{\Sigma^{-1}}}\right) \\
&\leqslant \sqrt{d_\Sigma + |\widehat{S}_t|}\sqrt{\frac{2\beta(t-1,\delta/2)}{c(t-1,\delta)}}
\end{aligned}$$

where the last inequality follows on the event $\mathcal{E}_{t,\delta/2}$. Combining these displays with $c(t,\delta) \coloneqq \frac{\beta(t,\delta/2)}{\log(1/\delta)}$, we have on the event $\mathcal{E}_{t,\delta/2}$,

$$\begin{aligned}
(**) &\geqslant \det(\bar{\sigma}\Sigma)^{-1/2}\delta\widetilde{R}\left(\frac{\sqrt{d_\Sigma + |\widehat{S}_t|}}{d + |\widehat{S}_t|}\sqrt{2\log(1/\delta)}\right) \\
&= \det(\bar{\sigma}\Sigma)^{-1/2}\delta r(\delta^{\frac{d_\Sigma+|\widehat{S}_t|}{d+|\widehat{S}_t|}}, d + |\widehat{S}_t|).
\end{aligned}$$

Following the same reasoning, it is simple to show that under the event $\mathcal{E}_{t,\delta/2}$, we have

$$\mathbb{P}_{Y\sim\mathcal{N}(0_d,\Sigma)}\left(Y \geqslant \sqrt{\left(\frac{1}{N_{t,i}}+\frac{1}{N_{t,j}}\right)^{-1}\frac{1}{c(t-1,\delta)}}\left((\hat{\mu}_{t,i}-\hat{\mu}_{t,j})-(\mu_i-\mu_j)\right)\right)$$
$$\geqslant \det(\bar{\sigma}\Sigma)^{-1/2}\delta\widetilde{R}(\sqrt{2\log(1/\delta)d_\Sigma}/d) = \det(\bar{\sigma}\Sigma)^{-1/2}\delta r(\delta^{\frac{d_\Sigma}{d}},d).$$

Combining these, we have

$$\mathbb{1}\left(\mathcal{E}_{t,\delta/2}\right)\mathbb{1}\big(\widehat{S}_t\neq\mathcal{S}^\star\big)\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\widehat{S}_t)) \geqslant \delta\det(\bar{\sigma}\Sigma)^{-1/2}\max\left\{\mathbb{1}\left(\mu\in\mathrm{Alt}^-(\widehat{S}_t)\right)r(\delta^{\frac{d_\Sigma}{d}},d),\right.$$
$$\left.\mathbb{1}\left(\mu\in\mathrm{Alt}^+(\widehat{S}_t)\right)r(\delta^{\frac{d_\Sigma+|\widehat{S}_t|}{d+|\widehat{S}_t|}},d+|\widehat{S}_t|)\right\}$$
$$\geqslant \delta\det(\bar{\sigma}\Sigma)^{-1/2}\min\left\{r(\delta^{\frac{d_\Sigma}{d}},d), r(\delta^{\frac{d_\Sigma+|\widehat{S}_t|}{d+|\widehat{S}_t|}},d+|\widehat{S}_t|)\right\}$$

which follows similarly to the Σ diagonal case, since $\widehat{S}_t\neq\mathcal{S}^\star \Longrightarrow \mu\in\mathrm{Alt}(\widehat{S}_t)=\mathrm{Alt}^+(\widehat{S}_t)\cup\mathrm{Alt}^-(\widehat{S}_t)$. Note that when $\Sigma=\sigma I_d$, we recover the results of section 5.5.1.

Proof of Lemma 5.5.3. Σ is a $d\times d$ covariance matrix and f_Σ is the density function of $\mathcal{N}(0_d,\Sigma)$. We let $\mathcal{R}_\Sigma$ denote the Mills ratio of the distribution $\mathcal{N}(0_d,\Sigma)$, which is defined for a vector $x\in\mathbb{R}^d$ as

$$\mathcal{R}_\Sigma(x) \coloneqq \frac{\mathbb{P}\left(\mathcal{N}(0_d,\Sigma)\geqslant x\right)}{f_\Sigma(x)} \tag{5.11}$$

where for two vectors x,y, the notation $x\leqslant y$ should be understood component-wise. Expanding (5.11) gives

$$\mathcal{R}_\Sigma(x) = \exp(\|x\|^2_{\Sigma^{-1}}/2)\int_{(x,\infty)}\exp\left(-\|u\|^2_{\Sigma^{-1}}/2\right)du,$$

with $(x,\infty)=(x(1),\infty)\times\ldots\times(x(d),\infty)$. By a simple translation, it follows that

$$\mathcal{R}_\Sigma(x) = \int_{(0_d,\infty)}\exp\left(-x^\intercal\Sigma^{-1}u-\|u\|^2_{\Sigma^{-1}}/2\right)du.$$

Since $(V-\Sigma^{-1})$ is psd, we have

$$\begin{aligned}\mathcal{R}_\Sigma(x) &\geqslant \int_{(0_d,\infty)}\exp\left(-x^\intercal\Sigma^{-1}u-u^\intercal Vu/2\right)du\\ &= \det(V)^{-1/2}\int_{(0_d,\infty)}\exp\left(-x^\intercal\Sigma^{-1}V^{-1/2}u-u^\intercal u/2\right)du\end{aligned}$$

which follows by change of variable. Thus, letting $z=V^{-1/2}\Sigma^{-1}x$, we have

$$\begin{aligned}\mathcal{R}_\Sigma(x) &\geqslant \det(V)^{-1/2}\int_{(0_d,\infty)}\exp\left(-u^\intercal z-u^\intercal u/2\right)du\\ &= \det(V)^{-1/2}\prod_{c\in[d]}R(z(c))\end{aligned}$$

then using Lemma 5.5.2, it follows that

$$\begin{aligned}\mathcal{R}_\Sigma(x) &\geqslant \det(V)^{-1/2} R\left(1_d^\intercal V^{-1/2}\Sigma^{-1}x/d\right)^d \\ &\geqslant \det(V)^{-1/2} R\left(\|1_d\|_{V^{-1/2}\Sigma^{-1}V^{-1/2}}\|x\|_{\Sigma^{-1}}/d\right)^d.\end{aligned}$$

□

5.5.2 Sample complexity

We recall the following result on the sampling rule, proven in Kone, Jourdan, et al. 2025.

Theorem 5.5.4 (Theorem 3 of Kone, Jourdan, et al. 2025)**.** *There exists events $(\Xi_t)_{t\geqslant 1}$ and $T_0 \in \mathbb{N}$ such that for all $t \geqslant T_0$,*

$$2GLR(t) \geqslant (t-1)T^\star(\nu)^{-1} - f(t), \tag{5.12}$$

with $f(t) =_\infty o(t)$ and $\mathbb{P}_\nu(\Xi_t) \geqslant 1 - \mathcal{O}(1/t^2)$. In particular for $t \geqslant T_0$, $\widehat{S}_t = \mathcal{S}^\star$.

The next lemma is proven in section 5.3.1

Lemma 5.5.5 (Posterior probability bound via GLR)**.** *At each round t, it holds that*

$$\mathbb{P}_{\widehat{\Pi}_t|\mathcal{H}_t}(Alt(\widehat{S}_t)) \leqslant \alpha_t \exp\left(-\frac{GLR(t)}{c(t-1,\delta)}\right),$$

where $\alpha_t := \frac{1}{2}n(p_t)$ with $p_t := |\widehat{S}_t|$ and $n(p_t) = p_t(p_t-1) + (K-p_t)d^{p_t}$.

Theorem 5.3.1. *For any budget M and inflation c such that $\limsup_{\delta\to 0} \frac{c(t,\delta)\log M(t,\delta)}{\log(1/\delta)} \leqslant 1$, PSIPS satisfies that*

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_\nu[\tau_\delta^{PS}]}{\log\frac{1}{\delta}} \leqslant T^\star(\nu) \quad \textit{and} \quad \mathbb{P}\left(\limsup_{\delta\to 0} \frac{\tau_\delta^{PS}}{\log\frac{1}{\delta}} \leqslant T^\star(\nu)\right) = 1, \quad \forall \nu \in \mathcal{D}^K.$$

Proof. In this section, $(\Xi_t)_{t\geqslant 1}$ denotes the sequence of events Theorem 5.5.4 and $T_0 \in \mathbb{N}$.

We have $\tau_\delta^{\mathrm{PS}} = 1 + \sum_{t\geqslant 1} \mathbb{1}(\tau_\delta^{\mathrm{PS}} > t)$ then

$$\begin{aligned}\mathbb{E}_\nu[\tau_\delta^{\mathrm{PS}}] &= 1 + \mathbb{E}\left[\sum_{t\geqslant 1} \mathbb{P}_\nu\left(\tau_\delta^{\mathrm{PS}} > t | \mathcal{H}_t\right)\right] \\ &= 1 + \mathbb{E}_\nu\left[\sum_{t\geqslant 1} \mathbb{P}_\nu\left(\exists m \in [M(t,\delta)] : \theta_t^m \in \mathrm{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)\right] \\ &\leqslant \mathbb{E}_\nu\left[\sum_{t\geqslant T_0} \mathbb{1}\left(\Xi_t\right) M(t,\delta) \mathbb{P}_\nu\left(\theta_t \in \mathrm{Alt}(\widehat{S}_t) | \mathcal{H}_t\right)\right] + \left[\sum_{t\geqslant 1} \mathbb{P}_\nu(\Xi_t^{\mathsf{c}})\right] + (T_0+1),\end{aligned}$$

where $\theta_t, \theta_t^1 \ldots, \ldots, \theta_t^m$ are *i.i.d.* given $\mathcal{H}_t$. Moreover, by Lemma 5.5.5, we have

$$\mathbb{P}_\nu\left(\theta_t \in \mathrm{Alt}(\widehat{S}_t)|\mathcal{H}_t\right) \leqslant \alpha_t \exp\left(-\frac{\mathrm{GLR}(t)}{c(t-1,\delta)}\right)$$

therefore,

$$\mathbb{E}_\nu[\tau_\delta^{\mathrm{PS}}] \leqslant \underbrace{\mathbb{E}_\nu\left[\sum_{t\geqslant t_3} \mathbb{1}\left(\Xi_t\right) \alpha_0 M(t,\delta) \exp\left(-\frac{\mathrm{GLR}(t)}{c(t-1,\delta)}\right)\right]}_{L_1(\delta)} + \underbrace{\left[\sum_{t\geqslant 1} \mathbb{P}_\nu(\Xi_t^{\mathsf{c}})\right]}_{L_2} + T_0,$$

then, since $\mathbb{P}_\nu(\Xi_t^{\mathsf{c}}) \leqslant \mathcal{O}(1/t^2)$ we immediately have

$$L_2 \leqslant 5\pi^2/6. \tag{5.13}$$

It remains to bound $L_1(\delta)$. Using the saddle-point convergence property on the event Ξ_t ensures that (Theorem 5.5.4) for $t \geqslant T_0$,

$$\mathrm{GLR}(t) \geqslant tT^\star(\nu)^{-1} - f(t), \tag{5.14}$$

which further results in

$$\begin{aligned} L_1(\delta) &\leqslant \sum_{t\geqslant t_3} \alpha_0 M(t,\delta) \exp\left(-\frac{t}{T^\star(\nu)c(t,\delta)} + f(t)/c(t,\delta)\right) \\ &= \sum_{t\geqslant t_3} \exp\left(-\frac{t}{T^\star(\nu)c(t,\delta)} + f(t)/c(t,\delta) + \log\left(\alpha_0 M(t,\delta)\right)\right). \end{aligned}$$

To bound the above quantity, let us introduce

$$T(\delta) \coloneqq \sup\left\{t \mid \frac{t}{T^\star(\nu)c(t,\delta)} - \log(M(t,\delta)) - f(t)/c(t,\delta) - \log(\alpha_0) \leqslant \log(t\log(t))\right\}, \tag{5.15}$$

then it follows that

$$L_1(\delta) \leqslant T(\delta) + \sum_{t\geqslant t_3} (t\log(t))^{-1}. \tag{5.16}$$

Further observe that $T(\delta)$ can be rewritten as

$$T(\delta) \coloneqq \sup\left\{t | t \leqslant T^\star(\nu)c(t,\delta)\left(\log(M(t,\delta)) + f(t)/c(t,\delta) + \log(\alpha_0) + \log(t\log(t))\right)\right\}.$$

Bounding $T(\delta)$. Since f is sub-linear in t, there exists $\varepsilon \in (0,1)$ such that $f(t) = o_{t\to\infty}(t^\varepsilon)$ then $f(\log(1/\delta)^{1/\varepsilon}) = o(\log(1/\delta))$. Further observe that for $t_\delta = \log(1/\delta)^{1/\alpha}$,

$$\underbrace{c(t_\delta,\delta)\log(M(t_\delta,\delta)) + f(\log(1/\delta)^{1/\varepsilon}) + c(t_\delta,\delta)\left(\log(t_\delta\log(t_\delta)) + \alpha_0\right)}_{b(t_\delta)} \underset{\delta\to 0}{\sim} \log(1/\delta).$$

Let $\delta_{\min} \in (0,1)$ be defined as

$$\delta_{\min} \coloneqq \inf\left\{\delta \in (0,1) | b(t_\delta) > \log(1/\delta)^{1/\varepsilon} T^\star(\nu)^{-1}\right\}$$

which is well defined as $b(t_\delta) \underset{\delta\to 0}{\sim} \log(1/\delta)$ and $\varepsilon \in (0,1)$. Letting $T_{\max} = \log(1/\delta_{\min})^{1/\varepsilon}$, further note that for all $t \geqslant T_{\max}$, there is $(0,1) \ni \delta' \leqslant \delta_{\min}$ $t'_\delta = t$ and $b(t_{\delta'}) < t_{\delta'}$. Therefore, for all $\delta \leqslant \delta_{\min}$

$$T(\delta) \leqslant \log(1/\delta)^{1/\varepsilon} \tag{5.17}$$

and further noting that by definition $T(\delta) \leqslant T^\star(\nu) b(T(\delta))$ and b is increasing, it follows that

$$T(\delta) \leqslant T^\star(\nu) b(\log(1/\delta)^{1/\varepsilon}). \tag{5.18}$$

Combining (5.13), (5.16) and (5.18), it follows that for $\delta \leqslant \delta_{\min}$,

$$\mathbb{E}_\nu[\tau_\delta^{\mathrm{PS}}] \;\leqslant\; T^\star(\nu) b(\log(1/\delta)^{1/\varepsilon}) + 5\pi^2/6 + (T_0+1) + \sum_{t \geqslant T_0} (t\log(t))^{-1}. \tag{5.19}$$

Finally, noting that $b(\log(1/\delta)^{1/\varepsilon}) \underset{\delta\to 0}{\sim} \log(1/\delta)$, we have proved that

$$\limsup_{\delta\to 0} \frac{\mathbb{E}[\tau_\delta^{\mathrm{PS}}]}{\log \frac{1}{\delta}} \leqslant T^\star(\nu). \tag{5.20}$$

Almost-sure upper bound. The almost-sure bound on the sample complexity is derived from an application of Borel-Cantelli's lemma. Indeed, introducing for any T the event

$$\mathcal{E}(T) \coloneqq \left\{ \sum_{t=1}^{T} \mathbb{1}\left(\tau_\delta^{\mathrm{PS}} > t\right) - \sum_{t=1}^{T} \mathbb{P}(\tau_\delta > t | \mathcal{H}_{t-1}) \leqslant \sqrt{2T\log T^2} \right\},$$

we have when $\mathcal{E}(T)$ holds

$$\begin{aligned}
\min\left\{\tau_\delta^{\mathrm{PS}}, T\right\} &\leqslant 1 + \sum_{t=1}^{T} \mathbb{1}\left(\tau_\delta^{\mathrm{PS}} > t\right) \\
&\leqslant 1 + \sum_{t=1}^{T} \mathbb{P}(\tau_\delta^{\mathrm{PS}} > t | \mathcal{H}_{t-1}) + \sqrt{2T\log T^2} \\
&\leqslant T(\delta) + \sqrt{2T\log T^2} + T_0\,,
\end{aligned}$$

where the last inequality follows from the derivations above on the event Ξ_T and holds with probability larger than $1 - \mathcal{O}(1/T^2)$. Next, assume $\mathcal{E}(T)$ holds. Then, if

$$T \geqslant T'(\delta) \coloneqq \sup\left\{ t \geqslant 1 : t < T_0 + T(\delta) + \sqrt{2t\log t^2} \right\},$$

and $\mathcal{E}(T) \cap \Xi(T)$ holds then the stopping rule would trigger before episode T. Thanks to Azuma-Hoeffding $\mathcal{E}(T)$ holds with probability at least $1 - \mathcal{O}(1/T^2)$. Thus $\Xi(T)^{\mathsf{c}}$ holds with probability at most $\mathcal{O}(1/T^2)$, similarly for $\mathcal{E}(T)^{\mathsf{c}}$ and, as $\sum_T \frac{1}{T^2} < \infty$, we invoke Borel-Cantelli's lemma to justify that with probability 1, there exists $\widetilde{T}$ (possibly random) such that for $T \geqslant \widetilde{T}$, $\Xi(T) \cap \mathcal{E}(T)$ holds.

From the calculations above, we observe that the leading term in $T'(\delta)$ is at most of order $\log(1/\delta)$. Let $\delta'_{\min}$ be such that $\forall\ \delta \leqslant \delta'_{\min}$,

$$\lceil \log(1/\delta)^{3/2} \rceil > T'(\delta),$$

which is well-defined. Let $\delta \leqslant \delta'_{\min}$ such that additionally $\lceil \log(1/\delta)^{3/2} \rceil > \widetilde{T}$. We then have

$$\begin{aligned} \tau_\delta^{\mathrm{PS}} &\leqslant T_0 + T(\delta) + \sqrt{2\lceil \log(1/\delta)^{3/2} \rceil \log \lceil \log(1/\delta)^{3/2} \rceil^2} \\ &\leqslant T_0 + T(\delta) + \log(1/\delta)^{3/4} \sqrt{8 \log(1 + \log(1/\delta)^{3/2})}\,. \end{aligned}$$

Finally recalling that

$$\limsup_{\delta \to 0} \frac{T(\delta)}{\log \frac{1}{\delta}} \leqslant T^\star(\nu).$$

We conclude by noting that

$$\begin{aligned} \limsup_{\delta \to 0} \frac{\tau_\delta^{\mathrm{PS}}}{\log \frac{1}{\delta}} &\leqslant \limsup_{\delta \to 0} \frac{T(\delta)}{\log \frac{1}{\delta}} + \lim_{\delta \to 0} \frac{T_0 + \log(1/\delta)^{3/4} \sqrt{8 \log(1 + \log(1/\delta)^{3/2})}}{\log \frac{1}{\delta}} \\ &\leqslant T^\star(\nu). \end{aligned}$$

□

5.5.3 Posterior convergence

We prove that the posterior contraction rate of PSIPS is unimprovable.

In BAI, Russo 2016 and Z. Li et al. 2024 proved a similar result for truncated Gaussian (restricted to a bounded domain). We prove it more generally in the unbounded setting for Gaussian distributions, thanks to the novel Lemma 5.5.7.

Theorem 5.3.2. *Let $\widetilde{\Pi}_t := \bigotimes_{k=1}^K \mathcal{N}(\hat{\mu}_{t,k}, \Sigma/N_{t,k})$ be the posterior distribution under a flat Gaussian prior (without inflation). For any bandit instance $\nu \in \mathcal{D}^K$, it holds with probability 1 that*

$$\limsup_{t \to +\infty} -\frac{1}{t} \log \mathbb{P}_{\widetilde{\Pi}_t | \mathcal{H}_t}(\mathit{Alt}(\mathcal{S}^\star)) \leqslant T^\star(\nu)^{-1},$$

for any algorithm, and PSIPS almost surely satisfies that

$$\liminf_{t \to +\infty} -\frac{1}{t} \log \mathbb{P}_{\widetilde{\Pi}_t | \mathcal{H}_t}(\mathit{Alt}(\mathcal{S}^\star)) \geqslant T^\star(\nu)^{-1}\,.$$

Proof. We first prove the upper bound. Similarly to Lemma 5.5.5, we can derive (using Lemma 5.3.4 and Lemma 5.3.5),

$$\mathbb{P}_{\widetilde{\Pi}_t | \mathcal{H}_t}(\mathrm{Alt}(\mathcal{S}^\star)) \leqslant \alpha_0 \exp\left(- \inf_{\lambda \in \mathrm{Alt}(\mathcal{S}^\star)} \sum_{k=1}^K \frac{N_{t,k}}{2} \|\hat{\mu}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}} \right)\,.$$

From Theorem 5.5.4, there exists events $(\Xi_t)_{t\geqslant}$ and $T_0 \in \mathbb{N}$ such that for $t \geqslant T_0$, $\widehat{S}_t = S^\star$, and

$$\mathrm{GLR}(t) \quad \geqslant \quad (t-1)T^\star(\nu)^{-1} - f(t), \tag{5.21}$$

with $f(t) =_\infty o(t)$ and $\mathbb{P}_\nu(\Xi_t) \geqslant 1 - \mathcal{O}(1/t^2)$. Since $\sum_{t\geqslant 1} \mathbb{P}(\Xi_t^\mathsf{c}) < \infty$, by Borel-Cantelli's lemma, with probability 1, there exists a finite time $\widetilde{T}_0$ possibly random such that for $t \geqslant \widetilde{T}_0$, Ξ_t holds. So for $t \geqslant \max(T_0, \widetilde{T}_0)$, we have $\widehat{S}_t = S^\star$ and

$$\begin{aligned}\inf_{\lambda\in \mathrm{Alt}(\mathcal{S}^\star)} \frac{N_{t,k}}{2}\|\lambda_k - \hat{\mu}_{t,k}\|^2_{\Sigma^{-1}} = \mathrm{GLR}(t) \quad &\geqslant \quad (t-1)T^\star(\nu)^{-1} - f(t), \\ &\geqslant \quad tT^\star(\nu)^{-1} - o(t)\end{aligned}$$

then

$$\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\mathcal{S}^\star)) \leqslant \alpha_0 \exp\left(-tT^\star(\nu)^{-1} + o(t)\right),$$

so that

$$-\frac{1}{t}\log(\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\mathcal{S}^\star))) \geqslant T^\star(\nu)^{-1} - (1/t)\log(\alpha_0) - o(t)/t\,.$$

Put together, the above displays show that with probability 1,

$$\liminf_{t\to\infty} -\frac{1}{t}\log(\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\mathcal{S}^\star))) \geqslant T^\star(\nu)^{-1}.$$

The proof of the lower bound uses Lemma 5.5.6.

Let $\mathbb{B}_\varepsilon$ be the ball centered on μ and with radius ε as in Lemma 5.5.7. We have $-\mathbb{P}(\mathrm{Alt}(\mathcal{S}^\star)) \leqslant -\mathbb{P}(\mathrm{Alt}(\mathcal{S}^\star)\cap\mathbb{B}_\varepsilon)$, then $O = \mathrm{Alt}(\mathcal{S}^\star)\cap\mathbb{B}_\varepsilon$ is countably convex (Lemma 5.3.4 and $\mathbb{B}_\varepsilon$ is convex) and bounded. Applying Lemma 5.5.6, we obtain

$$\begin{aligned}\limsup_{t\to\infty} -\frac{1}{t}\log\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\mathcal{S}^\star)) \quad &\leqslant \quad \limsup_{t\to\infty} -\frac{1}{t}\log\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(\mathrm{Alt}(\mathcal{S}^\star)\cap\mathbb{B}_\varepsilon) \\ &\leqslant \quad \frac{1}{2}\sup_{w\in\triangle_K}\inf_{\lambda\in\mathrm{Alt}(\mathcal{S}^\star)\cap\mathbb{B}_\varepsilon}\left[\sum_{k=1}^K \frac{w_k}{2}\|\mu_k-\lambda_k\|^2_{\Sigma^{-1}}\right] \\ &= \quad \frac{1}{2}\sup_{w\in\triangle}\inf_{\lambda\in\mathrm{Alt}(\mathcal{S}^\star)}\left[\sum_{k=1}^K \frac{w_k}{2}\|\mu_k-\lambda_k\|^2_{\Sigma^{-1}}\right] \qquad \text{(Lemma 5.5.7)} \\ &= \quad T^\star(\nu)^{-1},\end{aligned}$$

which concludes the proof. □

Lemma 5.5.6. *Let $O \subset \mathcal{I}^K$ be a countably convex bounded set with a non-empty interior. With probability 1, it holds that*

$$\limsup_{t\to\infty} -\frac{1}{t}\log\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(O) \leqslant \frac{1}{2}\sup_{w\in\triangle_K}\inf_{\lambda\in O}\left[\sum_{k=1}^K \|\mu_k-\lambda_k\|^2_{\Sigma^{-1}}\right].$$

Proof. Since O is countably there exists convex sets $C_1 \dots C_n$ such that $O = \cup_{i\in[n]} C_i$.

We have

$$\mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(O) = (2\pi)^{-dh/2} D_t^{-1/2} \int_O \exp\left(-\sum_{k=1}^K \frac{N_{t,k}}{2}\|\hat{\mu}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}}\right) d\lambda\,,$$

where $D_t = \prod_{k=1}^K \det(\Sigma/N_{t,k})$.

Let $\gamma > 0$ (to be defined) and

$$\widetilde{\lambda}_t \in \operatorname*{argmin}_{\lambda\in\overline{O}} \left[\sum_{k=1}^K \frac{N_{t,k}}{2}\|\lambda_k - \hat{\mu}_{t,k})\|^2_{\Sigma^{-1}}\right]\,.$$

Since O is a union of convex sets, there exists a convex set $\mathcal{C} \subset O$ such that $\widetilde{\lambda}_t \in \mathcal{C}$. Then letting $\mathcal{N}_\gamma \coloneqq \{(1-\gamma)\widetilde{\lambda}_t + \gamma\lambda, \lambda \in \mathcal{C}\} = (1-\gamma)\widetilde{\lambda}_t + \gamma\mathcal{C}$ we have $\mathcal{N}_\gamma \subset \mathcal{C} \subset O$ and

$$\begin{aligned}
I_t &\geqslant \int_{\mathcal{N}_\gamma} \exp\left(-\sum_{k=1}^K \frac{N_{t,k}}{2}\|\hat{\mu}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}}\right) d\lambda \\
&= \int_{\gamma\mathcal{C}} \exp\left(-\sum_{k=1}^K \frac{N_{t,k}}{2}\|\hat{\mu}_{t,k} - (1-\gamma)\widetilde{\lambda}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}}\right) d\lambda \\
&= \int_{\mathcal{C}} \gamma^{Kd} \exp\left(-\sum_{k=1}^K \frac{N_{t,k}}{2}\|(1-\gamma)(\hat{\mu}_{t,k} - \widetilde{\lambda}_{t,k}) + \gamma(\hat{\mu}_{t,k} - \lambda_k)\|^2_{\Sigma^{-1}}\right) d\lambda \\
&\geqslant \int_{\mathcal{C}} \gamma^{Kd} \exp\left(-\sum_{k=1}^K \frac{N_{t,k}}{2}\left[(1-\gamma)\|\hat{\mu}_{t,k} - \widetilde{\lambda}_{t,k}\|^2_{\Sigma^{-1}} + \gamma\|\hat{\mu}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}}\right]\right) d\lambda
\end{aligned}$$

which follows thanks to the convexity of the squared norm. Therefore

$$\begin{aligned}
\log I_t \geqslant Kd\log(\gamma) - (1-\gamma)\sum_{k=1}^K \frac{N_{t,k}}{2}\|\hat{\mu}_{t,k} - \widetilde{\lambda}_{t,k}\|^2_{\Sigma^{-1}} - \\
\frac{\gamma\mathbb{E}_{\lambda\sim\text{unif}(\mathcal{C})}\left[\sum_{k=1}^K \frac{N_{t,k}}{2}\|\lambda_k - \hat{\mu}_{t,k}\|^2_{\Sigma^{-1}}\right]}{2} + \log(\text{vol}(\mathcal{C}_*))
\end{aligned} \tag{5.22}$$

where $\mathcal{C}_*$ is the set of minimum positive volume among $\mathcal{C}_1, \dots, \mathcal{C}_n$. We introduce the event

$$\Xi_{1,t} \coloneqq \left\{\forall s \leqslant t, \sum_{k=1}^K \frac{N_{s,k}}{2}\|\mu_k - \hat{\mu}_{s,k}\|^2_{\Sigma^{-1}} \leqslant \beta(t, 1/t^2) =: f_1(t)\right\},$$

where $\beta(t,\delta)$ is defined as in Lemma 5.2.1, and $\mathbb{P}_\nu(\Xi_{1,t}) \geqslant 1 - 1/t^2$.

Moreover, $f_1(t)$ is logarithmic in t. Since $\sum_{t\geqslant 1}\mathbb{P}(\Xi_t^\mathsf{c}) < \infty$, by Borel-Cantelli's lemma, with probability 1, there exists a finite time $\widetilde{T}$ possibly random such that for $t \geqslant \widetilde{T}$, $\Xi_{1,t}$ holds.

Taking $\gamma = 1/t$ in Equation (5.22), and for $t \geqslant \widetilde{T}$, we get (after simplification)

$$\log I_t \;\geqslant\; -dh\log(t) - \sum_{k=1}^{K} \frac{N_{t,k}}{2}\|\hat{\mu}_{t,k} - \widetilde{\lambda}_{t,k}\|^2_{\Sigma^{-1}} - \mathcal{O}(\sqrt{tL(O)} + \sqrt{f_1(t)})^2/t + \log(\mathrm{vol}(\mathcal{C}_*)),$$

where $L(O) = \max_{\lambda\in O}\sum_{k=1}^{K}\|\lambda_k - \mu_k\|^2_{\Sigma^{-1}}$. Now, by convexity, we have for any $\lambda \in O$,

$$\begin{aligned}\sum_{k=1}^{K} N_{t,k}\|\hat{\mu}_{t,k} - \lambda_k\|^2_{\Sigma^{-1}} &\leqslant \sum_{k=1}^{K} N_{t,k}\|\mu_k - \lambda_k\|^2_{\Sigma^{-1}} + 2\sum_{k=1}^{K} N_{t,k}\|\hat{\mu}_{t,k} - \lambda_k\|_{\Sigma^{-1}}\|\hat{\mu}_{t,k} - \mu_k\|_{\Sigma^{-1}} \\ &\leqslant \sum_{k=1}^{K} N_{t,k}\|\mu_k - \lambda_k\|^2_{\Sigma^{-1}} + 4f_1(t) + 4\sqrt{f_1(t)}\sqrt{tL(O)}\,,\end{aligned}$$

thus we have

$$\sum_{k=1}^{K} N_{t,k}\|\hat{\mu}_{t,k} - \widetilde{\lambda}_{t,k}\|^2_{\Sigma^{-1}} \;\leqslant\; \inf_{\lambda\in O}\sum_{k=1}^{K} N_{t,k}\|\mu_k - \lambda_k\|^2_{\Sigma^{-1}} + 4f_1(t) + 4\sqrt{f_1(t)}\sqrt{tL(O)}.$$

In all cases, we have proved that

$$-\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(O) \leqslant \;\tfrac{1}{2}\inf_{\lambda\in O}\sum_{k=1}^{K} N_{t,k}\|\mu_{t,k} - \lambda_k\|^2_{\Sigma^{-1}} + o(t) + D_t.$$

Next, due to the initialization, we have

$$D_t = \prod_{k=1}^{K}\det(\Sigma/N_{t,k}) \leqslant \det(\Sigma)^K\,.$$

Combining the displays above yields,

$$-\log \mathbb{P}_{\widetilde{\Pi}_t|\mathcal{H}_t}(O) \leqslant \inf_{\lambda\in O}\sum_{k=1}^{K}\frac{N_{t,k}}{2}\|\mu_k - \lambda_k\|^2_{\Sigma^{-1}} + o(t),$$

which yields the claimed result by dividing by t and taking the limit. □

The result below shows that the best response always exists as an $\inf$ over a compact subset of the alternative. Although its closed form is unknown in PSI, we show that it belongs to a ball centered at θ and whose radius also depends on θ.

Lemma 5.5.7. *Let $w \in \mathbb{R}^K_+$. For any $\theta \coloneqq (\theta_1, \ldots, \theta_K) \in \mathcal{I}^K$, the following statement holds*

$$\inf_{\lambda\in Alt(S^\star(\theta))}\left[\sum_{k=1}^{K}\frac{w_k}{2}\|\theta_k - \lambda_k\|^2_{\Sigma^{-1}}\right] = \inf_{\lambda\in Alt(\mathcal{S}^\star(\theta))\cap\{\lambda|\ \max_i\|\theta_i-\lambda_i\|_{\Sigma^{-1}}<\varepsilon\}}\left[\sum_{k=1}^{K}\frac{w_k}{2}\|\theta_k - \lambda_k\|^2_{\Sigma^{-1}}\right], \tag{5.23}$$

where $\varepsilon(\theta) \coloneqq \max\{2\max_{i\notin S}\max_{j\in S}\|\theta_i - \theta_j\|_{\Sigma^{-1}}, \max_{i,j\in S^2}\|\theta_i - \theta_j\|_{\Sigma^{-1}}\}$, and $S = S^\star(\theta)$.

Proof. The idea of the proof is to show that when there is a ball $\mathbb{B}$ such that if an alternative parameter of θ does not belong to $\mathbb{B}$, then there is a parameter in $\mathrm{Alt}(\mathcal{S}^\star(\theta)) \cap \mathbb{B}$ for which the transportation cost will be smaller. Let $\theta \coloneqq (\theta_1, \dots, \theta_K)^\intercal$, where θ_i denotes the vector mean of arm i and we let $\lambda \coloneqq (\lambda_1, \dots, \lambda_K)^\intercal$ where $\lambda_i \in \mathbb{R}^d$. To ease notation, we let $S = \mathcal{S}^\star(\theta)$. Introducing

$$\mathrm{Alt}^-(S) \coloneqq \bigcup_{i,j \in S^2 : i \neq j} W_{i,j} \quad \text{and} \quad \mathrm{Alt}^+(S) \coloneqq \bigcup_{i \in S^\mathsf{c}} V_i, \tag{5.24}$$

where $\mathcal{I}^K \coloneqq \mathbb{R}^{K \times d}$, we define

$$W_{i,j} \coloneqq \left\{ \lambda \coloneqq (\lambda_1 \dots \lambda_K)^\intercal \in \mathcal{I}^K \mid \lambda_i \leqslant \lambda_j \right\} \quad \text{and}$$
$$V_i \coloneqq \left\{ \lambda \coloneqq (\lambda_1 \dots \lambda_K)^\intercal \in \mathcal{I}^K \mid \exists\, i \in (S)^\mathsf{c} : \forall\, j \in S, \lambda_i \not\prec \lambda_j \right\}.$$

Let us introduce

$$D(w, \lambda; \theta) \coloneqq \sum_{i=1}^K w_i \|\lambda_i - \theta_i\|^2_{\Sigma^{-1}}.$$

Let $(i,j) \in S^2$ be fixed and $\lambda \in W_{i,j}$. Let $\alpha_{i,j} = \|\theta_i - \theta_j\|_{\Sigma^{-1}}$. If $\|\lambda_i - \theta_i\|_{\Sigma^{-1}} > \alpha_{i,j}$ then define the instance $\widetilde{\lambda}$ as

$$\widetilde{\lambda}_k \coloneqq \begin{cases} \theta_k & \text{if } k \notin \{i,j\}, \\ \theta_j & \text{else}, \end{cases}$$

which satisfies $\widetilde{\lambda}_i \leqslant \widetilde{\lambda}_j$, so $\widetilde{\lambda} \in W_{i,j}$, and

$$\begin{aligned} D(w, \widetilde{\lambda}; \theta) &= w_i \|\theta_i - \theta_j\|^2_{\Sigma^{-1}} + \sum_{k \notin \{i,j\}} w_k \|\lambda_k - \theta_k\|^2_{\Sigma^{-1}}, \\ &= w_i \|\theta_i - \theta_j\|^2_{\Sigma^{-1}} < w_i \|\lambda_i - \theta_i\|^2_{\Sigma^{-1}} \\ &< D(w, \lambda, \theta), \end{aligned}$$

and further observe that $\max_{k \in [K]} \|\widetilde{\lambda}_k - \theta_k\|_{\Sigma^{-1}} \leqslant \alpha_{i,j}$. We proceed similarly if $\|\lambda_j - \theta_j\|_{\Sigma^{-1}} > \alpha_{i,j}$, by defining

$$\widetilde{\lambda}_k \coloneqq \begin{cases} \theta_k & \text{if } k \notin \{i,j\}, \\ \theta_i & \text{else}, \end{cases}$$

and the same conclusion follows. So we have proved that there exists $\widetilde{\lambda} \in W_{i,j}$ with $\max_{k \in [K]} \|\widetilde{\lambda}_k - \theta_k\|_{\Sigma^{-1}} \leqslant \alpha_{i,j}$ and for which the transportation cost is not larger than that of λ. We prove a similar property for $\mathrm{Alt}^+(S)$.

Now, fix $i \notin S$ and take $\lambda \in V_i$. Let $b_i = \max_{k \in S} \|\theta_i - \theta_k\|_{\Sigma^{-1}}$. If $\|\lambda_i - \theta_i\|_{\Sigma^{-1}} > b_i$ then it suffices to define $\widetilde{\lambda}$ as

$$\widetilde{\lambda}_p \coloneqq \begin{cases} \theta_p & \text{if } p \neq i, \\ \theta_{\widetilde{i}} & \text{else}, \end{cases}$$

for $\widetilde{i} \in S$, to ensure that $\widetilde{\lambda} \in V_i$ and

$$\begin{aligned} D(w, \widetilde{\lambda}; \theta) &= w_i \|\theta_{\widetilde{i}} - \theta_i\|^2_{\Sigma^{-1}} + \sum_{k \neq i} w_k \|\lambda_k - \theta_k\|^2_{\Sigma^{-1}} \\ &\leqslant w_i b_i^2 \\ &< w_i \|\lambda_i - \theta_i\|^2_{\Sigma^{-1}} \\ &< D(w, \lambda; \theta), \end{aligned}$$

where the second line also uses the fact that $\widetilde{i} \in S$. Assume $\|\lambda_i - \theta_i\|_{\Sigma^{-1}} < b_i$ and that $H_i \coloneqq \{k \in S : \|\lambda_k - \theta_k\|_{\Sigma^{-1}} > 2b_i\}$ is non-empty. Let us define the instance $\widetilde{\lambda}$:

$$\widetilde{\lambda}_k \coloneqq \begin{cases} \lambda_i & \text{if } k \in H_i, \\ \lambda_k & \text{if } k \in (S \cup \{i\}) \backslash H_i, \\ \theta_k & \text{else.} \end{cases}$$

Since $\lambda \in V_i$, $\widetilde{\lambda}$ as defined above satisfies $\widetilde{\lambda}_i \not\prec \widetilde{\lambda}_k, k \in S$, that is $\widetilde{\lambda} \in V_i$ and

$$\begin{aligned} D(w, \widetilde{\lambda}; \theta) &= \sum_{k \in (S \cup \{i\})} w_k \|\widetilde{\lambda}_k - \theta_k\|^2_{\Sigma^{-1}} \\ &= \sum_{k \in (S \cup \{i\}) \backslash H_i} w_k \|\lambda_k - \mu_k\|^2_{\Sigma^{-1}} + \sum_{k \in H_i} w_k \|\lambda_i - \theta_k\|^2_{\Sigma^{-1}} \\ &< \sum_{k \in (S \cup \{i\}) \backslash H_i} w_k \|\lambda_k - \theta_k\|^2_{\Sigma^{-1}} + \sum_{k \in H_i} w_k 4b_i^2, \end{aligned}$$

which follows since $\|\lambda_i - \theta_k\|_{\Sigma^{-1}} \leqslant \|\lambda_i - \theta_i\|_{\Sigma^{-1}} + \|\theta_i - \theta_k\|_{\Sigma^{-1}} \leqslant 2b_i$. Recalling that for $k \in H_i$,

$$4b_i < \|\lambda_k - \theta_k\|^2_{\Sigma^{-1}},$$

it follows that

$$D(w, \widetilde{\lambda}; \theta) < D(w, \lambda; \theta).$$

Combining these results, we have proved that for all $\lambda \in V_i$, there exists $\widetilde{\lambda} \in V_i$ whose transportation cost is not larger than that of λ and which additionally satisfies: $\max_{k \in [K]} \|\widetilde{\lambda}_k - \theta_k\|_{\Sigma^{-1}} \leqslant 2b_i$.

To conclude, let us define $\varepsilon \coloneqq \max\{2 \max_{i \notin S} b_i, \max_{i,j \in S^2} \alpha_{i,j}\}$. Using what precedes, we have proved that

$$\inf_{\lambda \in \mathrm{Alt}(S) \cap \{\lambda : \max_k \|\lambda_k - \theta_k\|_{\Sigma^{-1}} < \varepsilon\}} D(w, \lambda; \theta) \leqslant \inf_{\lambda \in \mathrm{Alt}(S)} D(w, \lambda; \theta),$$

which proves the claimed result as we have $\mathrm{Alt}(S) \cap \{\lambda : \max_k \|\lambda_k - \theta_k\|_{\Sigma^{-1}} < \varepsilon\} \subset \mathrm{Alt}(S)$. □

Chapter 6

Pareto Set Identification with Linear Constraints

This chapter studies the problem of *Pareto Set Identification* (PSI) under feasibility constraints in a multi-objective bandit framework. Given a K-armed bandit with unknown mean vectors $\mu_1, \dots, \mu_K \in \mathbb{R}^d$, we aim to identify, with high confidence, the arms that are both *feasible*, *i.e.*, satisfying a known set of linear constraints, and *non-dominated* with respect to all objectives. This constrained variant of PSI is particularly relevant in applications such as clinical trials, where treatment candidates must satisfy safety or efficacy thresholds before being compared for optimality.

To the best of our knowledge, we propose the first theoretically grounded algorithms for constrained PSI in the fixed-confidence setting. Our main algorithm, e-cAPE, jointly handles feasibility assessment and Pareto set identification by adaptively allocating samples based on upper and lower confidence bounds. We establish information-theoretic lower bounds showing that e-cAPE is near-optimal in sample complexity.

Thi chapter is based on joint work with Émilie Kaufmann and Laura Richert, published in the proceedings of *ICML 2025*.

6.1 Introduction

In previous chapters, we studied the *Pareto Set Identification* (PSI) problem in the fixed-confidence and fixed-budget settings. So far, no constraints have been imposed on admissible arms, allowing for arms that perform extremely poorly on one objective to still belong to the Pareto set if they excel on another. In many real-world applications, however, such unconstrained trade-offs are unacceptable.

A prime example arises in clinical trials, particularly in dose-finding or early-phase vaccine studies, where multiple biological markers, such as antibody titers, safety indicators, or biomarkers of immune response, must simultaneously meet acceptable thresholds. Here, the experimenter seeks to identify treatments that not only achieve a desirable balance between efficacy and safety but also satisfy explicit feasibility constraints, such as a minimum efficacy level or a maximum toxicity bound (Jennison & Turnbull 1993; Munro et al. 2021).

Similarly, in multi-objective recommender systems or material design optimization, one might be interested in finding the Pareto set of items that satisfy some minimal performance metrics or dimensional constraints (Afshari et al. 2019; Kumar et al. 2021).

This motivated us to introduce the *constrained Pareto Set Identification* (cPSI). Formally, given K arms with multivariate means $\mu_k \in \mathbb{R}^d$, we are provided with a set of known linear constraints defining a polyhedron

$$P \coloneqq \{x \in \mathbb{R}^d : Ax \leqslant b\}, \quad A \in \mathbb{R}^{m \times d},\ b \in \mathbb{R}^m,$$

encoding m feasibility requirements. Any arm k such that $\mu_k \in P$ is deemed *feasible*. The learner's objective is to identify the Pareto set of these feasible arms, denoted $\mathcal{S}_{\text{feas}}$.

We study this problem in the *fixed-confidence* setting, where the learner must output the correct constrained Pareto set with probability at least $1 - \delta$ while minimizing the expected number of samples. This formulation naturally extends the classical fixed-confidence PSI framework introduced in Chapter 3 to handle feasibility constraints.

From PSI to Constrained PSI. A straightforward approach might proceed in two stages: first, identify the feasible arms, and then apply a PSI algorithm on the remaining subset. However, as we demonstrate in Section 6.3, such a decoupled strategy can be arbitrarily inefficient, since feasibility and dominance information are statistically intertwined. This observation motivates algorithms that integrate both tasks adaptively.

Explainable Constrained PSI. Beyond identifying the constrained Pareto set, certain applications–again notably in clinical contexts–require explicit *justifications* for excluding arms. We therefore introduce an extended variant, the *explainable constrained PSI* (e-cPSI), in which the learner must classify every non-optimal arm as either (i) infeasible or (ii) dominated by a feasible arm. This provides interpretable reasoning for every decision, a property desirable in safety-critical domains.

Related work. As illustrated throughout Chapters 2 to 4, a growing body of work has addressed Pareto set identification in the fixed-budget, fixed-confidence, and linear settings (Zuluaga, Sergent, et al. 2013; Auer et al. 2016; Zuluaga, Krause, et al. 2016; Kone, Kaufmann, et al. 2023; Karagözlü et al. 2024; Kone, Kaufmann, et al. 2025a). Despite these advances, none of these studies consider the additional challenge of handling constraints on the Pareto set.

Feasibility detection in multi-objective bandit problems was independently studied by Katz-Samuels & Scott 2018, who considered identifying arms whose mean vectors lie within a known polyhedral constraint set $P \subset \mathbb{R}^d$. However, their goal was to determine feasibility rather than to compare arms with respect to one another.

A related constrained BAI setting was analyzed by Katz-Samuels & Scott 2019, in which the learner must identify the arm maximizing a weighted combination $w^\top \mu_k$ among feasible arms. While this formulation assumes a known preference vector w to rank the arms, it fundamentally differs from the constrained PSI setting studied here, where no such scalarization is available and dominance is defined in the original multi-objective space.

The recent work of Faizal & Nair 2022 further investigated constrained BAI in dimension $d = 2$, which can be viewed as a particular instance of Katz-Samuels & Scott 2019. The authors proposed specialized algorithms and a lower bound for this setting, but extending these results to general multi-objective problems and Pareto identification remains an open challenge.

Finally, beyond pure exploration, several works have examined *constrained regret minimization*, either in single-objective problems (Lattimore & Szepesvari 2020) or in multi-objective settings (Drugan & Nowe 2013; Amani et al. 2019; D. Li et al. 2024), where the learner must select arms satisfying feasibility constraints while maximizing cumulative reward.

In contrast, our focus here is on the fixed-confidence identification of the *constrained Pareto set*, for which no existing framework or algorithm currently provides theoretical guarantees.

Contributions. This chapter makes the following contributions. We first formalize the *constrained Pareto Set Identification* (cPSI) problem and its explainable variant (e-cPSI), extending the standard PSI framework introduced in Chapters 3–4 to settings with feasibility constraints and interpretability requirements.

We then derive information-theoretic lower bounds characterizing the optimal sample complexity of both problems, thereby establishing the fundamental limits of constrained multi-objective exploration. For cPSI, we propose a gradient-ascent procedure that matches the lower bound asymptotically with a computational complexity polynomial in the number of arms.

Yet such an approach is computationally expensive for the more challenging e-cPSI, and we devote the rest of our effort to proposing and analyzing an efficient algorithm for this setting.

We introduce e-cAPE, an adaptive algorithm that jointly handles feasibility detection and Pareto set identification through coupled upper and lower confidence bounds.

We provide upper bounds on its sample complexity and show that it can be significantly smaller than that of two-stage or racing-based baselines, while remaining near-optimal for a broad class of instances.

Learning model. We denote by K the number of arms, by $d \geqslant 1$ the number of objectives, and by $[K] \coloneqq \{1, \dots, K\}$ the index set of arms. Let $\mathcal{D}$ be a family of distributions over $\mathbb{R}^d$. Given a multivariate bandit instance $\nu \coloneqq (\nu_1, \dots, \nu_K) \in \mathcal{D}^K$, we denote its mean vectors by $\mu \coloneqq (\mu_1, \dots, \mu_K) \in \mathcal{I}^K$, where $\mathcal{I} \subset \mathbb{R}^d$ and $\mu_k \coloneqq \mathbb{E}_{X \sim \nu_k}[X]$. In the sequel, μ and ν will be used interchangeably.

We assume that $\mathcal{D}$ is either a family of marginally subgaussian distributions,[1] used for algorithmic analysis, or a parametric family of multivariate Gaussian distributions, used when deriving lower bounds. At each round $t = 1, 2, \dots$, the learner selects an arm $A_t \in [K]$ (based on past observations) and observes an outcome $Z_t \sim \nu_{A_t}$, independent of previous samples. We denote by $\mathcal{H}_t \coloneqq \sigma(A_1, Z_1, \dots, A_t, Z_t)$ the σ-algebra generated by the history up to time t, and by $\mathbb{P}_\nu$ and $\mathbb{E}_\mu$ the probability and expectation under instance ν.

In this chapter, we consider the case where the learner is additionally given a convex polyhedron

$$P \coloneqq \{x \in \mathbb{R}^d : Ax \leqslant b\},$$

where $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$, and m is the number of linear constraints. Given P, we study two closely related pure exploration problems:

Constrained PSI (cPSI). Identify $\mathcal{S}_{\text{feas}}$, the Pareto set of arms whose mean vectors lie within P.

Constrained PSI with explainability (e-cPSI). Identify $\mathcal{S}_{\text{feas}}$ and, in addition, provide a justification for rejecting every arm outside $\mathcal{S}_{\text{feas}}$. Each non-optimal arm must be classified as either *infeasible* (violating the constraints) or *dominated by a feasible arm*. When both conditions hold, both explanations are admissible.

While cPSI may suffice in some applications, e-cPSI is particularly relevant in domains such as clinical trials, where investigators must provide explicit reasons for excluding candidate treatments–distinguishing, for example, between infeasibility (*e.g.*, excessive toxicity) and domination by a safer or more effective alternative. This need for interpretability makes e-cPSI our primary focus in this chapter.

We note that Katz-Samuels & Scott 2019 considered a related "δ-PAC-Explanatory" framework for identifying a feasible arm maximizing a linear combination of objectives, though their approach assumes a fixed preference vector and thus does not extend to the Pareto setting considered here.

[1] Let $(X_i^c)_{c \leqslant d}$ denote a realization of arm i. It is marginally σ-subgaussian if for all $c \leqslant d$ and $\lambda \in \mathbb{R}$, $\mathbb{E}[e^{\lambda(X_i^c - \mu_i^c)}] \leqslant e^{\lambda^2 \sigma^2 / 2}$.

Additional notation. For any $x \in \mathbb{R}^d$, let x^c denote its c-th component, and define the distance to a set $\mathcal{X} \subset \mathbb{R}^d$ as $\mathrm{dist}(x, \mathcal{X}) := \inf\{\|x - y\|_2 : y \in \mathcal{X}\}$. We denote by $\mathrm{cl}(P)$ the closure of P, by $\mathrm{int}(P)$ its interior, and by $\partial P := \mathrm{cl}(P) \setminus \mathrm{int}(P)$ its boundary. For $a, b \in \mathbb{R}$, we write $(a)_+ := \max\{a, 0\}$, $a \wedge b := \min\{a, b\}$, and $a \vee b := \max\{a, b\}$.

6.2 On the complexity of constrained PSI

We now study the intrinsic difficulty of the constrained Pareto set identification problems introduced above. In particular, we establish information-theoretic lower bounds on the expected sample complexity of any δ-correct algorithm for both cPSI and e-cPSI. We begin by introducing basic definitions that formalize these two settings.

Given a polyhedron P and a bandit instance $\mu \in \mathcal{I}^K$, we define the *feasible set*

$$\mathrm{F}(\mu) := \{\, k \in [K] : \mu_k \in P \,\},$$

i.e., the set of arms satisfying all linear constraints.

For two arms i, j, we recall that arm i is (strictly) *dominated* by j (written $i \prec j$ or $\mu_i \prec \mu_j$) if $\mu_i^c < \mu_j^c$ for all $c \in [d]$. For any subset $S \subseteq [K]$ and parameter $\lambda \in \mathcal{I}^K$, let $\mathrm{Par}(S, \lambda)$ denote the (strict) Pareto set of $\{\lambda_i : i \in S\}$–the arms in S not dominated by any other in S under the instance λ. With this notation, the target Pareto set of feasible arms is

$$\mathcal{S}_{\text{feas}}(\mu) := \mathrm{Par}(\mathrm{F}(\mu), \mu),$$

and we further define the set of arms dominated by a feasible one as

$$\mathrm{SubOpt}(\mu) := \{\, i \in [K] : \exists j \in \mathrm{F}(\mu) \text{ such that } \mu_i \prec \mu_j \,\}.$$

When clear from context, we simply write F, $\mathcal{S}_{\text{feas}}$, and SubOpt for $\mathrm{F}(\mu)$, $\mathcal{S}_{\text{feas}}(\mu)$, and $\mathrm{SubOpt}(\mu)$, respectively.

6.2.1 Constrained PSI without explainability (cPSI)

In the cPSI problem, an algorithm with stopping time τ outputs a recommendation $R_\tau = O_\tau$, which is its estimate of $\mathcal{S}_{\text{feas}}(\mu)$.

Definition 6.2.1 (Correctness for cPSI)**.** An algorithm is said to be δ*-correct* for cPSI on $\mathcal{D}^K$ if, for any instance $\nu \in \mathcal{D}^K$ with mean parameter μ, it satisfies

$$\mathbb{P}_\nu(\tau < \infty,\, O_\tau \neq \mathcal{S}_{\text{feas}}(\mu)) \leqslant \delta.$$

We are interested in δ-correct algorithms with small expected sample complexity $\mathbb{E}_\mu[\tau]$. To correctly identify $\mathcal{S}_{\text{feas}}(\mu)$, an algorithm must be able to distinguish μ from any alternative instance λ for which the optimal set differs. We denote the set of such alternatives by

$$\mathrm{Alt}(S) := \{\, \lambda \in \mathcal{I}^K : \mathcal{S}_{\text{feas}}(\lambda) \neq S \,\}.$$

Proposition 6.2.2 (Lower bound for cPSI). *Let $\mathcal{D}$ be the class of multivariate Gaussian distributions with identity covariance, and let $\nu \in \mathcal{D}^K$ have mean vectors $\mu := (\mu_1, \dots, \mu_K) \in (\mathbb{R}^d)^K$. For any δ-correct algorithm for cPSI on $\mathcal{D}$ with stopping time τ, it holds that*

$$\mathbb{E}_\mu[\tau] \geqslant T^\star(\mu) \log\left(\frac{1}{2.4\delta}\right), \quad \textit{where} \quad T^\star(\mu)^{-1} := \max_{w \in \triangle_K} \inf_{\lambda \in \mathit{Alt}(\mathcal{S}_{\mathit{feas}})} \sum_{k=1}^{K} \tfrac{1}{2} w_k \|\mu_k - \lambda_k\|^2,$$
$$\textit{with } \mathit{Alt}(S) := \{\, \lambda \in \mathcal{I}^K : \mathcal{S}_{\mathit{feas}}(\lambda) \neq S \,\}. \tag{6.1}$$

The proof follows the general information-theoretic approach of Garivier & Kaufmann 2016, originally developed for best-arm identification. Algorithms that approach such bounds typically rely on iterative saddle-point or game-theoretic methods (Ménard 2019; Degenne, Ménard, et al. 2020; P.-A. Wang et al. 2021). However, they require efficient solvers for the inner optimization problem in (6.1), which is generally non-convex.

We show in Section 6.6.5 that subgradients of this inner inf function in Eq. 6.1 can be computed in polynomial time.

Then, applying the techniques in Ménard 2019 yields a computationally efficient algorithm for cPSI called Game-cPSI, whose analysis is given in Kone, Kaufmann, et al. 2025b.

6.2.2 Constrained PSI with Explainability (e-cPSI)

We now turn to the more challenging e-cPSI setting, in which the learner must not only identify the feasible Pareto set but also provide a justification for rejecting each non-optimal arm. At stopping time τ, an algorithm outputs a partition

$$R_\tau = (O_\tau, S_\tau, I_\tau)$$

of the arms such that

$$i)\ O_\tau = \mathcal{S}_{\text{feas}}, \qquad ii)\ S_\tau \subset \text{SubOpt}, \qquad iii)\ I_\tau \subset \text{F}^{\text{c}}.$$

Since some arms may be both infeasible and dominated by a feasible one, multiple correct classifications may exist. Accordingly, we define the set of valid *explanatory partitions* as

$$\mathcal{M}(P, \mu) := \big\{(S, I) \;\big|\; S \subset \text{SubOpt}(\mu),\ I \subset \text{F}(\mu)^{\text{c}},\ S \cap I = \emptyset,\ S \cup I = (\mathcal{S}_{\text{feas}}(\mu))^{\text{c}}\big\}.$$

When P and μ are clear from the context, we simply write $\mathcal{M}$. This setting thus corresponds to a pure exploration problem with multiple correct answers, a framework also studied by Degenne & W. Koolen 2019 for single-objective bandits.

Definition 6.2.3 (δ-correctness for e-cPSI). An algorithm is *δ-correct* for e-cPSI on $\mathcal{D}^K$ if, for any instance $\nu \in \mathcal{D}^K$ with mean parameter μ, it outputs a partition (O_τ, S_τ, I_τ) of $[K]$ such that

$$\mathbb{P}_\nu\left(\tau < \infty \text{ and } \neg(O_\tau = \mathcal{S}_{\text{feas}},\ (S_\tau, I_\tau) \in \mathcal{M})\right) \leqslant \delta.$$

We again focus on δ-correct algorithms with small expected sample complexity $\mathbb{E}_\mu[\tau]$. To derive a lower bound, we first define the set of alternative instances where a given partition (S, I) would not be a valid answer:

$$\mathrm{Alt}(S, I) \coloneqq \{\, \lambda \in \mathcal{I}^K : (S, I) \notin \mathcal{M}(P, \lambda) \,\}.$$

Proposition 6.2.4 (Lower bound for e-cPSI). *Let $\mathcal{D}$ be the class of multivariate normal distributions with identity covariance, and let $\nu \in \mathcal{D}^K$ have mean parameters $\mu \in \mathcal{I}^K$. For any δ-correct algorithm for e-cPSI on $\mathcal{D}$ with stopping time τ,*

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau]}{\log(1/\delta)} \geqslant T^\star_{\mathcal{M}}(\mu) \coloneqq \min_{(S,I)\in\mathcal{M}} T(\mu, S, I), \quad \textit{where}$$

$$T(\mu, S, I)^{-1} \coloneqq \max_{w \in \triangle_K} \inf_{\lambda \in Alt(S,I)} \sum_{k=1}^{K} \frac{w_k}{2} \|\mu_k - \lambda_k\|^2. \tag{6.2}$$

This asymptotic lower bound extends the game-theoretic technique introduced by Degenne & W. Koolen 2019 for pure exploration with multiple correct answers. For any valid partition $(S, I) \in \mathcal{M}$, the quantity $T(\mu, S, I)$ captures the information-theoretic cost required to identify (S, I) as the correct explanation. The bound implies that the minimal achievable sample complexity of any δ-correct algorithm is governed by the easiest valid partition to identify.

Even if μ were known, evaluating $T^\star_{\mathcal{M}}(\mu)$ remains computationally challenging, as it involves solving the inner $\inf$ problem in (6.2) for every $(S, I) \in \mathcal{M}$. The set $\mathcal{M}$ may have size up to 2^n with $n = |\mathrm{SubOpt} \cap \mathrm{F}^\mathsf{c}|$, which can grow exponentially with K. Consequently, directly implementing the optimal strategy is infeasible for large-scale problems.

Degenne & W. Koolen 2019 proposed an optimal algorithm for similar single-objective problems, but it requires enumerating all possible answers and solving the inner optimization at each step. In our setting, such enumeration over the power set of $\mathrm{SubOpt} \cap \mathrm{F}^\mathsf{c}$ would be computationally prohibitive. Instead, in the next section, we introduce a practical alternative: an adaptive confidence-bound algorithm, e-cAPE, which efficiently balances exploration between feasibility testing and Pareto set identification, while achieving near-optimal sample complexity in many instances.

6.3 Constrained Adaptive Pareto Exploration

We now introduce the main algorithm of this chapter, the *Explainable Constrained Adaptive Pareto Exploration* (e-cAPE), which efficiently balances the detection of feasible arms and the identification of the constrained Pareto set. Similar to the APE algorithm presented in Chapter 3, e-cAPE relies on confidence regions around pairwise dominance and feasibility quantities. Throughout this section, all arms are assumed to have marginally σ-subgaussian rewards.

Preliminaries. As in previous chapters, we characterize pairwise dominance using the operators

$$\mathrm{M}(i,j) \coloneqq \max_{c\leqslant d}\big[\mu_i^c - \mu_j^c\big], \qquad \mathrm{m}(i,j) \coloneqq \min_{c\leqslant d}\big[\mu_j^c - \mu_i^c\big], \tag{6.3}$$

introduced in Chapter 3. These quantities measure how far arm i is from being dominated by j: note that $\mathrm{M}(i,j) > 0$ if and only if $\mu_i \not\preceq \mu_j$. They play a central role in constructing empirical dominance tests.

In the constrained setting, we must also assess whether each arm satisfies the feasibility constraints. Following Katz-Samuels & Scott 2018, we define the signed feasibility margin:

$$\eta_i \coloneqq \begin{cases} \mathrm{dist}(\mu_i, P), & \text{if } \mu_i \notin P, \\ \mathrm{dist}(\mu_i, P^{\mathsf{c}}), & \text{if } \mu_i \in P, \end{cases} \tag{6.4}$$

which measures the distance of μ_i to the constraint boundary.

Empirical Estimates and Confidence Bounds. At each round t, we denote by $N_{t,i}$ the number of times arm i has been pulled and by $\hat{\mu}_{t,i}$ its empirical mean. We write $\mathrm{M}(i,j;t)$, $\mathrm{m}(i,j;t)$, and $\eta_i(t)$ for the empirical counterparts of (6.3) and (6.4). The empirical feasible set is

$$F_t \coloneqq \{\, i : \hat{\mu}_{t,i} \in P \,\},$$

and its empirical Pareto set is given by

$$O_t \coloneqq \mathrm{Par}(F_t, \hat{\mu}_t)\,.$$

To control uncertainty, we define a sequence of high-probability events:

$$\mathcal{E}_t \coloneqq \Big\{\forall i \in [K],\ \big\|\hat{\mu}_{t,i} - \mu_i\big\|_\infty \leqslant \beta_i(t,\delta) \text{ and } \big\|\hat{\mu}_{t,i} - \mu_i\big\|_2 \leqslant U_i(t,\delta)\Big\}, \quad \mathcal{E} \coloneqq \bigcap_{t\geqslant 1}\mathcal{E}_t,$$

with confidence bonuses of the form

$$\beta_i(t,\delta) = \sqrt{\frac{2\sigma^2 f(t,\delta)}{N_{t,i}}}, \qquad U_i(t,\delta) = \sqrt{\frac{2\sigma^2 g(t,\delta)}{N_{t,i}}},$$

where f and g are chosen so that $\mathbb{P}_\nu(\mathcal{E}) \geqslant 1-\delta$.

Lemma 6.3.1 (Concentration of empirical quantities). *Under $\mathcal{E}_t$, for all arms i, j:*

(i) $|\,\mathrm{M}(i,j) - \mathrm{M}(i,j;t)| \leqslant \beta_i(t,\delta) + \beta_j(t,\delta)$ *and* $|\,\mathrm{m}(i,j) - \mathrm{m}(i,j;t)| \leqslant \beta_i(t,\delta) + \beta_j(t,\delta)$*;*

(ii) for any $\mathcal{X} \subset \mathbb{R}^d$, $|\,\mathrm{dist}(\hat{\mu}_{t,i}, \mathcal{X}) - \mathrm{dist}(\mu_i, \mathcal{X})| \leqslant U_i(t,\delta)$.

This leads to the confidence intervals

$$\mathrm{M}^{\pm}(i,j;t) = \mathrm{M}(i,j;t) \pm (\beta_i(t,\delta) + \beta_j(t,\delta)), \qquad \mathrm{m}^{\pm}(i,j;t) = \mathrm{m}(i,j;t) \pm (\beta_i(t,\delta) + \beta_j(t,\delta)).$$

In particular, under $\mathcal{E}_t$, $\mathrm{M}^-(i,j;t) > 0$ guarantees that $\mu_i \not\preceq \mu_j$, while $\mathrm{M}^+(i,j;t) < 0$ implies that $\mu_i \prec \mu_j$.

Feasibility Detection. For any arm i, introducing the quantity $\gamma_i(t) \coloneqq \frac{1}{2\sigma^2} N_{t,i}\eta_i^2(t) - g(t,\delta)$, one can show using Lemma 6.3.1 that when $\gamma_i(t) > 0$, the vectors $\hat{\mu}_{t,i}$ and μ_i can either both belong to P or to P^c, with high confidence. We let $G_t \coloneqq \{i \in F_t^c \mid \gamma_i(t) < 0\}$ be the subset of F_t^c of empirically infeasible arms that cannot be confidently ruled out as infeasible at time t.

Stopping and Recommendation Rules. An algorithm for e-cPSI should stop as soon as any confidently valid answer has been identified. To introduce the stopping rule, we first define $Z_1^F(t) \coloneqq \min_{i\in O_t} \gamma_i(t)$ and $Z_1^{\mathrm{PS}}(t) \coloneqq \min_{i\in O_t} \min_{j\in O_t\setminus\{i\}} [\mathrm{M}^-(i,j;t)]$. We further define

$$Z_1(t) \coloneqq \min\{Z_1^F(t), Z_1^{\mathrm{PS}}(t)\}. \tag{6.5}$$

When $Z_1^F(t)$ is positive, arms in O_t can be proved to be feasible, and having $Z_1^{\mathrm{PS}}(t)$ positive is sufficient to guarantee that arms in O_t are not dominated by each other. Thus, when $Z_1(t) > 0$, arms in O_t are all confidently feasible and non-dominated by each other. However, this alone is not sufficient to ensure that $O_t = \mathcal{S}_{\text{feas}}$. We also define $\xi_i(t) \coloneqq \max_{j\in(F_t\cup G_t)\setminus\{i\}} \mathrm{m}^-(i,j;t)$ and note that $\xi_i(t) > 0$ implies that there exists $j \neq i \in (F_t \cup G_t)$ such that $\mu_i \prec \mu_j$. Therefore, $\mathbb{1}\big(i \notin F_t\big)((\gamma_i(t) \vee \xi_i(t))) > 0$ implies that either arm i is infeasible or it is dominated by an arm of $(F_t \cup G_t)$. We then introduce

$$Z_2(t) \coloneqq \min_{i\in[K]\setminus O_t} \left[\xi_i(t)\mathbb{1}\big(i \in F_t\big) + (\gamma_i(t) \vee \xi_i(t))\mathbb{1}\big(i \notin F_t\big)\right], \tag{6.6}$$

for which a positive value guarantees that each arm outside of O_t can be confidently classified as either infeasible or dominated by another arm of $(F_t \cup G_t)$. We will prove that having both $Z_1(t) \geqslant 0$ and $Z_2(t) \geqslant 0$ for some t is sufficient to prove that $O_t = \mathcal{S}_{\text{feas}}$ and to identify a correct answer $(S_t, I_t) \in \mathcal{M}$. Thus, we define the stopping time τ of our algorithm as

$$\tau \coloneqq \inf\{t \geqslant 1 \mid Z_1(t) \geqslant 0 \quad \text{and} \quad Z_2(t) \geqslant 0\}. \tag{6.7}$$

At stopping, the algorithm recommends the partition (O_τ, S_τ, I_τ) of $[K]$ with $O_\tau = \mathrm{Par}(F_\tau, \hat{\mu}_\tau)$, $S_\tau \coloneqq (F_\tau \cup G_\tau)\setminus O_\tau$, a set of arms that will be shown to be dominated by some arms in O_τ, and $I_\tau \coloneqq G_\tau^c \cap F_\tau^c$, a set of arms deemed infeasible.

Lemma 6.3.2 (Correctness). *On the event $\mathcal{E}$, if e-cAPE outputs (O_τ, S_τ, I_τ), we have $O_\tau = \mathcal{S}_{feas}$ and $(S_\tau, I_\tau) \in \mathcal{M}(P,\mu)$.*

The above result holds regardless of the sampling rule that is used, but the sampling rule is crucial to stop early.

Sampling Rule. The challenge in designing an efficient sampling rule for constrained PSI lies in efficiently balancing the information about feasibility and Pareto dominance.

Following the "top-two" philosophy of LUCB (Kalyanakrishnan et al. 2012) and its Pareto extensions from Chapter 3, we select at each round a *leader* arm b_t and a *challenger* c_t. The leader is chosen as the arm that currently limits the stopping condition:

$$b_t \coloneqq \begin{cases} \text{minimizer of (6.6)}, & \text{if } Z_2(t) < 0, \\ \text{minimizer of (6.5)}, & \text{otherwise.} \end{cases} \tag{6.8}$$

Intuitively, if $Z_2(t) < 0$, there remains uncertainty about an arm outside O_t being infeasible or dominated; if $Z_1(t) < 0$, some arm in O_t is not yet confidently verified as feasible or mutually non-dominated.

The challenger c_t is the arm most likely to invalidate the current status of b_t:

$$c_t \coloneqq \operatorname*{argmin}_{j \in (F_t \cup G_t) \setminus \{b_t\}} \mathrm{M}^-(b_t, j; t).$$

Pulling both b_t and c_t refines the dominance and feasibility estimates most relevant to the current uncertainty, thereby increasing $\min\{Z_1(t), Z_2(t)\}$ and accelerating convergence to the stopping condition.

Algorithm 6.1: e-cAPE: Explainable Constrained Adaptive Pareto Exploration

Require : risk parameter $\delta \in (0, 1)$

Initialize: Pull each arm once; set $t \leftarrow K + 1$

while *true* **do**

  // Feasibility and empirical Pareto set estimation

  Compute the empirical feasible set $F_t = \{\, i : \hat{\mu}_{t,i} \in P \,\}$;

  Compute

$$O_t \coloneqq \left\{\, i \in F_t \;\middle|\; \forall j \in F_t,\ \hat{\mu}_{t,i} \not\prec \hat{\mu}_{t,j} \,\right\}$$

  (empirical non-dominated arms).;

  // Select most ambiguous candidate

  if $Z_2(t) < 0$ **then**

    $b_t \leftarrow$ minimizer of (6.6);

  else

    $b_t \leftarrow$ minimizer of $\begin{cases} Z_1^F(t), & \text{if } Z_1^{\mathrm{PS}}(t) \geqslant Z_1^F(t), \\ Z_1^{\mathrm{PS}}(t), & \text{otherwise.} \end{cases}$;

  // Find the most conflicting arm

  $c_t \leftarrow \operatorname*{argmin}_{j \in (F_t \cup G_t) \setminus \{b_t\}} \mathrm{M}^-(b_t, j; t)$;

  if $\min\{Z_1(t), Z_2(t)\} \geqslant 0$ **then**

    return $(O_t,\ (F_t \cup G_t) \setminus O_t,\ F_t^c \cap G_t^c)$;

  // Sample selected arms

  Pull arms b_t and c_t; $t \leftarrow t + 1$;

6.4 Main theoretical results

We begin by introducing the complexity quantity that drives the analysis of e-cAPE. Beyond the feasibility margins η_i (Definition (6.4)), it depends on Pareto sub-optimality gaps already used for PSI in Chapters 2–3. For any nonempty $S \subset [K]$, let $\mathrm{Par}(S,\mu)$ be the Pareto set in the unconstrained setting and recall the dominance operators $\mathrm{M}(i,j) = \max_{c\in[d]}\big[\mu_i^c - \mu_j^c\big]$ and $\mathrm{m}(i,j) = \min_{c\in[d]}\big[\mu_j^c - \mu_i^c\big]$ (so $\mathrm{M}(i,j) = -\mathrm{m}(j,i)$). For $i \notin \mathrm{Par}(S,\mu)$, the (unconstrained) PSI gap is

$$\Delta_i(S) \;\coloneqq\; \Delta_i^\star(S) \;\coloneqq\; \max_{j\in\mathrm{Par}(S,\mu)} \mathrm{m}(i,j), \tag{6.9}$$

i.e., the minimal (coordinate-wise) increase that makes i appear Pareto-optimal against $S\setminus\{i\}$. For $i \in \mathrm{Par}(S,\mu)$, we use the standard "two-sided margin":

$$\Delta_i(S) \;\coloneqq\; \min\{\delta_i^+(S),\, \delta_i^-(S)\}, \tag{6.10}$$

where

$$\delta_i^+(S) \;\coloneqq\; \min_{j\in\mathrm{Par}(S,\mu)\setminus\{i\}} \min\big\{\,\mathrm{M}(i,j),\,\mathrm{M}(j,i)\big\}, \quad \delta_i^-(S) \;\coloneqq\; \min_{j\in S\setminus\mathrm{Par}(S,\mu)} \big[(\mathrm{M}(j,i))_+ + \Delta_j(S)\big],$$

with the convention $\min_\emptyset = +\infty$. Intuitively, δ_i^+ measures the proximity of i to other Pareto arms, whereas δ_i^- captures the margin separating i from sub-optimal arms.

A candidate-answer complexity. To certify a valid explainable answer $(S,I) \in \mathcal{M}$ (cf. Section 6.2), an algorithm must: (i) identify the Pareto set of $\mathcal{S}_{\text{feas}} \cup S$; (ii) certify feasibility of $\mathcal{S}_{\text{feas}}$; and (iii) certify infeasibility of I. We quantify this joint effort by

$$C(\mu,S,I) \;\coloneqq\; \sum_{i\in\mathcal{S}_{\text{feas}}} \frac{1}{\min\{\Delta_i^2(\mathcal{S}_{\text{feas}}\cup S),\,\eta_i^2\}} \;+\; \sum_{i\in S} \frac{1}{\Delta_i^2(\mathcal{S}_{\text{feas}}\cup S)} \;+\; \sum_{i\in I} \frac{1}{\eta_i^2},$$

and the "easiest" valid partition by

$$C_{\mathcal{M}}^\star(\mu) \;\coloneqq\; \min_{(S,I)\in\mathcal{M}(P,\mu)} C(\mu,S,I).$$

Theorem 6.4.1. *Let $f(t,\delta) = \log\big(\frac{4k_1Kdt^\alpha}{\delta}\big)$ and $g(t,\delta) = 4\log\big(\frac{4k_1K5^dt^\alpha}{\delta}\big)$ with $k_1 > 1+\frac{1}{\alpha-1}$ and $\alpha > 2$. Over the class $\mathcal{D}$ of marginally σ-subgaussian arms, e-cAPE is δ-correct for any $\delta\in(0,1)$ and, for any instance μ,*

$$\mathbb{E}_\mu[\tau_\delta] \;\leqslant\; 256\sigma^2\, C_{\mathcal{M}}^\star(\mu)\, \log\Big(128\sigma^2 C_{\mathcal{M}}^\star(\mu)\,\big(\tfrac{4k_1Kd}{\delta}\big)^{1/\alpha}\Big) \;+\; \Lambda_\alpha \;+\; 4H(\mu,F^c),$$

where

$$\Lambda_\alpha \;\leqslant\; \frac{2^{\alpha-1}\delta}{4k_1}\sum_{T\geqslant 1}(\log T+1)\,\frac{f\big(T,\delta\cdot d/5^d\big)+f(T,\delta)}{T^{\alpha-1}}, \quad H(\mu,F^c) \;\coloneqq\; \sum_{a\in F^c\cup\mathcal{S}_{feas}} \frac{32\sigma^2\log(5^d/d)}{\eta_a^2}.$$

This result shows that the dominant term of the sample complexity scales as $C^\star_{\mathcal{M}}(\mu)\ \log(\frac{K\, C^\star_{\mathcal{M}}(\mu)}{\delta})$. While a general, instance-wise comparison to the lower bound in Proposition 6.2.4 is delicate, the theorem below shows that there exist families of instances for which the optimal sample complexity must scale with $C^\star_{\mathcal{M}}(\mu)$, implying near-optimality of e-cAPE in a worst-case sense.

Theorem 6.4.2. *There exists a class $\widetilde{\mathcal{D}}$ of multivariate Gaussian instances such that any δ-correct algorithm for e-cPSI satisfies, for all μ,*

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\log(1/\delta)} \ \geqslant\ \frac{C^\star_{\mathcal{M}}(\mu)}{4}.$$

Comparison to two-stage approaches. Since $(\mathrm{F}\setminus \mathcal{S}_{\text{feas}}, \mathrm{F}^c)$ is always a valid explanation, we have

$$\begin{aligned} C^\star_{\mathcal{M}}(\mu) \ &\leqslant\ C(\mu, \mathrm{F}\setminus\mathcal{S}_{\text{feas}}, \mathrm{F}^c) \ =\ \sum_{i\in\mathcal{S}_{\text{feas}}} \frac{1}{\Delta_i^2(\mathrm{F})\wedge\eta_i^2} \ +\ \sum_{i\in\mathrm{F}\setminus\mathcal{S}_{\text{feas}}} \frac{1}{\Delta_i^2(\mathrm{F})} \ +\ \sum_{i\notin\mathrm{F}} \frac{1}{\eta_i^2} \\ &\leqslant\ \underbrace{\sum_{i\in\mathrm{F}} \frac{1}{\Delta_i^2(\mathrm{F})}}_{H_{\text{PSI}}(\mu)} + \underbrace{\sum_{i\in[K]} \frac{1}{\eta_i^2}}_{H_F(\mu)}. \end{aligned}$$

Here H_F is the feasibility-detection complexity (Katz-Samuels & Scott 2018), and H_{PSI} the Pareto-identification complexity on F, as studied in Chapter 3. Thus the leading term of e-cAPE is *always* no larger than that of a naive two-stage method that first learns F and then identifies $\mathrm{Par}(\mathrm{F},\mu)$: e-cAPE pays a PSI cost only for $\mathrm{SubOpt}\cap\mathrm{F}$, and for $\mathrm{SubOpt}\cap\mathrm{F}^c$ it pays the cheaper of infeasibility or dominance. Figure 6.1 illustrates an instance where this yields an arbitrarily large advantage.

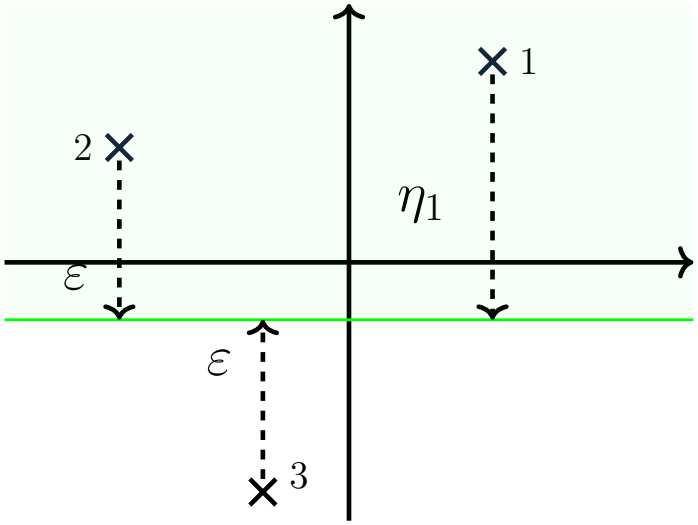


Figure 6.1: A constrained PSI instance. A two-stage approach scales as $1/\eta_1^2 + 2/\varepsilon^2 + \sum_{i=1}^2 1/\Delta_i(\{1,2\})^2$, whereas e-cAPE scales as $1/\eta_1^2 + \sum_{i=1}^3 1/\Delta_i(\{1,2,3\})^2$ for $\varepsilon \ll 1$.

When all arms are well inside/outside P (*i.e.*, $\eta_i \gg 1$), our bound reduces to the PSI complexities on F. Symmetrically, when PSI is easy (large Δ_i), we match feasibility-only detection rates.

Algorithms based on uniform sampling and rejection for cPSI. Racing-style methods eliminate an arm only when (i) it is confidently infeasible; (ii) it is confidently dominated by an *already* confidently feasible arm; or (iii) it is confidently optimal and does not dominate any other active arm. Condition (ii) is the bottleneck: to discard a dominated arm, one must first certify the feasibility of the dominator. Consider an instance with $\mathrm{F} = [K]$, $\mathcal{S}_{\text{feas}} = \{1\}$, and $\mathrm{Par}([K] \setminus \{1\}, \mu) = [K] \setminus \{1\}$. A racing algorithm must pull *every* arm until arm 1 is certified feasible before discarding any $k \neq 1$, yielding a complexity on the order of

$$\frac{K}{\eta_1^2} + \sum_{i \neq 1} \frac{1}{\Delta_i([K])^2}.$$

In contrast, e-cAPE scales as

$$\frac{1}{\eta_1^2} + \sum_{i \neq 1} \frac{1}{\Delta_i([K])^2},$$

removing the multiplicative factor K on the feasibility term. Our experiments in the next section corroborate this gap.

6.4.1 Sample complexity

We prove the correctness and refer to the end of the chapter for the sample complexity bound.

Proof. We show that on $\mathcal{E}$, when cAPE stops and returns a partition (O_τ, S_τ, I_τ), it is a valid partition. We recall that P is the polyhedron of feasible arms, F is the set of indices of the feasible arms and

$$\mathcal{S}_{\text{feas}} \coloneqq \{i \mid \mu_i \in P \text{ and } \forall j \text{s.t. } \mu_j \in P, \mu_i \not\prec \mu_j\}$$

is the Pareto-optimal feasible set; $\mathrm{SubOpt} \coloneqq \{i \in [K] \mid \exists j \in \mathcal{S}_{\text{feas}} \backslash \{i\} : \mu_i \prec \mu_j\}$ which can be rewritten as

$$\mathrm{SubOpt} \coloneqq \{i \in [K] \mid \exists j \in \mathrm{F} \backslash \{i\} : \mu_i \prec \mu_j\}. \tag{6.11}$$

To see this, note that if $\mu_i \prec \mu_j$ and $j \in \mathrm{F}$, then, as $j \in \mathrm{F}$, either $j \in \mathcal{S}_{\text{feas}}$ or there exists $j' \in \mathcal{S}_{\text{feas}}$ such that $\mu_j \prec \mu_{j'}$ and so $\mu_i \prec \mu_{j'}$. Thus

$$\exists j \in \mathrm{F} \backslash \{i\} : \mu_i \prec \mu_j \Longrightarrow \exists j \in \mathcal{S}_{\text{feas}} \backslash \{i\} : \mu_i \prec \mu_j \tag{6.12}$$

and the reverse inclusion is trivial, which justifies (6.11). Finally, it is simple to check that for the polyhedron P,

$$\mathrm{dist}(\mu, P^c) = \mathrm{dist}(\mu, \partial P) \quad \forall \mu \in P. \tag{6.13}$$

To prove the correctness, we have to show that the returned partition (O_t, S_t, I_t) at $t = \tau$ satisfies $O_t = \mathcal{S}_{\text{feas}}$, $S_t \subset \text{SubOpt}$ and $I_t \subset \text{F}^c$. In this proof, we assume $\mathcal{E}$ holds and we let $t = \tau$.

Step 1: Proving that $S_t \subset \text{SubOpt}$. Following (6.12), we have to show that

$$\forall i \in S_t, \exists j \in \text{F} \text{ such that } \mu_i \prec \mu_j. \tag{6.14}$$

By design, $S_t \coloneqq (F_t \cup G_t) \backslash O_t$ and at stopping

$$Z_2(t) \coloneqq \min_{i \in [K] \backslash O_t} \left[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \backslash \{i\}} \text{m}^-(i, j; t) \right) + \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \backslash \{i\}} \text{m}^-(i, j; t) \right) \right]$$

satisfies $Z_2(t) \geqslant 0$ so that

$$\forall i \in S_t, \mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \backslash \{i\}} \text{m}^-(i, j; t) \right) + \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \backslash \{i\}} \text{m}^-(i, j; t) \right) \geqslant 0. \tag{6.15}$$

Note that if $i \in S_t \cap G_t$, then $i \notin F_t$ and $\gamma_i(t) \leqslant 0$ so by (6.15),

$$\max_{j \in (F_t \cup G_t) \backslash \{i\}} \text{m}^-(i, j; t) \geqslant 0,$$

which also holds if $i \in F_t$. Thus, it holds that

$$\forall i \in S_t, \max_{j \in (F_t \cup G_t) \backslash \{i\}} \text{m}^-(i, j; t) \geqslant 0 \tag{6.16}$$

which on the event $\mathcal{E}$, using Lemma 6.3.1 and by the definition of $\text{m}(i, j)$ yields

$$\forall i \in S_t, \ \exists j \in (F_t \cup G_t) \backslash \{i\} \text{ such that } \mu_i \prec \mu_j. \tag{6.17}$$

However, this does not directly imply (6.14) as the stopping rule only guarantees that $O_t \subset \text{F}$ and at this point, some arms in $(F_t \cup G_t)$ could be infeasible (for the actual means). Thus, we shall prove that (6.17) holds for some $j \in O_t$. Note that by their definition, we have $F_t \cup G_t = S_t \cup O_t$. Let

$$H_t \coloneqq \text{Par}(S_t, \mu) = \{i \in S_t \,|\, \forall j \in S_t \backslash \{i\}, \mu_i \not\prec \mu_j\},$$

the Pareto set of S_t based on the true means.

As $H_t \subset S_t$, (6.17) applies and

$$\forall i \in H_t, \exists j \neq i, j \in (F_t \cup G_t) = (S_t \cup O_t) \text{ such that } \mu_i \prec \mu_j,$$

then, as H_t is the Pareto set of S_t, arms in H_t cannot be dominated by another arm of S_t. Then, the equation above implies

$$\forall i \in H_t, \exists j \neq i, j \in O_t \text{ such that } \mu_i \prec \mu_j,$$

which as $O_t \subset \mathrm{F}$ yields

$$\forall i \in H_t, \exists j \in \mathrm{F}\backslash\{i\} \text{ such that } \mu_i \prec \mu_j. \tag{6.18}$$

Now, for $i \in S_t \backslash H_t$, there exists (by definition of H_t), $j \in H_t$ such that $\mu_i \prec \mu_j$ and by (6.18) there exists $j_2 \in \mathrm{F}$ such that $\mu_j \prec \mu_{j_2}$ so $\mu_i \prec \mu_{j_2}$. Combining these results, we have proved that

$$\forall i \in S_t,\ \exists j \in \mathrm{F}\backslash\{i\} \text{ such that } \mu_i \prec \mu_j,$$

hence $S_t \subset \mathrm{SubOpt}$.

Step 2: Proving that $I_t \subset \mathbf{F}^c$. The proof of this claim simply follows from the definition of I_t and Lemma 6.3.1. Indeed, as $I_t := G_t^c \cap F_t^c \subset O_t^c$ and by definition of G_t, it holds that for all arm $i \in I_t, \gamma_i(t) \geqslant 0$. Then by definition of γ_i, the previous yields $\eta_i(t) \geqslant \sqrt{\frac{2\sigma^2 g(t,\delta)}{N_{t,i}}} := U(t,\delta)$ and since $i \in F_t^c$ we invoke Lemma 6.3.1, and

$$\begin{aligned} \mathrm{dist}(\mu_i, P) &> \mathrm{dist}(\hat{\mu}_{t,i}, P) - U_i(t,\delta) \\ &= \eta_i(t) - U_i(t,\delta) \geqslant 0\,, \end{aligned}$$

so $\mu_i \notin P$ that is $i \in \mathrm{F}^c$ which completes the proof that $I_t \subset \mathrm{F}^c$.

Step 3: Proving that $O_t = \mathcal{S}_{\mathbf{feas}}$. First, note that (O_t, S_t, I_t) is a partition of $[K]$. Since we have proved that $S_t \subset \mathrm{SubOpt}$ and $I_t \subset \mathrm{F}^c$, and by definition $\mathcal{S}_{\mathrm{feas}} \cap (\mathrm{SubOpt} \cup \mathrm{F}^c) = \emptyset$, thus it holds that $\mathcal{S}_{\mathrm{feas}} \subset O_t$.

Then, noting that as $Z_1(t) \geqslant 0$, proceeding similarly to Step 2, we have $O_t \subset \mathrm{F}$, so at this step it holds that

$$\mathcal{S}_{\mathrm{feas}} \subset O_t \text{ and } O_t \subset \mathrm{F}. \tag{6.19}$$

Moreover, from $Z_1(t) \geqslant 0$, we derive

$$\forall i \in O_t,\ \forall j \in O_t\backslash\{i\},\ \mathrm{M}(i,j) \overset{\mathcal{E}}{>} \mathrm{M}^-(i,j;t) \geqslant 0$$

which from the definition of $\mathrm{M}(i,j)$ translates to

$$\forall i \in O_t,\ \forall j \in O_t\backslash\{i\}, \mu_i \not\prec \mu_j. \tag{6.20}$$

Therefore, as $\mathcal{S}_{\mathrm{feas}} \subset O_t$ (6.19), we have

$$\forall i \in O_t,\ \forall j \in \mathcal{S}_{\mathrm{feas}}\backslash\{i\}, \mu_i \not\prec \mu_j$$

which implies (using the transitivity of the Pareto dominance and the fact that any sub-optimal element in F is dominated by an element in $\mathcal{S}_{\mathrm{feas}}$) that

$$\forall i \in O_t,\ \forall j \in \mathrm{F}\backslash\{i\}, \mu_i \not\prec \mu_j\ .$$

Moreover $O_t \subset \mathrm{F}$, so $O_t \subset \mathcal{S}_{\mathrm{feas}}$ hence the conclusion follows as $O_t = \mathcal{S}_{\mathrm{feas}}$.

Combining these results, we have shown that the recommendation is a valid partition of $[K]$, which concludes the proof of the correctness of Algorithm 6.1 for e-cPSI (which also implies correctness of cPSI). □

Remark 6.1. The proof of the correctness can be adapted to generic time-uniform confidence bounds on $\mathrm{M}^{\pm}, \mathrm{m}^{\pm}, \eta^{\pm}$ using an event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$ where

$$\mathcal{E}_1 \coloneqq \Bigg\{ \forall t \geqslant 1, \forall (i,j) \in [K]^2,\ \mathrm{M}(i,j) \in \big(\mathrm{M}^-(i,j;t),\, \mathrm{M}^+(i,j;t)\big), \\ \text{and} \quad \mathrm{m}(i,j) \in \big(\mathrm{m}^-(i,j;t),\, \mathrm{m}^+(i,j;t)\big) \Bigg\},$$

and

$$\mathcal{E}_2 \coloneqq \Big(\forall t \geqslant 1, \forall i \in [K], \eta_i \in \big(\eta_i^-(t), \eta_i^+(t)\big) \Big)$$

such that $\mathcal{E}$ holds with probability at least $1 - \delta$.

6.5 Numerical study and discussion

Experimental setup. We now evaluate the performance of e-cAPE on instances inspired by real-world data, with a particular focus on clinical trial applications where feasibility and interpretability are essential. Throughout, we compare e-cAPE against three baselines: (**A-A**) the two-stage algorithm that combines the feasible-arm identification procedure of Katz-Samuels & Scott 2018 with the Pareto set identification algorithm of Kone, Kaufmann, et al. 2023; (**U**) a uniform sampling policy executed with the same stopping rule as e-cAPE; and (**R-CP**) an adaptation of the racing algorithm of Auer et al. 2016 to the constrained setting (see Appendix F of Kone, Kaufmann, et al. 2025b). All algorithms use identical confidence bonuses $\beta_i(t,\delta) = \sqrt{2\sigma_i^2 \log(\log(t)/\delta)/N_{t,i}}$ and pairwise bounds $\beta_{i,j} = \sqrt{\beta_i^2 + \beta_j^2}$ with $\delta = 0.1$, yielding negligible empirical errors.

Clinical trial datasets. We begin with two experiments derived from published biomedical studies, each representing a typical multi-objective dose-finding or vaccine evaluation problem. The first dataset originates from the phase 2 clinical trial of Mark C et al. 2013, which assessed the safety and efficacy of various doses (25–300mg) of Secukinumab for rheumatoid arthritis. Each arm corresponds to a dosage with measured efficacy and toxicity probabilities, and the feasible region is defined by a minimum efficacy of 40% and acceptable toxicity. This formulation mirrors early-stage dose-escalation studies in which the goal is to identify all doses that are both effective and safe, rather than a single optimum. Figure 6.2 shows the feasible region and mean responses of each dosage level.

The second dataset is based on the CovBoost study of Munro et al. 2021, which compared 20 COVID-19 booster strategies across multiple immunological indicators. Following Crepon et al. 2024, we retain three outcomes—neutralizing antibody titre (NT_{50}), immunoglobulin G (IgG), and cellular response—forming a three-dimensional response space. The feasible region corresponds to vaccine strategies achieving IgG levels above 8.25 titre. Figure 6.3 illustrates the feasible region and the mean response of each vaccine candidate.

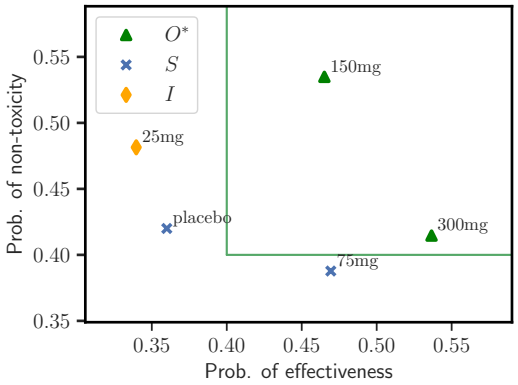


Figure 6.2: Average response in the Secukinumab dose-finding trial. The feasible region corresponds to efficacy $\geqslant 40\%$ and acceptable toxicity.

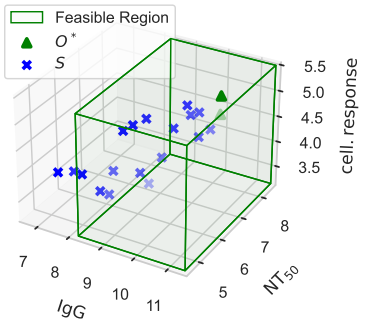


Figure 6.3: Average response in the CovBoost vaccine trial. The feasible region is defined by an IgG response above 8.25 titer.

Results. Figure 6.4 reports the empirical sample complexity over 500 runs for both datasets (the uniform sampling algorithm was omitted for CovBoost due to excessive runtime). In these experiments, e-cAPE consistently achieves lower sample complexity than its competitors. In the Secukinumab case, the 75mg dose lies near the feasibility boundary yet is dominated by the 150mg arm. While the two-stage algorithm (**A-A**) incurs additional cost verifying feasibility for such borderline arms, e-cAPE integrates feasibility and dominance testing, discarding them earlier and more efficiently. This behavior is fully consistent with the theoretical analysis of Section 6.2 which shows that decoupling feasibility from dominance can lead to sub-optimal sample allocation.

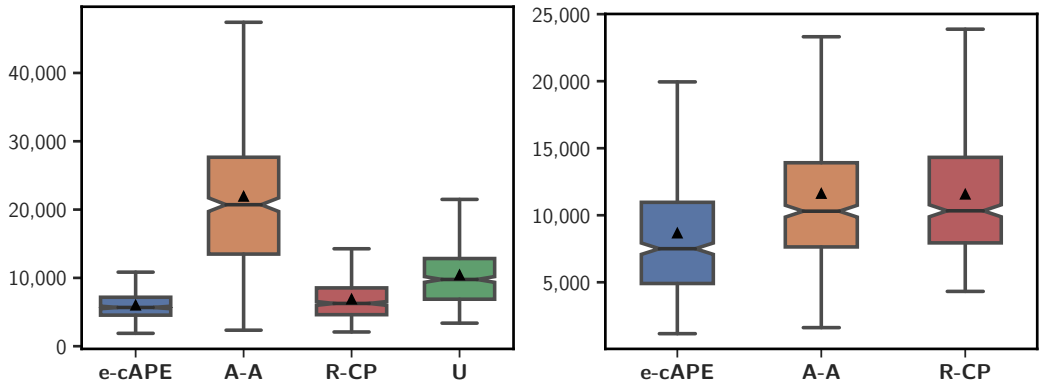


Figure 6.4: Empirical sample complexity averaged over 500 runs for the Secukinumab (left) and CovBoost (right) experiments.

Synthetic benchmarks. To further assess robustness, we test e-cAPE on synthetic three-dimensional instances defined over different feasible regions: an ordered polyhedron $P_1 = \{x \in \mathbb{R}^3 : x_i \leqslant x_{i+1}, i \in [2]\}$, a simplex $P_2 = \{x \in \mathbb{R}^3 : x_i \geqslant 0, \sum_i x_i \leqslant 1\}$, and a cube $P_3 = \{x \in \mathbb{R}^3 : 0.15 \leqslant x_i \leqslant 1, i \in [3]\}$. Figure 6.5 shows the geometry of these

instances, and Table 6.1 reports the sample complexity averaged over 500 runs. In all cases, e-cAPE achieves the best performance, often reducing the number of required samples by up to a factor 3 compared to the two-stage baseline. Interestingly, the uniform sampling strategy combined with our stopping rule sometimes approaches the two-stage baseline's performance, confirming that some of the gain of e-cAPE stems from its improved stopping condition rather than aggressive sampling alone.

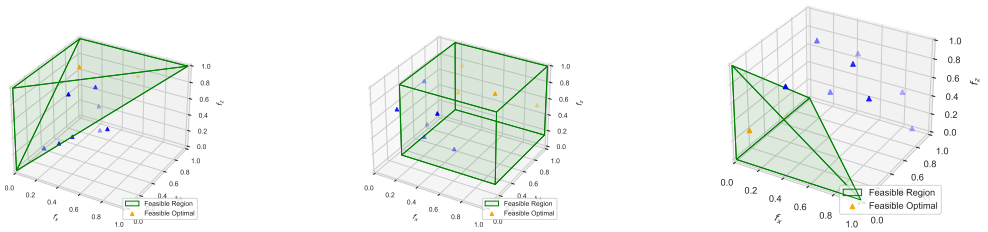


Figure 6.5: Synthetic constrained PSI instances with (left to right): ordered polyhedron, cube, and simplex. The green region denotes feasibility.

Table 6.1: Empirical sample complexity averaged over 500 runs.

Experiment	**e-cAPE**		MD-APT+APE (**A-A**)		Uniform (**U**)	
	Mean	Std	Mean	Std	Mean	Std
Simplex	**2679**	517	4886	873	12156	2428
Hypercube	**19034**	3158	59841	10186	56184	14392
Ordered polyhedron	**4313**	693	7183	792	11895	1932

Challenging instances for racing algorithms. To further highlight the limitations of elimination-based methods, we constructed a "hard" constrained PSI instance designed to challenge racing algorithms. In this two-dimensional Bernoulli setting, all arms are feasible but only one is Pareto-optimal, while the feasible region is defined by a half-space aligned with the y-axis. We considered $K \in \{5, 10\}$ arms, set $\delta = 0.01$, and averaged results over 250 runs with negligible empirical error. As anticipated from our theoretical discussion, racing algorithms must first certify that the dominating arm is feasible before discarding any dominated arm, which leads to a sample complexity that scales roughly linearly with K. In contrast, e-cAPE avoids this redundancy by jointly evaluating feasibility and dominance, resulting in substantially smaller sample requirements. This gap becomes even more pronounced as the number of arms increases, as illustrated in Figure 6.6 and Figure 6.7.

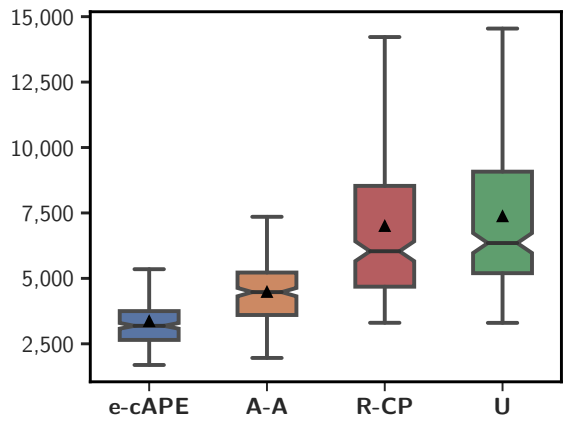


Figure 6.6: Empirical distribution of the sample complexity on the difficult constrained 5-armed instance.

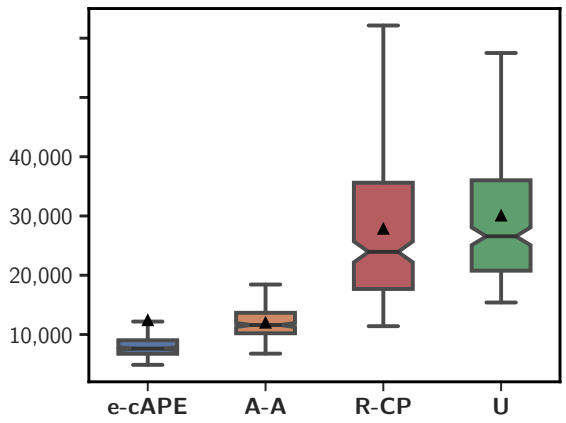


Figure 6.7: Empirical distribution of the sample complexity on the difficult constrained 10-armed instance.

Discussion. The advantage of e-cAPE is most pronounced in clinical scenarios where some candidate treatments lie close to safety or efficacy thresholds. By balancing feasibility testing and dominance exploration within a unified confidence-based framework, e-cAPE efficiently identifies all clinically acceptable strategies while explicitly explaining why others are rejected.

This feature of explainability is crucial in medical and regulatory settings, where investigators must justify exclusion decisions. In contrast to traditional racing or two-stage procedures, e-cAPE produces interpretable partitions of candidate arms into feasible, dominated, and infeasible subsets. Such interpretability supports transparent and data-driven decision-making during early-stage trials.

Conclusion. This chapter introduced the problem of constrained Pareto Set Identification (cPSI) and its explainable variant (e-cPSI), established information-theoretic lower bounds, and proposed e-cAPE, an algorithm achieving near-optimal performance in both theory and practice. Empirical results on both real and synthetic datasets confirm its efficiency and interpretability, especially in clinically motivated scenarios. Looking ahead, an exciting avenue for future research lies in designing adaptive exploration strategies that tolerate limited constraint violations during learning. Such extensions would be particularly relevant for ethical clinical trial design, where safety must be ensured throughout the study while still enabling rapid, data-efficient identification of promising treatment candidates.

6.6 Additional proofs

6.6.1 Stopping time

We upper bound the expected stopping time of Algorithm 6.1. We recall that P is fixed and known to the algorithm,

$$\mathcal{M}(P,\mu) \coloneqq \{(S,I) : S \subset \mathrm{SubOpt}, I \subset \mathrm{F}^c \text{ and } S \cap I = \emptyset, S \cup I = (\mathcal{S}_{\text{feas}})^c\}$$

is the set of correct answers. The idea of the proof is to show that if the algorithm has not stopped after round t and $\mathcal{E}_t$ holds, then at least one of b_t, c_t has not been sufficiently pulled *w.r.t.* the cost of identifying any correct response of $\mathcal{M}$. More precisely, we introduce the sets

$$W_t^1(S) \coloneqq \{i \in \mathcal{S}_{\text{feas}} : \Delta_i(\mathcal{S}_{\text{feas}} \cup S) \leqslant 4\beta_i(t,\delta) \text{ or } \eta_i \leqslant 2U_i(t,\delta)\}, \tag{6.21}$$

$$W_t^2(I) \coloneqq \{i \in I : \eta_i \leqslant 2U_i(t,\delta)\}, \tag{6.22}$$

$$W_t^3(S) \coloneqq \{i \in S : \Delta_i(\mathcal{S}_{\text{feas}} \cup S) \leqslant 4\beta_i(t,\delta)\}, \tag{6.23}$$

and finally,

$$W_t(S,I) = W_t^1(S) \cup W_t^2(I) \cup W_t^3(S).$$

We may omit the dependency in t when it is clear from the context. At round t, any arm that belongs to $W_t(S,I)$ is called under-explored *w.r.t.* (S,I). In the sequel, as $\mathcal{S}_{\text{feas}}$ is fixed and uniquely determined, we simplify notation: given $S \subset \mathrm{SubOpt}$, we write $\Delta_i(S)$ to denote $\Delta_i(S \cup \mathcal{S}_{\text{feas}})$.

Proposition 6.6.1. *If Algorithm 6.1 has not stopped at time t and $\mathcal{E}_t$ holds then for all $(S,I) \in \mathcal{M}$, $\{b_t, c_t\} \cap W_t(S,I)$ is non-empty.*

Proof of Proposition 6.6.1

In the remaining of this proof, we fix a correct answer $(S,I) \in \mathcal{M}$ and we show that if the algorithm has not stopped at time t and $\mathcal{E}_t$ holds then one of b_t, c_t is under-explored *i.e.*, $\{b_t, c_t\} \cap W_t(S,I) \neq \emptyset$. First, we address the case $(F_t \cup G_t)\backslash\{b_t\} = \emptyset$. Note that $(F_t \cup G_t)\backslash\{b_t\} = \emptyset$, implies that arms in $[K]\backslash\{b_t\}$ can all be classified as confidently infeasible at round t. Then,

$$\mathcal{S}_{\text{feas}} = \begin{cases} \{b_t\} & \text{if } b_t \in \mathcal{S}_{\text{feas}} \\ \emptyset & \text{else.} \end{cases}$$

In both cases, the gap of b_t involves the quantity η_{b_t}. Assume $b_t \in \mathrm{F}$, then $b_t \in \mathcal{S}_{\text{feas}}$.

Case 1.a If $b_t \in F_t$ then, as $G_t = \emptyset$ and $F_t \cup G_t = \{b_t\}$, we have

$$\begin{aligned} Z_2(t) &\coloneqq \min_{i \in [K] \setminus O_t} \left[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) + \right. \\ &\qquad \left. \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) \right] \\ &= \min_{i \neq b_t} \left[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) + \right. \\ &\qquad \left. \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) \right] \\ &= \min_{i \neq b_t} \left[\mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) \right] \quad \text{(since } F_t = \{b_t\}\text{)} \\ &\geqslant 0 \end{aligned}$$

which follows by definition of G_t, as $G_t = \emptyset$ and $F_t = \{b_t\}$. So if the algorithm has not stopped, $Z_1(t) < 0$. In this case, as O_t is a singleton, we simply have

$$Z_1(t) = \gamma_{b_t}(t) < 0$$

(the reader can check the formula of $Z_1(t)$ as $F_t = \{b_t\}$, we recall $\min_\emptyset = \infty$). $\gamma_{b_t}(t) < 0$ then implies $\eta_i(t) \leqslant U_i(t,\delta)$.

Next, assume $b_t \in \mathrm{F}$, then $b_t \in \mathcal{S}_{\text{feas}}$ and

$$\begin{aligned} \eta_{b_t} = \mathrm{dist}(\mu_{b_t}, P^c) &\overset{\mathcal{E}_t}{\leqslant} \mathrm{dist}(\hat{\mu}_{t,b_t}, P^c) + U_{b_t}(t,\delta) \\ &= \eta_i(t) + U_{b_t}(t,\delta) \\ &\leqslant 2U_{b_t}(t,\delta) \end{aligned}$$

where the first inequality follows from Lemma 6.3.1. Thus, if $b_t \in \mathrm{F}$ we have $b_t \in W_t^1(S)$. Now we assume $b_t \notin \mathrm{F}$. We observe that

$$\begin{aligned} \eta_{b_t} = \mathrm{dist}(\mu_{b_t}, P) &\overset{\mathcal{E}_t}{\leqslant} \mathrm{dist}(\hat{\mu}_{t,b_t}, P) + U_{b_t}(t,\delta) \\ &= 0 + U_{b_t}(t,\delta) \end{aligned}$$

which follows as by assumption $b_t \in F_t$. So we have $b_t \in W_t^2(I)$.

Case 1.b If $b_t \in G_t$, then by definition $\gamma_{b_t}(t) = \mathrm{dist}(\hat{\mu}_{t,i}, P) - U_i(t,\delta) \leqslant 0$. By Lemma 6.3.1, either $b_t \notin \mathrm{F}$ and

$$\begin{aligned} \eta_{b_t} = \mathrm{dist}(\mu_{b_t}, P) &\overset{\mathcal{E}_t}{\leqslant} \mathrm{dist}(\hat{\mu}_{t,b_t}, P) + U_{b_t}(t,\delta) \\ &\leqslant 2U_{b_t}(t,\delta) \end{aligned}$$

or $b_t \in \mathrm{F}$ (so $\mathcal{S}_{\text{feas}} = \{b_t\}$) and

$$\begin{aligned} \eta_{b_t} = \mathrm{dist}(\mu_{b_t}, P^c) &\overset{\mathcal{E}_t}{\leqslant} \mathrm{dist}(\hat{\mu}_{t,b_t}, P^c) + U_{b_t}(t,\delta) \\ &= 0 + U_{b_t}(t,\delta). \end{aligned}$$

Therefore, either $b_t \notin \mathrm{F}$ and $b_t \in W_t^2(I)$ or $b_t \in \mathrm{F} = \{b_t\} = \mathcal{S}_{\text{feas}}$ and $b_t \in W_t^1(S)$. This concludes the analysis for the case $(F_t \cup G_t) = \{b_t\}$.

In the sequel, we assume at round t, $(F_t \cup G_t) \backslash \{b_t\} \neq \emptyset$. For this proof, we treat different cases for b_t, c_t in a series of lemmas summarized in the Table 6.2 which covers all the possible cases that could happen regarding b_t, c_t.

Case	Reference
$b_t \in S,$ and $c_t \in \mathcal{S}_{\text{feas}}$	Lemma 6.6.5
$b_t \in \mathcal{S}_{\text{feas}}$ and $c_t \in \mathcal{S}_{\text{feas}}$	Lemma 6.6.4
$b_t \in I$ or $c_t \in I$	Lemma 6.6.2
$c_t \in S$	Lemma 6.6.3

Table 6.2: References to the exhaustive list of cases analyzed

The following lemmas are proved under the condition that cAPE has not stopped and $\mathcal{E}_t$ holds.

Lemma 6.6.2. *If the cAPE has not stopped and $b_t \in I$ or $c_t \in I$ then one of them satisfies $\eta_i \leqslant 2U_i(t,\delta)$,* i.e., *$W_t^2(I) \cap \{b_t, c_t\} \neq \emptyset$.*

Proof. By design of b_t it holds that

$$\mathrm{dist}(\hat{\mu}_{t,b_t}, P) \leqslant U_i(t,\delta). \tag{6.24}$$

To see this, observe that if $Z_2(t) < 0$, then, as in this case

$$b_t \coloneqq \operatorname*{argmin}_{i \in O_t} \left[\gamma_i(t) \wedge \min_{j \in O_t} \mathrm{M}^-(i,j;t) \right]$$

and $O_t \subset F_t$ we have $\hat{\mu}_{t,b_t} \in P$, so (6.24) trivially holds. If otherwise $Z_2(t) > 0$ and $Z_1(t) \leqslant 0$ then, recalling that in this case

$$b_t = \operatorname*{argmin}_{i \in [K] \backslash O_t} \left[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i,j;t) \right) + \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i,j;t) \right) \right]$$

it holds that either i) $b_t \in F_t$ so $\hat{\mu}_{t,b_t} \in P$ and (6.24) holds or ii) $b_t \notin F_t$ and $\gamma_{b_t}(t) \leqslant 0$ (otherwise $Z_2(t)$ would be non-negative). which by definition of γ_{b_t} yields Equation (6.24).

Next, if $b_t \in I$, then, as the event $\mathcal{E}_t$ holds, Lemma 6.3.1 yields

$$\eta_{b_t} \coloneqq \mathrm{dist}(\mu_{b_t}, P) \leqslant \mathrm{dist}(\hat{\mu}_{t,b_t}, P) + U_{b_t}(t,\delta)$$

and combining with (6.24) yields $\eta_{b_t} \leqslant 2U_{b_t}(t)$.

Assume instead $c_t \in I$. In this case, as $c_t \in F_t \cup G_t$ we either have $\text{dist}(\hat{\mu}_{t,c_t}, P) = 0$ (when $c_t \in F_t$) or $\gamma_{b_t}(t) \leqslant 0$ (when $c_t \in G_t \subset F_t^c$) *i.e.*, $\text{dist}(\hat{\mu}_{t,b_t}, P) \leqslant U_{b_t}(t,\delta)$ so that (6.24) holds. Thus, the proof follows as in the previous case, leading to the conclusion that

$$\eta_{c_t} \leqslant 2U_{c_t}(t,\delta),$$

and so $W_t^2(I) \cap \{b_t, c_t\} \neq \emptyset$. □

Lemma 6.6.3. *If cAPE has not stopped and $c_t \in S$ then one of b_t, c_t is under-explored.*

Proof. If $c_t \in S \subset \text{SubOpt}$, then there exists $c_t^\star \in \mathcal{O}^\star$ such that

$$\Delta_{c_t}(S) = \mathrm{m}(c_t, c_t^\star).$$

As $c_t^\star \in \text{F}$, it is easy to see that, on the event $\mathcal{E}_t$, $c_t^\star$ can not be declared as confidently infeasible at time t, so $c_t^\star \in F_t \cup G_t$. If $b_t \neq c_t^\star$ then by definition of

$$c_t \coloneqq \underset{j \in (F_t \cup G_t)\setminus\{b_t\}}{\text{argmax}} \ \mathrm{m}^+(b_t, j; t)$$

it holds that $\mathrm{m}^+(b_t, c_t; t) \geqslant \mathrm{m}^+(b_t, c_t^\star; t)$, which by definition of $\mathrm{m}^+(i, j; t)$ yields

$$\exists c \in [d] : \hat{\mu}^c_{t,c_t} - \hat{\mu}^c_{t,b_t} + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta) \geqslant \hat{\mu}^c_{t,c_t^\star} - \hat{\mu}_{t,b_t} + \beta_{b_t}(t,\delta) + \beta_{c_t^\star}(t,\delta)$$

thus when $\mathcal{E}_t$ holds this implies that

$$\exists c \in [d] : \mu^c_{c_t^\star} - \mu^c_{c_t} \leqslant 2\beta_{c_t}(t,\delta)$$

which in turn, implies

$$\Delta_{c_t}(S, I) \leqslant 2\beta_{c_t}(t,\delta). \tag{6.25}$$

If otherwise $b_t = c_t^\star \in \mathcal{S}_{\text{feas}}$ then as the algorithm has not stopped, either $Z_2(t) < 0$ or $Z_1(t) < 0$ holds. We analyze both cases below.

Case 1: $Z_2(t) < 0$**.** In this case, as by design $b_t \in O_t^c$, either

a) $b_t \in F_t$ and there exists $j \in O_t$ such that $\hat{\mu}_{t,b_t} \prec \hat{\mu}_{t,j}$, that is b_t is empirically feasible sub-optimal or

b) $b_t \notin F_t$ holds.

In the case *a)*, for the arm $j \in O_t$, as $\hat{\mu}_{t,b_t} \prec \hat{\mu}_{t,j}$, it holds that $\mathrm{M}(b_t, j; t) \leqslant 0$ so $\mathrm{M}^-(b_t, j; t) \leqslant 0$. So, as $j \in (F_t \cup G_t)$ and by definition of c_t it holds that $\mathrm{M}^-(b_t, c_t; t) \leqslant 0$ *i.e.*,

$$\mathrm{M}(b_t, c_t; t) \leqslant \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta), \tag{6.26}$$

then, note that as $b_t = c_t^\star$, $\Delta_{b_t}(S) \leqslant \Delta_{c_t}(S)$ (this follows from (6.10)) so

$$\begin{aligned}\max(\Delta_{b_t}(S), \Delta_{c_t}(S)) &\leqslant \Delta_{c_t}(S) \coloneqq \mathrm{m}(c_t, c_t^\star) = \mathrm{m}(c_t, b_t) \\ &\leqslant \mathrm{M}(b_t, c_t) \\ &\overset{(i)}{\leqslant} \mathrm{M}(b_t, c_t; t) + \beta_{b_t}(t, \delta) + \beta_{c_t}(t, \delta) \\ &\overset{(6.26)}{\leqslant} 2(\beta_{b_t}(t, \delta) + \beta_{c_t}(t, \delta))\end{aligned}$$

where (i) follows from Lemma 6.3.1. Thus, for a_t, the least explored among b_t, c_t, it holds that

$$\Delta_{a_t}(S) \leqslant 4\beta_{a_t}(t, \delta),$$

which implies that $a_t \in W_t(S, I)$ in the case $a)$. In the sub-case $b)$, we have $b_t \notin F_t$, as $b_t = c_t^\star\star \in \mathcal{S}_{\text{feas}}$, we have

$$\begin{aligned}\eta_{b_t} &= \mathrm{dist}(\mu_{b_t}, P^c) \\ &\leqslant \mathrm{dist}(\hat{\mu}_{t,b_t}, P^c) + U_{b_t}(t, \delta) \quad \text{(by Lemma 6.3.1 on } \mathcal{E}_t) \\ &\leqslant U_{b_t}(t, \delta),\end{aligned}$$

$b_t \in W_t(S, I)$ and this concludes the analysis of Case 1.

Case 2: $Z_1(t) < 0$ and $Z_2(t) \geqslant 0$. In this case, by design, $b_t \in O_t$. As

$$Z_1(t) \coloneqq \min_{i \in O_t} \left[\gamma_i(t) \wedge \min_{j \in O_t \setminus \{i\}} [\mathrm{M}^-(i, j; t)]\right]$$

is negative, and b_t is its minimizer, either $a)$: there exists $j \in O_t$ such that $\mathrm{M}^-(b_t, j; t) \leqslant 0$, which further yields

$$\mathrm{M}^-(b_t, c_t, t) \leqslant 0$$

or b): $\gamma_{b_t}(t) < 0$. In the case $a)$, proceeding similarly to Case $1.a)$ will lead to the conclusion that a_t, the least explored arm among b_t, c_t belongs to $W(S, I)$. The latter sub-case $b)$ leads to $\mathrm{dist}(\hat{\mu}_{t,b_t}, P^c) \leqslant U_{b_t}(t, \delta)$ so, as $b_t = c_t^\star \in \mathcal{O}^\star$, again we have

$$\begin{aligned}\eta_{b_t} &= \mathrm{dist}(\mu_{b_t}, P^c) \\ &\leqslant \mathrm{dist}(\hat{\mu}_{t,b_t}, P^c) + U_{b_t}(t, \delta) \quad \text{(by Lemma 6.3.1 on } \mathcal{E}_t) \\ &\leqslant 2U_{b_t}(t, \delta),\end{aligned}$$

that is $b_t \in W_t(S, I)$. Therefore, we conclude that one of b_t or c_t is under-explored. □

Lemma 6.6.4. *If $b_t \in \mathcal{S}_{feas}$ and $c_t \in \mathcal{S}_{feas}$, then one of them is under-explored.*

Proof. Note that in this case, by definition of the gaps as $b_t, c_t \in \mathcal{S}_{\text{feas}}$,

$$\Delta_{b_t}(S) \leqslant \mathrm{M}(b_t, c_t) \text{ and } \Delta_{c_t}(S) \leqslant \mathrm{M}(b_t, c_t). \tag{6.27}$$

We analyze first the case $Z_1(t) < 0, Z_2(t) \geqslant 0$. By design,

$$b_t \coloneqq \operatorname*{argmin}_{i \in O_t} \left[\gamma_i(t) \wedge \min_{j \in O_t \setminus \{i\}} \mathrm{M}^-(i,j;t) \right] ,$$

so, as $Z_1(t) < 0$, either b_t is not confidently feasible (*i.e.*, $\gamma_{b_t}(t) \leqslant 0$ in which case the proof is similar to Lemma 6.6.2) or there exists $j \in O_t \setminus \{b_t\}$ such that $\mathrm{M}^-(b_t, j; t) \leqslant 0$. Which as $j \in F_t$ (since $\mathcal{O}_t \subset F_t$) and by design

$$c_t \coloneqq \operatorname*{argmin}_{j \in (F_t \cup G_t) \setminus \{b_t\}} \left[\mathrm{M}^-(b_t, j; t) \right] ,$$

it follows that

$$\mathrm{M}^-(b_t, c_t; t) \leqslant 0. \tag{6.28}$$

Then, using concentration properties of $\mathrm{M}(i,j;t)$ (cf Lemma 6.3.1) and from (6.27), it follows

$$\begin{aligned} \max(\Delta_{b_t}, \Delta_{c_t}) &\leqslant \mathrm{M}^+(b_t, c_t, t) \\ &\leqslant \mathrm{M}(b_t, c_t; t) + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta) \\ &\leqslant 2(\beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta)) \end{aligned}$$

where the last inequality follows from $\mathrm{M}^-(b_t, c_t; t) \leqslant 0$. For a_t the least-explored among b_t, c_t, the latter inequality implies

$$\Delta_{a_t}(S) \leqslant 4\beta_{a_t}(t,\delta),$$

that is $a_t \in W_t(S, I)$, which concludes our proof in the case $Z_1(t) < 0, Z_2(t) \geqslant 0$.

In the case $Z_2(t) < 0$, we recall that

$$b_t = \operatorname*{argmin}_{i \in [K] \setminus O_t} \left[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) + \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \setminus \{i\}} \mathrm{m}^-(i,j;t) \right) \right] .$$

Assume $b_t \notin F_t$ (*i.e.*, $\hat{\mu}_{t,b_t} \notin P$). Then by design, since $Z_2(t) < 0$ we have $\gamma_{b_t}(t) < 0$ and using concentration properties of Lemma 6.3.1, and as $b_t \in \mathcal{S}_{\text{feas}}$,

$$\begin{aligned} \eta_{b_t} &= \mathrm{dist}(\mu_{b_t}, P^c) \\ &\leqslant \mathrm{dist}(\hat{\mu}_{t,b_t}, P^c) + U_{b_t}(t,\delta), \quad \text{Lemma 6.3.1 on } \mathcal{E}_t \\ &\leqslant U_{b_t}(t,\delta) , \end{aligned}$$

which follows as $\hat{\mu}_{t,b_t} \in P^c$. Thus $b_t \in W_t(S, I)$.

If $b_t \in F_t$ (*i.e.*, $\hat{\mu}_{t,b_t} \in P$), as by design $b_t \in O_t^c$, then $b_t \in O_t^c \cap F_t$ so b_t is empirically feasible sub-optimal and

$$\exists j \in O_t \text{ such that } \hat{\mu}_{t,b_t} \prec \hat{\mu}_{t,j}$$

that is there exists $\exists j \in F_t \backslash \{b_t\}$ such that $\mathrm{M}(b_t, j; t) \leqslant 0$, which as $\mathrm{M}^-(b_t, j; t) \leqslant \mathrm{M}(b_t, j; t)$ for all j yields

$$\exists j \in F_t \backslash \{b_t\} \text{ such that } \mathrm{M}^-(b_t, j; t). \tag{6.29}$$

Therefore, by design of c_t,

$$\mathrm{M}^-(b_t, c_t; t) \leqslant 0 \tag{6.30}$$

so that we have with (6.27),

$$\begin{aligned} \max(\Delta_{b_t}(S), \Delta_{c_t}(S)) &\leqslant \mathrm{M}^+(b_t, c_t, t) \\ &\overset{\mathcal{E}_t}{\leqslant} \mathrm{M}(b_t, c_t; t) + \beta_{b_t}(t, \delta) + \beta_{c_t}(t, \delta) \\ &\overset{(6.30)}{\leqslant} 2(\beta_{b_t}(t, \delta) + \beta_{c_t}(t, \delta)) \end{aligned}$$

so, for a_t, the least explored arm among b_t, c_t

$$\Delta_{a_t}(S) \leqslant 4\beta_{a_t}(t, \delta) \,,$$

that is $a_t \in W_t(S, I)$. □

Lemma 6.6.5. *If $b_t \in S$ and $c_t \in \mathcal{S}_{feas}$ then either b_t or c_t is under explored.*

Proof. If $Z_2(t) \leqslant 0$ then by design, as

$$\begin{aligned} b_t = \operatorname*{argmin}_{i \in [K] \backslash O_t} \Bigg[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i, j; t) \right) + \\ \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i, j; t) \right) \Bigg] \,, \end{aligned}$$

and

$$\begin{aligned} Z_2(t) \coloneqq \min_{i \in [K] \backslash O_t} \Bigg[\mathbb{1}\big(i \in F_t\big) \left(\max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i, j; t) \right) + \\ \mathbb{1}\big(i \notin F_t\big) \left(\gamma_i(t) \vee \max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i, j; t) \right) \Bigg] \end{aligned}$$

plus $c_t \in (F_t \cup G_t) \backslash \{b_t\}$, we have $\mathrm{m}^-(b_t, c_t; t) \leqslant 0$. We invoke concentration properties on $\mathcal{E}_t$ to have

$$\Delta_{b_t}(S) = \mathrm{m}(b_t, b_t^\star) \leqslant \mathrm{m}^+(b_t, b_t^\star; t)$$

where $b_t^\star := \operatorname{argmax}_{j \in \mathrm{F}} \mathrm{m}(b_t, j)$ and note that on the event $\mathcal{E}_t$, it holds that $b_t^\star \in (F_t \cup G_t)$ so by definition of c_t, $\Delta_{b_t} \leqslant \mathrm{m}^+(b_t, c_t; t)$ and replacing by its expression, we further get

$$\Delta_{b_t}(S) \leqslant 2(\beta_{b_t}(t, \delta) + \beta_{c_t}(t, \delta)) \,.$$

On the other side, observe that by definition of the gaps and as $b_t \in S$,

$$\Delta_{c_t}(S) \leqslant \mathrm{M}(b_t, c_t)_+ + \Delta_{b_t}(S) \tag{6.31}$$

which then yields

$$\begin{aligned}\Delta_{c_t}(S) &\leqslant (-\mathrm{m}(b_t, c_t; t) + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta))_+ + \mathrm{m}(b_t, c_t; t) + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta) \\ &\leqslant \max\left(\mathrm{m}(b_t, c_t; t) + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta), 2(\beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta))\right)\end{aligned}$$

thus

$$\Delta_{c_t}(S) \leqslant 2(\beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta)) \text{ and } \Delta_{b_t}(S) \leqslant 2(\beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta)),$$

so we conclude in the case $Z_2(t) \leqslant 0$.

Now, we assume $Z_2(t) > 0$; then $Z_1(t) < 0$ (otherwise, the algorithm would have stopped). We still have

$$\Delta_{b_t}(S) = \mathrm{m}(b_t, b_t^\star) \leqslant \mathrm{m}^+(b_t, b_t^\star; t) \leqslant \mathrm{m}^+(b_t, c_t; t)\ ,$$

then, the idea is to show that $\mathrm{m}^-(b_t, c_t, t) \leqslant 0$. We will prove that for any $i \in (F_t \cup G_t), \mathrm{m}^-(b_t, i, t) \leqslant 0$. To see this, first, note that, as $b_t \in O_t$, for any $i \in F_t, \hat{\mu}_{t,b_t} \not\prec \hat{\mu}_{t,i}$, thus for any $i \in F_t, \mathrm{m}(b_t, i, t) \leqslant 0$.

Now let $i \in G_t$. By Lemma 6.6.6, there exists $j \in O_t$ such that $\hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}$. Having $\hat{\mu}_{t,b_t} \prec \hat{\mu}_{t,i}$ would yield either $\hat{\mu}_{t,b_t} \prec \hat{\mu}_{t,i} \prec \hat{\mu}_{t,b_t}$ (if $j = b_t$) or by transitivity

$$\hat{\mu}_{t,b_t} \prec \hat{\mu}_{t,j}\ ,$$

with $j \neq b_t$, which is not possible for $b_t, j \in O_t$. Therefore, for any $i \in F_t$, $\mathrm{m}(b_t, i; t) \leqslant 0$, so combined with the previous display,

$$\forall i \in (F_t \cup G_t), \mathrm{m}(b_t, i; t) \leqslant 0\ .$$

In particular, $\mathrm{m}(b_t, c_t; t) \leqslant 0$. Moreover, recalling that on $\mathcal{E}_t$,

$$\Delta_{b_t}(S) \leqslant \mathrm{m}^+(b_t, c_t; t)\ ,$$

we have $\Delta_{b_t}(S) \leqslant \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta)$. On the other side, as in (6.31),

$$\Delta_{c_t}(S) \leqslant \mathrm{M}(b_t, c_t)_+ + \Delta_{b_t}(S),$$

therefore,

$$\begin{aligned}\Delta_{c_t}(S) &\leqslant (-\mathrm{m}(b_t, c_t; t) + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta))^+ + \mathrm{m}(b_t, c_t; t) + \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta) \\ &\leqslant 2(\beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta))\end{aligned}$$

Combined, we have proved that

$$\Delta_{c_t}(S) \leqslant \beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta) \text{ and } \Delta_{b_t}(S) \leqslant 2(\beta_{b_t}(t,\delta) + \beta_{c_t}(t,\delta))$$

which again makes it possible to conclude, as letting a_t the least explored among b_t and c_t,

$$\Delta_{a_t}(S) \leqslant 4\beta_{a_t}(t,\delta)$$

that is $a_t \in W_t(S, I)$ □

Lemma 6.6.6. *If $Z_2(t) > 0$, then, for any $i \in G_t$, there exists $j \in O_t$ such that $\hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}$.*

Proof. Recall that arms in G_t cannot yet be confidently classified as feasible or infeasible.

As $Z_2(t) > 0$, and $i \in G_t \cap O_t^c$, from the definition of $Z_2(t)$, it holds that

$$\max_{j \in (F_t \cup G_t) \backslash \{i\}} \mathrm{m}^-(i, j; t) > 0,$$

so there exists $j \in j \in (F_t \cup G_t)\backslash\{i\}$ such that $\mathrm{m}^-(i, j; t) > 0$, in particular, $\hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}$. Introducing

$$\mathfrak{J}_i \coloneqq \{j \in F_t \backslash \{i\} : \hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}\},$$

we will show that $\mathfrak{J}_i \neq \emptyset$. Let $\Omega_i \coloneqq \{j \in (F_t \cup G_t)\backslash\{i\} : \hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}\}$ and define $H_i \coloneqq \operatorname{Par} \Omega_i, \hat{\mu}_t$, the Pareto set of Ω_i based on the empirical means. By the result above, Ω_i is non-empty, so H_i is also non-empty.

We claim that $H_i \subset F_t$. Indeed, for $k \in H_t \cap F_t^c$ we have $k \in G_t \cap O_t^c$, then, as $Z_2(t) > 0$, as justified above for i, there exists $l \in (F_t \cup G_t)\backslash\{k\}$ such that $\hat{\mu}_{t,k} \prec \hat{\mu}_{t,l}$, however, as $k \in H_t \coloneqq \operatorname{Par} \Omega_i, \hat{\mu}_t$, the above is only possible if $l \notin \Omega_i$. Recalling that $l \in (F_t \cup G_t)\backslash\{k\}$ and $l \neq b_t$ we have $l \in (F_t \cup G_t)\backslash\{b_t\}$ so putting the above displays together, if $H_t \cap F_t^c \neq \emptyset$, there exists l and k such that

$$\hat{\mu}_{t,i} \prec \hat{\mu}_{t,k}; \hat{\mu}_{t,k} \prec \hat{\mu}_{t,l}$$

but $\hat{\mu}_{t,i} \not\prec \hat{\mu}_{t,l}$, which is not possible as Pareto dominance is transitive. Therefore, $H_t \subset F_t$ and since $H_t \neq \emptyset$, $\mathfrak{J}_i \neq \emptyset$. Thus, there exists $j \in F_t \backslash \{b_t\}$ such that $\hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}$. Then, either $j \in O_t$ or there exists $j_2 \in O_t$ such that, $\hat{\mu}_{t,j} \prec \hat{\mu}_{t,j_2}$. In any case, there exists $j_3 \in O_t$ (either j or j_2) such that $\hat{\mu}_{t,i} \prec \hat{\mu}_{t,j_3}$. □

6.6.2 Sample complexity: Proof of Theorem 6.4.1

Proof. Let $(S, I) \in \mathcal{M}$ be fixed. Let $T > 0$ be fixed and observe that

$$\min\left(\tau_\delta, T\right) \leqslant \left\lceil \frac{T}{2} \right\rceil + \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(t > \tau_\delta).$$

Then, assuming

$$\mathcal{E}^T \coloneqq \bigcap_{\lceil T/2 \rceil \leqslant t \leqslant T} \mathcal{E}_t \tag{6.32}$$

holds, and using Proposition 6.6.1 we have

$$\begin{aligned}
\min(\tau_\delta, T) &\leqslant \left\lceil \frac{T}{2} \right\rceil + \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(t > \tau_\delta) \\
&\leqslant \left\lceil \frac{T}{2} \right\rceil + \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(b_t \in W_t(S,I) \vee c_t \in W_t(S,I)) \\
&\leqslant \left\lceil \frac{T}{2} \right\rceil + \sum_{t=\lceil T/2 \rceil}^{T} \sum_{a=1}^{K} \mathbb{1}(b_t = a \vee c_t = a)\mathbb{1}(a \in W_t(S,I)) \\
&= \left\lceil \frac{T}{2} \right\rceil + \sum_{a=1}^{K} \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(b_t = a \vee c_t = a)\mathbb{1}(a \in W_t(S,I)) \,,
\end{aligned}$$

then, we decompose the latter sum into three terms by defining the function R for $U \subset [K]$ as

$$R(U) \coloneqq \sum_{a \in U} \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(b_t = a \vee c_t = a)\mathbb{1}(a \in W_t(S,I))$$

we have for the set $S \subset \mathrm{SubOpt}$, using the definition of $W_t^3(S)$ for $a \in S$:

$$\begin{aligned}
R(S) &= \sum_{a \in S} \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(b_t = a \vee c_t = a)\mathbb{1}(\Delta_a(S) \leqslant 4\beta_a(t,\delta)) \\
&\leqslant \sum_{a \in S} \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(b_t = a \vee c_t = a)\mathbb{1}(N_{t,a} \leqslant 32\sigma^2 \Delta_a^{-2}(S) f(t,\delta)) \\
&\leqslant \sum_{a \in S} \sum_{t=\lceil T/2 \rceil}^{T} \mathbb{1}(b_t = a \vee c_t = a)\mathbb{1}(N_{t,a} \leqslant 32\sigma^2 \Delta_a^{-2}(S) f(T,\delta)) \quad \text{(as } f \text{ is increasing)}
\end{aligned}$$

Therefore,

$$R(S) \leqslant \sum_{a \in S} \frac{32\sigma^2}{\Delta_a^2(S)} f(T,\delta). \tag{6.33}$$

Proceeding similarly for I, we obtain

$$R(I) \leqslant \sum_{a \in I} \frac{8\sigma^2}{\eta_a^2} g(T,\delta), \tag{6.34}$$

and for $\mathcal{O}^\star$, we obtain

$$R(\mathcal{O}^\star) \leqslant \sum_{a \in \mathcal{O}^\star} \max\left(\frac{8\sigma^2 g(T,\delta)}{\eta_a^2}, \frac{32\sigma^2 f(T,\delta)}{\Delta_a^2(S)} \right). \tag{6.35}$$

Therefore, recalling the expressions of f and g

$$f(T,\delta) = \log\left(\frac{4k_1KdT^\alpha}{\delta}\right) \text{ and } g(T,\delta) = 4\log\left(\frac{4k_1K5^dT^\alpha}{\delta}\right),$$

we have

$$\begin{aligned} R(\mathcal{O}^\star) &\leqslant \sum_{a\in\mathcal{S}_{\text{feas}}} 32\sigma^2 \max\left\{\frac{\log\left(\frac{4k_1K5^dT^\alpha}{\delta}\right\}}{\eta_a^2}, \frac{\log\left(\frac{4k_1KdT^\alpha}{\delta}\right)}{\Delta_a^2(S)}\right\} \\ &\leqslant \sum_{a\in\mathcal{S}_{\text{feas}}} 32\sigma^2 \max\left\{\frac{1}{\eta_a^2}, \frac{1}{\Delta_a^2(S)}\right\} \log\left(\frac{4k_1KdT^\alpha}{\delta}\right) + \sum_{a\in\mathcal{S}_{\text{feas}}} \frac{32\sigma^2}{\eta_a^2}\log\left(\frac{5^d/d}{\delta}\right). \end{aligned}$$

Combining (6.35), (6.34) and (6.33) yields

$$\begin{aligned} R([K]) &\leqslant 32\sigma^2\left(\sum_{a\in\mathcal{O}^\star}\frac{1}{\min(\eta_a,\Delta_a(S))^2} + \sum_{a\in S}\frac{1}{\Delta_a^2(S)} + \sum_{a\in I}\frac{1}{\eta_a^2}\right) f(T,\delta) + \\ &= 32\sigma^2 C(S,I) f(T,\delta) + H(\mu,I), \end{aligned}$$

where $H(\mu,I) = \sum_{a\in I\cup O^\star} 32\sigma^2\frac{\log(5^d/d)}{\eta_a^2}$. Therefore, assuming (6.32), we have proved that

$$\min\left(\tau_\delta, T\right) < \left\lceil\frac{T}{2}\right\rceil + 32\sigma^2 C(S,I) f(T,\delta) + H(\mu,I)$$

Then, taking T such that the RHS is smaller than T would yield that the algorithm stops before T. Applying Lemma 6.6.7 with $a = 128\sigma^2 C(S,I)\alpha$ and $b = (4k_1Kd/\delta)^{1/\alpha}$ yields

$$T \geqslant 2a\log(ab) \Longrightarrow a\log(bT) < T,$$

that is, letting $\widetilde{C}(S,I) = 128\sigma^2 C(S,I)\alpha$,

$$\begin{aligned} T \geqslant 2\widetilde{C}(S,I)\log\left(\widetilde{C}(S,I)(4k_1Kd/\delta)^{1/\alpha}\right) &\Longrightarrow 128\sigma^2 C(S,I)\log((4k_1Kd/\delta)T) < T \\ &\Longrightarrow 32\sigma^2 C(S,I) f(T,\delta) < T/4. \end{aligned} \tag{6.36}$$

Moreover, we trivially have for $T > 4H(\mu,I)$, $H(\mu,I) < T/4$. Combining with (6.36), for $T > T^\star(S,I) \coloneqq 2\widetilde{C}(S,I)\log\left(\widetilde{C}(S,I)(4k_1Kd/\delta)^{1/\alpha}\right) + 4H(\mu,I)$, we have $\min(\tau_\delta, T) < T$ so the algorithm must have stopped before round T. Therefore,

$$\forall T > T^\star(S,I), \tau_\delta > T \Longrightarrow (\mathcal{E}^T)^c$$

that is $\mathbb{E}_\mu[\tau_\delta] \leqslant T^\star(S,I) + \sum_{t>T^\star(S,I)} \mathbb{P}_\nu((\mathcal{E}^t)^c)$. Letting $\Lambda_\alpha \coloneqq \sum_{T\geqslant 1}\mathbb{P}_\nu((\mathcal{E}^T)^c)$ and as S, I is arbitrary, we have

$$\mathbb{E}_\mu[\tau_\delta] \leqslant \min_{(S,I)\in\mathcal{M}} T^\star(S,I) + \Lambda_\alpha$$

and from Lemma D.3 of Kone, Kaufmann, et al. 2025b,

$$\Lambda_\alpha \leqslant \frac{2^{\alpha-1}\delta}{4k_1}\sum_{T\geqslant 1}(\log(T)+1)\left(\frac{f(T,\delta\cdot d/5^d)+f(T,\delta)}{T^{\alpha-1}}\right).$$

Finally, as $x \mapsto x\log(x)$ is increasing on $(1,\infty)$, we have

$$\mathbb{E}_\mu[\tau_\delta] \leqslant 256\sigma^2 C^\star_{\mathcal{M}}(\mu)\log(128\sigma^2 C^\star_{\mathcal{M}}(\mu)(4k_1Kd/\delta)^{1/\alpha}) + 4H(\mu,\mathrm{F}^c) + \Lambda_\alpha\,.$$

□

6.6.3 Technical results

We state some concentration inequalities that are used in our proofs.

Proof. Letting $\mathcal{X}\subset\mathbb{R}^d$ non-empty we have for all arm i

$$\begin{aligned}
|\mathrm{dist}(\hat\mu_{t,i},\mathcal{X}) - \mathrm{dist}(\mu_i,\mathcal{X})| &= \left|\sup_{x\in\mathcal{X}}[-\|\mu_i - x\|_2] - \sup_{x\in\mathcal{X}}[-\|\hat\mu_{t,i}-x\|_2]\right| \\
&\overset{(i)}{\leqslant} \sup_{x\in\mathcal{X}}|\|\hat\mu_{t,i}-x\|_2 - \|\mu_i - x\|| \\
&\overset{(ii)}{\leqslant} \|\hat\mu_{t,i}-\mu_i\| \\
&\leqslant U_i(t,\delta)
\end{aligned}$$

where the last inequality follows from $\mathcal{E}_t$, (i) follows from the inequality

$$|\sup f - \sup g| \leqslant \sup|f-g|$$

and (ii) follows from the $|\|x\| - \|y\|| \leqslant \|x-y\|$. □

Lemma 6.6.7. *Let $a,b \geqslant e$. The following statement holds:*

$$t \geqslant 2a\log(ab) \Longrightarrow a\log(bt) \leqslant t.$$

Proof. First, note that $t \mapsto \log(bt)/t$ is decreasing on $[e/b,+\infty)$ and observe for $t_0 = 2a\log(ab)$, it holds that

$$\begin{aligned}
a\log(bt_0) &= a\log(2ab\log(ab)) \\
&= a\log(ab) + a\log(2\log(ab) \\
&< 2a\log(ab) \quad (\text{as } \log(x) < x/2) \\
&= t_0
\end{aligned}$$

therefore, $a\frac{\log(t_0)}{t_0} < 1$ and as the function is decreasing

$$t \geqslant t_0 \Longrightarrow a\frac{\log(bt)}{t} \leqslant a\frac{\log(bt_0)}{t_0} < 1.$$

□

6.6.4 Lower bound of e-cPSI

In this section, we prove a lower bound on the sample complexity of any δ-correct algorithm for the constrained PSI problem. As this is a pure exploration problem with multiple correct answers, it might be tempting to apply Theorem 1 of Degenne & W. Koolen 2019. However, their result does not directly apply to our setting as the authors studied the regime $\delta \to 0$ and considered answers that are singletons of $[K]$ while we consider partitions of $[K]$.

We state below the result we will prove in this section. We denote by $\widetilde{\mathcal{D}}$ a family of bandit instances that will become explicit by the end of the section.

Theorem 6.6.8. *Let $P \coloneqq \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$ and let $\nu \in \widetilde{\mathcal{D}}^K$ with parameter μ. The following holds for any δ-correct algorithm*

$$\mathbb{E}_\mu[\tau] \min_{(S',I')\in\mathcal{M}} C(S', I')\frac{g(\delta)}{2}, \text{ where } g(\delta) \geqslant (1-\delta)\log(1/\delta) - \log(4).$$

In particular,

$$\liminf_{\delta\to 0} \frac{\mathbb{E}_\mu[\tau]}{\log(1/\delta)} \geqslant \frac{C^\star_{\mathcal{M}}(\mu)}{4}.$$

In what follows, we consider $\mathcal{A}_\delta$, a δ-correct algorithm (cf Definition 6.2.3) and μ, a bandit instance, fixed and unknown to the algorithm. We assume that $\tau < \infty$ almost surely (otherwise the lower bound trivially holds). Let $\widehat{O}, \widehat{S}, \widehat{I}$ be the recommendation of the algorithm after τ (random) steps and $\mathcal{H}_\tau$ the natural filtration of the stochastic process (A_t, X_t) stopped at round τ. Since $\mathcal{A}$ is δ-correct, the following statement holds

$$\mathbb{P}_\mu\left(\widehat{O} = \mathcal{S}^\star \text{ and } (\widehat{S}, \widehat{I}) \in \mathcal{M}\right) \geqslant 1-\delta. \tag{6.37}$$

We define

$$\widetilde{S} \coloneqq \{i \in [K] \mid \mu_i \in P \text{ and } \exists j \in \mathcal{S}^\star \mid \mu_i \prec \mu_j\}$$

the set of arms that are feasible but sub-optimal. Note that alternatively, $\widetilde{S} = \mathrm{SubOpt} \cap \mathrm{F}$. Similarly, we introduce

$$\widetilde{I} \coloneqq \{i \in [K] \mid \mu_i \notin P \text{ and } \nexists j \in \mathcal{S}^\star \mid \mu_i \prec \mu_j\},$$

the set of arms that are unsafe but not uniformly worse than any other optimal arm, which rewrites as $\widetilde{I} = \mathrm{F}^c \cap \mathrm{SubOpt}^c$. In particular, $\widetilde{S} \cap \widetilde{I} = \emptyset$ and it can be observed that for any valid answer $(S, I) \subset \mathcal{M}$, $\widetilde{S} \subset S$ and $\widetilde{I} \subset I$. Thus, for any δ-correct algorithm it holds that

$$\forall i \in \widetilde{I},\ \mathbb{P}_\mu(i \in \widehat{I}) \geqslant 1-\delta \quad \text{and} \quad \forall i \in \widetilde{S},\ \mathbb{P}_\mu(i \in \widehat{S}) \geqslant 1-\delta. \tag{6.38}$$

Then by definition, every arm of $\mathcal{B} \coloneqq (\mathcal{S}^\star)^c \cap (\widetilde{S} \cup \widetilde{I})^c$ can be (correctly) classified either in $\widehat{S}$ or $\widehat{I}$, resulting in $2^{|\mathcal{B}|}$ correct answers, which can be up to 2^{K-1} correct answers.

A high probability correct answer for $\mathcal{A}_\delta$. We describe a particular correct answer under bandit μ that is very likely to be recommended by $\mathcal{A}_\delta$.

We build on $(\widetilde{S}, \widetilde{I})$ to construct a partition with (S, I) for which a property similar to Eq. (6.38) holds for every arm of (S, I). To construct such instance, note that as $\mathcal{A}_\delta$ is δ-correct, for any $i \in \mathcal{B} = (\mathcal{S}^\star)^c \cap (\widetilde{S} \cup \widetilde{I})^c$, it holds that $\mathbb{P}_\mu(i \in \widehat{S} \cup \widehat{I}) \geqslant 1 - \delta$, which yields

$$\max\left(\mathbb{P}_\mu(i \in \widehat{S}), \mathbb{P}_\mu(i \in \widehat{I})\right) \geqslant \frac{1-\delta}{2}. \tag{6.39}$$

Then, we define S, I as

$$S \coloneqq \widetilde{S} \cup \left\{i \in C : \mathbb{P}_\mu(i \in \widehat{S}) > \mathbb{P}_\mu(i \in \widehat{I})\right\}$$

and we similarly define the set

$$I \coloneqq \widetilde{I} \cup \{i \in C \mid \mathbb{P}_\mu(i \in \widehat{S}) \leqslant \mathbb{P}_\mu(i \in \widehat{I})\}.$$

By construction, (S, I) satisfies $S \cap I = \emptyset$ and $S \cup I = (\mathcal{S}^\star)^c$, *i.e.*, (S, I) is a valid answer. Moreover, one can verify that

a) $\mathbb{P}_\mu(\widehat{O} = \mathcal{S}^\star) \geqslant 1 - \delta$

b) for all $i \in S$, $\mathbb{P}_\mu(i \in \widehat{S}) \geqslant \frac{1-\delta}{2}$ and

c) for all $i \in I$, $\mathbb{P}_\mu(i \in \widehat{I}) \geqslant \frac{1-\delta}{2}$;

that is this particular $(\mathcal{S}^\star, S, I)$ is a likely response for the algorithm $\mathcal{A}_\delta$. Using a change of distribution lemma will allow us to derive a lower bound on the number of pulls of each arm for $\mathcal{A}_\delta$ to identify (S, I) as a correct answer. In the sequel, we restrict ourselves to bandits with multivariate normal arms and identity covariance. Therefore, each arm will be identified with its mean vector.

Lemma 6.6.9 (Kaufmann, Cappé, et al. 2016)**.** *For all bandit models μ, λ and for any $\mathcal{H}_\tau$-measurable event $\mathcal{E}$,*

$$\sum_{i=1}^{K} \mathbb{E}_\mu[N_{\tau,i}] \operatorname{KL}(\mu_i, \lambda_i) \geqslant \operatorname{kl}(\mathbb{P}_\mu(\mathcal{E}), \mathbb{P}_\lambda(\mathcal{E})),$$

where $\operatorname{KL}(\mu_i, \lambda_i)$ is the Kullback-Leibler divergence between distributions identified by parameters μ_i, λ_i and kl is the relative binary entropy.

Change of distribution. The idea now is to apply the classical change of distribution lemma. We define instances with means $\lambda^1, \ldots, \lambda^K$ such that λ^i differs from μ only in the mean of arm i *i.e.*, the bandit μ and λ^i are identical except $\lambda^i_i \neq \mu_i$. For such instances, applying Lemma 6.6.9 yields

$$\mathbb{E}_\mu[N_{\tau,i}] \operatorname{KL}(\mu_i, \lambda^i_i) \geqslant \sup_{\mathcal{E} \in \mathcal{H}_\tau} \operatorname{kl}(\mathbb{P}_\mu(\mathcal{E}), \mathbb{P}_{\lambda^i}(\mathcal{E})) \tag{6.40}$$

where we recall that $\mathrm{KL}(\mu_i, \lambda_i^i)$ is the Kullback-Leibler divergence between multivariate normals of means μ_i, λ_i^i and identity covariance. $\mathrm{kl}(x, y) \coloneqq x \log(x/y) + (1-x)\log((1-x)/(1-y))$, and noting that $x \mapsto \mathrm{kl}(x, y)$ is increasing on $\{x > y\}$ and decreasing on $\{x < y\}$, making an event likely under μ unlikely under λ^i will increase the RHS of Eq. (6.40). Going back to the instance (S, I) constructed earlier, we introduce the event

$$\mathcal{E}_i \coloneqq \begin{cases} \{i \in \widehat{S}\} & \text{if } i \in S \\ \{i \in \widehat{I}\} & \text{if } i \in I, \\ \{\widehat{O} = \mathcal{S}^\star\} & \text{if } i \in \mathcal{S}^\star. \end{cases}$$

We recall that $(\mathcal{O}^\star, S, I)$ forms a partition of $[K]$. From the definition of (S, I) we have shown that for any δ-correct algorithm,

$$\mathbb{P}_\mu(\mathcal{E}_i) \geqslant \begin{cases} 1-\delta & \text{if } i \in \mathcal{O} \\ \frac{1-\delta}{2} & \text{else,} \end{cases}$$

that is, such an event is very likely under bandit μ. We will choose each instance λ^i such that $\mathcal{E}_i$ is unlikely under bandit λ^i, *i.e.*, we assume that λ^i is chosen such as for all i

$$\mathbb{P}_{\lambda^i}(\mathcal{E}_i) \leqslant \delta. \tag{6.41}$$

We now discuss the choice of λ^i depending on i to ensure such a property. Thus, assuming this property holds, we have for all arms i

$$\mathbb{E}_\mu[N_{\tau,i}]\, \mathrm{KL}(\mu_i, \lambda_i^i) \geqslant \mathrm{kl}\left(\frac{1-\delta}{2}, \delta\right)$$

which, using the KL formula for multivariate normal, yields

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\|\mu_i - \lambda_i^i\|^2} \underbrace{\mathrm{kl}\left(\frac{1-\delta}{2}, \delta\right)}_{g(\delta)}.$$

Case 1: $i \in I$ **an unsafe arm.** In this case, $\mu_i \notin P$. Since P is a closed convex set, by Hilbert projection onto closed convex sets, there exists (unique) $z_i \in P$ such that $\|\mu_i - z_i\|_2 = \mathrm{dist}(\mu_i, P)$. We now define $\lambda_i^i = z_i$. In this bandit λ_i, i is now a safe arm so, as $\mathcal{A}_\delta$ is a δ-correct algorithm it holds that

$$\mathbb{P}_{\lambda_i}(i \in \widehat{I}) \leqslant \delta \quad \textit{i.e.}, \quad \mathbb{P}_{\lambda_i}(\mathcal{E}_i) \leqslant \delta,$$

which by further noting that $\|\mu_i - \lambda_i^i\|_2 = \mathrm{dist}(\mu_i, P) = \eta_i$ results in

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\eta_i^2} g(\delta).$$

Case 2: $i \in S$ **a sub-optimal arm.** Note that in this case, there exists a subset $\Omega_i \subset \mathcal{S}^\star$ such that $\mu_i \prec \mu_j$ for all $j \in \Omega_i$: it is the set of optimal arms that dominate μ_i. For any $j \in \Omega_i$ we define

$$c_j = \operatorname*{argmin}_c \left[\mu_j(c) - \mu_i(c)\right]$$

the coordinate for which the margin between i and j is the lowest. In particular, $\mathrm{m}(i,j) = \mu_j(c_j) - \mu_i(c_j)$. Moreover, for any $\varepsilon > 0$, the vector defined by

$$\widetilde{\mu}_i(c) = \begin{cases} \mu_i(c) + (1+\varepsilon)\,\mathrm{m}(i,j) & \text{if } c = c_j \\ \mu_i(c) & \text{else} \end{cases}$$

is not dominated by μ_j. Now let us define the vector s_i such that for any $c \in [d]$,

$$s_i(c) \coloneqq \max_{j \in \Omega_i : c_j = c} \mathrm{m}(i,j) \tag{6.42}$$

with the convention that $\max_\emptyset = 0$. We argue that defining

$$\lambda^i_i \coloneqq \mu_i + (1+\varepsilon)s_i,$$

we have

$$\mathbb{P}_{\lambda^i}(i \in \widehat{S}) \leqslant \delta.$$

To see this, note that as only the mean of arm i has changed between bandits μ and λ^i, the set of safe and unsafe arms of $[K]\backslash\{i\}$ under μ and λ^i are the same. Moreover, the change of distribution ensures that under the bandit λ^i, i is not dominated by any safe arm. Therefore, $\mathbb{P}_{\lambda^i}(i \in \hat{S}) \leqslant \delta$. So applying what precedes and letting $\varepsilon \to 0$ yield

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\|s_i\|^2} g(\delta). \tag{6.43}$$

Case 3: $i \in \mathcal{S}^\star$ the optimal set. We recall that arms in $\mathcal{S}^\star$ are both feasible and Pareto optimal among feasible arms. We then build alternative instances where arm $i \in \mathcal{S}^\star$ is either made infeasible or sub-optimal among feasible arms under bandit λ^i.

Letting $\eta_i = \mathrm{dist}(\mu_i, P^c) = \mathrm{dist}(\mu_i, \partial P)$ (as $\mu_i \in P$, cf (6.13)) and since ∂P is a closed, using Lemma F2 of Katz-Samuels & Scott 2019, there exists $z_i \in \partial P$ such that $\mathrm{dist}(\mu_i, \partial P) = \|\mu_i - z_i\|_2 = \eta_i$. Then, note that, as $z_i \notin P^\circ$ (the interior of P), for all $\varepsilon > 0, \exists z_i^\varepsilon \in P^c$ such that $\|z_i^\varepsilon - z_i\| \leqslant \varepsilon$. Then, letting $\lambda^i_i = z_i^\varepsilon$ (for some ε fixed). It comes that i is infeasible under λ^i, so that $\mathbb{P}_{\lambda^i}(\mathcal{E}_i) \leqslant \delta$. Therefore, we have

$$\begin{aligned} \mathbb{E}_\mu[N_{\tau,i}] &\geqslant \frac{2}{\|\mu_i - z_i^\varepsilon\|^2} g(\delta) \\ &= \frac{2}{\|\mu_i - z_i + (z_i - z_i^\varepsilon)\|^2} g(\delta). \end{aligned}$$

Letting $\varepsilon \to 0$ yields

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\|\mu_i - z_i\|^2} g(\delta) = \frac{2}{\eta_i^2} g(\delta).$$

Alternative change of distribution will prove the dependency on the other quantities involved in the gaps. Assume the gap of i is attained for an arm $j \in \mathcal{S}^\star$ with $\Delta_i = \mathrm{M}(i,j)$. In this case, j is close to dominating i so that decreasing i will make it dominated by j, as result, i becomes either feasible but dominated or non-feasible, *i.e.*, $i \notin \mathcal{S}^\star$ in both cases. We define in this case

$$\lambda^i_i \coloneqq \mu_i - (1+\varepsilon)(\mu_i - \mu_j)_+$$

where $(x)_+$ is defined component-wise for $x \in \mathbb{R}^d$. Then it can be easily checked that under bandit λ^i, arm i is now dominated by j and j is still feasible. Therefore, $\mathbb{P}_{\lambda^i}(\mathcal{E}_i) \leqslant \mathbb{P}_{\lambda^i}(i \in \widehat{O}) \leqslant \delta$. So that proceeding as before we prove that

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\|(\mu_i - \mu_j)_+\|^2} g(\delta). \tag{6.44}$$

Similarly, if we had $\Delta_i = \mathrm{M}(j,i)$ for some $j \in \mathcal{S}^\star$. Then, increasing i will make it dominate j. Thus, defining

$$\lambda^i_i \coloneqq \mu_i + (1+\varepsilon)(\mu_j - \mu_i)_+,$$

on can observe that arm j is now dominated by i under bandit λ^i. Note that j is still feasible in bandit λ^i (since its mean has not changed). So two things may occur. Either i is still feasible under λ^i in which case j is feasible and sub-optimal in bandit λ^i or i is now infeasible under bandit λ^i so that i is no longer an optimal arm. In both cases, the optimal set has changed, so that $\mathbb{P}_{\lambda^i}(\mathcal{E}_i) \leqslant \delta$. Reasoning as in the previous cases,

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\|(\mu_j - \mu_i)_+\|^2} g(\delta). \tag{6.45}$$

Therefore, by defining the quantity we have proved that

$$\mathbb{E}_\mu[N_{\tau,i}] \geqslant \frac{2}{\min(\widetilde{\delta}^+_i, \eta_i)^2} g(\delta),$$

where

$$\widetilde{\delta}^+_i = \min_{j \in \mathcal{S}^\star \setminus \{i\}} \min(\|(\mu_j - \mu_i)_+\|_2, \|(\mu_i - \mu_j)_+\|_2). \tag{6.46}$$

Note that this is a strictly positive quantity as $i \in \mathcal{S}^\star$ and from the definition of non-dominance for all $j \in \mathcal{S}^\star$.

Relation to the gaps in PSI. We recall the expressions of the PSI gaps (*w.r.t.* $\mathcal{S}^\star$) as introduced in Auer et al. 2016. For a sub-optimal arm $i \in S$,

$$\Delta_i \coloneqq \Delta_{i\star} \coloneqq \max_{j \in \mathcal{S}^\star} \mathrm{m}(i,j), \tag{6.47}$$

For an optimal arm $i \in \mathcal{S}^\star$,

$$\Delta_i(S \cup \mathcal{S}^\star) \coloneqq \min(\delta^+_i, \delta^-_i(S))$$

where

$$\delta^+_i \coloneqq \min_{j \in \mathcal{S}^\star \setminus \{i\}} \min(\mathrm{M}(i,j), \mathrm{M}(j,i)) \text{ and } \delta^-_i(S) \coloneqq \min_{j \in S}[(\mathrm{M}(j,i))^+ + \Delta_j],$$

and $\mathrm{M}(i,j) = \max_{c \in [d]}[\mu_i(c) - \mu_j(c)]$. Since $\mathcal{S}^\star$ is fixed, we ease notation and write $\Delta_i(S \cup \mathcal{S}^\star) = \Delta_i(S)$ for all $i \in \mathcal{S}^\star$.

As justified in Remark 18 of Auer et al. 2016, $\delta^-_i(S)$, can be compensated by both $\delta^+_i(S)$ and $\Delta_j\star$ for some arms $j \in \mathrm{SubOpt}$ (in our case). So we focus on matching the gaps in δ^+_i

for $i \in \mathcal{S}^\star$. Moreover, from the definition of $\mathrm{M}(i,j)$ for $i,j \in \mathcal{S}^\star$ (and from the definition of Pareto dominance), it can be easily checked that for any $i \in \mathcal{S}^\star$,

$$\delta_i^+ \coloneqq \min_{j \in \mathcal{S}^\star \setminus \{i\}} \min(\|(\mu_j - \mu_i)_+\|_\infty, \|(\mu_i - \mu_j)_+\|_\infty). \tag{6.48}$$

Intuitively, (in the case of independent objectives), (6.46) suggests to measure the distance between Pareto optimal arms in Euclidean whereas (6.48) measures in $\sup$ norm. As these quantities are computed over Pareto optimal arms, it can be checked that for $d = 2$, both terms are equal. In the worst case, the discrepancy between δ_i^+ and $\widetilde{\delta}_i^+$ is at most $\sqrt{d}$. However, as we discuss later, there are classes of PSI/constrained PSI instances where these quantities are equal.

Similarly, one can check that for a sub-optimal arm, the discrepancy between $\Delta_i^\star$ and $\|s_i\|^2$ (6.42) is at most $\min(\sqrt{|\Omega_i|}, \sqrt{d})$, where we recall that

$$\Omega_i \coloneqq \{j \in \mathcal{S}^\star : \mu_i \prec \mu_j\}.$$

Thus, in instances where Ω_i is singleton, that is each of SubOpt is dominated by a unique feasible optimal arm, $\|s_i\|_2$ and $\Delta_i^\star$ are equal. To summarize, we have been justifying that under the following,

$$\forall i \in \mathcal{S}^\star, \widetilde{\delta}_i^+ = \delta_i^+ \quad \text{and} \quad \forall i \in \mathrm{SubOpt}, \|s_i\|_2 = \Delta_i\star \tag{6.49}$$

combined with the results of sub-section 6.6.4 we would have

$$\mathbb{E}_\mu[N_{\tau,i}]/(2g(\delta)) \gtrsim \begin{cases} \frac{1}{\eta_i^2} & \text{if } i \in I \\ \frac{1}{(\Delta_i\star)^2} & \text{if } i \in S \\ \frac{1}{\min(\eta_i, \Delta_i(S))^2} & \text{if } i \in \mathcal{S}^\star. \end{cases} \tag{6.50}$$

Then, as $\mathbb{E}_\mu[\tau] = \sum_{i=1}^K \mathbb{E}_\mu[N_{\tau,i}]$,

$$\mathbb{E}_\mu[\tau] \geqslant 2C(S,I)g(\delta).$$

where we recall that

$$C(S,I) = \sum_{i \in \mathcal{S}^\star} \frac{1}{\min(\Delta_i(\mathcal{S}^\star \cup S), \eta_i)^2} + \sum_{i \in S} \frac{1}{(\Delta_i(\mathcal{S}^\star \cup S))^2} + \sum_{i \in I} \frac{1}{\eta_i^2}$$

Finally observing that (S,I) is an arbitrary correct response yields

$$\mathbb{E}_\mu[\tau] \gtrsim \min_{(S',I') \in \mathcal{M}} C(S',I')2g(\delta) \tag{6.51}$$

and the conclusion follows by observing that $2g(\delta) = 2\,\mathrm{kl}((1-\delta)/2, \delta) \geqslant (1-\delta)\log(1/\delta) - 2\log(2)$ (as $\mathrm{kl}(x,y) \geqslant x\log(1/y) - \log(2)$).

Conclusion. We now give examples of families of instances constructed to satisfy (6.49). This includes, *e.g.*, instances where the means vectors are close in many objectives except a few (2 or 3). To give more intuition, we build some examples below. We let $\widetilde{\mathcal{D}}$ be the family of multivariate normals with identity covariance and whose mean vector $\mu_1, \dots, \mu_K$ is as below.

Let $\alpha > 0$ and let $\mu_0 \in P$, and $d \geqslant 1$ (arbitrary). Define $\mu_1 = \mu_0$ and for any $1 < i \leqslant K$, $\mu_k^1 = \mu_{k-1}^1 + \alpha$ and $\mu_k^2 = \mu_{k-1}^2 - \alpha$ and $\mu_k^c = \mu_0^c$ for all $c \notin \{1, 2\}$. Such sequences of vectors are non-dominated by each other and direct computation yields for any $i \in [K]$

$$\widetilde{\delta}_i^+ = 2\alpha \quad \text{and} \quad \delta_i^+ = \alpha.$$

Thus, for $i \in \mathcal{S}^\star$ the discrepancy between $\widetilde{\delta}_i^+$ and δ_i^+ is 2, irrespective of the dimension. As for such instances, for any P, we either have $i \in \mathcal{S}^\star$ or $\mu_i \notin P$, so, (6.51) is satisfied with an extra multiplicative constant 1/4.

Together, these imply that on this set of instances, we have

$$\mathbb{E}_\mu[\tau] \gtrsim \min_{(S',I') \in \mathcal{M}} C(S', I') \frac{g(\delta)}{2} \tag{6.52}$$

and $2g(\delta) = 2\,\mathrm{kl}((1-\delta)/2, \delta) \geqslant (1-\delta)\log(1/\delta) - 2\log(2)$ which gives the claimed result,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau]}{\log(1/\delta)} \geqslant \frac{C^\star_{\mathcal{M}}(\mu)}{4}.$$

6.6.5 Complexity of Best response for cPSI

We focus on describing the procedure for computing the "best response"

$$\lambda_t \coloneqq \operatorname*{argmin}_{\lambda \in \mathrm{Alt}(O_t)} \sum_i N_{t,i} \|\hat{\mu}_{t,i} - \lambda_i\|_2^2$$

, which is the most critical part in the implementation of Game-cPSI. The algorithm itself is described and analyzed in Kone, Kaufmann, et al. 2025b.

In a recent work, Crepon et al. 2024 proposed an algorithm to compute the best response λ_t in the case of unconstrained PSI (*i.e.*, for $A = 0, b \in \mathbb{R}_+^d$) . The computational cost of their algorithm is $O\left((K(p+d) + d^3 p)\binom{p+d-1}{d-1}\right)$ where p is the size of the Pareto set. We will show that their algorithm can be adapted to the case of constrained PSI. We prove the following result, which allows us to decompose Alt into two simpler sets $\mathrm{Alt}^+(O)$ and $\mathrm{Alt}^-(O)$.

Lemma 6.6.10. *It holds that*

$$\mathit{Alt}(O) = \mathit{Alt}^-(O) \cup \mathit{Alt}^+(O) \text{ , where}$$

$$\begin{aligned} \mathit{Alt}^-(O) &\coloneqq \bigcup_{i,j \in O} \left(\{\lambda \in \mathcal{I}^K \mid A\lambda_i \not\leqslant b\} \cup \{\lambda \mid \lambda_i \leqslant \lambda_j\}\right) \\ \mathit{Alt}^+(O) &\coloneqq \bigcup_{i \notin O} \{\lambda \in \mathcal{I}^K \mid A\lambda_i \leqslant b \text{ and } \forall j \in O, \lambda_i \not\leqslant \lambda_j\}. \end{aligned}$$

Proof. We have

$$\begin{aligned}\lambda \in \text{Alt}(O) &\iff \mathcal{S}^\star(\lambda) \neq O \\ &\iff (a) : \exists i \in O \backslash \mathcal{S}^\star(\lambda) \text{ or } (b) : \exists i \notin O \text{ such that } i \in \mathcal{S}^\star(\lambda) \\ &\iff (a) : \exists (i,j) \in O^2, i \neq j \textit{s.t. } \lambda_i \notin P \text{ or } \lambda_i \prec \lambda_j \text{ or} \\ &\quad\ (b) : \exists i \notin O \textit{s.t. } \lambda_i \in P \text{ and } \quad \forall j \in O, \lambda_i \not\prec \lambda_j\end{aligned}$$

To see the direct inclusion, let $\lambda \in \text{Alt}(O)$. If $O \not\subset \text{F}(\lambda)$ then (a) follows. Next, we assume $O \subset \text{F}(\lambda)$.

As $\lambda \in \text{Alt}(O)$, either $O \backslash \mathcal{S}^\star(\lambda) \neq \emptyset$ or $\mathcal{S}^\star(\lambda) \backslash O \neq \emptyset$. Assume there exists $i \in O$ such that $i \notin \mathcal{S}^\star(\lambda)$.

If $\lambda_i \notin P$, then, as in the case above, the inclusion follows. Assume $\lambda_i \in P$, *i.e.*, i is still a feasible arm in λ. In this case, there exists $j \in \mathcal{S}^\star(\lambda)$ such that $\lambda_i \prec \lambda_j$, otherwise, we would have $i \in \mathcal{S}^\star(\lambda)$.

If $j \in O$, then the inclusion (a) follows. If $j \notin O$, then, as $j \in \mathcal{S}^\star(\lambda)$, we have: $\lambda_j \in P$ and $\forall k \in \text{F}(\lambda), \lambda_j \not\prec \lambda_k$. In particular, as $O \subset \text{F}(\lambda)$ it holds that for $k \in O$: $\lambda_j \not\prec \lambda_k$ and $j \in \mathcal{S}^\star(\lambda) \subset \text{F}(\lambda)$, so (b) follows.

Now we assume there exists $i \in \mathcal{S}^\star(\lambda)$ such that $i \notin O$, as we have $i \in \mathcal{S}^\star(\lambda)$, it holds that $\lambda_i \in P$ and $\forall j \in \text{F}(\lambda), \lambda_i \not\prec \lambda_j$, in particular, as $O \subset \text{F}(\lambda)$, we have $\forall j \in O, \lambda_i \not\prec \lambda_j$, then (b) follows.

For the reverse inclusion, assume (a) holds. Then, it follows directly that we cannot have $O = \mathcal{S}^\star(\lambda)$. Similarly, suppose (b) holds and $O = \mathcal{S}^\star(\lambda)$. Then $\exists i \notin O = \mathcal{S}^\star(\lambda)$ such that $i \in \text{F}(\lambda)$ and $\forall j \in O = \mathcal{S}^\star(\lambda), \lambda_i \not\prec \lambda_j$, *i.e.*, i is feasible in the instance λ, it does not belong to the optimal set $\mathcal{S}^\star(\lambda)$ and it is not dominated by any arm of $\mathcal{S}^\star(\lambda)$, which is impossible if $O = \mathcal{S}^\star(\lambda)$. Therefore, when (a) or (b) holds, we have $\lambda \in \text{Alt}(O)$. □

Lemma 6.6.11. *It holds that*

$$\inf_{\lambda \in Alt^-(O)} \sum_i \frac{1}{2} w_i \|\mu_i - \lambda_i\|_2^2 = \frac{1}{2} \min(\phi_1, \phi_2) \textit{ where}$$

$$\begin{aligned}\phi_1 &:= \min_{i \in O} w_i \, \textit{dist}(\mu_i, P^c)^2, \\ \phi_2 &:= \min_{i,j \in O^2, i \neq j} \frac{w_i w_j}{w_i + w_j} \sum_{c \leqslant d} (\mu_i^c - \mu_j^c)_+^2,\end{aligned}$$

and a minimizer can be computed in time $O(p^2 d + pdm)$ where m is the number of constraints and $p = |O|$.

Proof. We have

$$\text{Alt}^-(O) = \left(\bigcup_{i \in O} \Gamma_i \right) \bigcup \left(\bigcup_{\substack{i,j \in O^2 \\ i \neq j}} \Lambda_{i,j} \right),$$

with

$$\begin{aligned}\Gamma_i &:= \{\lambda \in \mathcal{I}^K \mid \lambda_i \notin P\}\\ \Lambda_{i,j} &:= \{\lambda \in \mathcal{I}^K \mid \lambda_i \prec \lambda_j\}.\end{aligned}$$

Therefore, letting $D_w(\lambda) := \sum_i \frac{1}{2} w_i \|\mu_i - \lambda_i\|_2^2$ we have

$$\inf_{\lambda \in \mathrm{Alt}^-(O)} D_w(\lambda) = \left(\min_{i \in O} \inf_{\lambda \in \Gamma_i} D_w(\lambda)) \wedge \min_{i,j \in O^2, i \neq j} \inf_{\lambda \in \Lambda_{i,j}} D_w(\lambda) \right) .$$

Next, we observe that

$$\inf_{\lambda \in \Gamma_i} D_w(\lambda) = \frac{1}{2} w_i \,\mathrm{dist}(\mu_i, P^c)^2$$

and a minimizer of this quantity can be computed in $O(md)$. Indeed, since $\mu_i \in P$ (as $i \in O$),

$$\mathrm{dist}(\mu_i, P^c) = \mathrm{dist}(\mu_i, \partial P)$$

and the latter can be computed in $O(md)$ for a polyhedron with matrix of constraints $A \in \mathbb{R}^{m,d}$ (cf Lemma H.1 of Katz-Samuels & Scott 2018). Finally Lemma 2 of Crepon et al. 2024 shows that the value of $\inf_{\lambda \in \Lambda_{i,j}} D_w(\lambda)$ and its minimizer can be computed in time $\mathcal{O}(p^2 d)$. □

Lemma 6.6.12. *There exists a polynomial-time algorithm (in $|O|$) that computes the value and a minimizer of*

$$\min_{\lambda \in Alt^+(O)} \sum_i \frac{1}{2} w_i \|\mu_i - \lambda_i\|_2^2.$$

Proof. In the instances of $\lambda \in \mathrm{Alt}^+(O)$ a novel arm is added to the Pareto set of feasible arms. From its definition in Lemma 6.6.10, we have

$$\mathrm{Alt}^+(O) = \bigcup_{i \in O^c} \mathrm{Alt}_i^+(O)$$

where

$$\mathrm{Alt}_i^+(O) := \{\lambda \in \mathcal{I}^K \mid \lambda_i \in P \text{ and } \forall j \in O, \lambda_i \not\prec \lambda_j\}.$$

Then, note that to guarantee that arm i is not dominated by j while minimizing the transportation cost, it is sufficient to move the arm (or both arms) along one objective. Therefore, to minimize the transportation cost $D(w, \lambda)$, it is sufficient to move arm i to a novel point $\lambda \in \mathbb{R}^d$ (satisfying the constraints) and move each arm $j \in O$ only along the coordinate with minimal cost to make λ not dominated by the $(\lambda_j)_{j \in O}$. This amounts to solving the following optimization problem:

$$\inf_{\lambda \in \mathrm{Alt}_i^+(O)} D(w, \lambda) = \inf_{\substack{\lambda \in \mathbb{R}^d \\ A\lambda \leqslant b}} \frac{1}{2} w_i \|\mu_i - \lambda\|_2^2 + \sum_{j \in O} \frac{1}{2} w_j \min_{c \leqslant d} (\mu_j^c - \lambda^c)_+^2 . \tag{6.53}$$

The rightmost problem is related to the optimization problem studied by Crepon et al. 2024, except that now we have additional linear constraints $A\lambda \leqslant b$ due to the constrained PSI

setting. However, we will show that their algorithm can still be adapted to efficiently solve (6.53). To see this, we let ϕ be a mapping from O to $[d]$ and introduce $h_i^\phi : \mathbb{R}^d \to \mathbb{R}_+$, defined as

$$h_i^\phi(\lambda) \coloneqq \frac{1}{2} w_i \|\mu_i - \lambda\|_2^2 + \sum_{j \in O} \frac{1}{2} w_j (\mu_j^{\phi(j)} - \lambda^{\phi(j)})_+^2.$$

Then note that the optimization problem in Equation 6.53 rewrites as

$$\min_{\phi \in [d]^p} \inf_{\substack{\lambda \in \mathbb{R}^d \\ A\lambda \leqslant b}} h_i^\phi(\lambda)\,, \tag{6.54}$$

where the notation $[d]^p$ denotes the set of mapping from O to $[d]$ and $p = |O|$. However, as shown by Crepon et al. 2024, not all such maps are valid, and there is no need to optimize h_i^ϕ for an invalid map. The authors proposed an efficient graph-based algorithm to enumerate all the valid maps. They showed that the number of valid maps is bounded by $\binom{p+d-1}{d-1}$. Now, it remains to minimize h_i^ϕ under the linear constraints for each valid map. Note that up to a reordering of the quantities, $(\mu_j^{\phi(j)})_{j \in O}$, h_i^ϕ is piecewise quadratic, and the problem can be decomposed into at most p^d convex quadratic problems with constraints $\widetilde{A}\lambda \leqslant \widetilde{b}$ where $\widetilde{A} \in \mathbb{R}^{m+2d,d}$ and $\widetilde{b} \in \mathbb{R}^{m+2d}$. It is known that solving a convex QP can be done in polynomial time (Ye & Tse 1989). Thus, when the number of arms is large and d is small, the overall complexity of problem 6.54 is $\mathcal{O}(p^{2d} s(m+d, d))$, where $s(m+2d, d)$ is the time complexity of a convex QP with $m + 2d$ constraints in dimension d. □

Chapter 7

Conclusion and Perspectives

Summary of Contributions

This dissertation investigated the design and analysis of algorithms for *pure exploration* in stochastic multi-objective decision problems, with a particular focus on identifying the *Pareto set* of optimal arms. Our goal was to develop theoretically grounded and computationally tractable strategies that achieve near-optimal performance in both the fixed-confidence and fixed-budget settings, while accommodating correlations and structural assumptions relevant to real-world applications such as early-stage clinical trials and multi-criteria recommender systems.

Unstructured Pareto Set Identification. The first part of the dissertation introduced the Pareto Set Identification (PSI) problem in its general, unstructured form. We analyzed PSI in the fixed-budget regime and proposed the Empirical Gap Elimination (EGE) algorithm, which estimates sub-optimality gaps associated with the difficulty of certifying each arm and progressively eliminates arms with larger gaps. We established an exponentially decreasing upper bound on its error probability. We proved that two specific variants, EGE-SH and EGE-SR, achieve near-optimal decay rates, matching the theoretical lower bounds up to logarithmic factors. Empirical evaluations confirmed these theoretical results, demonstrating that adaptive elimination markedly improves the efficiency of Pareto set identification.

In the fixed-confidence setting, we proposed Adaptive Pareto Exploration (APE), the first fully adaptive LUCB-type algorithm for PSI. We formalized several practical relaxations of the PSI objective, such as $(\varepsilon_1, \varepsilon_2)$-PSI and ε_1-PSI-k, and proved that they offer substantial reductions in sample complexity while fitting practical purposes.

Applications to clinical-trial-inspired datasets illustrated how adaptive exploration can efficiently identify promising treatment strategies across multiple efficacy criteria.

Linear PSI. Building on these foundations, we extended PSI to the linear setting, where arm means are linearly related through a shared but unknown parameter matrix θ and known feature vectors. We developed the G-optimal Empirical Gap Elimination (GEGE) algorithm, which integrates G-optimal design exploration with gap-based elimination. Analyzed under both the fixed-budget and fixed-confidence regimes, GEGE achieves tight

instance-dependent sample complexity depending only on the h smallest sub-optimality gaps, where h is the feature dimension, leading to substantial gains when the number of arms is large compared to h. Extensive simulations corroborated these theoretical insights, showing that exploiting linear structure dramatically improves sample efficiency compared to unstructured approaches.

Asymptotic Optimality and Correlated Objectives. The fifth chapter of the dissertation addressed *asymptotic optimality* in PSI, particularly under *correlated objectives*. We introduced PSIPS (Posterior Sampling for PSI), a posterior sampling-based algorithm that leverages posterior draws for both sampling and stopping. PSIPS removes the need to predefine the confidence level δ, avoids heavy computations required by Chernoff stopping rules, and can exploit correlated objectives to reduce the sample complexity. We proved that PSIPS attains *information-theoretic optimality*, with sample complexity matching the asymptotic lower bound. Empirically, PSIPS not only achieved comparable or superior accuracy to confidence-based methods but also exhibited large computational speedups compared to existing asymptotically optimal gradient-based approaches. In correlated settings, PSIPS may efficiently exploit the dependency structure to reduce sampling costs.

Constrained Pareto Set Identification. Finally, we extended PSI to account for feasibility constraints, where only non-dominated arms satisfying some linear constraints should be identified. We studied two variants of this problem: the non-explainable setting, where the goal is solely to identify the feasible Pareto set, and the explainable setting, where the learner must additionally provide reasons for excluding each non-selected arm (infeasible or Pareto-dominated). We proposed e-cAPE, a constrained extension of APE adapted to the explainable case, and proved its near-optimality for identifying the feasible Pareto-optimal set.

Perspectives and Open Problems

The results presented in this dissertation open several promising directions for future research in multi-objective pure exploration and beyond. While the algorithms developed here provide nearly optimal solutions in terms of sample complexity, several extensions remain to be explored to make these methods more adaptive, scalable, and applicable to richer decision-making environments.

Structured Correlation Models and Beyond Gaussianity. The current framework assumes a known covariance matrix. Extending PSI to the case of unknown or arm-dependent correlations, possibly learned online, would be of high practical value. Such extensions would enable the algorithms to adapt not only to arm means but also to the structure of dependencies among objectives, which could further reduce sample complexity in applications where correlation patterns differ across arms and are difficult to estimate upfront. Also, while convenient in theory, the Gaussian assumption is too restrictive, as many real-world outcomes (especially clinical or biological data) exhibit heavy tails, discrete distributions, or

log-normal distributions. Deriving tight guarantees for PSI under heavy-tailed distributions could be an interesting direction.

Scalability and Computational Efficiency. While PSIPS is more efficient than previous methods, sampling-based stopping still involves repeated posterior draws, which can become burdensome for large K or high-dimensional objectives. Developing low-variance approximations or variance reduction techniques will improve scalability and make these methods more practical for large-scale and high-dimensional applications.

Cost-Efficient Multi-Objective Pure Exploration. A natural continuation of this work is to explicitly account for the *cost of observations*. In many applications, observing all marginals of a multi-dimensional outcome is neither feasible nor desirable, as each marginal may incur a distinct computational, financial, or ethical cost. Developing cost-efficient PSI algorithms where the learner adaptively selects both arms and marginals to observe, aiming to minimize the total sampling cost while preserving δ-correctness, leads to a novel class of algorithms. This setting raises several open questions: How can one design *adaptive marginal selection strategies* that optimally balance exploration and cost under correlated objectives? Can one derive *instance-dependent cost lower bounds* analogous to the information-theoretic complexity introduced in this work? How does correlation reduce the overall cost of identifying the Pareto set?

A deeper understanding of such cost-aware algorithms could substantially impact applications such as adaptive clinical trials and personalized recommendation systems, where each measurement carries a distinct burden, whether financial or logistical.

Extension to Reinforcement Learning. Extending multi-objective pure exploration principles to reinforcement learning (RL) remains an exciting research frontier. Many practical tasks require reasoning over trajectories and state transitions, for example, long-term treatment administration to a patient. An important next step is to formulate *multi-objective best policy identification* in finite-horizon Markov decision processes, where one seeks to identify policies that are Pareto-optimal with respect to multiple criteria (*e.g.*, reward, risk, energy consumption). Recent advances in pure exploration for MDPs provide strong methodological foundations (A. J. Wagenmaker et al. 2022), yet no work has fully solved this problem. The challenge will lie in designing algorithms that combine efficient exploration with appropriate extensions of the sub-optimality gaps studied in this dissertation to some notion of "policy gaps." The algorithms developed here could provide a starting point for such extensions to RL settings.

Towards Projection-Free Pure Exploration. While PSIPS was introduced with a focus on simplicity and computational tractability, it demonstrates that the posterior-resampling stopping rule is a viable alternative to the Chernoff stopping rule for PSI. Nevertheless, several challenges remain. Recent advances (Kaufmann & W.-M. Koolen 2021) have considerably simplified and improved the analysis of correctness for Chernoff stopping rules, making them increasingly popular, yet they remain hardly tractable except for problems with simple

or no structure. The posterior-resampling stopping rule offers a computationally tractable alternative to the GLR stopping rule. However, its correct calibration is problem-specific and relies on non-trivial anti-concentration results. Extending the GLR stopping rule to more general pure exploration problems and combining it with posterior-based sampling rules, in particular, top-two sampling strategies, will provide a general framework for computationally efficient and projection-free pure exploration applicable to a broader class of problems.

List of Datasets

CovBoost Dataset

Source and scope. We use immunogenicity outcomes from the Cov-Boost clinical trial of Munro et al. 2021. For each vaccine strategy ("arm") and for each of three immunogenicity indicators—anti-spike IgG, neutralizing antibody titre (NT_{50}), and wild-type cellular response—the paper reports: (i) the geometric mean response, (ii) cohort size, and (iii) a 95% confidence interval (CI) around the geometric mean, under a log-normal measurement model.

Modeling assumptions. Consistent with Munro et al. 2021, we assume that each immunogenicity indicator is log-normally distributed. Consequently, the *log*-responses are approximately Gaussian. Empirically, the three indicators are weakly correlated; we therefore model them as *independent* and take a diagonal covariance matrix.

Recovering means on the log scale. Let $x_1, \ldots, x_n$ denote the raw (positive) biological measurements for a given arm and indicator, assumed log-normal. The reported geometric mean is

$$\bar{x}_{\mathrm{geo}} = \Big(\prod_{i=1}^{n} x_i\Big)^{1/n}.$$

On the log scale, the empirical mean is

$$\bar{x} \;=\; \frac{1}{n}\sum_{i=1}^{n} \log x_i \;=\; \log(\bar{x}_{\mathrm{geo}}),$$

which we use as the (Gaussian) mean parameter for that arm/indicator.

Recovering variances on the log scale. The paper provides a 95% CI for the geometric mean. On the log scale, this induces a CI for $\bar{x}$ of the form

$$\bar{x} \pm 1.96 \cdot \mathrm{SE}_{\log}, \qquad \text{with } \mathrm{SE}_{\log} \;=\; \frac{\log U - \log L}{2 \cdot 1.96},$$

where $[L, U]$ is the reported CI in the original (positive) scale. From $\mathrm{SE}_{\log}$ and the cohort size n, we estimate the sample variance on the log scale via $s^2 \approx n \cdot \mathrm{SE}_{\log}^2$. To reduce heteroscedasticity across arms, we pool per-indicator variances across all arms using the

usual unbiased pooled estimator

$$s^2_{\text{pooled}} \;=\; \frac{\sum_{a=1}^{K}(n_a-1)\,s_a^2}{\sum_{a=1}^{K}(n_a-1)}.$$

The resulting pooled log-variances (one per indicator) are reported in Table 7.2.

Bandit abstraction. We thus obtain a $K = 20$-armed, $d = 3$-dimensional Gaussian bandit on the *log* scale: for arm i, $X_i \sim \mathcal{N}(\mu_i, \Sigma)$ with

$$\mu_i \in \mathbb{R}^3 \text{ given by the three log-means in Table 7.1}, \quad \Sigma = \operatorname{diag}\big(\, s^2_{\text{IgG}},\, s^2_{\text{NT}_{50}},\, s^2_{\text{cell}} \,\big).$$

Sampling an arm simulates measuring the (log) response of a new patient on the three indicators.

Arm nomenclature and grouping. Arms correspond to *three-dose* strategies and are grouped by the first two doses ("Prime BNT/BNT" or "Prime ChAd/ChAd"). Within each group, the row label denotes the third (booster) dose. For example, in the "Prime BNT/BNT" group, the arm labeled "ChAd" means BNT as dose 1 and 2, and ChAd as the booster.

Table 7.1: Table of the empirical arithmetic mean of the log-transformed immune response for three immunogenicity indicators. Each acronym corresponds to a vaccine. There are two groups of arms corresponding to the first 2 doses: one with prime BNT/BNT (BNT as first and second dose) and the second with prime ChAd/ChAd (ChAd as first and second dose). Each row in the table gives the values of the 3 immune responses for an arm (*i.e.*, a combination of three doses).

Dose 1/Dose 2	Dose 3 (booster)	Immune response		
		Anti-spike IgG	NT_{50}	cellular response
Prime BNT/BNT	ChAd	9.50	6.86	4.56
	NVX	9.29	6.64	4.04
	NVX Half	9.05	6.41	3.56
	BNT	10.21	7.49	4.43
	BNT Half	10.05	7.20	4.36
	VLA	8.34	5.67	3.51
	VLA Half	8.22	5.46	3.64
	Ad26	9.75	7.27	4.71
	m1273	10.43	7.61	4.72
	CVn	8.94	6.19	3.84
Prime ChAd/ChAd	ChAd	7.81	5.26	3.97
	NVX	8.85	6.59	4.73
	NVX Half	8.44	6.15	4.59
	BNT	9.93	7.39	4.75
	BNT Half	8.71	7.20	4.91
	VLA	7.51	5.31	3.96
	VLA Half	7.27	4.99	4.02
	Ad26	8.62	6.33	4.66
	m1273	10.35	7.77	5.00
	CVn	8.29	5.92	3.87

Table 7.2: Pooled variance of each group.

	Immune response		
	Anti-spike IgG	NT_{50}	cellular response
Pooled sample variance	0.70	0.83	1.54

List of Notation

Global Notation

Acronyms and abbreviations

i.i.d.	independent and identically distributed
i.e.	*id est* (that is)
s.t.	such that
w.r.t.	with respect to
BAI	Best Arm Identification
PSI	Pareto Set Identification
FB	fixed-budget
FC	fixed-confidence
UCB	upper confidence bound
LCB	lower confidence bound
KL	Kullback–Leibler
GLR	generalized likelihood ratio
PAC	Probably Approximately Correct
APE	Adaptive Pareto Exploration
EGE	Empirical Gap Elimination
GEGE	G-optimal Empirical Gap Elimination
PSIPS	Pareto Set Identification with Posterior Sampling
e-cAPE	*explainable constrained* Adaptive Pareto Exploration

Multi-Armed Bandits

$K \in \mathbb{N}$	number of arms (candidate strategies/ treatments)
$d \in \mathbb{N}$	number of objectives (coordinates)
$[K] \coloneqq \{1, \dots, K\}$	set of arms; indices $i, j, k, a \in [K]$
$[d] \coloneqq \{1, \dots, d\}$	set of objectives; index $c \in [d]$
$\Sigma \in \mathbb{R}^{d \times d}$	covariance matrix of the d-dimensional outcomes (when applicable)
$\nu_i \in \mathcal{D}$	outcome distribution of arm i (supported on $\mathbb{R}^d$ in the multi-objective case; scalar if $d = 1$)
$\nu := (\nu_1, \dots, \nu_K)$	bandit instance (collection of arm distributions)

$\mu_i \coloneqq \mathbb{E}_{X\sim\nu_i}[X]$	mean outcome of arm i (vector in $\mathbb{R}^d$ or scalar, μ_k^c its c-th coordinate)
$\mu \coloneqq (\mu_1, \dots, \mu_K) \in \mathcal{I}^K$	mean parameter (with $\mathcal{I} \subseteq \mathbb{R}^d$ the set of feasible mean vectors)
$A_t \in [K]$	arm pulled at round t
$N_{t,i}$	number of pulls of arm i up to time $t-1$
$Z_t \in \mathbb{R}^d$	observed outcome at round t; $Z_t \mid A_t \sim \nu_{A_t}$
$\hat{\mu}_{t,i}$	empirical mean of arm i based on $N_{t,i}$ samples
$\hat{\mu}_t$	stacked vector $(\hat{\mu}_{t,i})_{i\in[K]}$
$\mathcal{H}_t \coloneqq \sigma(A_1, Z_1, \dots, A_t, Z_t)$	history / filtration up to round t
σ^2	variance proxy for subgaussian arm distributions

Probability theory

$X \sim \nu$	random variable/vector X has distribution ν
$\mathbb{P}(E)$	probability of event E
$\mathbb{E}[X]$	expectation of X
$\mathbb{P}_\nu$	probability under bandit instance ν
$\mathbb{E}_\nu$	expectation under $\mathbb{P}_\nu$
$\mathbb{1}(E)$	indicator of event E
$\mathrm{KL}(P\|Q)$	Kullback–Leibler divergence between distributions P and Q
$\mathrm{kl}(p, q)$	binary relative entropy (Bernoulli KL)
$\triangle_K$	$(K-1)$-dimensional probability simplex $\{w \in \mathbb{R}_+^K : \sum_{k=1}^K w_k = 1\}$

Pure exploration

$\mathcal{S}^\star : \mathcal{I}^K \to 2^{[K]}$	query task; $\mathcal{S}^\star(\mu)$ is the set of acceptable answers
$(A_t)_{t\geqslant 1}$	sampling rule (predictable *w.r.t.* $(\mathcal{H}_t)_{t\geqslant 1}$)
$(\widehat{S}_t)_{t\geqslant 1}$	recommendation rule (answer at time t, may be an index, a subset of $[K]$, or another object)
τ	stopping time (with respect to $(\mathcal{H}_t)$)
$\widehat{S}_\tau$	final output
$\delta \in (0, 1)$	fixed-confidence risk parameter
δ-PAC	$\mathbb{P}_\nu(\widehat{S}_\tau \neq \mathcal{S}^\star(\mu)) \leqslant \delta$ (single correct answer convention)

T	fixed-budget horizon (typically $\tau = T$)

Algebra and asymptotics

$(x)_+$	$\max\{x, 0\}$
$a \wedge b,\ a \vee b$	$\min\{a, b\},\ \max\{a, b\}$
$\mathcal{O}(\cdot), \Omega(\cdot), o(\cdot)$	Landau notation
$\log,\ \log_2$	natural and base-2 logarithms
$\langle x, y\rangle,\ x^\top y$	inner product
$\|x\|_2, \|x\|$	Euclidean norm
$\|x\|_\infty$	ℓ_∞ norm: $\max_{c\in[d]} \|x^c\|$
I_d	$d \times d$ identity matrix
$\mathbf{1}/1_d, \mathbf{0}/0_d$	all-ones and all-zeros vectors in $\mathbb{R}^d$
$A, A^\mathsf{c}, \partial A$	a set, its complement, and its boundary

Pareto dominance and Pareto set

$\mu_i \leqslant \mu_j$	componentwise order: $\mu_i^c \leqslant \mu_j^c$ for all $c \in [d]$
$\mu_i \preceq \mu_j$	Pareto dominance: $\mu_i \leqslant \mu_j$ and strict in at least one coordinate
$\mu_i \prec \mu_j$	strict Pareto dominance: $\mu_i^c < \mu_j^c$ for all $c \in [d]$
$i \preceq j,\ i \prec j$	shorthand for $\mu_i \preceq \mu_j$ and $\mu_i \prec \mu_j$
$\mathcal{S}^\star(\mu)$	Pareto set: $\{i \in [K] : \nexists j \in [K] \text{ s.t. } \mu_i \prec \mu_j\}$
$\mathrm{Alt}(S^\star(\mu)) := \{\lambda : S^\star(\lambda) \neq S^\star(\mu)\}$	alternative instances with a different Pareto set
$\mathrm{m}(i,j) := \min_c[\mu_j^c - \mu_i^c]$	margin of non-domination of i by j
$\mathrm{M}(i,j) := \max_c[\mu_i^c - \mu_j^c]$	margin by which i leads j
$\Delta_i^\star := \max_{j\in\mathcal{S}^\star} m(i,j)$	gap for sub-optimal i (distance to Pareto front)
$\delta_i^+ := \min_{j\notin\mathcal{S}^\star} \min\{\mathrm{M}(i,j), \mathrm{M}(j,i)\}$	margin-to-being-dominated for optimal arm i
$\delta_i^- := \min_{j\in\mathcal{S}^\star\setminus\{i\}} \left[\mathrm{M}(j,i)_+ + \Delta_j^\star\right]$	margin-to-sub-optimality for optimal arm i
$\Delta_i := \begin{cases}\Delta_i^\star, & i \notin \mathcal{S}^\star,\\ \min\{\delta_i^+, \delta_i^-\}, & i \in \mathcal{S}^\star,\end{cases}$	combined suboptimality gap
$H(\nu) := \sum_{i=1}^K \Delta_i^{-2}$	instance complexity (governs fixed-confidence sample complexity)
$H_2(\nu) := \max_{k\in[K]} k\,\Delta_{(k)}^{-2}$	fixed-budget complexity ($\Delta_{(1)} \leqslant \cdots \leqslant \Delta_{(K)}$ ordered gaps)

Fixed-Budget PSI

Fixed-budget objective

$e_T(\nu) := \mathbb{P}_\nu(\widehat{S}_T \neq S^\star(\nu))$	misidentification probability (exact PSI)
$e_{T,k}(\nu) := \mathbb{E}_\nu\big[L(\widehat{S}_T, k)\big]$	k-relaxed expected loss (PSI-k objective)
$\mathcal{L}(\widehat{S}, k) := \begin{cases} \mathbb{1}\big(\widehat{S} \not\subset S^\star\big), & \lvert\widehat{S}\rvert = k, \\ \mathbb{1}\big(\widehat{S} \neq S^\star\big), & \text{otherwise}, \end{cases}$	loss function for PSI-k

EGE algorithm structure

R	total number of elimination rounds
λ_r	active set size threshold at the start of round r
t_r	number of samples per arm in round r
$\mathcal{A}_r \subseteq [K]$	active arm set at round r
$\mathcal{B}_r,\ \mathcal{D}_r$	arms classified optimal / sub-optimal through round r
$S_r := \mathrm{EmpPareto}(\mathcal{A}_r)$	empirical Pareto set of active arms at round r

Empirical primitives and unified gap

$\hat{\mu}_{r,i}$	empirical mean of arm i at the end of round r
$\mathrm{m}(i,j;r) := \min_{c\in[d]} \left[\hat{\mu}^c_{r,j} - \hat{\mu}^c_{r,i}\right]$	empirical version of $m(i,j)$ at round r
$\mathrm{M}(i,j;r) := \max_{c\in[d]} \left[\hat{\mu}^c_{r,i} - \hat{\mu}^c_{r,j}\right]$	empirical version of $M(i,j)$ at round r
$\widehat{\Delta}^\star_{i,r} := \max_{j\in\mathcal{A}_t} \mathrm{m}(i,j;r)$	gap for sub-optimal i
$\widehat{\delta}^\star_{i,r} := \min_{j\in\mathcal{A}_r\setminus\{i\}} [\mathrm{M}(i,j;r) \vee (\mathrm{M}(j,i;r)_+ + (\widehat{\Delta}^\star_{j,r})_+)]$	gap for optimal arm k
$\widehat{\Delta}_{i,r} := \max\{\widehat{\Delta}^\star_{i,r}, \widehat{\delta}^\star_{i,r}\}$	unified gap (estimable without prior knowledge of $S^\star$)
$\widetilde{T}^{R,\mathbf{t},\lambda}(\nu)$	generic EGE error exponent

PSI-k relaxation

$k \in [K]$	target number of Pareto-optimal arms to return
$\omega(k) := \max^{(k)}_{i\in\mathcal{S}^\star} \Delta_i$	k-th largest gap among optimal arms
$\Delta^{(k)}_i := \begin{cases} \max\{\Delta_i, \omega(k)\}, & i \in \mathcal{S}^\star, \\ \Delta_i, & \text{otherwise}, \end{cases}$	k-relaxed gap
$H^{(k)}_2(\nu) := \max_{i\in[K]} i\,(\Delta^{(k)}_{(i)})^{-2}$	k-relaxed fixed-budget complexity

Metrics

$\mathrm{HV}(S)$	hypervolume indicator of arms in S (empirical evaluation metric)

Fixed-Confidence PSI

Relaxed Pareto identification

$\varepsilon_1, \varepsilon_2 \geqslant 0$	ε_1: Pareto-margin relaxation; ε_2: domination relaxation
$\mathcal{S}^\star_{\varepsilon_1} := \{i \in [K] : \nexists j,\ \mu_i + \varepsilon_1 \mathbf{1} \prec \mu_j\}$	ε_1-Pareto set (additive relaxation of the Pareto set)
$k \in [K]$	size parameter for k-PSI: return at most k Pareto-optimal arms

Empirical primitives

$\mathrm{m}(i,j;t),\ \mathrm{M}(i,j;t)$	empirical m, M computed from $\hat{\mu}_t$
$\mathrm{M}^+(i,j;t),\ \mathrm{M}^-(i,j;t)$	upper / lower confidence bounds on $\mathrm{M}(i,j)$
$\mathrm{m}^+(i,j;t),\ \mathrm{m}^-(i,j;t)$	upper / lower confidence bounds on $\mathrm{m}(i,j)$
$\beta_{i,j}(t,\delta)$	confidence radius for pair (i,j) (LIL-type bound)
$L_i(t) := \hat{\mu}_{t,i} - \beta_i(t,\delta), \quad U_i(t) := \hat{\mu}_{t,i} + \beta_i(t,\delta)$	lower / upper confidence bounds for arm i

APE stopping statistics

$S(t) := \{i : \nexists j,\ \hat{\mu}_{t,i} \prec \hat{\mu}_{t,j}\}$	empirical Pareto set at time t
$\mathrm{OPT}_{\varepsilon_1}(t) := \{i : \forall j \neq i,\ \mathrm{M}^-(i,j;t) + \varepsilon_1 > 0\}$	ε_1-empirically optimal set at time t
$g_i^{\varepsilon_2}(t) := \max_j \big[\mathrm{m}^-(i,j;t) + \varepsilon_2 \mathbb{1}\big(j \in \mathrm{OPT}_{\varepsilon_1}(t)\big)\big]$	stopping score for sub-optimal arm i
$h_i^{\varepsilon_1}(t) := \min_j \big[\mathrm{M}^-(i,j;t) + \varepsilon_1\big]$	stopping score for optimal arm i
$Z_1^{\varepsilon_1}(t) := \min_i h_i^{\varepsilon_1}(t)$	aggregate optimality statistic
$Z_2^{\varepsilon_1}(t) := \min_{i \in S(t)^c} \max\{g_i(t), h_i^{\varepsilon_1}(t)\}$	aggregate sub-optimality statistic

Stopping times

$\tau_{\varepsilon_1} := \inf\{t \geqslant K :\ Z_1^{\varepsilon_1}(t) > 0\ \wedge\ Z_2^{\varepsilon_1}(t) > 0\}$	ε_1-PSI stopping time
$\tau_{\varepsilon_1,\varepsilon_2}$	$(\varepsilon_1, \varepsilon_2)$-PSI stopping time
$\tau^k := \inf\{t \geqslant K :\ \lvert\mathrm{OPT}_{\varepsilon_1}(t)\rvert \geqslant k\}$	k-PSI stopping time
$\tau^k_{\varepsilon_1} := \min(\tau_{\varepsilon_1}, \tau^k)$	ε_1-PSI-k stopping time

PSI-k gaps and complexity

ω_i	gap of optimal arm i to the k-th best
$\omega^k := \min_{i \in S^{\star,k}} \omega_i$	k-th worst gap among optimal arms
$\widetilde{\Delta}^k_a := \max\{\omega^k, \varepsilon_1, \Delta_a\}$	modified gap for PSI-k
$H(\nu) := \sum_a (\widetilde{\Delta}^k_a)^{-2}$	sample complexity for APE with relaxation

Linear PSI

Multi-output linear bandit model

h	feature dimension ($h \leqslant K$)
$x_a \in \mathbb{R}^h$	feature vector of arm a
$\theta \in \mathbb{R}^{h \times d}$	unknown parameter matrix (shared regression model)
$\mu_a = \theta^\top x_a \in \mathbb{R}^d$	linear mean model
$X := (x_1, \dots, x_K)^\top \in \mathbb{R}^{K \times h}$	feature matrix (all arms)
$y_t = \theta^\top x_{A_t} + \eta_t$	observation model (signal + noise) at round t
η_t	noise vector at round t (centered, subgaussian marginals)

Least-squares estimation

$X_n \in \mathbb{R}^{n \times h},\ Y_n \in \mathbb{R}^{n \times d}$	arm-selection / observation matrices after n pulls
$V_n := X_n^\top X_n = \sum_i N_{n,i}\, x_i x_i^\top$	information / Gram matrix after n samples
$V_n^\dagger$	Moore–Penrose pseudoinverse of V_n
$\hat{\theta}_n := \arg\min_A \lVert X_n A - Y_n \rVert_F^2$	least-squares estimate of θ
$\hat{\mu}_{i,r} := \hat{\theta}_r^\top x_i$	estimated mean of arm i at round r (via linear model)
$\lVert x \rVert_A^2 := x^\top A x$	(Mahalanobis) squared norm with respect to matrix A

Optimal experimental design

$w \in \triangle_K :\ V(w) := \sum_{a=1}^K w_a\, x_a x_a^\top$	design matrix for allocation w
$\lVert x \rVert_{V(w)^{-1}}^2$	statistical prediction variance of an arm under allocation w
G-optimal design	$\arg\min_{w \in \triangle_K} \max_{i \in [K]} \lVert x_i \rVert_{V(w)^{-1}}^2$ (minimizes worst-case prediction variance)
$X_S := (x_i : i \in S)^\top$	feature sub-matrix restricted to arm set S
$h_S := \dim(\operatorname{span}\{x_i : i \in S\})$	effective dimension of active arms S

Linear PSI complexity

$H_{1,\mathrm{lin}}(\nu)$	fixed-confidence complexity: $\sum_{i \leqslant h} \Delta_{(i)}^{-2}$
$H_{2,\mathrm{lin}}(\nu)$	fixed-budget complexity: $\max_{i \leqslant h}\ i\, \Delta_{(i)}^{-2}$

GEGE algorithm

$S_r := \{i \in \mathcal{A}_r : \nexists j \in \mathcal{A}_r,\ \hat{\mu}_{i,r} \prec \hat{\mu}_{j,r}\}$	empirical Pareto set of active arms at round r

Bayesian PSI

Gaussian learning model

$\nu_i = \mathcal{N}(\mu_i, \Sigma)$	Gaussian arm distribution (shared, known covariance)
$\Sigma \in \mathbb{R}^{d\times d}$	known covariance matrix
$\|x\|_{\Sigma^{-1}} := \sqrt{x^\top \Sigma^{-1} x}$	Mahalanobis norm
$U_t \sim \mathrm{Uniform}([0,1])$	exogenous randomness (for randomized sampling rules)
$\mathcal{H}_t := \sigma(U_1, A_1, Z_1, \ldots, U_t)$	augmented filtration (includes exogenous noise)

Bayesian notation

$\Pi_t := \bigotimes_k \mathcal{N}\big(\hat{\mu}_{t,k}, \Sigma/N_{t,k}\big)$	product posterior (Gaussian, flat prior)
$\widehat{\Pi}_t := \bigotimes_k \mathcal{N}\big(\hat{\mu}_{t,k}, c(t-1,\delta)\,\Sigma/N_{t,k}\big)$	inflated posterior for posterior-sampling stopping
$\theta_t^1, \theta_t^2, \ldots \overset{\text{i.i.d.}}{\sim} \widehat{\Pi}_t \vert \mathcal{H}_{t-1}$	posterior samples at round t

PS stopping rule

$c(t,\delta)$	inflation / coverage factor
$M(t,\delta)$	maximum number of posterior draws required at round t
$\tau_\delta^{\mathrm{PS}} := \inf\{t : \forall m \in [M(t-1,\delta)],\ S^\star(\theta_t^m) = \widehat{S}_t\}$	posterior-sampling stopping time
$m_{t,\delta} := \inf\{m : \theta_t^m \in \mathrm{Alt}(\widehat{S}_t)\}$	first posterior draw contradicting the current answer

Game-based sampling rule

$w_t := N_t/t$	empirical allocation at round t
$w_{\exp}$	forced-exploration allocation
$\widetilde{w}_t := (1-\gamma_t)w_t + \gamma_t\, w_{\exp}$	mixed allocation (optimal + forced exploration)

Additional notation

$d_\Sigma := d_\Sigma = \|1_d\|^2_{(\bar{\sigma}\Sigma)^{-1}}, \bar{\sigma} = \|\|\Sigma^{-1}\|\|$	covariance-dependent constant in tail bounds
$R_\Sigma(x)$	multivariate Mills ratio under Σ
$W^{-1}(x) := -W_{-1}(-e^{-x})$	Lambert W (negative branch), used in stopping analysis

Constrained PSI

Constrained PSI / feasibility

$P := \{x \in \mathbb{R}^d : Ax \leqslant b\}$	feasibility polyhedron ($A \in \mathbb{R}^{m\times d},\ b \in \mathbb{R}^m$)
m	number of linear constraints
$\mathrm{dist}(x, \mathcal{X}) := \inf_{y\in\mathcal{X}} \|x - y\|_2$	distance from point x to set $\mathcal{X}$
$\mathrm{cl}(P),\ \mathrm{int}(P),\ \partial P := \mathrm{cl}(P) \setminus \mathrm{int}(P)$	closure, interior, and boundary of P
$\mathrm{F}(\mu) := \{k : \mu_k \in P\}$	set of (truly) feasible arms
$\mathcal{S}^\star_{\mathrm{feas}}(\mu) := \mathrm{Par}(\mathrm{F}(\mu), \mu)$	feasible Pareto set: $\{i : \mu_i \in P$ and no feasible j dominates $i\}$
$\mathrm{SubOpt}(\mu) := \{i : \exists j \in \mathrm{F}(\mu),\ \mu_i \prec \mu_j\}$	arms dominated by a feasible arm
$\eta_i := \begin{cases} \mathrm{dist}(\mu_i, P^c), & \mu_i \in P, \\ \mathrm{dist}(\mu_i, \partial P), & \mu_i \notin P, \end{cases}$	feasibility margin of arm i

e-cPSI problem formulation

$\mathrm{Alt}(S) := \{\lambda : S^\star_{\mathrm{feas}}(\lambda) \neq S\}$	alternatives for constrained PSI (cPSI)
$\mathrm{Alt}(S, I) := \{\lambda : (S, I) \notin \mathcal{M}(\mathcal{P}, \lambda)\}$	alternatives for explainable cPSI (e-cPSI)
$\mathcal{M}(\mathcal{P},\mu) := \{(S,I) : S \subset \mathrm{SubOpt},\ I \subset \mathrm{F}^c,\ S\cap I = \emptyset, S\cup I = \mathcal{S}^c_{\mathrm{feas}}\}$	set of correct partition answers for e-cPSI
$T(\mu, S, I)$	e-cPSI characteristic time for answer (S, I)
$C^\star(\nu) := \min_{(S,I)} C(\mu, S, I)$	finite-time e-cPSI sample complexity coefficient
$C(\mu, S, I)$	sample complexity measure for partition (S, I)

Constrained gaps and complexity

$\Delta^\star_i(S) := \max_{j\in\mathrm{Par}(S)} m(i, j)$	gap of suboptimal arms in S
$\delta^\star_i(S) := \min\{\delta^+_i(S), \delta^-_i(S)\}$	gap of arm i relative to feasible set S
$\delta^+_i(S) := \min_{j\notin\mathrm{Par}(S)} \min\big(\mathrm{M}(i, j), \mathrm{M}(j, i)\big)$	margin-to-being-dominated (constrained to S)
$\delta^-_i(S) := \min_{j\in\mathrm{Par}(S)\setminus\{i\}} \big(\mathrm{M}(j, i)_+ + (\Delta^\star_j(S))_+\big)$	margin-to-sub-optimality (constrained)
$\Delta_i(S) := \max(\delta^\star_i(S), \Delta^\star_i(S))$	(Pareto) sub-optimality gap of arm i, restricted to S

List of Figures

List of Tables

List of Algorithms

Bibliography

Mills, J. P. (1926). "Table of the Ratio: Area to Bounding Ordinate, for Any Portion of Normal Curve". *Biometrika*.

Thompson, W. R. (1933). "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples". *Biometrika*.

Birnbaum, Z. W. (1942). "An Inequality for Mill's Ratio". *The Annals of Mathematical Statistics*.

Robbins, H. E. (1952). "Some aspects of the sequential design of experiments". *Bulletin of the American Mathematical Society* 58.

Kiefer, J. & J. Wolfowitz (1960). "The Equivalence of Two Extremum Problems". *Canadian Journal of Mathematics*.

Kung, H. T., F. Luccio & F. P. Preparata (1975). "On Finding the Maxima of a Set of Vectors". *J. ACM*.

Lai, T. L. (1976). "On Confidence Sequences". *The Annals of Statistics* 4.

Lai, T. & H. Robbins (1985). "Asymptotically efficient adaptive allocation rules". *Advances in Applied Mathematics* 6.1.

Ye, Y. & E. Tse (July 1989). "An extension of Karmarkar projective algorithm for convex quadratic programming". *Math. Program.* 44, pp. 157–179.

Jennison, C. & B. W. Turnbull (1993). "Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety endpoints." *Biometrics*.

Miettinen, K. (1998). "Nonlinear multiobjective optimization". *International Series in Operations Research and Management Science*.

Even-Dar, E., S. Mannor & Y. Mansour (2002). "PAC Bounds for Multi-armed Bandit and Markov Decision Processes". *Annual Conference Computational Learning Theory*.

Auer, P. (2003). "Using Confidence Bounds for Exploitation-Exploration Trade-Offs".

Pukelsheim, F. (2006). *Optimal Design of Experiments (Classics in Applied Mathematics) (Classics in Applied Mathematics, 50)*. Society for Industrial and Applied Mathematics.

Baricz, Á. (2008). "Mills' ratio: Monotonicity patterns and functional inequalities". *Journal of Mathematical Analysis and Applications*.

Bubeck, S., R. Munos & G. Stoltz (2009). "Pure exploration in multi-armed bandits problems". *Proceedings of the 20th International Conference on Algorithmic Learning Theory*. ALT'09. Springer-Verlag.

Lu, D. & W. V. Li (2009). "A note on multivariate Gaussian estimates". *Journal of Mathematical Analysis and Applications*.

Audibert, J.-Y. & S. Bubeck (2010). "Best Arm Identification in Multi-Armed Bandits". *COLT - 23th Conference on Learning Theory - 2010*.

Almer, O., N. Topham & B. Franke (2011a). "A learning-based approach to the automated design of MPSoC networks". *Proceedings of the 24th International Conference on Architecture of Computing Systems*. Springer-Verlag.

— (2011b). "A learning-based approach to the automated design of MPSoC networks". *Proceedings of the 24th International Conference on Architecture of Computing Systems*. ARCS'11. Springer-Verlag.

Garivier, A. & O. Cappé (2011). "The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond". *Proceedings of the 24th Annual Conference on Learning Theory*. Vol. 19. Proceedings of Machine Learning Research. PMLR.

Agrawal, S. & N. Goyal (2012). "Analysis of Thompson Sampling for the Multi-armed Bandit Problem". *Proceedings of the 25th Annual Conference on Learning Theory*. Vol. 23. Proceedings of Machine Learning Research. PMLR.

Gabillon, V., M. Ghavamzadeh & A. Lazaric (2012). "Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Kalyanakrishnan, S., A. Tewari, P. Auer & P. Stone (2012). "PAC Subset Selection in Stochastic Multi-Armed Bandits". *Proceedings of the 29th International Conference on Machine Learning*. Omnipress.

Kaufmann, E., N. Korda & R. Munos (2012). "Thompson sampling: An asymptotically optimal finite-time analysis". *International conference on algorithmic learning theory*.

Plotkin, S. A. & P. B. Gilbert (2012). "Nomenclature for Immune Correlates of Protection After Vaccination". *Clinical Infectious Diseases* 54.11.

Zuluaga, M., P. Milder & M. Püschel (2012). "Computer generation of streaming sorting networks". *DAC Design Automation Conference 2012*.

Bubeck, S., T. Wang & N. Viswanathan (2013). "Multiple Identifications in Multi-Armed Bandits". *Proceedings of the 30th International Conference on Machine Learning*. PMLR.

Drugan, M.-M. & A. Nowe (2013). "Designing multi-objective multi-armed bandits algorithms: A study". *The 2013 International Joint Conference on Neural Networks (IJCNN)*.

Karnin, Z., T. Koren & O. Somekh (2013). "Almost Optimal Exploration in Multi-Armed Bandits". *Proceedings of the 30th International Conference on International Conference on Machine Learning*. JMLR.

Kaufmann, E. & S. Kalyanakrishnan (2013). "Information Complexity in Bandit Subset Selection". *Conference On Learning Theory*. JMLR: Workshop and Conference Proceedings.

Mark C, G., D. Patrick, R. Hanno B, S. Jerzy, D. Eva, M. Vadim, A. Jacob A, L. Sang-Heon, C. Christine E, K. Herbert, I. Takashi, H. Sophie & M. Shephard (2013). "Efficacy and safety of secukinumab in patients with rheumatoid arthritis: a phase II, dose-finding, double-blind, randomised, placebo controlled study". *Annals of the rheumatic diseases*.

Zuluaga, M., G. Sergent, A. Krause & M. Püschel (2013). "Active Learning for Multi-Objective Optimization". *Proceedings of the 30th International Conference on Machine Learning*. PMLR.

Chen, S., T. Lin, I. King, M. R. Lyu & W. Chen (2014). "Combinatorial Pure Exploration of Multi-Armed Bandits". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

De Rooij, S., T. Van Erven, P. D. Grünwald & W. M. Koolen (2014). "Follow the leader if you can, hedge if you must". *J. Mach. Learn. Res.*

Jamieson, K., M. Malloy, R. Nowak & S. Bubeck (2014). "lil' UCB: An Optimal Exploration Algorithm for Multi-Armed Bandits". *Proceedings of The 27th Conference on Learning Theory*. PMLR.

Jamieson, K. & R. Nowak (2014). "Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting". *2014 48th Annual Conference on Information Sciences and Systems (CISS)*.

Soare, M., A. Lazaric & R. Munos (2014). "Best-arm identification in linear bandits". *Advances in Neural Information Processing Systems*.

Auer, P., C.-K. Chiang, R. Ortner & M.-M. Drugan (2016). "Pareto Front Identification from Stochastic Bandit Feedback". *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. PMLR.

Carpentier, A. & A. Locatelli (2016). "Tight (Lower) Bounds for the Fixed Budget Best Arm Identification Bandit Problem". *29th Annual Conference on Learning Theory*. PMLR.

Degenne, R. & V. Perchet (2016). "Combinatorial semi-bandit with known covariance". *Advances in Neural Information Processing Systems*.

Garivier, A. & E. Kaufmann (2016). "Optimal Best Arm Identification with Fixed Confidence". *29th Annual Conference on Learning Theory*. PMLR.

Kaufmann, E., O. Cappé & A. Garivier (2016). "On the Complexity of Best-Arm Identification in Multi-Armed Bandit Models". *Journal of Machine Learning Research*.

Locatelli, A., M. Gutzeit & A. Carpentier (2016). "An optimal algorithm for the Thresholding Bandit Problem". *Proceedings of The 33rd International Conference on Machine Learning*. PMLR.

Russo, D. (2016). "Simple Bayesian Algorithms for Best Arm Identification". *29th Annual Conference on Learning Theory*. Vol. 49. Proceedings of Machine Learning Research. PMLR.

Zuluaga, M., A. Krause & M. Püschel (2016). "e-PAL: An Active Learning Approach to the Multi-Objective Optimization Problem". *Journal of Machine Learning Research*.

Allen-Zhu, Z., Y. Li, A. Singh & Y. Wang (2017). "Near-optimal design of experiments via regret minimization". *Proceedings of the 34th International Conference on Machine Learning*. JMLR.org.

Simchowitz, M., K. Jamieson & B. Recht (2017). "The Simulator: Understanding Adaptive Sampling in the Moderate-Confidence Regime". *Proceedings of the 2017 Conference on Learning Theory*. PMLR.

Howard, S. R., A. Ramdas, J. D. McAuliffe & J. S. Sekhon (2018). "Time-uniform, nonparametric, nonasymptotic confidence sequences". *The Annals of Statistics*.

Katz-Samuels, J. & C. Scott (2018). "Feasible Arm Identification". *Proceedings of the 35th International Conference on Machine Learning*. PMLR.

Tao, C., S. Blanco & Y. Zhou (2018). "Best Arm Identification in Linear Bandits with Linear Dimension Dependency". *Proceedings of the 35th International Conference on Machine Learning*. PMLR.

Xu, L., J. Honda & M. Sugiyama (2018). "A fully adaptive algorithm for pure exploration in linear bandits". *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*. PMLR.

Afshari, H., W. Hare & S. Tesfamariam (2019). "Constrained multi-objective optimization algorithms: Review and comparison with application in reinforced concrete structures". *Applied Soft Computing*.

Amani, S., M. Alizadeh & C. Thrampoulidis (2019). "Linear Stochastic Bandits Under Safety Constraints". *Advances in Neural Information Processing Systems*.

Degenne, R. & W. Koolen (2019). "Pure Exploration with Multiple Correct Answers". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Degenne, R. (2019). "Impact of structure on the design and analysis of bandit algorithms". PhD thesis.

Degenne, R., W. M. Koolen & P. Ménard (2019). "Non-asymptotic pure exploration by solving games". *Advances in Neural Information Processing Systems*.

Fiez, T., L. Jain, K. G. Jamieson & L. Ratliff (2019). "Sequential Experimental Design for Transductive Linear Bandits". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Garivier, A., P. Ménard & G. Stoltz (2019). "Explore first, exploit next: The true shape of regret in bandit problems". *Mathematics of Operations Research*.

Katz-Samuels, J. & C. Scott (2019). "Top Feasible Arm Identification". *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. PMLR.

Lu, S., G. Wang, Y. Hu & L. Zhang (2019). "Multi-Objective Generalized Linear Bandits". *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press.

Ménard, P. (2019). *Gradient Ascent for Active Exploration in Bandit Problems*.

Roy Chaudhuri, A. & S. Kalyanakrishnan (2019). "PAC Identification of Many Good Arms in Stochastic Multi-Armed Bandits". *Proceedings of the 36th International Conference on Machine Learning*. PMLR.

Daulton, S., M. Balandat & E. Bakshy (2020). "Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization". *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Curran Associates Inc.

Degenne, R., P. Ménard, X. Shang & M. Valko (2020). "Gamification of pure exploration for linear bandits". *International Conference on Machine Learning*. PMLR.

Jedra, Y. & A. Proutiere (2020a). "Optimal Best-arm Identification in Linear Bandits". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

— (2020b). "Optimal Best-arm Identification in Linear Bandits". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Katz-Samuels, J., L. Jain, z. karnin zohar & K. G. Jamieson (2020). "An Empirical Process Approach to the Union Bound: Practical Algorithms for Combinatorial and Linear Bandits". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Lattimore, T. & C. Szepesvari (2020). *Bandit Algorithms*. Cambridge University Press.

Mehrotra, R., N. Xue & M. Lalmas (2020). "Bandit Based Optimization of Multiple Objectives on a Music Streaming Platform". *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery.

Russo, D., B. V. Roy, A. Kazerouni, I. Osband & Z. Wen (2020). *A Tutorial on Thompson Sampling*.

Shang, X., R. de Heide, P. Menard, E. Kaufmann & M. Valko (2020). "Fixed-confidence guarantees for Bayesian best-arm identification". *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR.

Alieva, A., A. Cutkosky & A. Das (2021). "Robust Pure Exploration in Linear Bandits with Limited Budget". *Proceedings of the 38th International Conference on Machine Learning*. PMLR.

Camilleri, R., K. Jamieson & J. Katz-Samuels (2021). "High-dimensional Experimental Design and Kernel Bandits". *Proceedings of the 38th International Conference on Machine Learning*. PMLR.

Garivier, A. & E. Kaufmann (2021). *Non-Asymptotic Sequential Tests for Overlapping Hypotheses and application to near optimal arm identification in bandit models*.

Kaufmann, E. & W.-M. Koolen (2021). "Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals". *Journal of Machine Learning Research*.

Kumar, A., G. Wu, M. Z. Ali, Q. Luo, R. Mallipeddi, P. N. Suganthan & S. Das (2021). "A Benchmark-Suite of real-World constrained multi-objective optimization problems and some baseline results". *Swarm and Evolutionary Computation*.

Munro, A.-P.-S., L. Janani, V. Cornelius & et al. (2021). "Safety and immunogenicity of seven COVID-19 vaccines as a third dose (booster) following two doses of ChAdOx1 nCov-19 or BNT162b2 in the UK (COV-BOOST): a blinded, multicentre, randomised, controlled, phase 2 trial". *The Lancet*.

Wang, P.-A., R.-C. Tzeng & A. Proutiere (2021). "Fast Pure Exploration via Frank-Wolfe". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Azizi, M., B. Kveton & M. Ghavamzadeh (2022). "Fixed-Budget Best-Arm Identification in Structured Bandits". *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization.

Camilleri, R., A. Wagenmaker, J. Morgenstern, L. Jain & K. Jamieson (2022). "Active learning with safety constraints". *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS '22. Curran Associates Inc.

Faizal, F. Z. & J. Nair (2022). *Constrained Pure Exploration Multi-Armed Bandits with a Fixed Budget*.

Jourdan, M., R. Degenne, D. Baudry, R. de Heide & E. Kaufmann (2022). "Top Two Algorithms Revisited". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Karpov, N. & Q. Zhang (2022). "Collaborative Best Arm Identification with Limited Communication on Non-IID Data". *arXiv preprint, eprint:2207.08015*.

Ramdas, A., J. Ruf, M. Larsson & W. Koolen (2022). *Admissible anytime-valid sequential inference must rely on nonnegative martingales*.

Wagenmaker, A. J., M. Simchowitz & K. Jamieson (2022). "Beyond No Regret: Instance-Dependent PAC Reinforcement Learning". *Proceedings of Thirty Fifth Conference on Learning Theory*. Vol. 178. Proceedings of Machine Learning Research. PMLR.

Wang, S. & J. Zhu (2022). "Thompson sampling for (combinatorial) pure exploration". *International Conference on Machine Learning*. PMLR.

Yang, J. & V. Tan (2022). "Minimax Optimal Fixed-Budget Best Arm Identification in Linear Bandits". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Ararat, C. & C. Tekin (2023). "Vector Optimization with Stochastic Bandit Feedback". *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*. PMLR.

Degenne, R. (2023). "On the Existence of a Complexity in Fixed Budget Bandit Identification". *Proceedings of Thirty Sixth Conference on Learning Theory*. PMLR.

Emmenegger, N., M. Mutny & A. Krause (2023). "Likelihood Ratio Confidence Sets for Sequential Decision Making". *Thirty-seventh Conference on Neural Information Processing Systems*.

Jourdan, M., R. Degenne & E. Kaufmann (2023). "An ε-Best-Arm Identification Algorithm for Fixed-Confidence and Beyond". *Thirty-Seventh Conference on Neural Information Processing Systems*.

Kim, W., G. Iyengar & A. Zeevi (2023). *Pareto Front Identification with Regret Minimization*.

Kone, C., E. Kaufmann & L. Richert (2023). "Adaptive Algorithms for Relaxed Pareto Set Identification". *Thirty-seventh Conference on Neural Information Processing Systems*.

You, W., C. Qin, Z. Wang & S. Yang (2023). "Information-Directed Selection for Top-Two Algorithms". *Proceedings of Thirty Sixth Conference on Learning Theory*. Vol. 195. Proceedings of Machine Learning Research. PMLR.

Crepon, É., A. Garivier & W. M Koolen (2024). "Sequential learning of the Pareto front for multi-objective bandits". *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR.

Kanarios, K., Q. Zhang & L. Ying (2024). "Cost Aware Best Arm Identification". *Reinforcement Learning Conference*.

Karagözlü, E. M., Y. C. Yıldırım, C. Ararat & C. Tekin (2024). "Learning the Pareto Set Under Incomplete Preferences: Pure Exploration in Vector Bandits". *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*.

Kone, C., E. Kaufmann & L. Richert (2024). "Bandit Pareto Set Identification: the Fixed Budget Setting". *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR.

Li, D., F. Zhang, C. Liu & Y. Chen (2024). *Constrained Multi-objective Bayesian Optimization through Optimistic Constraints Estimation*.

Li, Z., K. Jamieson & L. Jain (2024). "Optimal Exploration is no harder than Thompson Sampling". *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. Proceedings of Machine Learning Research. PMLR.

Poiani, R., R. Degenne, E. Kaufmann, A. M. Metelli & M. Restelli (2024). "Optimal Multi-Fidelity Best-Arm Identification". *Advances in Neural Information Processing Systems*. Curran Associates, Inc.

Shukla, A. & D. Basu (2024). "Preference-based Pure Exploration". *Advances in Neural Information Processing Systems*. Vol. 37. Curran Associates, Inc.

Das, U., A. Shukla & D. Basu (2025). *FraPPE: Fast and Efficient Preference-based Pure Exploration*.

Kejriwal, K., N. Karamchandani & J. Nair (2025). "On the Asymptotic Optimality of Confidence Interval Based Algorithms for Fixed Confidence MABs". *Proceedings of the AAAI Conference on Artificial Intelligence* 39.17.

Kirschner, J., A. Krause, M. Meziu & M. Mutny (2025). *Confidence Estimation via Sequential Likelihood Mixing*.

Kone, C., M. Jourdan & E. Kaufmann (2025). "Pareto Set Identification With Posterior Sampling". *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*. Vol. 258. Proceedings of Machine Learning Research. PMLR.

Kone, C., E. Kaufmann & L. Richert (2025a). "Bandit Pareto Set Identification in a Multi-Output Linear Model". *The 28th International Conference on Artificial Intelligence and Statistics*.

— (2025b). "Constrained Pareto Set Identification with Bandit Feedback". *Forty-second International Conference on Machine Learning*.