Bandit Optimal Transport

Lorenzo Croissant
CREST, ENSAE, & INRIA FairPlay Team
Palaiseau, France
lorenzo.croissant@ensae.fr

Abstract

Despite the impressive progress in statistical Optimal Transport (OT) in recent years, there has been little interest in the study of the *sequential learning* of OT. Surprisingly so, as this problem is both practically motivated and a challenging extension of existing settings such as linear bandits. This article considers (for the first time) the stochastic bandit problem of learning to solve generic Kantorovich OT problems from repeated interactions when the marginals are known but the cost is unknown. By exploiting the intrinsic regularity of the OT problem, we show that this problem satisfies classical Hilbert space bandit regret guarantees ($\tilde{\mathcal{O}}(\sqrt{T})$ multiplied by log-determinant terms) for both problems. To deal with learning in infinite dimension, we provide a functional regression method which can exploit intrinsic regularity of the cost to obtain complete regret bounds interpolating between $\tilde{\mathcal{O}}(\sqrt{T})$ (finite and parametric cases) and $\mathcal{O}(T)$ (unlearnable costs).

1 Introduction

Originally, Optimal Transport (OT) was developed as a mathematical theory to optimise the transportation and logistics of goods (Monge, 1781; Kantorovich, 2006). Over the last two decades, however, this theory has experienced a meteoric rise in applied mathematics due to a sustained series of major breakthroughs (Villani, 2003, 2009). The optimisation problem aims to find the most efficient allocation (given a cost function) of ressources from sources (supply) to sinks (demand).[1]

Historically, this economic interpretation has been the main application for the theory, see e.g. (Galichon, 2021; Kreinovich et al., 2024), but the recent theoretical progress has renewed interest for new domains of applications, such as machine learning. Indeed, many have noticed that the ability to quantify and minimise "distances" between probability measures parallels key questions in problems such as generative modelling (Arjovsky et al., 2017) or domain adaptation (Courty et al., 2017). As these developments have matured, they have percolated into statistical learning theory to create the rich literature of statistical optimal transport, recently surveyed by Chewi et al. (2024).

In spite of this ongoing activity, the question of *sequential* learning remains a blind spot of this emerging field. Barring a handful of exceptions, all existing works consider static (*batch*) i.i.d. datasets and traditional statistical estimation. This is despite the many applications of optimal transport that are naturally sequential. Classical examples include assignment of kidney donors to recipients (Glorie et al., 2014), doctors to hospitals (Hatfield and Milgrom, 2005), etc. In these examples the number of assignments is finite, but, as it becomes large, the OT problem is best modelled by an infinite-dimensional problem, see e.g. Cao et al. (2024); Carlier (2010). Optimal transport finds countless other naturally sequential applications across economics and operations research, which motivate the study of sequential learning of OT.

[1] From a formal standpoint, this problem concerns the minimisations of functionals of measures under constraints imposed by their marginals.

18th European Workshop on Reinforcement Learning (EWRL 2025).

In sequential learning tasks, samples are highly correlated which introduces significant new complexities relative to the batch setting. Moreover, sequential learning tasks are more naturally evaluated during the learning process, rather than at the end. This *online* evaluation creates a trade-off between *exploration* (statistical efficiency) and *exploitation* (online performance).

Consequently, this paper sets out to investigate the question of the online learnability of the general OT problem in a stochastic partial feedback setting known as a *stochastic bandit*.

In this setting (see Section 2 for details), an agent is given the constraints of an optimal transport problem, but not the cost function. It must partake in a repeated game in which it submits a transport plan (i.e. an admissible point) at each round, and receives a noisy reward estimate of the cost of the submitted plan. Importantly, this feedback is *bandit*: it gives no information about the outcome of any plan other than the one played. We measure the performance of the agent by its *regret*, i.e. its cumulative loss compared to the optimal plan.

This setting raises intriguing connections to classical work in bandit problems. First, since optimal transport functionals are linear functionals, this problem appears an extension of linear bandits (Auer, 2003). Closer inspection however reveals that classical tools break down because the cost function which must be learned does not live in the same space as the actions. Second, the infinite-dimensionality of the cost function draws a connection to kernel bandits (Valko et al., 2013). In kernel bandits, the regularity of the hypothesis space is what allows transformation to a linear problem. In contrast, we will see that the regularity of the OT problem is intrinsic to its geometry and we can thus work with much larger hypothesis spaces despite this problem not being a linear bandit.

Contributions. As a result of our investigation, we establish the first regret bounds for learning the general stochastic bandit OT problem. We construct a modified optimistic algorithm which exploits the intrinsic regularity of the OT problem to construct a coherent action sequence which maintains valid confidence sets in the style of (Abbasi-Yadkori, 2012). This algorithm incurs a regret of $\tilde{\mathcal{O}}(\sqrt{T})$ up to learning-dependent log-determinant terms, which is the same order as the regret of linear bandits (Auer, 2003). This isolates the statistical sub-problem of estimating the cost function, for which we propose a functional regression method which is adaptative to the regularity of the cost function, obtaining a regret that interpolates between $\tilde{\mathcal{O}}(\sqrt{T})$ for discrete or parametric problems and $\mathcal{O}(T)$ for unlearnable instances directly from regularity conditions on the cost.

Organisation. We devote Section 2 to clearly defining the technically intricate Bandit Optimal Transport (BOT) setting. Then, in Section 3, we discuss high-level insights, related work regarding learning of OT problems, and detail our contributions. Thereafter, we focus Section 4 on the technicalities of our solution to the BOT problem, detailing algorithms and regret bounds. We conclude by touching on some promising open directions in Section 5. Appendices extend these discussion and contain rigourous details of technical contributions and proofs.

2 Setting

2.1 The decision problem of optimal transport

Consider a pair of probability measures $(\mu, \nu) \in \mathscr{P}(\mathcal{M}_\mu \times \mathcal{M}_\nu)$ on two topological measurable spaces $(\mathcal{M}_\mu, \mathcal{F}_\mu)$ and $(\mathcal{M}_\nu, \mathcal{F}_\nu)$, as well as a cost function $c : \mathcal{M}_\mu \times \mathcal{M}_\nu \to \mathbb{R}$. For ease of exposition, we consider $\mathcal{X} := \mathcal{M}_\mu \times \mathcal{M}_\nu \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, but the problems below are also defined on highly esoteric spaces $(\mathcal{M}_\mu, \mathcal{M}_\nu)$ such as a graph or a space of curves.

The Kantorovich formulation of the OT problem (Kantorovich, 2006) asks for the optimal way to transport all the mass from μ to ν, where the cost of moving an infinitesimal unit of mass from $x \in \mathcal{M}_\mu$ to $y \in \mathcal{M}_\nu$ is captured by $c(x, y)$. If $c(x, y) = \|x - y\|$, the cost of transporting x to y is just the distance between the source and the destination. However, the ability to roll arbitrarily complex considerations into c is what makes OT highly versatile in applications.

Formally, the *Kantorovich* (optimal transport) problem is defined as

$$\text{Kant.}(\mu, \nu, c) := \inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y) \mathrm{d}\pi(x, y) \tag{1}$$

in which $\Pi(\mu, \nu) := \{\pi \in \mathscr{P}(\mathcal{M}_\mu \times \mathcal{M}_\nu) : \pi(\cdot, \mathcal{M}_\nu) = \mu, \pi(\mathcal{M}_\mu, \cdot) = \nu\}$ is the set of all *couplings* of μ and ν, i.e. any joint distribution whose marginals over $\mathcal{M}_\mu$ and $\mathcal{M}_\nu$ are μ and ν, respectively. Importantly, the Kantorovich problem allows mass located at x to be split and sent to several y, and vice-versa, but a set $S \in \mathcal{F}_\mu$ may not give more mass that $\mu(S)$, just as $S' \in \mathcal{F}_\nu$ may only receive $\nu(S')$ mass. In fact Π imposes that they give and receive *exactly* this amount of mass.

Divisibility of mass was absent in the original formulation of OT, which rendered the problem highly difficult (see Appendix G.2). In contrast, the Kantorovich problem is a linear program and is solvable when c is lower semi-continuous and bounded below, see (Villani, 2009, Thm. 4.1). The generality of this result[2] explains its adoption as the core problem of OT theory.

The *linearity* of the optimal transport functional functional refers to the fact that the map $\pi \mapsto \int c(x,y)\mathrm{d}\pi(x,y)$ is linear in π. In fact, this functional is a bilinear form which can be represented as a duality pairing $\langle c|\pi\rangle$ (see Section 4.1) so that (1) can be rewritten as

$$\text{Kant.}(\mu,\nu,c) = \inf_{\pi\in\Pi(\mu,\nu)} \langle c|\pi\rangle\,. \tag{2}$$

This pairing is not an inner product however, as c is a function while π is a measure.

Nevertheless, linearity speaks in favour of the regularity of Equation (1). Intuitively, it behaves like an infinite-dimensional linear program. Indeed, $\langle c|\cdot\rangle$ is linear and $\Pi(\mu,\nu)$ is defined by linear (integral) constraints, and, in fact, is convex and compact (Ambrosio et al., 2021, Cor. 2.9). However, unlike in finite-dimensional linear programs, the optimisation domain $\Pi(\mu,\nu)$ is neither a vector space nor flat. This is the source of significant technical difficulties in the resolution of (1), which appears a difficult roadblock to the application of standard learning methods.

2.2 The learning problem

Formally, we consider the following learning game: at each round $t \in \mathbb{N} := \{1,2,\ldots\}$, the agent submits a transport plan $\pi_t \in \Pi(\mu,\nu)$, and receives a noisy cost feedback

$$C_t := \int c^*(x,y)\mathrm{d}\pi_t(x,y) + \xi_t\,,$$

in which $(\xi_t)_{t\in\mathbb{N}}$ is a sequence of random variables and c^* is the *unknown* true cost function. Henceforth, we work on a suitable probability space filtered by the natural filtration of $(\xi_t)_{t\in\mathbb{N}}$. In our formulation, we consider that $(\mathcal{M}_\mu, \mathcal{M}_\nu, \mu, \nu)$ are known ahead of time in order to ensure the constraints are respected.

In order to assess its performance, we assimilate the algorithm of any learning agent to its action sequence $\boldsymbol{\pi} := (\pi_t)_{t\in\mathbb{N}}$. We evaluate the quality of $\boldsymbol{\pi}$ *online* (i.e. during the learning rather than at the end) using the classical tool of regret

$$\mathscr{R}_T(\boldsymbol{\pi}) := \sum_{t=1}^{T} C_t - \text{Kant.}(\mu,\nu,c^*) \qquad \text{for} \qquad T\in\mathbb{N}\,. \tag{3}$$

Low (sub-linear) regret requires performance during learning, which is not the case in classical learning settings. This is due to the appearance of an *exploration-exploitation* trade-off, as the agent must balance between exploring to learn c^* and exploiting its current knowledge to minimise its cost.

Note that regret is a decision-theoretic criterion: it measures the quality of the decision π_t in terms of the OT problem, not the quality of any estimation of c^*. Achieving low regret thus requires only learning the structure of c^* that is relevant to finding its minimum over $\Pi(\mu,\nu)$. The structure of the OT problem itself can thus facilitate learning even with minimal assumptions on c^*.

Our goal is to design a learning algorithm $\boldsymbol{\pi}$ which achieves the slowest regret growth (as a function of T) as possible, in a high-probability sense. This is the standard approach in stochastic bandit problems, with any sub-linear in T regret growth implying convergence to the optimal value of the problem, and a regret of $\tilde{\mathcal{O}}(\sqrt{T})$ (i.e. growing slower than $\sqrt{T}\mathrm{polylog}(T)$) being the standard parametric rate, see e.g. (Lattimore and Szepesvári, 2020, Thm. 9.1, Thm. 19.2, Thm. 38.6).

[2] One might notice that the provided reference in fact uses even weaker conditions.

3 Challenges, related work, and contributions

Giving an exhaustive account of the vast literature of Optimal Transport would be outside the scope of this article. As this article focuses on aspects of online learning, we will limit our attention to this segment. Nevertheless, we provide the curious reader a modest bibliography in Appendix H.

3.1 Online learning and optimal transport

Online learning to transport The first paper to take interest in online learning of OT appears to have been (Guo et al., 2022). In this article, the authors take an Online Convex Optimisation (OCO) approach to the problem, meaning that an adversary chooses a cost function c_t at each round t from a class of suitably regular (convex) functions. The learner aims to choose a sequence of transport plans π_t which has a small regret with respect to the best fixed transport plan in hindsight.

While this work pioneered the study of online (repeated) optimal transport, there are no direct reductions between this paper and their work. Indeed, most of the work of Guo et al. (2022) is done under a full-information adversarial setting (as is typical in OCO): the transport problem changes at each round and is completely revealed after a coupling π_t is played. However, in Section 3, the authors provide a 0-order semi-bandit scheme based on a discretisation of $\mathcal{X}$. In contrast, our work is directed at a stochastic setting under complete bandit feedback (only C_t is observed).

Due to the use of OCO techniques, as well as PDE-based optimal transport tools based on the work of Brenier (1989), the results of Guo et al. (2022) are only valid under strong assumptions on the regularity of the cost functional (and thus the cost function) and the marginals. In contrast, we work without specific assumptions on the cost function and marginals, beyond the minimal ones for (1) to be well-defined. This difference arises because they consider general functionals on the Wasserstein space, while our work focuses on the specific regularity of OT functionals.

This work was followed by Zhu and Ryzhov (2023) which considers the first online learning problem in semi-discrete optimal transport (i.e. μ discrete, ν continuous). They construct a semi-myopic algorithm with forced exploration which can learn to behave as the optimal plan from samples of the continuous marginal. Unfortunately, they do not study a general problem but rather only the case in which the cost c^* is a linear parametric model. This choice obfuscates a large part of the complexity of the general problem and dilutes any insights about the geometry of the problem. Moreover, Zhu and Ryzhov (2023) do not provide direct regret bounds, but rather performance metrics which may be converted into regret bounds. Sadly, these metrics fail to generalise to the continuous marginal case, and their analysis breaks down in the general setting.

3.2 On bandit algorithms

As Section 3.1 shows, bandits and optimal transport have been in peripheral contact for some time. Nevertheless, despite its interest in many optimisation problems, the bandit literature has remained uninterested in the general optimal transport problem. Still, let us highlight the key elements of this theory on which we can build to solve the BOT problem.

Multi-armed bandits. The classical bandit problem (Thompson, 1933; Lai and Robbins, 1985; Auer et al., 2002) considered the issue of choosing the best amongst a finite set of arms based on bandit feedback about arm rewards. Since then, bandit theorists have taken some interest in higher-dimensional optimisation problems either linear or non-linear. For instance, Tran-Thanh and Yu (2014) show regret bounds for learning a general functional using bandit feedback, but sadly still considers only finitely many arms. While a general theory of bandits for functionals remains elusive (Wang et al., 2022), bandits under weaker assumptions on the set of arms have been studied.

Lipschitz bandits. Several papers (Bubeck et al., 2011b; Magureanu et al., 2014; Kleinberg et al., 2019) have leveraged Lipschitz reward functions to provide regret bounds and algorithms, even on arbitrary metric spaces. Unfortunately, the bounds for general Lipschitz functions using these methodology are of the order of $\Theta(T^{(d+1)/(d+2)})$, in dimension $d \in \mathbb{N}$ (Kleinberg et al., 2019). In the case of the continuous optimal transport problem, this dimension is infinite, and the regret bounds become vacuous. The infinite dimensional nature of our problem also prevents the practical usability of most discretisations, even sophisticated ones like the tree-based scheme of Bubeck et al. (2011a).

Linear bandits. In the hope of circumventing this problem, we can take inspiration from Kantorovich and recall that (1) is linear program. Indeed, linear functions have much stronger global regularity than Lipschitz ones, meaning that linear bandits may escape vacuity even when $d = +\infty$.

The setting of linear bandits was introduced by Auer (2003), and refined by many subsequent works (Abeille and Lazaric, 2017; Vernade et al., 2020; Hao et al., 2020), most notably for us Abbasi-yadkori et al. (2011). In his doctoral thesis, Y. Abbasi-Yadkori (2012) includes a version of this article in which the technical results are given not just for $\mathbb{R}^d$, but for an arbitrary Hilbert space. These works all use the celebrated Optimism in the Face of Uncertainty (OFU) principle to tackle the previously mentioned exploration-exploitation dilemma.

Nevertheless, in spite of its generality, Abbasi-Yadkori (2012) is not sufficient to solve the bandit optimal transport problem, because the action space of our bandit is not a Hilbert space, and in fact the actions do not live in the same space as c^*. This fundamentally breaks the assumptions of this work, in spite of the fact that the duality product $\langle c|\cdot\rangle$ defining Kant. is a linear form.

Kernel bandits. Kernel methods intrinsically consider infinite-dimensional linear rewards, and may appear, at first, an ideal solution for solving bandit optimal transport. Kernel bandits have seen extensive work (Chowdhury and Gopalan, 2017; Janz et al., 2020; Takemori and Sato, 2021), including Valko et al. (2013) which comes closest to our approach by introducing a kernelised OFU algorithm. These methods posit a particular structure for the reward function c^*, and then use the representer theorem to reduce the problem to a linear problem. Our problem, in contrast, is already linear so it should not require any such assumptions.

One place where kernel methods shine is in making infinite-dimensional problems computationally tractable. While they can be used for this purpose in our setting, we will show that we can obtain similar bounds directly from the regularity of c^* without assuming an RKHS structure.

3.3 Challenges and contributions

Challenges. There are three main challenges to the general BOT problem.

A) The actions of this bandit problem are probability measures. In the discrete optimal transport (matching) problems previously studied in the literature, probability measures remain finite dimensional and can be represented using an inner product. This hides the true complexity of the general case in which one must confront a continuum of infinite-dimensional actions which require sophisticated tools to analyse. Moreover, this is compounded by the fact that the space of probability measures has a difficult geometry.

B) The cost function c^*, which plays the role of a "parameter" to estimate, is a continuous function. Since the optimal transport problem only requires minimal integrability assumptions on c^*, the natural hypothesis classes for c^* will be large function spaces[3] such as L^2. This creates a significant difficulty for estimation and thus for bandit algorithms. The construction of estimators and confidence sets that permit the use of OFU algorithms is challenged by the infinite dimensionality of c^*.

C) Even if estimators for c^* can be constructed, they must face the infinite-dimensionality of c^*. This raises the challenge of efficient approximation of infinite-dimensional estimators under weak assumptions, and of their associated regrets.

Contributions. This paper is the first study of the general stochastic bandit optimal transport problem. It provides a general framework for further work in this area, by showing that the problem is learnable under weak assumptions. Beyond this, the technical contributions can be summarised as follows.

1) To overcome challenge **A**, we construct a phase-space representation of the optimal transport problem which allows us to transform the problem into a linear bandit on a Hilbert space. By regularising optimism by entropy, and using the dual form of the resulting entropic OT problem (see (6)), we are able to ensure our algorithm maintains the validity of the phase-space representation as it learns.

[3] Circumventing this difficulty by parametrising c^* as in Zhu and Ryzhov (2023) would dilute any insight about the geometry of the problem.

2) Combining **1** with the framework of Abbasi-Yadkori (2012) we are able to construct the necessary confidence sets and least-squares estimators to estimate c^* and address challenge **B**. In the resulting regret analysis, we leverage the regularity of the entropic problem to prove bounds on the regret.

3) To face the infinite-dimensional quantities which arise in the learning terms of the regret bounds, we construct a general non-parametric estimation method based on the regularity of the cost function. This method addresses challenge **C** by allowing us to obtain regret bounds of order $\tilde{\mathcal{O}}(\sqrt{T})$ in simple cases, and an interpolation up to $\tilde{\mathcal{O}}(T)$ dependent on the regularity of c^*.

4 Solving the BOT problem

In this section, we detail the core elements of our contributions in a solution to the Bandit Optimal Transport (BOT) problem. In Section 4.1, we detail contribution **1** of Section 3, i.e. the construction of a self-coherent procedure within an optimistic algorithm which reduces the BOT problem to a linear bandit problem in a Hilbert space. In Section 4.2, we detail contribution **2** by constructing confidence sets in the phase space of the problem, in the style of (Abbasi-Yadkori, 2012). We give a complete algorithm in Section 4.3 which combines these two contributions, and finally in Section 4.4 we focus on the estimation aspect of the BOT problem, i.e. contribution **3** of Section 3, by providing a functional regression method which can exploit the intrinsic regularity of the cost function to obtain regret bounds interpolating between $\tilde{\mathcal{O}}(\sqrt{T})$ and $\mathcal{O}(T)$.

4.1 A self-coherent reduction procedure to a Hilbert-space bandit

The technical issue in challenge **A** is that the bilinear form $\langle\cdot|\cdot\rangle$ of (2) is not an inner product. This prevents us from applying standard linear bandit tools. Recall that, formally, $\langle\cdot|\cdot\rangle$ is the duality pairing between continuous functions vanishing at infinity and finite measures (see Appendix A for notations). To reconcile these two types of objects, we leverage Fourier analysis and represent them both in phase-space. To ensure the Fourier transform and the transport problems are well defined, let us assume Assumption 1, in which $L^p(\mathbb{R}^d;\varrho)$ for $p \geq 1$ is the Lebesgue space associated with a reference measure $\varrho \in \mathscr{P}(\mathbb{R}^d)$. Note that $L^2(\mathbb{R}^d;\varrho)$ is a Hilbert space.

Assumption 1. The true cost function c^* is continuous and belongs to $L^2(\mathbb{R}^d;\varrho)$.

Let F denote the Fourier transform operator that acts on $L^2(\mathbb{R}^d;\varrho)$ or on measures, using the formulæ

$$\mathsf{F} : \phi \in L^2(\mathbb{R}^d;\varrho) \mapsto \int \phi(x) e^{-2\pi i \langle x|\cdot\rangle_2} \mathrm{d}\varrho(x) \text{ and } \mathsf{F} : \gamma \in \mathscr{P}(\mathbb{R}^d) \mapsto \int e^{-2\pi i \langle x|\cdot\rangle_2} \mathrm{d}\gamma(x)\,.$$

when these are well defined (see Appendix B for rigourous constructions of this section). In particular, since a coupling π is a finite measure, $\mathsf{F}\pi$ is (ϱ-a.e.) bounded and thus $L^2(\mathbb{R}^d;\varrho)$. The operator F is an isometry on $L^2(\mathbb{R}^d;\varrho)$, and thus we can write

$$\langle c^*|\pi\rangle = \langle \mathsf{F}c^*(-\cdot)|\mathsf{F}\pi\rangle_{L^2(\mathbb{R}^d;\varrho)} := \int \mathsf{F}c^*(-z)\overline{\mathsf{F}\pi}(z)\mathrm{d}\varrho(z)\,. \tag{4}$$

The representation in (4) thus expresses the Kantorovich OT problem as an inner product in the Hilbert space $L^2(\mathbb{R}^d;\varrho)$, which allows us to use the standard linear bandit tools. The subtle detail lies in the fact that this representation of $\langle c^*|\pi\rangle$ only holds if $\mathsf{F}\pi \in L^2(\mathbb{R}^d;\varrho)$, which is equivalent to requiring that π have a density $\mathrm{d}\pi/\mathrm{d}\varrho$ with respect to ϱ which is itself in $L^2(\mathbb{R}^d;\varrho)$ (see Lemma B.3).

The challenge of a "coherent" representation is to force the algorithm to only ever play plans $\pi_t \in \Pi(\mu,\nu)$ such that $\mathsf{F}\pi_t \in L^2(\mathbb{R}^d;\varrho)$, which is highly non-trivial. The regularity of the OT problem can be exploited here too, as we will see next.

The entropic OT problem is the regularisation of the linear problem of Kantorovich by relative entropy (a.k.a. Kullback-Leibler divergence) of π with respect to a reference measure $\varrho \in \mathscr{P}(\mathcal{X})$,

$$\mathscr{H}(\pi|\varrho) := \begin{cases} \displaystyle\int \log \frac{\mathrm{d}\pi}{\mathrm{d}\varrho}\mathrm{d}\pi \text{ if } \pi \ll \varrho \\ +\infty \text{ if } \pi \not\ll \varrho \end{cases}\,, \tag{5}$$

in which $\pi \ll \varrho$ means that π is absolutely continuous with respect to ϱ, which is sufficient for the density $\mathrm{d}\pi/\mathrm{d}\varrho$ to exist (by Radon-Nikodym). This functional is strictly convex on $\mathscr{P}(\mathcal{X})$, and the *entropic* optimal transport problem is then formally defined as

$$\text{Ent.}(\mu, \nu, c, \varepsilon) := \inf_{\pi \in \Pi(\mu,\nu)} \langle c|\pi\rangle + \varepsilon \mathscr{H}(\pi|\varrho) \,. \tag{6}$$

In (6), relative entropy penalises concentration of measure on sets to which ϱ assigns low mass, which one can interpret as forcing the transport to be more spread out on the support of ϱ. For example, if $\varrho = \mu \otimes \nu$, the independent coupling[4] of μ and ν, then the entropic regularisation forbids the mass from any x from being sent wholly to a single y (and vice-versa).

An admissible point of (6) has an $L^\infty(\mathbb{R}^d; \varrho) \subset L^2(\mathbb{R}^d; \varrho)$ density. The problem can be solved using its dual formulation of the entropic problem

$$\text{Ent.}(\mu, \nu, c, \varepsilon) = \sup_{(\varphi,\psi)\in\Xi} \left\{ \int \varphi \mathrm{d}\mu + \int \psi \mathrm{d}\nu - \varepsilon \int e^{\varepsilon^{-1}(\varphi+\psi-c)} \mathrm{d}(\varrho) + \varepsilon \right\} , \tag{7}$$

wherein $\Xi := \{(\varphi, \psi) \in L^1(\mathbb{R}^d; \mu), L^1(\mathbb{R}^d \mu) : \varphi \oplus \psi \leq c\}$ with $\varphi \oplus \psi : (x, y) \mapsto \varphi(x) + \psi(y)$. From a dual solution $(\varphi^*, \psi^*) \in \Xi$, one may recover (see e.g. Nutz (2022, Thm. 4.2)) a primal solution π^* with density

$$\frac{\mathrm{d}\pi^*}{\mathrm{d}\varrho} = e^{\frac{\varphi^* \oplus \psi^* - c}{\varepsilon}} < +\infty \,.$$

By (7), a solution to the entropic problem is a transport plan (an action) with an $L^\infty(\mathbb{R}^d; \varrho) \subset L^2(\mathbb{R}^d; \varrho)$ density. Consequently, an optimistic algorithm which chooses a belief-action pair

$$(\tilde{\pi}_t, \tilde{c}_t) \in \operatorname{argmin} \left\{ \langle c|\pi\rangle + \varepsilon_t \mathscr{H}(\pi|\varrho) : \pi \in \Pi(\mu, \nu) \,, \, c \in \mathsf{F}^{-1}\mathcal{C}_t(\delta) \right\} \,, \tag{8}$$

guarantees the coherence of the construction of Section 4.1. Moreover, for ε_t small, the solution of (8) is close to the unregularised optimistic point, see e.g. (Carlier et al., 2023).

4.2 Confidence sets in the phase space

With our reduction onto the Hilbert space $\mathscr{F} := L^2(\mathbb{R}^d; \varrho)$, we can construct an optimistic algorithm by following the general methodology of Abbasi-Yadkori (2012). However, this time the confidence sets are not on the function c^* itself, but in a phase space representation of the problem. The validity of this methodology relies on the standard Assumption 2.

Assumption 2. An *a priori* scale estimate $C \geq \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}$ is known. The sequence $(\xi_t)_{t\in\mathbb{N}}$ is σ^2-sub-Gaussian for some $\sigma \in (0, +\infty)$.

Given a history $(a_s, C_s)_{s\leq t}$ of "actions" ($a_s := \mathsf{F}\pi_s \in \mathscr{F}$) and costs, a strongly convex regulariser Λ (e.g. $\|\cdot\|^2_{L^2(\mathbb{R}^d;\varrho)}$), and $\lambda > 0$, the Regularised Least-Squares (RLS) estimator $\hat{f}_t^\lambda$ of $\mathsf{F}c^*$ in $\mathscr{F}$ is

$$f_t^\lambda \in \operatorname*{argmin}_{f\in\mathscr{F}} \sum_{s=1}^t \left\| C_s - \langle f|a_s\rangle_{L^2(\mathbb{R}^d;\varrho)} \right\|_2^2 + \lambda\Lambda[f] \,. \tag{9}$$

One can also characterise f_t^λ through a closed form expression, see Proposition C.1. Like in the finite dimensional problem, the confidence sets requires the definition of the feature operator

$$M_t : f \in \mathscr{F} \mapsto \left(\langle f|a_s\rangle_{L^2(\mathbb{R}^d;\varrho)}\right)_{s=1}^t \in \mathbb{R}^t \,, \tag{10}$$

its adjoint M_t^*, and the covariance operator $D_t^\lambda := M_t^* M_t + \lambda \mathrm{D}\Lambda$ (D denoting Fréchet differentiation). Given $\delta \in [0, 1]$ the confidence set is then defined as

$$\mathcal{C}_t(\delta) := \left\{ f \in \mathscr{F} : \left\| f - \hat{f}_t^\lambda \right\|_{D_t^\lambda} \leq \beta_t(\delta) \right\} \,, \tag{11}$$

with its width $\beta_t(\delta)$ is chosen as

$$\beta_t(\delta) := \sigma \sqrt{\log\left(\frac{4 \det(\mathrm{D}\Lambda + \lambda^{-1} M_t M_t^*)}{\delta^2} \right)} + \left(\frac{\lambda}{\|\mathrm{D}\Lambda\|_{\mathrm{op}}} \right)^{\frac{1}{2}} C \,. \tag{12}$$

We defer the proofs of the validity of these confidence sets to Appendix C. Performing least-squares in the phase space is a novel technique, but the arguments remain standard.

[4] So called because it is the joint law of random variables $X \sim \mu$ and $Y \sim \nu$ which are independent.

4.3 Solving the decisional aspect of the problem: EntUCB

The combination of these three technical elements (phase-space representation of the problem, phase-space least-squares confidence sets, and entropy regularised optimism) into the OFU framework yields Algorithm 1. We underline that the key contribution is the repeated exploitation of the geometry of the entropic OT problem, whose strong regularity properties allow us to ensure that the algorithm preserves the validity of the phase-space representation it uses to learn. By leveraging the regularity of the OT problem, we show in Theorem 4.1 that Algorithm 1 achieves $\tilde{\mathcal{O}}(\sqrt{T})$ regret, up to learning terms.

```
Data: Confidence $\delta$, regularization level $\lambda$, entropy penalisation $(\varepsilon_t)_{t\in\mathbb{N}}$.
for $t \in \mathbb{N}$ do
    Compute the RLS estimator $\hat{f}_t^\lambda$ using (9) or (20);
    Construct the confidence set $\mathcal{C}_t(\delta)$ using (11) and (12);
    Optimism: pick $(\tilde{\pi}_t, \tilde{c}_t)$ according to (8);
    Play $\pi_t = \tilde{\pi}_t$ if $t > 0$, else $\pi_0 = \mu \otimes \nu$; receive feedback $R_t$;
end
```

Algorithm 1: `EntUCB`

Theorem 4.1. *Under Assumptions 1 and 2, if c^* is L-Lipschitz on supp(μ) × supp(ν) $\subset \mathbb{R}^d$, then for any $\delta > 0$, $\lambda > 0$, $\alpha \in (0,1)$, and $T \in \mathbb{N}$, the regret of Algorithm 1 with $(\varepsilon_t)_{t\in\mathbb{N}} = (\alpha t^{-\alpha})_{t\in\mathbb{N}}$, denoted $\mathcal{B}$, satisfies*

$$\mathscr{R}_T(\mathcal{B}) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + 2C\beta_T(\delta)\sqrt{T\log\det\left(\mathrm{Id} + \frac{1}{2\lambda C}M_T(\mathrm{D}\Lambda)^{-1}M_T^*\right)} + \frac{\kappa\alpha}{1-\alpha}\left(T^{1-\alpha}\log(T) + \frac{\alpha}{2^\alpha}\log(6)\right)$$

with probability at least $1-\delta$, in which κ depends only on (C, L, μ, ν).

Proof sketch. *Having done the technical work to ensure that the phase space construction is valid, the proof now follows the standard OFU methodology. One first isolates the noise of the estimations and controls it using concentration theory and Assumption 2. Next, one uses coherence, modified optimism, and the high-probability validity of $(\mathcal{C}_t(\delta))_{t\in\mathbb{N}}$ to move from c^* to the beliefs $(\mathsf{F}\tilde{c}_t)_{t\in\mathbb{N}}$ in the phase space. Then, one uses the width of the confidence sets to control the regret. The error due to the modified optimism is controlled by $(\varepsilon_t)_{t\in\mathbb{N}}$ and the approximation results for the Kantorovich problem by the entropic one, see Lemma D.2. The full proof is given in Appendix D.2.*

Theorem 4.1 matches the regret rate of Abbasi-Yadkori (2012), showing the problem is learnable in the same way as a linear bandit. However, one must exercise care in controlling the determinant term, as M_T is infinite-dimensional so the confidence sets may be unbounded.

4.4 Solving the statistical aspect of BOT

In order to control the growth with T of the determinant term in Theorem 4.1, two approaches are possible, both aiming to control the spectrum of the feature operator M_T. On the one hand, one can apply structural assumptions to the problem, such as a parametric model for c^* or finite support for μ and ν. On the other hand, we will modify the least-squares estimator to characterise the learning complexity of the problem directly in terms of Assumption 3, a general regularity assumption on c^*. A complete treatment is deferred to Appendix E.

Assumption 3. There is a known orthonormal basis $(\phi_i)_{i\in\mathbb{N}}$ of $L^2(\mathbb{R}^d;\varrho)$ in which we write $\mathsf{F}c^* := \sum_{i=1}^{+\infty}\gamma_i^*\phi_i$ and $\zeta : \mathbb{R}_+ \to [0,1]$, a known monotonically increasing continuous function satisfying

$$\inf_{n\in\mathbb{N}} \frac{\sum_{i=1}^n |\gamma_i^*|^2}{\zeta(n)} \geq \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}^2 .$$

We now replace infinite-dimensional least-squares estimation of c^* in Algorithm 1 by a finite-dimensional approximation on the truncation of the basis $(\phi_i)_{i\in\mathbb{N}}$ to some order $n_t \in \mathbb{N}$ at time $t \in \mathbb{N}$. The resulting Algorithm 2 is given in Appendix E, and its regret is given in Theorem 4.2.

Theorem 4.2 (Theorem E.8)**.** *Assume Assumptions 1 to 3 and $\zeta(n) = 1 - n^{-q}$ for some $q > 0$. For any $\delta \in (0, 1)$, $\lambda > 0$, $\varepsilon > 0$, let $\tilde{\mathcal{B}}$ denote Algorithm 2 with $(n_t)_{t \in \mathbb{N}} = (\lceil t^{\frac{1}{q+1}} \rceil)_{t \in \mathbb{N}}$, $\Lambda_n = \frac{1}{2} \|\cdot\|_2^2$, for all $n \in \mathbb{N}$, and $(\varepsilon_t)_{t \in \mathbb{N}} = (\alpha t^{-\alpha})_{t \in \mathbb{N}}$. For any $T \in \mathbb{N}$,*

$$\mathscr{R}_T(\tilde{\mathcal{B}}) \leq \sigma \sqrt{2T \log\left(\frac{2}{\delta}\right)} + C\left(1 + \frac{2qT^{\frac{q+2}{2q+2}}}{q+1}\right) + \kappa(1 + \sqrt{T}\log(T))$$

$$+ 2C\sigma T^{\frac{q+2}{2q+2}} \left(\sqrt{2 \log\left(\frac{\lambda^{-1} + 2T^{q+2}C^2}{\delta}\right)} + \sqrt{\lambda} C\right) \sqrt{\log\left(1 + \frac{2T^{q+2}}{C^2}\right)}.$$

with probability at least $1 - \delta$, in which κ depends only on (C, L, μ, ν).

Theorem 4.2 shows that the regret of Algorithm 2 is controlled by the regularity of the cost function c^*, which varies from $\mathcal{O}(\sqrt{T})$ for $q \to +\infty$ down to $\mathcal{O}(T)$ as $q \downarrow 0$. Note that $q = 0$ corresponds to c^* being an indicator function, in which learning the optimal plan is clearly infeasible. In this manner the learning complexity is captured directly as a function of the regularity of the cost function.

In particular, the abstract form of Assumption 3 actually allows it to encompass a broad range of structural assumptions on c^*. For example, in parametric and discrete problems, it leads to Corollary 4.3, yielding $\tilde{\mathcal{O}}(\sqrt{KK'T})$ on discrete problems, in which $|\text{supp}(\mu)| = K$ and $|\text{supp}(\nu)| = K'$, and $\tilde{\mathcal{O}}(\sqrt{pT})$ on p-dimensional parametric models. At the same time, Theorem 4.2 also yields a non-parametric rate of $\tilde{\mathcal{O}}(T^{\frac{q+2}{2q+2}})$ when $q > 0$, showing the flexibility of Assumption 3.

Corollary 4.3 (Proposition E.5)**.** *Under Assumptions 1 to 3, if $\zeta(n) = \mathbb{1}_{\{n \geq N\}}$ for some $N \in \mathbb{N}$, then Algorithm 2 can achieve a regret of $\tilde{\mathcal{O}}(\sqrt{NT})$ with $n_t = N$ for all $t \in \mathbb{N}$.*

Theorem 4.2 thus extends Theorem 4.1 by covering the full spectrum of learning complexities from finite and parametric problems to non-parametric ones, with a smooth interpolation of sub-linear regret rates. We detail in Appendix E.6 a Kernel method estimator which can be used instead of the functional regressor of Algorithm 2, but this approach doesn't yield this smooth interpolation.

5 Conclusion and open directions

This article provided the first regret analysis of the bandit optimal transport problem. We have shown (Theorems 4.2 and D.1) that the Kantorovich formulations of the problem can be solved by a modified optimistic algorithm. This algorithm intrinsically exploits the regularity of the optimal transport problem and its entropic regularisation to maintain a "coherent" phase-space representation of the cost function, enforced by a penalised optimism step. Using confidence sets in the phase space and (non-parametric) functional regression allows us to obtain a smooth interpolation of regret bounds from $\tilde{\mathcal{O}}(\sqrt{T})$ in parametric and finite problems, matching the classical bounds of Abbasi-yadkori et al. (2011) for linear bandits, and non-parametric rates up to unlearnable problems.

Our exploration of Bandit Optimal Transport raises several open questions for bandit theorists. First, one wonders what other non-Hilbertian problems can be solved via a "coherent" representation (phase space or otherwise) of the cost function? Are all convex regularisations of optimism possible? Can we simply reduce an ambient space to a Hilbert sub-space by penalising optimism by the subspace's norm? Are there some other complicated spaces or functional problems in applications which would benefit from this approach?

Our analysis also raises several questions in OT theory. In order to implement the optimism step (8), one would need a numerical algorithm which outputs an ϵ-optimal transport plan after finitely many steps. This appears to be absent from the literature, as Sinkhorn's algorithm does *not* output a valid plan in finite time, only in the limit. This also raises the more general question of the regularity properties of this entropy-regularised bilinear problem.

References

Yasin Abbasi-Yadkori. *Online Learning for Linearly Parametrized Control Problems*. PhD thesis, University of Alberta, Department of Computing Science, 2012.

Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved Algorithms for Linear Stochastic Bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Marc Abeille and Alessandro Lazaric. Linear Thompson Sampling Revisited. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 176–184. PMLR, April 2017. ISSN: 2640-3498.

David Abensur, Ivan Balashov, Shaked Bar, Ronny Lempel, Nurit Moscovici, Ilan Orlov, Danny Rosenstein, and Ido Tamir. Productization Challenges of Contextual Multi-Armed Bandits, July 2019. arXiv:1907.04884 [cs].

M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, December 1984. ISSN 0209-9683, 1439-6912. doi: 10.1007/BF02579135.

Noga Alon, Richard Beigel, Simon Kasif, Steven Rudich, and Benny Sudakov. Learning a Hidden Matching. *SIAM Journal on Computing*, 33(2):487–501, January 2004. ISSN 0097-5397, 1095-7111. doi: 10.1137/S0097539702420139.

Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on Optimal Transport*, volume 130 of *UNITEXT*. Springer International Publishing, Cham, 2021. ISBN 978-3-030-72161-9 978-3-030-72162-6. doi: 10.1007/978-3-030-72162-6.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 214–223. PMLR, July 2017. ISSN: 2640-3498.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3 (null):397–422, March 2003. ISSN 1532-4435.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2):235–256, May 2002. ISSN 1573-0565. doi: 10.1023/A: 1013689704352.

Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *Journal of the American Mathematical Society*, 2(2):225–255, 1989. ISSN 0894-0347, 1088-6834. doi: 10.1090/S0894-0347-1989-0969419-8.

Haim Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *J. Mach. Learn. Res.*, 12(null):1655–1695, July 2011a. ISSN 1532-4435.

Sébastien Bubeck, Gilles Stoltz, and Jia Yuan Yu. Lipschitz Bandits without the Lipschitz Constant. In Jyrki Kivinen, Csaba Szepesvári, Esko Ukkonen, and Thomas Zeugmann, editors, *Algorithmic Learning Theory*, volume 6925, pages 144–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011b. ISBN 978-3-642-24411-7 978-3-642-24412-4. doi: 10.1007/978-3-642-24412-4_14.

Haoyang Cao, Xin Guo, and Mathieu Laurière. Connecting GANs, Mean-Field Games, and Optimal Transport. *SIAM Journal on Applied Mathematics*, 84(4):1255–1287, August 2024. ISSN 0036-1399, 1095-712X. doi: 10.1137/22M1499534.

Guillaume Carlier. Optimal Transportation and Economic Applications. 2010.

Guillaume Carlier, Paul Pegon, and Luca Tamanini. Convergence rate of general entropic optimal transport costs. *Calculus of Variations and Partial Differential Equations*, 62(4):116, May 2023. ISSN 0944-2669, 1432-0835. doi: 10.1007/s00526-023-02455-0.

Sinho Chewi, Jonathan Niles-Weed, and Philippe Rigollet. Statistical optimal transport, July 2024. arXiv:2407.18163 [math, stat].

Sayak Ray Chowdhury and Aditya Gopalan. On Kernelized Multi-armed Bandits. In *Proceedings of the 34th International Conference on Machine Learning*, pages 844–853. PMLR, July 2017. ISSN: 2640-3498.

Adrian Constantin. *Fourier Analysis*. Cambridge University Press, 1 edition, May 2016. ISBN 978-1-107-35850-8 978-1-107-04410-4 978-1-107-62035-3. doi: 10.1017/CBO9781107358508.

Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Nabarun Deb, Promit Ghosal, and Bodhisattva Sen. Rates of Estimation of Optimal Transport Maps using Plug-in Estimators via Barycentric Projections. In *Advances in Neural Information Processing Systems*, volume 34, pages 29736–29753. Curran Associates, Inc., 2021.

Stephan Eckstein and Marcel Nutz. Quantitative Stability of Regularized Optimal Transport and Convergence of Sinkhorn's Algorithm, July 2022. arXiv:2110.06798 [math].

Gerald B. Folland. *Fourier analysis and its applications*. Wadsworth & Brooks/Cole mathematics series. Wadsworth & Brooks/Cole advanced books & software, Pacific Grove (Calif.), 1992. ISBN 978-0-534-17094-3.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, August 2015. ISSN 0178-8051, 1432-2064. doi: 10.1007/s00440-014-0583-7.

Nicolas Fournier and Jacques Printems. Absolute continuity for some one-dimensional processes. *Bernoulli*, 16(2):343–360, 2010. ISSN 13507265.

Alfred Galichon. The unreasonable effectiveness of optimal transport in economics, July 2021. arXiv:2107.04700 [econ].

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample Complexity of Sinkhorn Divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 1574–1583. PMLR, April 2019. ISSN: 2640-3498.

Kristiaan Glorie, Bernadette Haase-Kromwijk, Joris van de Klundert, Albert Wagelmans, and Willem Weimar. Allocation and matching in kidney exchange programs. *Transplant International*, 27(4): 333–343, 2014. ISSN 1432-2277. doi: 10.1111/tri.12202.

Alberto Gonzalez-Sanz, Jean-Michel Loubes, and Jonathan Niles-Weed. Weak limits of entropy regularized Optimal Transport; potentials, plans and divergences, June 2024. arXiv:2207.07427 [math].

Florian F. Gunsilius. On the convergence rate of potentials of Brenier maps. *Econometric Theory*, 38 (2):381–417, April 2022. ISSN 0266-4666, 1469-4360. doi: 10.1017/S0266466621000037.

Wenxuan Guo, YoonHaeng Hur, Tengyuan Liang, and Christopher Thomas Ryan. Online Learning to Transport via the Minimal Selection Principle. In *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178, pages 4085–4109. PMLR, 2022.

Botao Hao, Tor Lattimore, and Mengdi Wang. High-Dimensional Sparse Linear Bandits. In *Advances in Neural Information Processing Systems*, volume 33, pages 10753–10763. Curran Associates, Inc., 2020.

John William Hatfield and Paul R. Milgrom. Matching with Contracts. *American Economic Review*, 95(4):913–935, September 2005. ISSN 0002-8282. doi: 10.1257/0002828054825466.

Joseph Horowitz and Rajeeva L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. *Journal of Computational and Applied Mathematics*, 55(3):261–273, November 1994. ISSN 03770427. doi: 10.1016/0377-0427(94)90033-7.

Haichen Hu, Rui Ai, Stephen Bates, and David Simchi-Levi. Contextual Online Decision Making with Infinite-Dimensional Functional Regression, January 2025. arXiv:2501.18359 [stat].

Jan-Christian Hütter and Philippe Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166–1194, April 2021. ISSN 0090-5364, 2168-8966. doi: 10.1214/20-AOS1997.

Meena Jagadeesan, Alexander Wei, Yixin Wang, Michael Jordan, and Jacob Steinhardt. Learning Equilibria in Matching Markets from Bandit Feedback. In *Advances in Neural Information Processing Systems*, volume 34, pages 3323–3335. Curran Associates, Inc., 2021.

David Janz, David Burt, and Javier Gonzalez. Bandit optimisation of functions in the Matérn kernel RKHS. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 2486–2495. PMLR, June 2020. ISSN: 2640-3498.

Ramesh Johari, Vijay Kamble, and Yash Kanoria. Matching While Learning. *Operations Research*, 69(2):655–681, March 2021. ISSN 0030-364X. doi: 10.1287/opre.2020.2013.

Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. Fairness in Learning: Classic and Contextual Bandits. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Leonid Vitaliyevich Kantorovich. On the Translocation of Masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006. Originally published in Dokl. Akad. Nauk SSSR, 37, No. 7-8, 227–229 (1942); translation by A.N. Sobolevskiĭ.

Robert Kleinberg, Alexandru Niculescu-Mizil, and Yogeshwer Sharma. Regret bounds for sleeping experts and bandits. *Machine Learning*, 80(2):245–272, September 2010. ISSN 1573-0565. doi: 10.1007/s10994-010-5178-7.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and Experts in Metric Spaces. *J. ACM*, 66(4):30:1–30:77, 2019. ISSN 0004-5411. doi: 10.1145/3299873.

Andrej N. Kolmogorov. On the best approximation of functions of a given class. In Vladimir M. Tikhomirov, editor, *Selected Works of A. N. Kolmogorov*, volume Volume I Mathematics and Mechanics, chapter 28, pages 202–205. Kluwer Acad. Publ., Dordrecht, 1991. Translated from: Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. —Ann. Math., 1936, vol.37, p.107–110.

Vladik Kreinovich, Woraphon Yamaka, and Supanika Leurcharusmee, editors. *Applications of Optimal Transport to Economics and Related Topics*, volume 556 of *Studies in Systems, Decision and Control*. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-67769-4 978-3-031-67770-0. doi: 10.1007/978-3-031-67770-0.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, March 1985. ISSN 01968858. doi: 10.1016/0196-8858(85) 90002-8.

Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, July 2020. ISBN 978-1-108-57140-1 978-1-108-48682-8. doi: 10.1017/9781108571401.

Xin Liu, Bin Li, Pengyi Shi, and Lei Ying. An efficient pessimistic-optimistic algorithm for stochastic linear bandits with general constraints. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, pages 24075–24086, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 978-1-71384-539-3.

George G. Lorentz. *Approximation of functions*. AMS Chelsea Publ, Providence, RI, 2. ed., repr edition, 2005.

Flavien Léger. A Gradient Descent Perspective on Sinkhorn. *Applied Mathematics & Optimization*, 84(2):1843–1855, October 2021. ISSN 1432-0606. doi: 10.1007/s00245-020-09697-w.

Stefan Magureanu, Richard Combes, and Alexandre Proutiere. Lipschitz Bandits: Regret Lower Bound and Optimal Algorithms. In *Proceedings of The 27th Conference on Learning Theory*, pages 975–999. PMLR, May 2014. ISSN: 1938-7228.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *The Annals of Statistics*, 52(3), June 2024. ISSN 0090-5364. doi: 10.1214/24-AOS2379.

Yifei Min, Tianhao Wang, Ruitu Xu, Zhaoran Wang, Michael Jordan, and Zhuoran Yang. Learn to Match with No Regret: Reinforcement Learning in Markov Matching Markets. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 19956–19970. Curran Associates, Inc., 2022.

Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

Jeffrey S. Morris. Functional Regression. *Annual Review of Statistics and Its Application*, 2(Volume 2, 2015):321–359, May 2015. ISSN 2326-8298, 2326-831X. doi: 10.1146/annurev-statistics-010814-020413.

Marcel Nutz. Introduction to Entropic Optimal Transport. 2022.

Michaël Perrot, Nicolas Courty, Rémi Flamary, and Amaury Habrard. Mapping Estimation for Discrete Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Allan Pinkus. *n-Widths in Approximation Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1985. ISBN 978-3-642-69896-5 978-3-642-69894-1. doi: 10.1007/978-3-642-69894-1. URL `http://link.springer.com/10.1007/978-3-642-69894-1`.

Philippe Rigollet and Austin J. Stromme. On the sample complexity of entropic optimal transport, June 2022. arXiv:2206.13472 [math].

Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. Improving gans using optimal transport. *arXiv preprint arXiv:1803.05573*, 2018.

Flore Sentenac. *Learning and Algorithms for Online Matching*. PhD thesis, Institut Polytechnique de Paris, July 2023.

Flore Sentenac, Jialin Yi, Clement Calauzenes, Vianney Perchet, and Milan Vojnovic. Pure Exploration and Regret Minimization in Matching Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, pages 9434–9442. PMLR, July 2021. ISSN: 2640-3498.

Mathieu Seurin, Philippe Preux, and Olivier Pietquin. "I'm Sorry Dave, I'm Afraid I Can't Do That" Deep Q-Learning from Forbidden Actions. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2020. doi: 10.1109/IJCNN48605.2020.9207496. ISSN: 2161-4407.

Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, May 1967. ISSN 0030-8730. Publisher: Mathematical Sciences Publishers.

Austin J. Stromme. Minimum Intrinsic Dimension Scaling for Entropic Optimal Transport. In Jonathan Ansari, Sebastian Fuchs, Wolfgang Trutschnig, María Asunción Lubiano, María Ángeles Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Combining, Modelling and Analyzing Imprecision, Randomness and Dependence*, pages 491–499, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-65993-5. doi: 10.1007/978-3-031-65993-5_60.

Sho Takemori and Masahiro Sato. Approximation Theory Based Methods for RKHS Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, pages 10076–10085. PMLR, July 2021. ISSN: 2640-3498.

M. Talagrand. The Transportation Cost from the Uniform Measure to the Empirical Measure in Dimension ≥ 3. *The Annals of Probability*, 22(2):919–959, 1994. ISSN 0091-1798.

Carla Tameling, Max Sommerfeld, and Axel Munk. Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *The Annals of Applied Probability*, 29(5): 2744–2781, October 2019. ISSN 1050-5164, 2168-8737. doi: 10.1214/19-AAP1463. Publisher: Institute of Mathematical Statistics.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Luis Caicedo Torres, Luiz Manella Pereira, and M. Hadi Amini. A Survey on Optimal Transport for Machine Learning: Theory and Applications, June 2021. arXiv:2106.01963.

Long Tran-Thanh and Jia Yuan Yu. Functional Bandits, May 2014. arXiv:1405.2432 [stat].

A.B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2008.

Michal Valko, Nathan Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits, 2013.

Claire Vernade, Alexandra Carpentier, Tor Lattimore, Giovanni Zappella, Beyza Ermis, and Michael Brückner. Linear bandits with Stochastic Delayed Feedback. In *Proceedings of the 37th International Conference on Machine Learning*, pages 9712–9721. PMLR, November 2020. ISSN: 2640-3498.

C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. ISBN 0-8218-3312-X.

Cédric Villani. Optimal Transport, old and new. volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-71049-3 978-3-540-71050-9. doi: 10.1007/978-3-540-71050-9.

Yifei Wang, Tavor Baharav, Yanjun Han, Jiantao Jiao, and David Tse. Beyond the Best: Distribution Functional Estimation in Infinite-Armed Bandits. *Advances in Neural Information Processing Systems*, 35:9262–9273, December 2022.

Larry Wasserman. *All of Nonparametric Statistics*. Springer Science & Business Media, September 2006.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, November 2019. ISSN 1350-7265. doi: 10.3150/18-BEJ1065. Publisher: Bernoulli Society for Mathematical Statistics and Probability.

Yinchu Zhu and Ilya O. Ryzhov. Semidiscrete optimal transport with unknown costs, November 2023. arXiv:2310.00786 [econ].

Appendices

A Preliminaries

A.1 Organisation of Appendices

The following appendices are organised thematically and are mostly independent completions of various parts of the text. Appendix A.2 contains notations and clarifications that are shared across them.

Appendix B provides a rigourous treatment of necessary Fourier analysis notions, which allow for a rigourous outlining of the schema detailed in Section 4.1.

Appendices C to E contains the majority of the technical contributions of this work, including the major lemmata used in the proofs of the main text. Appendix C is dedicated to the details of the constructions in Section 4, while Appendix D focuses on the general regret proofs of Section 4.3, specifically the proof of Theorem 4.1. Finally Appendix E is dedicated to the details of Section 4.4 on specific regularity dependent regret. Some miscellaneous minor results, or reproductions of results from prior works are collected in Appendix F.

The remaining appendices contain complements to the text and discussion of topics not mentioned therein for the sake of brevity. Appendix G contains more detailed discussions of the open problems mentioned in Section 5. Appendix H contains bibliographical notes on statistical optimal transport which readers unfamiliar with the field might find of interest to understand the context of the paper. It is a complement to Section 3.

A.2 Notational precisions

Throughout the text, for a reference measure ϱ, let $L^p(\mathcal{X}, \mathbb{K}; \varrho)$, $p \in [1, \infty]$ and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, denote the space of functions $f : \mathcal{X} \to \mathbb{K}$ that are p-integrable. When $\mathcal{X}$, $\mathbb{K}$, or ϱ are clear from context we will drop them for brevity; by default $\mathbb{K} = \mathbb{C}$. We allow complex functions ($\mathbb{K} = \mathbb{C}$) to deal with the Fourier transforms, but this has no noticeable effect as it does not impact the Hilbertian structure of the space $L^2(\mathbb{R}^d, \mathbb{K}; \varrho)$.

In the following, let $\langle \cdot | \cdot \rangle_{L^2(\mathbb{R}^d, \varrho)}$ denote the inner product on $L^2(\mathbb{R}^d, \mathbb{K}; \varrho)$, $\langle \cdot | \cdot \rangle_{\ell_2(\mathbb{R}^d)}$ the one on $\ell^2(\mathbb{R}, \mathbb{K})$ (the space of square integrable $\mathbb{K}$-valued sequences) with $\|\cdot\|_{\ell_2(\mathbb{R}^d)}$ denoting its associated norm. On $\mathbb{R}^d$, $\langle \cdot | \cdot \rangle_2$ denotes the inner product, $\|\cdot\|_2$ the Euclidean norm. As before, let $\langle \cdot | \cdot \rangle$ denote the duality pairing between $\mathscr{M}(\mathbb{R}^d)$ (the space of finite Radon measures) and $\mathcal{C}_0(\mathbb{R}^d)$ (the space functions vanishing at infinity). The operator norm of a linear operator A (in finite or infinite dimension) is denoted by $\|A\|_{\mathrm{op}}$.

Throughout, all probabilistic statements are understood as holding in the filtered probability space $(\Omega, \mathcal{F}_\infty, \mathbb{F}, \mathbb{P})$, in which $\mathbb{F} := (\mathcal{F}_t)_{t \in \mathbb{N}}$ is the natural filtration of $(\xi_t)_{t \in \mathbb{N}}$, and $\mathcal{F}_\infty = \lim_{t \to \infty} \mathcal{F}_t$.

For two measures $(\gamma, \rho) \in \mathscr{M}(\mathbb{R}^d)$, $\gamma \ll \rho$ denotes that γ is absolutely continuous with respect to ρ, in which case we use $\mathrm{d}\gamma / \mathrm{d}\rho$ to denote the Radon-Nikodym derivative (a.k.a. the density) of γ with respect to ρ.

B Elements of Fourier Analysis

B.1 Formal definitions

To define the Fourier transform on $L^2(\mathbb{R}^d; \varrho)$, we will extend it from a dense subspace (see Definition 1) of $L^2(\mathbb{R}^d; \varrho)$ to the whole space. This technical construction arises as a consequence of the fact that $L^2(\mathbb{R}^d; \varrho) \not\subset L^1(\mathbb{R}^d; \varrho)$, meaning the right-hand side of (13) may not be defined and F is ill-posed on $L^2(\mathbb{R}^d; \varrho)$, despite the fact that (13) is well-posed for $f \in L^1(\mathbb{R}^d; \varrho)$. The following is summarised from (Constantin, 2016, Ch.5–6), refer therein for a more detailed treatment or, e.g., to (Folland, 1992).

Definition 1. The Schwartz space $\mathcal{S}(\mathbb{R}^d)$ is defined as

$$\left\{\phi \in \mathcal{C}^\infty(\mathbb{R}^d; \mathbb{C}) : \sup_{x\in\mathbb{R}^d} |x^\alpha \partial_\beta \phi(x)| < +\infty \text{ for any } \alpha, \beta \in \mathbb{N}^d\right\}$$

in which $\alpha, \beta \in \mathbb{N}^d$ are multi-indices so that $x^\alpha := (x_i^{\alpha_i})_{i=1}^d$, and $\partial_\beta := \partial_{x_1}^{\beta_1} \cdots \partial_{x_d}^{\beta_d}$.

Note that $\mathcal{S}(\mathbb{R}^d)$ is a dense subspace of $L^2(\mathbb{R}^d; \varrho)$ and $L^1(\mathbb{R}^d; \varrho)$ as it contains $\mathcal{C}_c^\infty(\mathbb{R}^d; \mathbb{C})$ the space of infinitely-differentiable compactly-supported (a.k.a. test) functions, which is dense in both $L^2(\mathbb{R}^d; \varrho)$ and $L^1(\mathbb{R}^d; \varrho)$.

Theorem B.1 ((Constantin, 2016, Thm. 6.1))**.** *Consider the Fourier transform operator* F *on the Schwartz space, with*

$$\mathsf{F} : \phi \in \mathcal{S}(\mathbb{R}^d) \mapsto \int \phi(x) e^{-2\pi i \langle x|\cdot\rangle_2} \mathrm{d}\varrho(x) \,. \tag{13}$$

This operator maps $\mathcal{S}(\mathbb{R}^d)$ *maps onto itself and is an isometric bijection. Moreover,*

$$\mathsf{F}^{-1} = \mathsf{F}\mathsf{R} \,, \tag{14}$$

in which $\mathsf{R} : \phi \in \mathcal{S}(\mathbb{R}^d) \mapsto \phi(-\cdot) \in \mathcal{S}(\mathbb{R}^d)$ *is the* reflection *operator.*

Theorem B.2 ((Constantin, 2016, Thm. 6.4))**.** *The fourier transform* F *can be extended to a unitary operator on* $L^2(\mathbb{R}^d; \varrho)$ *and* (14) *holds on* $L^2(\mathbb{R}^d; \varrho)$ *for this extension.*

The formal inversion property (14) is easily shown to recover the classical inversion formula

$$f(x) = \int \mathsf{F}f(\xi) e^{2\pi i \langle x|\xi\rangle} \mathrm{d}\varrho(\xi) \text{ for } \varrho\text{-a.e. } x \in \mathcal{X} \tag{15}$$

as soon as $f \in L^1(\mathbb{R}^d; \varrho) \cap L^2(\mathbb{R}^d; \varrho)$. In our case ϱ is a finite measure so $L^2(\mathbb{R}^d; \varrho) \subset L^1(\mathbb{R}^d; \varrho)$ and the inversion formula always holds. If ϱ is only σ-finite (e.g. the Lebesgue measure), one must take slightly higher care. Namely the difference between (14) and (15) is whether the integral in (15) is well defined for $f \in L^2(\mathbb{R}^d; \varrho)$, which is not guaranteed.

This technicality reflects the limits used in the definition of the extension which are hidden by the abstract statement of Theorem B.2. Nevertheless, since the Schwartz space $\mathcal{S}(\mathbb{R}^d)$ is dense in both $L^1(\mathbb{R}^d; \varrho)$ and $L^2(\mathbb{R}^d; \varrho)$, we can always take an arbitrarily close function in $\mathcal{S}(\mathbb{R}^d)$ and invert that, the result will remain arbitrarily close in $L^2(\mathbb{R}^d; \varrho)$.

The Schwartzian framework turns out to be a robust one for Fourier analysis more generally, and we can also use to extend F beyond $L^2(\mathbb{R}^d; \varrho)$. In particular, it can be used to unify the definitions we gave for the Fourier transform of a function and a measure, refer to (Constantin, 2016, § 6.1.2) for more details. Precisely, one extends to the topological dual of $\mathcal{S}(\mathbb{R}^d)$ (the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$), which includes $\mathscr{M}(\mathbb{R}^d)$ and $L^2(\mathbb{R}^d; \varrho)$ as sub-spaces.

A fundamental consequence of the various formulations of the Fourier transform is that measures whose transforms are in $L^2(\mathbb{R}^d; \varrho)$ are exactly those which have an $L^2(\mathbb{R}^d; \varrho)$ density with respect to ϱ. We will denote the density of a measure μ with respect to ϱ using the Radon-Nikodym notation $\mathrm{d}\mu/\mathrm{d}\varrho$, even when this tempered distribution can be identified with a function.

Lemma B.3. *Let* $\gamma \in \mathscr{M}(\mathcal{X})$ *be a finite Radon measure, if it has density with respect to* ϱ *and* $\mathrm{d}\gamma/\mathrm{d}\varrho \in L^2(\mathbb{R}^d; \varrho)$*, then*

$$\mathsf{F}\gamma = \mathsf{F}\frac{\mathrm{d}\gamma}{\mathrm{d}\rho} \in L^2(\mathbb{R}^d; \varrho) \,.$$

Conversely, if $\mathsf{F}\gamma \in L^2(\mathbb{R}^d; \varrho)$*, then* γ *has a density with respect to* ϱ *and* $\mathrm{d}\gamma/\mathrm{d}\varrho \in L^2(\mathbb{R}^d; \varrho)$.

Proof. The first part is a direct consequence of the definitions of the Fourier transforms of a measure and an $L^2(\mathbb{R}^d; \varrho)$ function. For the converse, the fact that $\mathsf{F}\gamma \in L^2(\mathbb{R}^d; \varrho)$ implies $\gamma \ll \varrho$ involves some technical minutiae due to the different topologies $\mathscr{M}(\mathcal{X})$ can be equipped with, which we won't reproduce for conciseness, refer to e.g. (Fournier and Printems, 2010, Lemma 1.1). That the density is then in $L^2(\varrho)$ is a simple consequence of Plancherel's theorem:

$$\left\|\frac{\mathrm{d}\gamma}{\mathrm{d}\rho}\right\|_{L^2(\mathbb{R}^d;\rho)} = \int_{\mathbb{R}^d} |F\gamma(\xi)|^2 \,\mathrm{d}\rho(\xi) = \|F\gamma\|_{L^2(\mathbb{R}^d;\rho)} \,.$$

□

B.2 Technical details of Section 4.1

Let $\mathcal{C}_0(\mathbb{R}^d, \mathbb{K})$ denote the space of continuous functions from $\mathbb{R}^d$ to $\mathbb{K} \in \{\mathbb{R}; \mathbb{C}\}$, $\mathscr{M}(\mathbb{R}^d)$ denote the space of finite Borel measures over $\mathbb{R}^d$, and let us define the Fourier operator on this space by using the same notation, i.e. $\mathsf{F} : \gamma \in \mathscr{M}(\mathbb{R}^d) \mapsto \mathsf{F}\gamma \in \mathcal{C}_0(\mathbb{R}^d; \mathbb{C})$ with

$$\mathsf{F}\gamma : \xi \in \mathbb{R}^d \mapsto \int e^{-2\pi i \langle x|\xi\rangle_2} \mathrm{d}\gamma(x)\,. \tag{16}$$

Note that we will eschew the standard notations $\hat{f}$ and $\hat{\gamma}$ in favour of $\mathsf{F}f$ and $\mathsf{F}\gamma$ to avoid confusion with the least-squares estimator, which we will denote using its standard hat.

The Riesz-Markov theorem shows that $(\mathscr{M}^*(\mathbb{R}^d), \|\cdot\|_\infty)$, the space of finite signed Borel measures on $\mathbb{R}^d$ (endowed with the total variation norm $\|\cdot\|_\infty$), is the topological dual of $(\mathcal{C}_0(\mathbb{R}^d), \|\cdot\|_\infty)$, the space of continuous functions which vanish at infinity (endowed with the supremum norm $\|\cdot\|_\infty$), refer e.g. to (Constantin, 2016, p. 242). This duality is characterised by the pairing

$$\langle\cdot|\cdot\rangle : (f, \gamma) \in \mathcal{C}_0(\mathbb{R}^d) \times \mathscr{M}^*(\mathbb{R}^d) \mapsto \int f \mathrm{d}\gamma \in \mathbb{R}\,.$$

This pairing applies in particular to all functions $f \in \mathcal{C}(\mathcal{X}; \mathbb{R})$ if $\mathcal{X}$ is compact and to all positive finite Borel measures $\gamma \in \mathscr{M}^+(\mathcal{X})$, and we will use the pairing notation in this case too. In general we will use the notation for arbitrary functions, understood that it will be well defined, see also Remark B.1. In particular:

$$\text{Kant.}(\mu, \nu, c) = \inf_{\pi \in \Pi(\mu,\nu)} \langle c|\pi\rangle\,.$$

Lemma B.4. *For any finite Borel measure $\rho \in \mathscr{M}(\mathbb{R}^d)$, any $\gamma \in \mathscr{M}(\mathbb{R}^d)$ finite and with $\mathrm{d}\mu/\mathrm{d}\rho \in L^2(\mathbb{R}^d; \rho)$, and any $f \in L^2(\mathbb{R}^d; \rho) \cap L^1(\mathbb{R}^d; \rho)$, we have*

$$\langle f|\gamma\rangle = \langle \mathsf{F}\mathsf{R}f|\mathsf{F}\gamma\rangle_{L^2(\mathbb{R}^d;\rho)}$$

and

$$|\langle f|\gamma\rangle| \le \|f\|_{L^2(\mathbb{R}^d;\rho)} \left|\rho(\mathbb{R}^d)\right| \left|\gamma(\mathbb{R}^d)\right|\,.$$

Proof. By (15),

$$\langle f|\gamma\rangle := \int f \mathrm{d}\gamma = \int\int \mathsf{F}f(\xi) e^{2\pi i\langle x|\xi\rangle} \mathrm{d}\rho(\xi)\mathrm{d}\gamma(x)\,. \tag{17}$$

Let $\varphi : (x, \xi) \mapsto e^{2\pi i\langle x|\xi\rangle}$. Using (17), since by the Cauchy-Schwartz inequality

$$\begin{aligned} |\langle f|\gamma\rangle| &\le \|Ff\|_{L^2(\mathbb{R}^d\times\mathbb{R}^d;\gamma\otimes\rho)} \|1\|_{L^2(\mathbb{R}^d\times\mathbb{R}^d;\gamma\otimes\rho)} \\ &= \|Ff\|_{L^2(\mathbb{R}^d;\rho)} \gamma(\mathbb{R}^d)^2 \rho(\mathbb{R}^d) < +\infty\,, \end{aligned} \tag{18}$$

the integrand in (17) is $\gamma \otimes \rho$-integrable, and thus we can apply the Fubini-Lebesgue theorem to obtain

$$\langle f|\gamma\rangle = \int \mathsf{F}f(\xi) e^{2\pi i\langle x|\xi\rangle} \mathrm{d}[\gamma \otimes \rho](\xi, x) = \langle Ff|\varphi\rangle_{L^2(\mathbb{R}^d\times\mathbb{R}^d;\gamma\otimes\rho)}\,.$$

Furthermore,

$$\begin{aligned} \langle f|\gamma\rangle &= \int \mathsf{F}f(\xi) \int e^{2\pi i\langle x|\xi\rangle} \mathrm{d}\gamma(x)\mathrm{d}\rho(\xi) \\ &= \langle \mathsf{F}\mathsf{R}f|\mathsf{F}\gamma\rangle_{L^2(\mathbb{R}^d;\rho)}. \end{aligned}$$

By (18), we have once more:

$$|\langle f|\gamma\rangle| = \left|\langle \mathsf{F}\mathsf{R}f|\mathsf{F}\gamma\rangle_{L^2(\mathbb{R}^d;\rho)}\right| \le \|Ff\|_{L^2(\mathbb{R}^d;\rho)} \gamma(\mathbb{R}^d)^2 \rho(\mathbb{R}^d)\,.$$

□

The benefit of Lemma B.4 may not be immediately apparent, but it is revealed when one notices that the $L^2(\mathbb{R}^d;\rho)$ inner products and norms considered on the right hand side depend only on the measure ρ and not on γ. Thus, we are able to assume only integrability of c^* only with respect to our reference measure ϱ (recall (5)) and still manipulate the duality product $\langle c^*|\gamma\rangle$ for any γ. In particular, by taking $\varrho = \mu\otimes\nu$ given marginals μ and ν and playing π_t such that $\Psi^{\varepsilon}_{\mu,\nu}(c^*,\pi_t) < +\infty$ (recall (6)) we can reduce $\langle c^*|\pi_t\rangle$ to a $L^2(\mathbb{R}^d;\varrho)$ inner product, moving our problem to a Hilbert space.

Remark B.1. Lemma B.4 opens the subject of discussing Assumption 1. Let us remark that if $S := \text{supp}(\mu\otimes\nu)$ is compact, continuity of c^* on the closure of S is sufficient to obtain these results. Similarly, if c^* is bounded. However, Assumption 1 allows for many more functions, for instance it allows $c^* : (x,y) = \|x-y\|_2^2$ if $(\mu,\nu)\in\mathscr{P}_2(\mathbb{R}^d)$, where $\mathscr{P}_2(\mathbb{R}^d)$ denotes measures with a finite second moment. This is of value as it covers the Wasserstein distances which are of broad interest. In general, one can develop finer assumptions based on (μ,ν) even if ϱ is not finite, but we do not detail this for brevity.

C Technical contributions in Bandit Theory

C.1 Confidence sets and Regularised least-squares

Recall that $C_t := \langle c^*|\pi_t\rangle + \xi_i = \langle \mathsf{F}\mathsf{R}c^*|\mathsf{F}\pi_t\rangle_{L^2(\mathbb{R}^d;\varrho)} + \xi_t$ (by Lemma B.4), in which by Assumption 2 we have $(\xi_i)_{i\in\mathbb{N}}$ a conditionally σ-sub-Gaussian sequence. Let $a_t := \mathsf{F}\pi_t \in L^2(\mathbb{R}^d;\varrho)$ for $t\in\mathbb{N}$, and $\vec{a}_t := (a_i)_{i=1}^t$ and $\vec{C}_t := (C_i)_{i=1}^t$.

Let us begin by defining the regularised least-squares estimator of c^*. Let $J_{\cdot}[\cdot] : \mathbb{N}\times\mathscr{F}\to\mathbb{R}$ be the (random) functional defined by

$$(t,f)\mapsto J_t[f] := \sum_{s=1}^{t} \left\| C_s - \langle f|a_s\rangle_{L^2(\mathbb{R}^d;\varrho)} \right\|_2^2 .$$

Consider $\Lambda : L^2(\mathbb{R}^d;\varrho)\to\mathbb{R}$, a strongly convex and continuously Fréchet-differentiable functional whose Fréchet derivative, denoted $\mathrm{D}\Lambda$, satisfies

$$\frac{1}{M_\Lambda}\|f\|_{L^2(\mathbb{R}^d;\varrho)} \le \mathrm{D}\Lambda[f] \le M_\Lambda \|f\|_{L^2(\mathbb{R}^d;\varrho)} \quad \text{for any} \quad f\in L^2(\mathbb{R}^d;\varrho) \tag{19}$$

for some $M_\Lambda > 0$, e.g. $\Lambda = \frac{1}{2}\|\cdot\|^2_{L^2(\mathbb{R}^d;\varrho)}$ with $M_\Lambda = 1$. Let us recall that the Fréchet derivative of a strongly convex Fréchet-differentiable functional is a (strongly) positive-definite operator denoted $\mathrm{D}\Lambda$. It is clear that $J_t + \lambda\Lambda$ is a strongly convex functional for any $\lambda > 0$ and $t\in\mathbb{N}^*$. Therefore, we can define the Λ-regularised least-squares estimator of c^* to be

$$\hat{f}_t^\lambda := \underset{f\in L^2(\mathbb{R}^d;\varrho)}{\operatorname{argmin}}\; J_t[f] + \lambda\Lambda[f] .$$

Proposition C.1. *Assume Assumption 1, then for any $\lambda > 0$, and $t\in\mathbb{N}^*$, we have*

$$\hat{f}_t^\lambda = (M_t^* M_t + \lambda\mathrm{D}\Lambda)^{-1} M_t^* \vec{C}_t , \tag{20}$$

in which, for every $t\in\mathbb{N}^*$, $M_t : L^2(\mathbb{R}^d;\varrho)\to\mathbb{R}^t$ is the linear a.s. bounded operator defined by

$$M_t : f\in L^2(\mathbb{R}^d;\varrho)\mapsto \left(\langle f|a_t\rangle_{L^2(\mathbb{R}^d;\varrho)}\right)_{i=1}^t \in\mathbb{R}^t , \tag{21}$$

and $M_t^* : \mathbb{R}^t\to L^2(\mathbb{R}^d;\varrho)$ is its adjoint, defined by

$$M_t^* : v\in\mathbb{R}^t \mapsto \sum_{s=1}^{t} v_s a_s \in L^2(\mathbb{R}^d;\varrho) . \tag{22}$$

Proof. This proof extends the standard arguments for finite-dimensional least-squares, we include it for completeness focusing on the differences owing to infinite dimensions, cf. e.g. (Abbasi-Yadkori, 2012, § 3.2). One first computes the Fréchet derivative of J_t, by studying a variation $\delta f \in L^2(\mathbb{R}^d;\varrho)$ and

$$J_t[f+\delta f] - J_t[f]\,.$$

One sees that the Fréchet derivative of J_t exists for all t and is given by

$$f \mapsto \sum_{s=1}^{t} \langle f | a_s \rangle_{L^2(\mathbb{R}^d;\varrho)} a_s - C_s a_s + \lambda \mathrm{D}\Lambda f = (M_t^* M_t + \lambda \mathrm{D}\Lambda) f - M_t^* \vec{C}_t\,.$$

Note that the right-hand side is easily checked by expanding the definition of M_t and M_t^*, and in doing so one easily checks that M_t^* is indeed the adjoint of M_t. Carrying on, by first order optimality, the normal equation is

$$(M_t^* M_t + \lambda \mathrm{D}\Lambda)\hat{f}_\lambda = M_t^* \vec{C}_t\,.$$

Since $M_t^* M_t$ is positive semi-definite and $\mathrm{D}\Lambda$ is positive definite, (20) follows. □

Let $D_t := M_t^* M_t$ and $D_t^\lambda := D_t + \lambda \mathrm{D}\Lambda$ denote the covariance and regularised covariance operators at time $t \in \mathbb{N}$. Let

$$\mathcal{E}_t(\delta) := \left\{ \left\| \hat{f}_t^\lambda - \mathsf{F}c^* \right\|_{D_t^\lambda} \le \sigma \sqrt{\log\!\left(\frac{4\det(\mathrm{D}\Lambda + \lambda^{-1} M_t M_t^*)}{\delta^2} \right)} + \left(\frac{\lambda}{\|\mathrm{D}\Lambda\|_{\mathrm{op}}} \right)^{\frac{1}{2}} \|\mathsf{F}c^*\|_{L^2(\mathbb{R}^d;\varrho)} \right\}. \tag{23}$$

for $t \in \mathbb{N}$.

Lemma C.2 ((Abbasi-Yadkori, 2012, Cor. 3.6))**.** *For every $\delta \in (0,1)$, $\lambda > 0$, under Assumptions 1 and 2 we have*

$$\mathbb{P}\left(c^* \in \bigcap_{t\in\mathbb{N}} \mathsf{F}^{-1}\mathcal{C}_t(\delta) \right) \ge \mathbb{P}\left(\bigcap_{t\in\mathbb{N}} \mathcal{E}_t(\delta/2) \right) \ge 1 - \frac{\delta}{2}\,.$$

Proof. Recall that F is an isometry on $L^2(\mathbb{R}^d;\varrho)$, and so is F^{-1}, so $\mathsf{F}^{-1}\mathcal{C}_t$ is a confidence set for c^* in $L^2(\mathbb{R}^d;\varrho)$, and it is an ellipsoid of identical radius $\beta_t(\delta)$ centred at $\mathsf{F}^{-1}\hat{f}_t^\lambda$. A direct combination of Assumption 2, (23), and (Abbasi-Yadkori, 2012, Cor. 3.6) yields

$$\mathbb{P}\left(\bigcap_{t\in\mathbb{N}} \mathcal{E}_t(\delta/2) \right) \ge 1 - \frac{\delta}{2}\,.$$

The second results follow by comparison of (23) and (12). □

Lemma C.3. *Under Assumptions 1 and 2, on the event $\{c^* \in \cap_{t\in\mathbb{N}} \mathsf{F}^{-1}\mathcal{C}_t(\delta)\}$, for any $T \in \mathbb{N}$ and $(c_t)_{t=1}^T$ with $c_t \in \mathsf{F}^{-1}\mathcal{C}_t(\delta)$ for $t \in [T]$, we have*

$$\sum_{t=1}^{T} \langle c^* - c_t | \tilde{\pi}_t \rangle \le 2C\beta_T(\delta) \sqrt{T \log\det\left(\mathrm{Id} + \frac{1}{2\lambda C} M_t (\mathrm{D}\Lambda)^{-1} M_t^* \right)}$$

Proof. Consider $t \ge 0$, $c_t \in \mathcal{C}_t(\delta)$, and let $\varphi_t := \langle c^* - c_t | \tilde{\pi}_t \rangle$. Recall that $a_t := \mathsf{F}\pi_t$. By Lemma C.2 and the Cauchy-Schwartz inequality, on the event $\{c^* \in \cap_{t\in\mathbb{N}} \mathsf{F}^{-1}\mathcal{C}_t(\delta)\}$, we have

$$|\varphi_t| \le \beta_t(\delta) \, \|a_t\|_{(D_t^\lambda)^{-1}}\ ,$$

while, by the Cauchy-Schwartz inequality, Assumption 1, and using the fact that F is an isometry on $L^2(\mathbb{R}^d;\varrho)$, we have

$$|\varphi_t| \le \|\mathsf{FR}c^* - \mathsf{FR}c\|_{L^2(\mathbb{R}^d;\varrho)} \|a_t\|_{L^2(\mathbb{R}^d;\varrho)} = \|c^* - c\|_{L^2(\mathbb{R}^d;\varrho)} \, \pi_t(\mathbb{R}^d) \le 2C\,.$$

Combining yields

$$|\varphi_t| \le \beta_t(\delta) \min\{\|a_t\|_{(D_t^\lambda)^{-1}}, 2C\} = 2C\beta_t(\delta) \left(\frac{1}{2C} \|a_t\|_{(D_t^\lambda)^{-1}} \wedge 1 \right).$$

Squaring and applying the inequality $u \leq 2\log(1+u)$, which holds on $[0,1]$, to the final term, yields

$$|\varphi_t|^2 \leq 8C^2 \beta_t(\delta)^2 \log\left(1 + \frac{1}{2C}\left\|a_t\right\|_{(D_t^\lambda)^{-1}}\right)$$

and, summing up,

$$\sum_{t=1}^T \varphi_t \leq \sqrt{T\sum_{t=1}^T |\varphi_t|^2} \leq 2C\beta_t(\delta)\sqrt{T\sum_{t=1}^T \log\left(1+\frac{1}{2C}\left\|a_t\right\|_{(D_t^\lambda)^{-1}}\right)}\,. \tag{24}$$

By definition of M_T and D_T^λ, we have

$$\begin{aligned}\sum_{t=1}^T \log\left(1+\frac{1}{2C}\left\|a_t\right\|_{(D_t^\lambda)^{-1}}\right) &= \log\left(\prod_{t=1}^T\left(1+\frac{1}{2C}\left\|a_t\right\|_{(D_t^\lambda)^{-1}}\right)\right)\\ &= \log\det\left(\mathrm{Id}+\frac{1}{2\lambda C}M_T(\mathrm{D}\Lambda)^{-1}M_T^*\right) \end{aligned}\tag{25}$$

as wanted. □

D Regret bounds

D.1 Entropic regret bounds

Let us introduce the notion of *entropic regret*, which is the regret relative to the entropic optimal transport problem (6), i.e.

$$\mathscr{R}_T^{\mathscr{H},\varepsilon}(\boldsymbol{\pi}) := \sum_{t=1}^T (C_t + \varepsilon\mathscr{H}(\pi_t|\varrho)) - \text{Ent.}(\mu,\nu,c^*,\varepsilon) \qquad \text{for} \quad T\in\mathbb{N}\,. \tag{26}$$

To facilitate the analysis and the presentation of results, recall the entropic transport functional

$$\Psi_{\mu,\nu}^\varepsilon : (c,\pi) \in L^2(\mathbb{R}^d;\varrho)\times\Pi(\mu,\nu) \mapsto \langle c|\pi\rangle + \varepsilon\mathscr{H}(\pi)\,.$$

Thus, $\text{Ent.}(\mu,\nu,c,\varepsilon) = \inf_{\pi\in\Pi(\mu,\nu)}\Psi_{\mu,\nu}^\varepsilon(c,\pi)$ and (26) becomes

$$\mathscr{R}_T^{\mathscr{H},\varepsilon}(\pi) := \sum_{t=1}^T \Psi_{\mu,\nu}^\varepsilon(c^*,\pi_t) + \xi_t - \text{Ent.}(\mu,\nu,c^*,\varepsilon)\,.$$

Theorem D.1. *Under Assumptions 1 and 2, for any $\varepsilon>0$, $\delta>0$, $\lambda>0$, and $T\in\mathbb{N}$, the regret of Algorithm 1 with $(\varepsilon_t)_{t\in\mathbb{N}} = (\varepsilon)_{t\in\mathbb{N}}$, denoted by $\mathcal{A}$, satisfies*

$$\mathscr{R}_T^{\mathscr{H},\varepsilon}(\mathcal{A}) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + 2C\beta_T(\delta)\sqrt{T\log\det\left(\mathrm{Id}+\frac{1}{2\lambda C}M_T(\mathrm{D}\Lambda)^{-1}M_T^*\right)}$$

with probability at least $1-\delta$. Note that M_T (thus also $\beta_T(\delta)$) depends implicitly on ε.

Proof. Recall the we identify $\mathcal{A}$ with the $\mathbb{F}$-adapted process $\boldsymbol{\pi} := (\pi_t)_{t\in\mathbb{N}} \subset \Pi(\mu,\nu)$ of transport plans played. The instantaneous regret of the algorithm at time $t\in\mathbb{N}$ is defined as

$$r_t := \Psi_{\mu,\nu}^\varepsilon(c^*,\pi_t) + \xi_t - \text{Ent.}(\mu,\nu,c^*,\varepsilon)\,.$$

It is clear that $\mathscr{R}_T^{\mathscr{H}}(\mathcal{A}) = \sum_{t=1}^T r_t$. Before pursuing further, let us apply Lemma F.1 to the sequence $(\xi_t)_{t\in\mathbb{N}}$, in view of Assumption 2, to obtain that for any $\delta>0$ we have

$$\mathbb{P}\left(\sum_{i=1}^T \xi_i \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)}\right) \geq 1-\frac{\delta}{2}\,. \tag{27}$$

Now, let $\bar{r}_t := \Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \text{Ent.}(\mu, \nu, c^*, \varepsilon)$ as we continue the decomposition. By definition of the algorithm, let

$$(\tilde{\pi}_t, \tilde{c}_t) \in \operatorname*{argmin}_{\substack{\pi \in \Pi(\mu,\nu) \\ c \in \mathcal{C}_t(\delta)}} \Psi^{\varepsilon}_{\mu,\nu}(c, \pi)\,,$$

where $\mathcal{C}_t(\delta)$ is the confidence set defined in (11).

Let us place ourselves on the event $\bigcap_{t=1}^{\infty} \left\{c^* \in \mathsf{F}^{-1}\mathcal{C}_t(\delta)\right\}$, an event which happens with probability at least $1 - \delta/2$ by Lemma C.2. By optimism, we have

$$\bar{r}_t \leq \Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \text{Ent.}(\mu, \nu, \tilde{c}_t, \varepsilon)$$

The instant regret can be decomposed as

$$\bar{r}_t = \Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \Psi^{\varepsilon}_{\mu,\nu}(\tilde{c}_t, \pi_t) + \Psi^{\varepsilon}_{\mu,\nu}(\tilde{c}_t, \pi_t) - \text{Ent.}(\mu, \nu, \tilde{c}_t, \varepsilon)$$

The first term $\Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \Psi^{\varepsilon}_{\mu,\nu}(\tilde{c}_i, \pi_t) = \langle c^* - \tilde{c}_t | \pi_t \rangle$ can be bounded by Lemma C.3, while the second term is 0 by definition of Algorithm 1. The proof is completed by taking a union bound over $\bigcap_{t=1}^{\infty} \left\{c^* \in \mathsf{F}^{-1}\mathcal{C}_t(\delta)\right\}$ and the event of (27). □

D.2 Kantorovich regret bounds

Let us begin by giving the requisite results on approximation of the Kantorovich problem by the entropic one. Let $d_{\mathscr{H}}(\gamma)$ (for $\gamma \in \{\mu, \nu\}$) denote the *upper Renyi dimension* of γ, defined by

$$d_{\mathscr{H}}(\gamma) := \limsup_{\epsilon \downarrow 0} \frac{H_{\varepsilon}(\gamma)}{\log(\varepsilon^{-1})}$$

in which $H_{\varepsilon}(\gamma)$ is the infimum (over all countable partitions of supp(γ) into Borel subsets of diameter at most ε) of the discrete entropy of γ with respect to the partition, see Carlier et al. (2023).

Lemma D.2 ((Carlier et al., 2023, Prop. 3.1))**.** *If c^* is L-Lipschitz on supp$(\mu) \times$ supp(ν), then*

$$\textit{Ent.}(\mu, \nu, c^*, \varepsilon) - \textit{Kant.}(\mu, \nu, c^*) \leq \varepsilon \left((d_{\mathscr{H}}(\mu) \wedge d_{\mathscr{H}}(\nu)) \log(\varepsilon^{-1}) + L \right)$$

as $\varepsilon \downarrow 0$.

Extensions of this result exist for more general absolute continuity conditions, see (Carlier et al., 2023, Rem. 3.4). This constant is sharp, but tighter bounds may be obtained under stronger regularity assumptions, see e.g. (Carlier et al., 2023, Prop. 3.7). In view of Lemma D.2, we can define $\kappa := (d_{\mathscr{H}}(\mu) \wedge d_{\mathscr{H}}(\nu)) + L$. In spite of its apparent complexity, upper Renyi dimension is a relatively well behaved object, and can be bounded in many common situations, see the following remarks.

Remark D.1 ((Carlier et al., 2023, Prop. 3.2)). If γ is a measure on $\mathbb{R}^d$ satisfying

$$\int 0 \vee \log(\|x\|_2) \mathrm{d}\gamma(x) < +\infty$$

then $d_{\mathscr{H}}(\gamma) \leq d$.

Remark D.2 ((Carlier et al., 2023, Rem. 3.5)). If γ is finitely supported, then $d_{\mathscr{H}}(\gamma) = 0$.

Theorem 4.1. *Under Assumptions 1 and 2, if c^* is L-Lipschitz on supp$(\mu) \times$ supp$(\nu) \subset \mathbb{R}^d$, then for any $\delta > 0$, $\lambda > 0$, $\alpha \in (0, 1)$, and $T \in \mathbb{N}$, the regret of Algorithm 1 with $(\varepsilon_t)_{t \in \mathbb{N}} = (\alpha t^{-\alpha})_{t \in \mathbb{N}}$, denoted $\mathcal{B}$, satisfies*

$$\mathscr{R}_T(\mathcal{B}) \leq \sigma \sqrt{2T \log\left(\frac{2}{\delta}\right)} + 2C\beta_T(\delta) \sqrt{T \log\det\left(\mathrm{Id} + \frac{1}{2\lambda C} M_T (\mathrm{D}\Lambda)^{-1} M_T^*\right)} + \frac{\kappa\alpha}{1-\alpha} \left(T^{1-\alpha} \log(T) + \frac{\alpha}{2^{\alpha}} \log(6)\right)$$

with probability at least $1 - \delta$, in which κ depends only on (C, L, μ, ν).

Proof. The proof follows the same lines as the proof of Theorem D.1. Again, we identify $\mathcal{B}$ with the transport plans $\boldsymbol{\pi} := (\pi_t)_{t\in\mathbb{N}} \subset \Pi(\mu,\nu)$ it plays. The instantaneous regret is different due to the change of objective, it is given by

$$r_t := C_t - \text{Kant.}(\mu,\nu,c^*) = \langle c^* | \pi_t - \pi^* \rangle + \xi_t \,.$$

As before, apply Lemma F.1 to the sequence $(\xi_i)_{i\in\mathbb{N}}$, and pass to $\bar{r}_t := \langle c^* | \pi_t - \pi^* \rangle$, which can be decomposed as

$$\begin{aligned}
\bar{r}_t &= \langle c^* | \pi_t \rangle - \langle c^* | \pi^* \rangle \\
&= \langle c^* | \pi_t \rangle - \text{Ent.}(\mu,\nu,c^*,\varepsilon) + \text{Ent.}(\mu,\nu,c^*,\varepsilon) - \text{Kant.}(\mu,\nu,c^*) \\
&\leq \langle c^* | \pi_t \rangle - \text{Ent.}(\mu,\nu,c^*,\varepsilon) + \varepsilon(d_{\mathscr{H}}(\mu) \wedge d_{\mathscr{H}}(\nu)) \log(\varepsilon^{-1}) + L\varepsilon
\end{aligned}$$

for any $\varepsilon > 0$, by Lemma D.2. In particular, for $\varepsilon = \varepsilon_t$ as used by Algorithm 1, we have

$$\sum_{t=1}^{T} \varepsilon_t (d_{\mathscr{H}}(\mu) \wedge d_{\mathscr{H}}(\nu)) \log(\varepsilon_t^{-1}) + L\varepsilon_t \leq \frac{\kappa\alpha}{1-\alpha} \left(T^{1-\alpha} \log(T) + \frac{\alpha}{2^\alpha} \log(6) \right) \,. \tag{28}$$

by Lemma F.2. Let us recall that optimism implies that

$$(\tilde{\pi}_t, \tilde{c}_t) \in \underset{\substack{\pi\in\Pi(\mu,\nu) \\ c\in\mathcal{C}_t(\delta)}}{\operatorname{argmin}} \; \Psi^{\varepsilon_t}_{\mu,\nu}(c,\pi) \,,$$

for $\varepsilon_t > 0$ as used by Algorithm 1, so that

$$\text{Ent.}(\mu,\nu,\tilde{c}_t,\varepsilon_t) \leq \text{Ent.}(\mu,\nu,c^*,\varepsilon_t) \text{ on } \mathcal{E}_t(\delta) \,.$$

Let us place ourselves on the event $\cap_{t=1}^{\infty} \left\{ c^* \in \mathsf{F}^{-1}\mathcal{C}_t(\delta) \right\} \supset \cap_{t=1}^{\infty} \mathcal{E}_t(\delta)$, which happens with probability at least $1 - \delta/2$ by Lemma C.2. On this event, we thus have

$$\langle c^* | \pi_t \rangle - \text{Ent.}(\mu,\nu,c^*,\varepsilon_t) \leq \langle c^* | \pi_t - \tilde{\pi}_t \rangle + \langle c^* - \tilde{c}_t | \tilde{\pi}_t \rangle \,,$$

since $\mathscr{H} \geq 0$. Applying $\pi_t = \tilde{\pi}_t$ we obtain the desired result, up to combining the $\bar{r}_t$ over $t \in \mathbb{N}$, recalling (28), and taking a union bound over the two events. □

E Controlling the infinite dimensional terms

The parametric and RKHS estimation methodologies are highly standard in bandit theory, because they seamlessly fit into the general Hilbert Space analysis of Abbasi-Yadkori (2012) while giving a control on the resulting regret bounds in terms of finite dimensional quantities. In Fourier analysis and fields which rely on it, such as functional regression (Morris, 2015), it is more natural to look for approximations by decomposing $\mathsf{F}c^*$ and a_t into an orthonormal basis and truncating it at some finite order. We detail this learning methodology below.

We being in Appendix E.1 by presenting the general concept of basis decomposition as an approximation method. Then, in Appendix E.2 we truncate at a fixed order and derive the regret bounds for this case. Before moving on to changing the truncation order with $t \in \mathbb{N}$ in Appendix E.4, we give a brief discussion in Appendix E.3 of some examples in which a finite basis is sufficient. Finally, we give a brief treatment of kernel methods in Appendix E.6 for completeness.

E.1 Intrinsic regularity and fourier basis decay

To simplify notation, let $f^* := \mathsf{F}c^*$. Recall the chosen orthonormal basis $(\phi_i)_{i\in\mathbb{N}}$ of the space $L^2(\mathbb{R}^d;\varrho)$, in which $(f^*, a_t) \in L^2(\mathbb{R}^d;\varrho)^2$, $t \in \mathbb{N}$, admit representations

$$f^* = \sum_{i=0}^{\infty} \gamma_i^* \phi_i \quad \text{and} \quad a_t = \sum_{i=0}^{\infty} \vartheta_i^{(t)} \phi_i \,, \quad \text{for some} \quad (\gamma^*, \vartheta^{(t)}) \in \ell_2(\mathbb{R})^2 \,.$$

Classical choices for $(\phi_i)_{i=\in\mathbb{N}}$ are wavelet systems such as the Haar or Hermitian systems, and the Fourier basis if supp$(\mu) \times$ supp(ν) is bounded. The choice of a specific basis is made *ad hoc* from knowledge of the structure of the problem; we present the general argument.

By definition of $(\phi_i)_{i\in\mathbb{N}}$ as an orthonormal basis, we have

$$\langle f|a_t\rangle_{L^2(\mathbb{R}^d;\varrho)} = \langle \gamma^*|\vartheta^{(t)}\rangle_{\ell_2(\mathbb{R})} = \sum_{n=1}^{+\infty} \gamma_n \vartheta_n^{(t)} \,.$$

Let $f^*|_n := \sum_{i=0}^{n} \gamma_i^* \phi_i$ be the truncation of the basis expansion of f at order $n \in \mathbb{N}$. By abuse of notation, and only when it is clear from context, we will override notation and denote $c^*|_n$ the result of applying the inverse fourier transform to $(f^*)|_n$, the basis truncation of $f^* := \mathsf{F}c^*$. A straightforward derivation yields the approximation bound of Lemma E.1.

Lemma E.1. *Let $(\phi_i)_{i\in\mathbb{N}}$ be an orthonormal basis of $L^2(\mathbb{R}^d;\varrho)$, and let $f \in L^2(\mathbb{R}^d;\varrho)$ with $f := \sum_{i=0}^{\infty} \gamma_i \phi_i$. Then, we have*

$$|\langle f - f|_n|g\rangle|_{L^2(\mathbb{R}^d;\varrho)} \leq \|g\|_{L^2(\mathbb{R}^d;\varrho)} \sqrt{\sum_{i=n+1}^{+\infty} |\gamma_i|^2} \quad \textit{for every } g \in L^2(\mathbb{R}^d;\varrho).$$

For our purpose, $g = a_t$ is bounded by 1 since $\|\mathsf{F}\pi_t\|_\infty \leq \pi_t(\mathbb{R}^d) = 1$ and $\varrho(\mathbb{R}^d) = 1$, so that the resulting approximation error of f^* by $(f^*|_n)_{n\in\mathbb{N}}$ is controlled entirely by the decay of the coefficients $(\gamma_i^*)_{i\in\mathbb{N}}$. Consequently, regret analysis can leverage Lemma E.1 to move the problem into a finite dimensional regression problem on the coefficients $(\gamma_i^*)_{i\in\mathbb{N}}$. We begin by setting the stage with a fixed order (i.e. n independent of t) methodology. Later, we will derive regret guarantees when n is allowed to grow with t in order to control the approximation error.

E.2 Fixed order basis truncation

In this section, let $n \in \mathbb{N}$ be fixed. One can approximately regress $\vec{C}_t$ against $\vec{a}_t$ up to order n by solving the n-dimensional Regularised Least-Squares (RLS) problem

$$\hat{\gamma}_t^{n,\lambda} := \operatorname*{argmin}_{\gamma\in\mathbb{R}^n} \sum_{s=1}^{t} \left\| C_s - \sum_{i=1}^{n} \gamma_i \vartheta_i^{(s)} \right\|_2^2 + \lambda \Lambda_n(\gamma) \,, \tag{29}$$

in which $\Lambda_n : \mathbb{R}^n \to [0,+\infty)$ is a strictly convex continuously Fréchet-differentiable regulariser such that its Fréchet derivative $\mathrm{D}\Lambda_n$ satisfies

$$\frac{1}{M_{\Lambda_n}} \,\mathrm{Id} \preceq \mathrm{D}\Lambda_n \preceq M_{\Lambda_n} \,\mathrm{Id} \;.$$

For clarity, let $\vartheta^{(s,n)}$ denote the truncation of $\vartheta^{(s)} \in \mathbb{R}^{\mathbb{N}}$ at order n, so that $\vartheta^{(s,n)} \in \mathbb{R}^n$ and $\vartheta_i^{(s,n)} = \vartheta_i^{(s)}$ for all $i \in [n]$. Following the standard arguments for online linear regression (omitted for brevity, see e.g. Abbasi-yadkori et al. (2011); Abbasi-Yadkori (2012)), one can construct the (valid, by Corollary E.2) confidence sets

$$\tilde{\mathcal{C}}_t^n(\delta) := \left\{ \gamma \in \mathbb{R}^n : \left\| \gamma - \hat{\gamma}_t^{n,\lambda} \right\|_{\tilde{D}_t^{\lambda,n}} \leq \tilde{\beta}_{t,n}(\delta) \right\} \,, \tag{30}$$

in which $\tilde{D}_t^{\lambda,n} := \lambda \mathrm{D}\Lambda_n + \sum_{s=1}^{t} \vartheta^{(s,n)} {\vartheta^{(s,n)}}^\top$ and

$$\tilde{\beta}_t^n(\delta) := \sigma \sqrt{\log\left(\frac{4\det\left(\mathrm{D}\Lambda_n + \lambda^{-1} \sum_{s=1}^{t} \vartheta^{(s,n)} {\vartheta^{(s,n)}}^\top \right)}{\delta^2} \right)} + \left(\frac{\lambda}{\|\mathrm{D}\Lambda_n\|_{\mathrm{op}}} \right)^{\frac{1}{2}} C \,. \tag{31}$$

Notice that $C > \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}$ implies that $C \geq \|\gamma^*\|_{\ell_2(\mathbb{R})}$ by definition of $(\phi_i)_{i\in\mathbb{N}}$, so that C is a valid upper bound on $\|\gamma^*\|_{\ell_2(\mathbb{R})}$. To verify the validity of the confidence sets (see Corollary E.2), let

$$\tilde{\mathcal{E}}_t^n(\delta) := \left\{ \left\| \gamma - \hat{\gamma}_t^{n,\lambda} \right\|_{\tilde{D}_t^{\lambda}} \leq \tilde{\beta}_t^n(\delta) \right\} \quad \text{for} \quad (t,n) \in \mathbb{N}^2 \,. \tag{32}$$

Corollary E.2. *Under Assumptions 1 and 2, for every $\delta > 0$, $\lambda > 0$, $n \in \mathbb{N}$,*

$$\mathbb{P}\left(\bigcap_{t=1}^{\infty} \tilde{\mathcal{E}}_t^n(\delta)\right) \geq 1 - \frac{\delta}{2}\,.$$

Proof. Follow the proof method of Lemma C.2 (or apply Lemma E.6 below). □

Applying this learning methodology to Algorithm 1 in place of the infinite-dimensional RLS, and with the optimistic choice of belief-action pairs

$$(\tilde{\pi}_t, \tilde{\gamma}_t^n) \in \operatorname*{argmin}_{\substack{\pi \in \Pi(\mu,\nu) \\ \gamma \in \tilde{\mathcal{C}}_t^n(\delta)}} \Psi^{\varepsilon}_{\mu,\nu}\left(\mu, \nu, \sum_{i=1}^{n} \gamma_i \phi_i, \varepsilon\right) \tag{33}$$

yields Algorithm 2 with $(n_t)_{t\in\mathbb{N}} = (n)_{t\in\mathbb{N}}$ and the regret bound of Corollary E.3.

```
Data: Confidence $\delta$, regularization level $\lambda$, entropy penalisation $(\varepsilon_t)_{t\in\mathbb{N}}$, orders $(n_t)_{t\in\mathbb{N}}$.
for $t \in \mathbb{N}$ do
    Compute the RLS estimator $\hat{\gamma}_t^{n_t,\lambda}$ using (29);
    Construct the confidence set $\tilde{\mathcal{C}}_t^{n_t}(\delta)$ using (30) and (31);
    Optimism: pick $(\tilde{\pi}_t, \tilde{\gamma}_t^{n_t})$ according to (33);
    Play $\pi_t = \tilde{\pi}_t$ if $t > 0$, else $\pi_0 = \mu \otimes \nu$; receive feedback $R_t$;
end
```

Algorithm 2: `Basis-truncation EntUCB`

Corollary E.3. *Under Assumptions 1 and 2, for any $\delta > 0$, $\lambda > 0$, $T \in \mathbb{N}$, using Algorithm 2 with $(\varepsilon_t)_{t\in\mathbb{N}} = (\varepsilon)_{t\in\mathbb{N}}$ and $(n_t)_{t\in\mathbb{N}} = (n)_{t\in\mathbb{N}}$ (denoted $\mathcal{A}_n$) yields*

$$\begin{aligned} \mathscr{R}_T^{\mathscr{H},\varepsilon}(\mathcal{A}_n) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + 2C\sqrt{nT}\left(\log\left(\frac{M_{\Lambda_n}}{\lambda} + \frac{tC^2}{n}\right) + \frac{n}{2(1\wedge\lambda C)}\log M_{\Lambda_n}\right) \\ + 2T\sum_{k=n+1}^{+\infty} |\gamma_k^*|\,, \end{aligned} \tag{34}$$

while using $(\varepsilon_t)_{t\in\mathbb{N}} = (\alpha t^{-\alpha})_{t\in\mathbb{N}}$ and $(n_t)_{t\in\mathbb{N}} = (n)_{t\in\mathbb{N}}$ (denoted $\mathcal{B}_n$) yields

$$\begin{aligned} \mathscr{R}_T(\mathcal{B}) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + 2C\sqrt{nT}\left(\log\left(\frac{M_{\Lambda_n}}{\lambda} + \frac{tC^2}{n}\right) + \frac{n}{2(1\wedge\lambda C)}\log M_{\Lambda_n}\right) \\ + \frac{\kappa\alpha}{1-\alpha}\left(T^{1-\alpha}\log(T) + \frac{\alpha}{2^{\alpha}}\log(6)\right) + 2T\sum_{k=n+1}^{+\infty} |\gamma_k^*|\,. \end{aligned} \tag{35}$$

Proof. The proof follows the usual decomposition up the following modifications which are the same for both (34) and (35). We give the modification for Theorem D.1, the same modifications need only be applied to Theorem 4.1 to complete the proof of the second bound.

At the second step of the proof, let $\bar{\pi}^{\epsilon}$ be an ϵ-minimiser of Ent.$(\mu, \nu, c^*, \varepsilon)$, for $\epsilon > 0$, and decompose $\bar{r}_t := \Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \text{Ent.}(\mu, \nu, c^*, \varepsilon)$ as

$$\begin{aligned} \bar{r}_t &\leq \epsilon + \Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \Psi^{\varepsilon}_{\mu,\nu}(c^*, \bar{\pi}^{\epsilon}) \\ &\leq \epsilon + \Psi^{\varepsilon}_{\mu,\nu}(c^*|_n, \pi_t) - \text{Ent.}(\mu, \nu, c^*|_n, \varepsilon) \\ &\quad + \Psi^{\varepsilon}_{\mu,\nu}(c^*, \pi_t) - \Psi^{\varepsilon}_{\mu,\nu}(c^*|_n, \pi_t) + \Psi^{\varepsilon}_{\mu,\nu}(c^*|_n, \bar{\pi}^{\epsilon}) - \Psi^{\varepsilon}_{\mu,\nu}(c^*, \bar{\pi}^{\epsilon}) \\ &\leq \epsilon + \Psi^{\varepsilon}_{\mu,\nu}(c^*|_n, \pi_t) - \text{Ent.}(\mu, \nu, c^*|_n, \varepsilon) + 2\sum_{k=n+1}^{+\infty} |\gamma_k^*|\,, \end{aligned}$$

by a double application of Lemma E.1 combined with the bound $\|a_t\|_{L^2(\mathbb{R}^2;\varrho)} \leq 1$. Sending $\epsilon \to 0$ allows one to then continue the proof, up to replacing the events $\mathcal{E}_t(\delta)$ by $\tilde{\mathcal{E}_t^n}(\delta)$, and Lemma C.2 by Corollary E.2.

Finally, let us introduce $\tilde{c}_t^n := \sum_{i=1}^n \tilde{\gamma}_{t,i}^n \phi_i$ for $t \in \mathbb{N}$, so that by Lemma C.3, we can directly derive

$$\sum_{t=1}^T \langle c^*|_n - \tilde{c}_t^n | \tilde{\pi}_t \rangle \leq 2C\tilde{\beta}_{T,n}(\delta)\sqrt{nT \log\det\left(\mathrm{I} + \frac{1}{2\lambda C}\sum_{t=1}^T \vartheta_t^{(t,n)} \mathrm{D}\Lambda_n^{-1} {\vartheta^{(t,n)}}^\top\right)}\,.$$

To obtain the stated bounds, it remains to bound

$$\det\left(\mathrm{D}\Lambda_n + \lambda^{-1}\sum_{t=1}^T \vartheta^{(t,n)} {\vartheta^{(t,n)}}^\top\right) \quad \text{and} \quad \det\left(\mathrm{I} + \frac{1}{2\lambda C}\sum_{t=1}^T \vartheta^{(t,n)} \mathrm{D}\Lambda_n^{-1} {\vartheta^{(t,n)}}^\top\right)$$

using Lemma E.4. □

Lemma E.4. *Under Assumptions 1 and 2, for $(n,t) \in \mathbb{N}^2$, we have*

$$\log\det\left(\mathrm{D}\Lambda_n + \lambda^{-1}\sum_{t=1}^T \vartheta^{(t,n)} {\vartheta^{(t,n)}}^\top\right) \leq \log\left(\frac{M_{\Lambda_n}}{\lambda} + \frac{tC^2}{n}\right) + n\log M_{\Lambda_n}\,. \tag{36}$$

$$\log\det\left(\mathrm{I} + \frac{1}{2\lambda C}\sum_{t=1}^T \vartheta_t^{(t,n)} \mathrm{D}\Lambda_n^{-1} {\vartheta^{(t,n)}}^\top\right) \leq \log\left(\frac{M_{\Lambda_n}}{\lambda} + \frac{tC^2}{n}\right) + \frac{n}{2\lambda C}\log M_{\Lambda_n}\,. \tag{37}$$

Proof. We take the two bounds in turn. First, apply the matrix determinant lemma to obtain

$$\det\left(\mathrm{D}\Lambda_n + \lambda^{-1}\sum_{t=1}^T \vartheta^{(t,n)} {\vartheta^{(t,n)}}^\top\right) \leq \det(\mathrm{D}\Lambda_n)\det\left(\mathrm{I} + \lambda^{-1}\sum_{t=1}^T \vartheta^{(t,n)} \mathrm{D}\Lambda_n^{-1} {\vartheta^{(t,n)}}^\top\right)\,,$$

which can be readily bounded as in (Abbasi-Yadkori, 2012, Lemma E.3) by noticing that $\|\vartheta^{(t,n)}\|_2 \leq \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}$ (with $\det(\mathrm{D}\Lambda_n) \leq M_{\lambda_n}^n \vee 1$) to obtain (36).

For the second bound, apply (Abbasi-Yadkori, 2012, Lemma E.3) directly to obtain

$$\det\left(\mathrm{I} + \frac{1}{2\lambda C}\sum_{t=1}^T \vartheta^{(t,n)} \mathrm{D}\Lambda_n^{-1} {\vartheta^{(t,n)}}^\top\right) \leq \left(\frac{2\lambda C\,\mathrm{Tr}(\mathrm{D}\Lambda_n) + tC^2}{n}\right)^n (2\lambda C\det(\mathrm{D}\Lambda_n))$$

wherefrom (37) follows. □

E.3 Finite order bases: matching and parametric models

At this point, let us recall Assumption 3 which provides the quantification of the regularity of c^* which we will use to set n. We will now discuss some examples in which a finite basis is sufficient to control the approximation error.

Assumption 3. There is a known orthonormal basis $(\phi_i)_{i\in\mathbb{N}}$ of $L^2(\mathbb{R}^d;\varrho)$ in which we write $\mathsf{F}c^* := \sum_{i=1}^{+\infty} \gamma_i^* \phi_i$ and $\zeta : \mathbb{R}_+ \to [0,1]$, a known monotonically increasing continuous function satisfying

$$\inf_{n\in\mathbb{N}} \frac{\sum_{i=1}^n |\gamma_i^*|^2}{\zeta(n)} \geq \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}^2\,.$$

Proposition E.5. *Under Assumptions 1 to 3, with $\zeta(n)\mathbb{1}_{\cdot\geq N}$ for some $N \in \mathbb{N}$ (i.e. if $\gamma_i^* = 0$ for every $i > N$), then under the conditions of Corollary E.3 with $n = N$, $\Lambda_n = \|\cdot\|_{L^2(\mathbb{R}^d;\varrho)}/2$, and $\alpha = 1/2$ the bounds of Corollary E.3 become*

$$\mathscr{R}_T^{\mathscr{H},\varepsilon}(\mathcal{A}_n) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + 2C\sqrt{NT}\log\left(\frac{1}{\lambda} + \frac{TC^2}{N}\right) \tag{38}$$

and

$$\mathscr{R}_T(\mathcal{B}_n) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + 2C\sqrt{NT}\log\left(\frac{1}{\lambda} + \frac{TC^2}{N}\right) + \kappa(1 + \sqrt{T}\log(T)) \tag{39}$$

Naturally, the assumption that $\gamma_i^* = 0$ for any $i > N$ is not satisfactory, but it is verified for several existing models and serves to demonstrate that some learning problems in BOT are learnable at the rate $\tilde{\mathcal{O}}(\sqrt{T})$ given only knowledge of an upper bound on N and $\|\gamma^*\|_{\ell_2(\mathbb{R})}$ and an appropriate basis $(\phi_i)_{i\in\mathbb{N}}$.

Consider a matching problem in which the measures μ and ν are supported on K and K' *loci* respectively. Let $\{x_1, \ldots, x_K\} = \operatorname{supp}(\mu)$ and $\{x'_1, \ldots, x'_{K'}\} = \operatorname{supp}(\nu)$ denote these *loci*. We can let c^* assume arbitrarily values outside of $\mathcal{X} = \{(x_i, x'_j) : (i,) \in [K] \times [K']\}$ without loss of generality. Let $\epsilon < \inf\{\|u - v\| : (u, v) \in \mathcal{X}^2\,,\ u \neq v\}$ and define the functions

$$\phi_{i,j} := \frac{6}{\pi\epsilon^3}\mathbb{1}_{\{B_2((x_i,x'_j)^\top,\epsilon/2)\}} \qquad \text{for } (i,j) \in [K] \times [K']\,.$$

Re-indexing the functions by $k \in [K \times K']$, and adding suitable functions for $k > KK'$, we obtain an orthonormal basis $(\phi_k)_{k\in\mathbb{N}}$ of $L^2(\mathbb{R}^d; \varrho)$, in which $c^* := \sum_{k=1}^{KK'} \gamma_k^*\phi_k$. Consequently, we can apply Proposition E.5 with $N = KK'$ to obtain a regret bound of $\tilde{\mathcal{O}}(\sqrt{KK'T})$ for the learning problem.

Alternatively, consider that there is a parametric model for c^*, i.e. there is $\theta^* \in \mathbb{R}^p$ such that

$$c^*(x, y) = \sum_{i=1}^{p} \theta_i^*\Phi_i(x, y)\,,$$

for some embedding function $\Phi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^p$. When the embedding function is known, one can construct a basis through the Gram-Schmidt process. Let $\phi_1 := \Phi_1/\left\|\Phi_1\right\|_{L^2(\mathbb{R}^d;\varrho)}$, and for $i \leq p$, define $S_i := \{\phi_k : k < i\}^\perp$ the orthogonal complement of the sequence this far. Now, repeatedly project the feature dimensions onto S_i to construct $\phi_i := P_{S_i}\Phi_i/\left\|P_{S_i}\Phi_i\right\|_{L^2(\mathbb{R}^d;\varrho)}$, wherein P_{S_i} denotes the projection onto S_i, $i \in \mathbb{N}$. For $i > p$, take any orthonormal basis of S_p to complete the basis, it will not be used anyway. Consequently, we can also apply Proposition E.5 with $N = p$ to obtain a regret bound of $\tilde{\mathcal{O}}(\sqrt{pT})$ for the learning problem.

These results are summarised in Corollary 4.3, but notice that higher order polynomial models can be readily considered as well, such as quadratic costs

$$c^*(x, y) = \Phi(x, y)^\top\Theta^*\Phi(x, y)\,,$$

for $\Theta^* \in \mathbb{R}^{p\times p}$, by simply reparametrising it as a linear model in dimension p^2 and applying the same construction. Many other models can be considered in this manner, and would benefit from further specialised investigation.

Corollary 4.3 (Proposition E.5)**.** *Under Assumptions 1 to 3, if $\zeta(n) = \mathbb{1}_{\{n\geq N\}}$ for some $N \in \mathbb{N}$, then Algorithm 2 can achieve a regret of $\tilde{\mathcal{O}}(\sqrt{NT})$ with $n_t = N$ for all $t \in \mathbb{N}$.*

E.4 Increasing order basis truncation

In this section, we will extend the results of Appendix E.2 to let n change with $t \in \mathbb{N}$ along the learning process. We will denote the corresponding sequence by $(n_t)_{t\in\mathbb{N}} \subset \mathbb{N}$. It is relatively simple to see that the proofs of the key properties of online least-squares estimation will extend, but we include the key proof sketches for completeness. We begin by diagonalising the validity of the confidence sets in Lemma E.6.

Lemma E.6. *Under Assumptions 1 and 2,*

$$\mathbb{P}\left(\bigcap_{t=1}^{\infty}\tilde{\mathcal{E}}_t^{n_t}(\delta)\right) \geq 1 - \frac{\delta}{2}\,.$$

Proof. The proof only requires diagonalisation of the standard stopping time construction. For $(\delta, t) \in (0,1) \times \mathbb{N}$, on the filtered probability space $(\Omega, \mathcal{F}_\infty, \mathbb{F}, \mathbb{P})$ define

$$B_t(\delta) := \left\{\omega \in \Omega : \left\|\gamma^* - \hat{\gamma}_t^{n_t,\lambda}\right\|_{\tilde{D}_t^{\lambda,n_t}} \leq \tilde{\beta}_{t,n_t}(\delta)\right\} \overset{\text{a.s.}}{=} \{\omega \in \Omega : c^*|_{n_t} \notin \tilde{\mathcal{C}}_t^{n_t}(\delta)\}\,,$$

be the t^{th} "bad event", and let $\tau_\delta : \omega \in \Omega \to \inf\{t \in \mathbb{N} : \omega \in B_t(\delta)\}$, which is a stopping time. We have

$$\{\tau < +\infty\} = \bigcup_{t\in\mathbb{N}} B_t(\delta)\,.$$

By construction, in the classical manner:

$$\mathbb{P}\left(\bigcup_{t\in\mathbb{N}} B_t(\delta)\right) = \mathbb{P}(\tau < +\infty, B_t(\delta)) \leq \mathbb{P}\left(\tilde{\mathcal{E}}_t^{n_t}(\delta)\right) \leq \frac{\delta}{2}\,.$$

□

The confidence sets using for non-constant $(n_t)_{t\in\mathbb{N}}$ are simply instantiations of (30) and (31) with n_t in place of n. This change of basis with time however requires a modification of the proof of Lemma C.3 as the steps summed up in (24) are no longer homogenous. In particular, (25) is no longer valid.

Lemma E.7. *Under Assumptions 1 and 2, if $\Lambda_n := \frac{1}{2}\left\|\cdot\right\|_2^2$ with the norm being on $\mathbb{R}^n$, then*

$$\sum_{t=1}^{T}\langle c^*|_{n_t} - \tilde{c}_t^{n_t}|\tilde{\pi}_t\rangle \leq 2C\sigma\left(\sqrt{2\log\left(\frac{\lambda^{-1}+\frac{TC^2}{n_T}}{\delta}\right)} + \sqrt{\lambda}C\right)\sqrt{n_T T\log\left(1+\frac{T}{n_T C^2}\right)}\,.$$

Proof. Recall the notation of Lemma C.3, which adapts to $\varphi_t := \langle c^*|_{n_t} - c_t^{n_t}|\tilde{\pi}_t\rangle$ for $t \in \mathbb{N}$ and $\tilde{c}_t^{n_t} := \sum_{i=1}^{n_t}\tilde{\gamma}_{t,i}^{n_t}\phi_i$. The proof of Lemma C.3 yields

$$\sum_{t=1}^{T}\varphi_t \leq 2C\beta_{T,n_T}(\delta)\sqrt{T\sum_{t=1}^{T}\log\left(1+\frac{1}{2C}\left\|\vartheta^{(t,n_t)}\right\|_{(\tilde{D}_t^{\lambda,n_t})^{-1}}\right)}\,. \qquad (40)$$

First, one can bound $\tilde{\beta}_{T,n_T}(\delta)$ by (37) in Lemma E.4.

It remains to adapt the logarithmic term into a log-determinant of the desired form by conforming the vectors $\vartheta^{(t,n_t)}$. To do so, let us define the block matrices

$$Z_t := \begin{pmatrix}(\tilde{D}_t^{\lambda,n_t})^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} \qquad \text{for } t \in \mathbb{N}\,,$$

so that we may use the rank one update formula to write

$$\prod_{t=1}^{T}\left(1+\frac{1}{2C}\left\|\vartheta^{(t,n_t)}\right\|_{(\tilde{D}_t^{\lambda,n_t})^{-1}}\right) = \frac{\det\left(\mathrm{D}\Lambda_n + \sum_{t=1}^{T}\vartheta^{(t,n_t)}Z_t{\vartheta^{(t,n_t)}}^\top\right)}{\det(\mathrm{D}\Lambda_n)}\,.$$

Taking $\Lambda_n = \frac{1}{2}\left\|\cdot\right\|_2^2$ as given, we can bound the determinant of the numerator by

$$\det\left(\mathrm{D}\Lambda_n + \sum_{t=1}^{T}\vartheta^{(t,n_t)}Z_t{\vartheta^{(t,n_t)}}^\top\right) \leq \left(1+\frac{T}{n_T C^2}\right)^{n_T}$$

as in (Abbasi-Yadkori, 2012, Lemma E.3). Combining with the bound on $\tilde{\beta}_{T,n_T}(\delta)$ completes the proof. □

Having established the technical lemmata, we now turn to the regret guarantees of the varying order basis truncation version of Algorithm 2. In particular, recall Assumption 3 to give a quantification of the regularity of c^*, which in turn will allow us to tune $(n_t)_{t\in\mathbb{N}}$ to obtain the best possible regret bounds in Theorem E.8.

Theorem E.8. *Assume Assumptions 1 to 3 and $\zeta(n) = 1 - n^{-q}$ for some $q > 0$. For any $\delta \in (0,1)$, $\lambda > 0$, $\varepsilon > 0$, let $\tilde{\mathcal{A}}$ (resp. $\tilde{\mathcal{B}}$) denote Algorithm 2 with $(n_t)_{t\in\mathbb{N}} = (\lceil t^{\frac{1}{q+1}}\rceil)_{t\in\mathbb{N}}$, $\Lambda_n = \frac{1}{2}\left\|\cdot\right\|_2^2$, for all*

$n \in \mathbb{N}$, and $(\varepsilon_t)_{t\in\mathbb{N}} = (\varepsilon)_{t\in\mathbb{N}}$ (resp. $(\varepsilon_t)_{t\in\mathbb{N}} = (\alpha t^{-\alpha})_{t\in\mathbb{N}}$). For any $T \in \mathbb{N}$, the following regret bounds hold:

$$\begin{aligned}\mathscr{R}_T^{\mathscr{H},\varepsilon}(\tilde{\mathcal{A}}) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + C\left(1 + \frac{2qT^{\frac{q+2}{2q+2}}}{q+1}\right) \\ + 2C\sigma T^{\frac{q+2}{2q+2}}\left(\sqrt{2\log\left(\frac{\lambda^{-1}+2T^{q+2}C^2}{\delta}\right)} + \sqrt{\lambda}C\right)\sqrt{\log\left(1+\frac{2T^{q+2}}{C^2}\right)},\end{aligned}$$

with probability at least $1-\delta$, and

$$\begin{aligned}\mathscr{R}_T(\tilde{\mathcal{B}}) \leq \sigma\sqrt{2T\log\left(\frac{2}{\delta}\right)} + C\left(1 + \frac{2qT^{\frac{q+2}{2q+2}}}{q+1}\right) + \kappa(1+\sqrt{T}\log(T)) \\ + 2C\sigma T^{\frac{q+2}{2q+2}}\left(\sqrt{2\log\left(\frac{\lambda^{-1}+2T^{q+2}C^2}{\delta}\right)} + \sqrt{\lambda}C\right)\sqrt{\log\left(1+\frac{2T^{q+2}}{C^2}\right)},\end{aligned}$$

with probability at least $1-\delta$, in which κ depends only on (C, L, μ, ν).

Proof. The proof requires only two steps from the one of Corollary E.3. First, we bound the approximation error term. Lemma E.1 readily implies that

$$|\langle c^* - c^*|_{n_t}|\pi_t\rangle| \leq \sqrt{\sum_{k=n_t+1}^{+\infty}|\gamma_i^*|^2}\,.$$

Summing over $t \in \mathbb{N}$, one obtains

$$\sum_{t=1}^{T}|\langle c^* - c^*|_{n_t}|\pi_t\rangle| \leq \sum_{t=1}^{T}\sqrt{\sum_{i=n_t+1}^{+\infty}|\gamma_i^*|^2}\,. \tag{41}$$

By Assumption 3, for any $n \in \mathbb{N}$, we have

$$\sum_{i=n_t+1}^{\infty}|\gamma_i^*|^2 = \|c^*\|^2_{L^2(\mathbb{R}^d;\varrho)} - \sum_{i=1}^{n_t}|\gamma_i^*|^2 \leq \|c^*\|^2_{L^2(\mathbb{R}^d;\varrho)}\left(1-\zeta(n_t)\right)$$

so that for any $u > 0$, the choice $n_t := \lceil\zeta^{-1}((1-t^{-u}))\rceil = \lceil t^{\frac{u}{q}}\rceil$ $(q > 0)$ yields

$$\sqrt{\sum_{i=n_t+1}^{\infty}|\gamma_i^*|^2} \leq \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}\, t^{-\frac{u}{2}}\,. \tag{42}$$

This follows from the fact that ζ can be made a bijection of $\mathbb{R}_+ \to (0,1]$, and that ζ is increasing. Injecting (42) into (41) yields

$$\sum_{t=1}^{T}\left|\langle c^* - c^*|_{n(t)}|\pi_t\rangle\right| \leq \|c^*\|_{L^2(\mathbb{R}^d;\varrho)}\left(1 + 2\frac{T^{1-\frac{u}{2}}}{u}\right)\,.$$

The second step simply involves applying Lemma E.7 for $n_T := \lceil T^{\frac{u}{q}}\rceil \leq 2T^{u/q}$ to obtain a bound of order $\mathcal{O}(T^{\frac{1}{2}+\frac{u}{2q}})$. Setting $u = \frac{q}{q+1}$ yields the stated bounds. □

E.5 On the choice of the basis

While Assumption 3 is natural from the theoretical perspective of functional regression, in practice one needs access to the basis $(\phi_i)_{i\in\mathbb{N}}$ to perform the regression. The ability to choose a specific basis in which the coefficients decay may also appear like a questionable characterisation of regularity. In this section, we will briefly discuss existing results related to the choice of basis.

It is intuitively obvious that judiciously choosing a specific basis is only possible if c^* has some additional regularity properties one can leverage. Thus, let us begin by assuming that $c^* \in \mathsf{F}^{-1}A$ for some known but generic set $A \subset L^2(\mathbb{R}; \varrho)$. Approximation theory has studied and characterised the optimal choice of a basis to approximate elements of A, and thus the coefficient decay conditions needed in Assumption 3. We give a brief summary below, but refer to (Lorentz, 2005; Pinkus, 1985) for a more thorough introduction to this area of research.

The Kolmogorov n-width (Kolmogorov, 1991) of A in $L^2(\mathbb{R}^d; \varrho)$, is defined as

$$d_n(A) := \inf_{\substack{V_n \subset L^2(\mathbb{R}^d;\varrho) \\ \dim(V_n) \leq n}} \sup_{f \in A} \inf_{v \in V_n} \|f - v\|_{L^2(\mathbb{R}^d;\varrho)} \, . \tag{43}$$

The sequence $(d_n(A))_{n\in\mathbb{N}}$ of Kolmogorov n-width captures the best possible approximations in $L^2(\mathbb{R}^d; \varrho)$ of A by subspaces of dimension n, for every $n \in \mathbb{N}$. The study of extremal subspaces, i.e. of the minimisers in (43) for any value of $n \in \mathbb{N}$, thus typically yields a basis for A which is optimal in the sense that it minimises the approximation error. Indeed, by definition, if $(\phi_i)_{i\in\mathbb{N}}$ is a sequence of minimisers of $(d_n(A))_{n\in\mathbb{N}}$ corresponding to a nested sequence of subspaces $(V_n)_{n\in\mathbb{N}}$, then it ought to form a basis of A. Thus, expressing in the basis $(\phi_i)_{i\in\mathbb{N}}$ and by the isometry of $\ell^2(\mathbb{R})$ and $L^2(\mathbb{R}^d; \varrho)$, we have

$$\sup_{f \in A} \inf_{v \in V_n} \|f - v\|_{L^2(\mathbb{R}^d;\varrho)} = \left(\sum_{i=n}^{+\infty} |\gamma_i^*|^2 \right)^{\frac{1}{2}}$$

so that characterising the decay of the n-width and identifying extremal subspaces is sufficient to specialise Assumption 3 to a specific choice of $\mathsf{F}^{-1}A$.

In general, finding the extremal functions of the Kolmogorov n-withs is difficult, but many cases are well known. In the following, we will briefly discuss one case which concerns a well known family of regularity classes of $L^2(\mathbb{R}^d; \varrho)$: the Sobolev spaces.

To recall essential definitions (see e.g. Brézis (2011) for a comprehensive treatment), let us introduce some standard notation. Let $d \in \mathbb{N}$, any multi-index $\alpha \in \mathbb{N}^d$ defines the differential operator D^α as $\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$. Let $|\alpha| := \|\alpha\|_2$ denote the order of the multi-index. For a domain $\Omega \subset \mathbb{R}^d$, we will denote by $H^m(\Omega)$ for $m \in \mathbb{N}$ the Sobolev space containing all $L^2(\Omega; \varrho)$ functions[5] which are m-times weakly differentiable and whose derivatives of order n are also in $L^2(\Omega; \varrho)$. This space can be rewritten in several different ways, and in particular:

$$H^m(\Omega) := \left\{ f \in L^2(\Omega; \varrho) : \|\mathrm{D}^\alpha f\|_{L^2(\Omega;\varrho)} < +\infty \text{ if } |\alpha| \leq m \right\} \, . \tag{44}$$

$$= \left\{ f \in L^2(\Omega; \varrho) : \|f\|_{H^m(\Omega)} < +\infty \right\} \tag{45}$$

$$\tag{46}$$

wherein

$$\|f\|_{H^m(\Omega)} := \sqrt{\sum_{|\alpha| \leq m} \|\mathrm{D}^\alpha f\|^2_{L^2(\Omega;\varrho)}} \, ,$$

with the sum being over all multi-indices α of order at most m. In particular, one can show from (45) that $(H^m(\Omega), \|\cdot\|_{H^m(\Omega)})$ is a Hilbert space.

It is known from the work of Kolmogorov (1991) that the Kolmogorov n-width of $H^m([0, 1])$ in $L^2([0, 1])$ is of order $\mathcal{O}(n^{-m})$ asymptotically. Furthermore, he provided a characterisation of the extremal functions (and thus of the optimal basis) as the eigenfunctions of the differential operator $(-1)^m \mathrm{D}^{2m}$, or equivalently to the solutions to an ordinary differential equation of order $2m$. This formation could be extended to the multi-dimensional case, but it would require more care to set up the differential operator. This connection to the spectrum of specific operators is reflected, e.g., in (Hu et al., 2025).

Instead, we will turn to characterising under what assumptions A is a Sobolev space, and more precisely, a Sobolev *class*. The Sobolev classes are defined as

$$\{W(m, L) : (m, L) \in \mathbb{N} \times [0, +\infty)\}$$

[5] For ease of exposition, we gloss over the distinction between functions and equivalence classes here.

in which each class $W(m, L)$ is defined as the ball of radius L centred at 0 in $H^m(\Omega)$. These spaces are a standard tool for characterising the difficulty of estimation in non-parametric statistics, see e.g. (Tsybakov, 2008; Wasserman, 2006). Let us finally introduce a regularity assumption on c^* with Assumption 4.

Assumption 4. Assume that $c^* \in L^2(\mathbb{R}^d; \varrho)$ satisfies the integrability (growth) condition:

$$\int [(1 + \|z\|)^m c^*(z)]^2 \mathrm{d}\varrho(z) < M^2 ,$$

for some $M > 0$ and that $\mathrm{supp}(\mu \otimes \nu) \subset \Omega$ for some bounded open domain $\Omega \subset \mathbb{R}^d$.

Under Assumption 4, we can use the Fourier transform's effect on differentials to write

$$\|\mathrm{D}^\alpha \mathsf{F} c^*\|^2_{L^2(\mathbb{R}^d;\varrho)} \leq C^2 \int [(1 + \|z\|)^m \left|\mathsf{F}^{-1}\mathsf{F}c^*(z)\right|]^2 \mathrm{d}\varrho(z) < C^2 M^2 ,$$

for some constant $C > 0$ and for every multi-index α. Consequently, $\mathsf{F}c^* \in W(m, CM) \subset H^m(\Omega)$, as wanted. In other words, higher order integrability of the cost function c^* directly translates to membership in a Sobolev class, and thus a Sobolev space, for its Fourier transform. This establishes Corollary E.9.

Corollary E.9. *Under Assumptions 1 to 3 and 4, the regret bounds of Theorem E.8 hold* $q = m$.

In contrast, one can also derive bases specialised to regularity conditions of the marginals rather than the cost function. Suppose that the reference measure is taken as $\varrho = \mu \otimes \nu$ and that this is a Gaussian measure (the standard one, for the sake of simplicity), then one might naturally consider the Hermite Polynomials as basis for $L^2(\mathbb{R}^d; \varrho)$. For a multi-index $\alpha \in \mathbb{N}^d$, the Hermite function of order α is defined as the product

$$H_\alpha := \prod_{i=1}^{d} h_{\alpha_i} ,$$

of the one-dimensional Hermite functions

$$h_k : x \in \mathbb{R} \mapsto \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{(-1)^j}{2^j j!(k-2j)!} x^{k-2j} \text{ for any } k \in \mathbb{N} .$$

It is easily shown that this is an orthonormal basis of $L^2(\mathbb{R}^d; \varrho)$. The decay of coefficients in this basis remains an *a priori* nebulous question but it can be related to similar arguments about cylindrical ellipsoids in $L^2(\mathbb{R}^d; \varrho)$ as in Kolmogorov (1991). This time, the decay rate will depend on the eigenvalues of the Ornstein-Uhlenbeck operator L, whose action is

$$Lf = \Delta f - \mathrm{Id} \cdot \nabla f ,$$

on smooth functions, for whom Hermite polynomials are eigenfunctions.

E.6 Tikhonov regularisation and RKHS theory

In this section, we will assume that $\Lambda = \frac{1}{2} \|\cdot\|_2^2$ for simplicity. In general any increasing positive function of $\|\cdot\|_2$ will suffice to use the representer theorem as per our argument. Suppose we are given $(\mathfrak{H}, K)$ a Reproducing Kernel Hilbert Space[6] (RKHS) such that $\mathfrak{H} \subset L^2(\mathbb{R}^d; \varrho)$. We may specialise the RLS estimator (see Proposition C.1) to this case by noting that $M_t := (K(a_0, \cdot), \ldots, K(a_{t-1}, \cdot))^\top$.

By the representer theorem, at any step $t \in \mathbb{N}$,the solution to the regularised least squares problem in $\mathfrak{H}$ is given by

$$\hat{f}_t^\lambda = \sum_{i=0}^{t-1} \upsilon_i K(a_i, \cdot) ,$$

[6] Understood, of course, up to the identifications necessary for the RKHS to be a space of functions. Recall that $L^2(\mathbb{R}^d; \varrho)$ is *not* an RKHS due to a subtlety of this nature.

for some $(\upsilon_i)_{i=0}^{t-1} \in \mathbb{R}^t$. The problem can therefore be reduced to the finite dimensional optimisation problem

$$\min_{\upsilon \in \mathbb{R}^t} \left\| \vec{C}_t - K_t \upsilon \right\|_2^2 + \lambda \upsilon^\top K_t \upsilon \,,$$

in which $K_t := [K(a_i, a_j)]_{i,j} \in \mathbb{R}^{t \times t}$ is the kernel (Grammian) matrix. The rest of the standard developments follow, and one arrives at the approximation bound

$$\sum_{t=1}^{T} \langle c^* - \tilde{c}_t | \tilde{\pi}_t \rangle \leq 2C\beta_T(\delta) \sqrt{2T \log\det\left(I + (\lambda C)^{-1} K_T\right)}$$

by Lemma C.3, and corresponding regret bounds easily follows via Theorems 4.1 and D.1. From here, one can easily recover bounds *ad hoc* or by following the general methodology of Appendix E.1.

One of the main benefits of kernel methods is that they can be used to learn in infinite-dimensional spaces efficiently. While they are inherently efficient thanks to the kernel trick, works in this field have suggested further efficiency refinements such as Takemori and Sato (2021) which uses approximation theory to reduce learning in an RKHS to a finite-dimensional approximation on a well chosen basis. This resembles the methodology used above, further developments in this direction appear an interesting avenue for research.

F Miscellaneous lemmas and proofs

F.1 Sub-Gaussian Analysis

Definition 2. A random variable $\xi : \Omega \to \mathbb{R}$ is σ^2-sub-Gaussian if

$$\mathbb{E}\left[\exp\left(t\xi\right)\right] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right) \qquad \text{for any } t \in \mathbb{R}\,.$$

A stochastic process $(\xi_i)_{i \in \mathbb{N}} : \Omega \to \mathbb{R}^{\mathbb{N}}$ is σ^2-conditionally sub-Gaussian if

$$\mathbb{E}\left[\exp\left(t\xi_i\right) \middle| \sigma\left((\xi_j)_{j<i}\right)\right] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right) \qquad \text{for all } i \in \mathbb{N} \text{ and any } t \in \mathbb{R}\,.$$

Lemma F.1. *Let $(\xi_i)_{i \in \mathbb{N}}$ be a σ^2-conditionally sub-Gaussian process,*

$$\mathbb{P}\left(\sum_{i=1}^{n} \xi_i \geq \sigma \sqrt{2n \log\left(\frac{1}{\delta}\right)}\right) \leq \delta \quad \textit{for any } (n, \delta) \in \mathbb{N} \times (0,1).$$

Proof. The proof follows Chernoff's method, by exponentiating $\sum_{i=1}^{n} \xi_i$ using $x \mapsto e^{tx}$, applying Markov's inequality, the tower rule accompanied by conditional sub-Gaussianity, and finally optimising the bound over the parameter $t > 0$. □

F.2 A common summation identity

Lemma F.2. *For $\alpha \in (0,1)$, let $\phi : u \in (0, +\infty) \to \alpha u^{-\alpha} \log(u) \in \mathbb{R}_+^*$, then for any $N \in \mathbb{N}$,*

$$\sum_{u=1}^{N} \phi(u) \leq \frac{\alpha}{1-\alpha} N^{1-\alpha} \log(N) + \frac{\alpha}{2^\alpha} \log(6)\,.$$

In particular, if $\alpha = 1/2$, then

$$\sum_{u=1}^{N} \phi(u) \leq \sqrt{N} \log(N) + \frac{1}{2\sqrt{2}} \log(6)\,.$$

Proof. Notice that ϕ is differentiable, with $\phi'(u) = \alpha u^{-(1+\alpha)}(1 - \alpha \log(u))$, so that it is decreasing on $(e^{1/\alpha}, +\infty)$. Since $\sup_{\alpha > 1} e^{1/\alpha} = e < 3$, comparison between the sum and the integral of ϕ yields

$$\sum_{u=1}^{N} \phi(u) \leq \phi(1) + \phi(2) + \phi(3) + \int_3^N \phi(u) \mathrm{d}u\,.$$

The remaining integral can be computed by parts, for $(a,b) \in \mathbb{R}_+^2$, $a < b$,

$$\begin{aligned}\int_a^b \phi(u)\mathrm{d}u &= \frac{\alpha}{1-\alpha}\left(\left[u^{1-\alpha}\log(u)\right]_a^b - \int_a^b u^{-\alpha}\mathrm{d}u\right)\\ &= \frac{\alpha}{1-\alpha}\left(\left[u^{1-\alpha}\left(\log(u) - \frac{1}{\alpha-1}\right)\right]_a^b\right)\\ &\leq \frac{\alpha}{1-\alpha}b^{1-\alpha}\log(b)\end{aligned}$$

for every $\alpha \in (0,1)$. Computing yields $\phi(1) = 0$, $\phi(2) = \alpha 2^{-\alpha}\log(2)$, and $\phi(3) = \alpha 3^{-\alpha}\log(3)$, so that $\phi(1) + \phi(2) + \phi(3) \leq \alpha 2^{-\alpha}\log(6)$. Combining the results yields the desired inequality. □

G Discussion of some open problems

G.1 Practical computation of actions and action-set violations

In Algorithms 1 and 2 we used a black-box solver for an entropic optimal transport problem. This is a computational abstraction and not implementable in practice. Implementing a computationally feasible resolution raises several questions.

G.1.1 Numerical resolution of the Kantorovich problem

Sinkhorn's algorithm is the standard method for solving entropic optimal transport problems. It relies on the dual formulation of the entropic problem, that is

$$\text{Ent.}(\mu,\nu,c,\varepsilon) = \sup_{\substack{\varphi\in L^1(\mu)\\ \psi\in L^1(\nu)\\ \varphi\oplus\psi\leq c}}\left\{\int\varphi\mathrm{d}\mu + \int\psi\mathrm{d}\nu - \varepsilon\int e^{\varepsilon^{-1}(\varphi+\psi-c)}\mathrm{d}(\mu\otimes\nu) + \varepsilon\right\}$$

in the case $c \in L^1(\mu\otimes\nu)$, see e.g. (Nutz, 2022, Thm. 4.7). The solution of the dual problem is given by the pair (φ^*, ψ^*) which satisfies the Schrödinger system

$$\begin{aligned}\varphi^* &= -\varepsilon\log\left(\int e^{\frac{\psi^*(y)-c(\cdot,y)}{\varepsilon}}\mathrm{d}\nu(y)\right) \qquad \mu\text{-a.s.}\\ \psi^* &= -\varepsilon\log\left(\int e^{\frac{\varphi^*(x)-c(x,\cdot)}{\varepsilon}}\mathrm{d}\mu(x)\right) \qquad \nu\text{-a.s.}\,.\end{aligned}$$

Sinkhorn's algorithm (Sinkhorn and Knopp, 1967), in its application to this problem (Cuturi, 2013), is a fixed-point iteration which improves one potential at a time. In other words, for $n \in \mathbb{N}$, it computes

$$\varphi_{2n+1} = -\varepsilon\log\left(\int e^{\frac{\psi_{2n}(y)-c(\cdot,y)}{\varepsilon}}\mathrm{d}\nu(y)\right)$$

and

$$\psi_{2n} = -\varepsilon\log\left(\int e^{\frac{\varphi_{2n-1}(x)-c(x,\cdot)}{\varepsilon}}\mathrm{d}\mu(x)\right)\,.$$

A primal solution to Ent.(μ,ν,c,ε) can be recovered from the optimal dual potentials (φ^*,ψ^*) via

$$\mathrm{d}\pi^* = e^{\frac{\varphi^*\oplus\psi^*-c}{\varepsilon}}\mathrm{d}[\mu\otimes\nu]\,,$$

in which $\varphi^*\oplus\psi^* : (x,y)\mapsto\varphi^*(x)+\psi^*(y)$. Through an analogue for $(\varphi_{2n+1},\psi_{2n})_{n\in\mathbb{N}}$, we can obtain iterates $(\varpi_n)_{n\in\mathbb{N}}$.

Lemma G.1 ((Eckstein and Nutz, 2022, Thm. 3.15)). *If c is Lipschitz on supp$(\mu)\times$ supp(ν), and μ,ν are sub-Gaussian measures, then the iterates $\{\varpi_n(c)\}_{n\in\mathbb{N}}$ of Sinkhorn's algorithm satisfy*

$$\Psi^{\varepsilon}_{\mu,\nu}(c^*,\varpi_n(c)) - \mathit{Ent.}(\mu,\nu,c,\varepsilon) \leq C_0\varepsilon n^{-\frac{1}{4}}\,,$$

for every $\varepsilon > 0$, in which C_0 is a numerical constant independent of n.

We omit the explicit dependencies in the constant C_0 as they are quite technical and require parsing a large part of Eckstein and Nutz (2022), which proceeds from within a highly general framework. We should note, however, that consequently their bound is valid under much weaker assumptions than the ones stated here, and that the rate can, in fact, be improved if c has sub-linear growth.

Unfortunately for regret minimisation, ϖ_n need not be a transport plan in $\Pi(\mu, \nu)$, meaning it is not a valid action. Removing the requirement that $\pi_t \in \Pi(\mu, \nu)$ entirely would render the problem meaningless, as the regret can be made negative by finding a single point such that $c(x, y) <$ Kant.(μ, ν, c), and playing $\delta_{(x,y)}$.

As an auxiliary remark, this problem is one of the main hurdle to adapting Algorithm 1 to unknown marginals, as there would be no conceivable way to pick valid transport plans, which renders the analysis a non-starter.

G.1.2 On action violations

Two possible directions appear to resolve this issue: one at the level of bandit design, and one at the level of numerical optimal transport. The former revolves around the idea of incorporating action-set violations to regret analysis, the latter around the idea of modifying Sinkhorn's algorithm to produce valid primal iterates at each step, e.g. by projecting onto $\Pi(\mu, \nu)$.

The question of violating action sets has been posed before in Bandit Theory and has also arisen in practical use-cases in Reinforcement Learning, see (Seurin et al., 2020). It is a staple topic in the context of fairness, see e.g. (Joseph et al., 2016) and of contextual bandits (including linear stochastic bandits) in which various other types constraint have also been considered, see e.g. (Liu et al., 2024). These types of constraints typically, in effect, disable certain arms at certain times, a generic setting which has been considered as well, e.g. by Kleinberg et al. (2010); Abensur et al. (2019).

These works adopt a range of strategies to formulate the problem in a meaningful way, but their perspectives don't really fit with the real challenge we have with the OT problem. The problem isn't so much that the constraints placed on the action set are complicated: $\Pi(\mu, \nu)$ is a convex, compact set defined by linear inequalities. The problem arises entirely from the facts that $\Pi(\mu, \nu)$ is infinite-dimensional, and that it is a subspace of $\mathscr{P}(\mathcal{X})$, whose geometry is far from straightforward.

A preliminary exploration of this topic would likely require a taxonomy of the different possible violations of $\Pi(\mu, \nu)$. Indeed, π_t could violate one or both marginal constraints, or it could even fail to be a probability measure through the total mass or positivity conditions. It appears likely that these will have quite different impacts both on the problem's geometry and on practical usefulness. Thereafter, one might consider whether guaranteeing finitely many violations, as Liu et al. (2024) do, or developing a penalised regret is more appropriate.

The alternative would be to design an algorithm which optimises the entropic or Kantorovich problems through while staying within the constraint set $\Pi(\mu, \nu)$ (either for all time, or once it reaches a desired precision). On the one hand, there are finite-dimensional intuitions for this to work as Sinkhorn's algorithm can be viewed as a form of gradient descent (Léger, 2021), which could be projected onto $\Pi(\mu, \nu)$ (which is convex and compact). On the other hand, the geometry of $\Pi(\mu, \nu)$ as an infinite-dimensional probability space is likely to make rigourously doing so (and deriving convergence rates) quite arduous work.

G.2 Extensions to the Monge problem

The *Monge* optimal transport problem associated to (μ, ν, c) is

$$\text{Monge}(\mu, \nu, c) := \inf_{T \in \mathscr{T}} \int c(x, T(x)) \mathrm{d}\mu(x) \,, \tag{47}$$

in which $\mathscr{T}$ is the set of all μ-measurable maps $T : \mathcal{M}_\mu \to \mathcal{M}_\nu$ such that $\mu(T^{-1}(\cdot)) = \nu$. Chronologically, this is in fact the original formulation of the OT problem (Monge, 1781).

The Monge problem is best approached through finite-dimensional practical applications such as *matchings* of students to universities, employees to employers, etc. The requirement that the map T be a function imposes an *indivisibility* of the mass T moves from μ to ν (i.e. one university per student). This makes the resolution of the problem much more difficult. For example, if μ and ν

each have two atoms with weights $(1/2, 1/2)$ and $(1/3, 2/3)$ respectively, then $\mathscr{T} = \emptyset$, meaning $\text{Monge}(\mu, \nu, \cdot) \equiv +\infty$, and the problem is never solvable.

If μ, ν are non-atomic, $\text{Monge}(\mu, \nu, c)$ can be interpreted as the cheapest way (w.r.t. c) to transport a μ-shaped pile of infinitesimally small things into a ν-shaped one, but its geometry remains complicated. The Kantorovich relaxation drastically simplified the geometry of the problem and remains one of the most effective tools to approach the Monge problem, which is why it is accepted as the standard in modern OT theory.

Note that the relaxation from $\text{Monge}(\mu, \nu, c)$ to $\text{Kant.}(\mu, \nu, c)$ is known to be exact in some cases, such as $c = \|\cdot - \cdot\|^2 / 2$ with $\mathcal{M}_\mu = \mathcal{M}_\nu = \mathbb{R}^d$, (μ, ν) having second-order moments and μ being absolutely continuous w.r.t. the Lebesgue measure (Ambrosio et al., 2021, Thm. 5.2). See also (Villani, 2009, Thm. 5.30) for weaker conditions. But it is also known (e.g. via the above example) that this relaxation is not without loss.

If we want to learn a Monge problem, we must, of course, make sufficient assumptions for it to be solvable, but more importantly we must face the issue that (47) is now a non-linear functional and that $\mathscr{T}$ is not as docile a set as $\Pi(\mu, \nu)$. Here, the recent work in statistical optimal transport on learning Monge maps (i.e. the solutions to (47)) is highly relevant, see e.g. (Chewi et al., 2024, Ch. 3) or the paragraph in Appendix H below. Though once again most work focuses on the batch sampling of marginals, not on online learning. This line of work would appear to also require more general results about the learning of minima of non-linear functionals, which are not yet available in the literature. Overall, it remains unclear if the Monge problem is on a similar or different level of difficulty to the Kantorovich problem as it is not clear that the techniques to reduce to online least-squares we used will transfer.

Beyond these statistical issues, one should also expect the problems of effective optimisation from Appendix G.1 to return with a vengeance as the Monge problem is a fully non-linear problem unlike the Kantorovich problem which is an (infinite-dimensional) linear program.

H Bibliographical complements on statistical optimal transport

An excellent detailed history of the development of OT as a mathematical theory, replete with bibliographical notes, can be found in (Villani, 2003, Ch. 3). Summarising this field's venerable history further would be of little value. Instead, we will expand on relevant research specifically about *learning* optimal transport problems. We touch on key aspects of the literature below, and refer to the forthcoming book Chewi et al. (2024), for a deeper longitudinal overview.

Estimation of OT functionals Much of the early work in statistical OT focused on estimating the value of the functional $\text{Kant.}(\mu, \nu, c)$ when (μ, ν) are unknown, but c is known and highly regular, e.g. (Horowitz and Karandikar, 1994; Weed and Bach, 2019). These regularity assumptions are motivated by the study of Wasserstein distances between probability measures (i.e. $c = \|\cdot - \cdot\|^p$, $p \geq 1$) via sampling. With the increased interest in the entropic OT problem, many works have asked the same questions about $\text{Ent.}(\mu, \nu, c, \varepsilon)$, e.g. (Rigollet and Stromme, 2022; Stromme, 2024).

This line of work is orthogonal to our investigation, as we know (μ, ν) but not c^*. The critical object in this line of work is the regularity structure of $\text{Kant.}(\mu, \cdot, c)$, when c is strongly regular. For our problem, the relevant geometry is that of the transport functional $\pi \in \Pi(\mu, \nu) \mapsto \langle c^* | \pi \rangle$.

Online matchings Concurrently, Matching (discrete marginal OT), has been actively studied by computer scientists and economists. These works, such as (Perrot et al., 2016), are often directly inspired by applications, and have yielded many creative extensions to the OT problem: Alon et al. (2004) aims to learn an optimal matching using queries to an oracle; Johari et al. (2021) to identify *types* of nodes; Min et al. (2022) to design a welfare-maximising social planner; etc.

The common thread amongst these works is the nature of the *market* on which they work: at each time t, a new supply becomes available to *match* (i.e. transport from), and the agent must decide to which of its available demands to transport it. This decision problem is fundamentally different from our repeated OT problem as mistakes in the matching are permanent, while we replay a whole matching at each step. Furthermore, the information structure is different. Jagadeesan et al. (2021); Sentenac et al. (2021); Sentenac (2023) (amongst others) have highlighted that this problem is a combinatorial semi-bandit problem, in which there is feedback about each connection made. In our problem the agent receives feedback only about the matching as a whole (full bandit). These two differences make the problems seem superficially similar, but they are fundamentally different.

Estimation of Wasserstein distances One of the most important contributions of optimal transport is a family of useful distances between probability measures: the Wasserstein metrics. The study of these distances has allowed major progress on the geometry of spaces of probability measures, and has been used in many applications. It is therefore natural that the estimation of these distances has been a major topic of interest in the learning of optimal transport.

The key question here is the convergence in Wasserstein distance of an empirical distribution to the true distribution. Pioneering work on this topic began in the 80s and 90s, see (Ajtai et al., 1984; Talagrand, 1994), with the study of *Matching* (i.e. discrete optimal transport). Key statistical analysis of this problem includes finite sample bounds, see (Horowitz and Karandikar, 1994) and more recently (Fournier and Guillin, 2015; Weed and Bach, 2019) among others, as well as distributional limits, see e.g. (Tameling et al., 2019) and references therein.

Sadly, most work has remained limited to Wasserstein distances rather than generic cost functions, owing to a reliance on the pleasant geometric properties that they enjoy.

Estimation of Entropic OT Motivated by the success of Entropic OT in designing numerical solution to OT problems, see (Cuturi, 2013), work on the Entropic problem has focused on estimating Ent.$(\mu, \nu, c, \varepsilon)$ using Ent.$(\hat{\mu}_n, \hat{\nu}_n, c, \varepsilon)$, for empirical measures $(\hat{\mu}_n, \hat{\nu}_n)$. This has often gone together with estimation for the Schrödinger potentials (φ, ψ) of (7).

While this is very much the same type of study as for the Kantorovich problem in Wasserstein metrics, it should be noted that the entropic problem exhibits qualitatively different behaviour. While learning the Kantorovich problem exhibits a curse of dimensionality, the entropic problem exhibits parametric-rate (dimension-free) convergence, as shown by Genevay et al. (2019); Rigollet and Stromme (2022). This was tempered by large dependencies in other problem quantities, which were reduced over time (Stromme, 2024) and were complemented by distributional limits, see e.g. (Gonzalez-Sanz et al., 2024).

Estimation of Monge maps While the estimation of Wasserstein distances is mostly motivated by statistical applications, the estimation of Monge maps is motived by effectively solving transport problems in an applied context. Here, one sees samples from two marginals μ and ν, and attempts to estimate T^* the minimiser of (47).

There has been a significant amount of machine learning and statistics literature on this topic, following on from (Hütter and Rigollet, 2021; Gunsilius, 2022). Various types of estimators have been constructed, either derived from optimal transport theory (Hütter and Rigollet, 2021), or from plug-in estimates using classical machine learning methods such as k-NN (Manole et al., 2024; Deb et al., 2021).

Optimal transport applied to learning While these bibliographical notes concern learning in optimal transport let us conclude by underline that the machine learning community has used optimal transport to impressive success in applications. One could highlight in particular Wassertein GANs (Arjovsky et al., 2017) and subsequent works, e.g. (Salimans et al., 2018) as well as the field of domain adaptation (Courty et al., 2017; Torres et al., 2021)