Feature-Based Dynamic Matching

Yilun Chen [*] Yash Kanoria [†] Akshit Kumar [‡] Wenxin Zhang [§]

Abstract

Motivated by matching platforms that match agents in a centralized manner, we study dynamic two-sided matching in a setting where both customers (demand) and service providers (supply) are heterogeneous and the pool of service providers is limited. We model heterogeneity on the two sides of the market by demand weight vectors drawn *i.i.d.* from some distribution, and supply feature vectors drawn *i.i.d.* from a (possibly) different distribution. The matching of a demand-supply pair generates a matching value that depends on their weight and feature vectors. We adopt a notion of regret, specifically the additive loss relative to the value (per match) achievable in the limiting hindsight optimum as our performance metric for matching policies. Simple myopic policies suffer non-vanishing $\Omega(1)$ regret in large markets. We propose a forward-looking supply-aware policy dubbed Simulate-Optimize-Assign-Repeat (SOAR) which balances between producing high match value for the current match and preserving valuable supply for future customers. We prove that SOAR achieves the optimal regret scaling under different assumptions on the demand and supply distributions. En-route to proving our guarantees we develop a novel framework for analyzing the performance of our SOAR policy which may be of broader interest. As a corollary of our techniques, we also resolve an open problem posed in Kanoria [2022].

Keywords: matching markets, dynamic matching, simulation-based policies, regret analysis

History: An extended abstract of this paper appeared in the Proceedings of 24th ACM Conference on Economics and Computation, EC'23

[*] The Chinese University of Hong Kong, Shenzhen Email: chenyilun@cuhk.edu.cn

[†] Columbia University, Graduate School of Business. Email: ykanoria@columbia.edu

[‡] Columbia University, Graduate School of Business. Email: ak4599@columbia.edu

[§] Columbia University, Graduate School of Business. Email: wx2574@columbia.edu

1 Introduction

Centralized matching platforms have transformed the way customers access services in sectors such as hospitality, transportation, and finding a job. These platforms facilitate on-demand connections between customers and service providers, based on customer needs and service provider attributes and availability. A common feature of many of these platforms is that the demand and supply sides are often *both* highly differentiated. For instance, on online home services platforms such as Handy.com, customers arrive and specify various service requests (e.g., plumbing, room cleaning, furniture repair) and personalized service preferences (location, time, price). Simultaneously, service providers on these platforms are differentiated by their diverse skill sets, locations, and availability schedules. The value of a match between a given customer and a service provider depends on both the specific preferences of the customer and the attributes of the service provider. Platforms are tasked with maximizing the overall matching value of customers, a composite measure that incorporates customer satisfaction, service provider value, and the platform's value, for those arriving over a finite horizon.

The varied nature of both demand and supply poses challenges to the platform's decision-making, especially when supply is constrained. With a limited pool of service providers and the arrival of customers over time, the platform faces a trade-off between (a) optimizing the value of the current match and (b) maintaining a diverse pool of service providers for future customers. While prioritizing (a) offers maximum immediate value, implementing (b) may bring the benefit of facilitating higher valued matches for future customers. Balancing these two objectives to ensure optimal outcomes is, in general, not straightforward. This motivates our central research question:

In a dynamic two-sided matching setting with heterogeneous customer and service providers, how should a centralized platform match customers arriving over time to maximize the overall value generated?

We capture the platform's problem through a *feature-based* dynamic matching model. In our base model, a pool of n service providers is initially available and n customer requests arrive sequentially. Upon the arrival of each customer request, the platform immediately and irrevocably assigns to it an available service provider, who then leaves the pool permanently. In our model, customers and service providers are characterized by multidimensional vectors, where each dimension numerically encodes an attribute observable to the platform, for instance, job location, listed prices, ratings, expertise (for service providers), and demographics in general. We capture market uncertainty and heterogeneity by assuming the demand weight vectors (that represent customers) and supply feature vectors (that represent service providers) are drawn *i.i.d.* from some demand distribution P and supply distribution Q, respectively. The match quality (or match value) between a customer and a service provider is modelled by a quality function $\varphi(X, Y)$ that depends on both of their weight and feature vectors X and Y respectively, one typical example being $\varphi(X, Y) = \langle X, Y \rangle$. The platform's goal is to repeatedly make matching decisions to maximize the expected average matching quality. In this paper, we also consider the setting of scarce supply where the number of customers is more than the number of service providers. We show that our modelling framework is flexible enough to incorporate such a setting (refer to Remark 1).

Modelling innovation. Much of the prior research on stochastic dynamic two-sided matching and resource allocation assumes a small number of demand and supply (or resource) types (Talluri et al. [2004], Bitran and Caldentey [2003], Braverman et al. [2022], Banerjee et al. [2022]). Such an assumption is critical in ensuring theoretical guarantees but leads to a limitation in modeling capability. In contrast, our feature-based modeling framework allows for many or even an infinite number of demand and supply types while remaining analytically tractable. This is achieved by exploiting proper continuity conditions such as a spatial structure on the feature vector spaces and the matching quality functions. In particular, we obtain algorithms that provide provably asymptotically optimal performance guarantees, i.e., our performance loss relative to the limiting hindsight optimal solution vanishes as n, the number of supply (demand) units, increases.

Within the modeling framework, we study and quantitatively compare different matching policies. While it may be tempting for practitioners to greedily assign service providers that maximize immediate matching quality, we observe that such myopic policies may end up incurring significant quality loss, and are thus highly sub-optimal (see Proposition 1, Proposition B.3 and Conjecture 1). The sub-optimality stems from overlooking future customers and exhausting a particular group of similar service providers too early. This observation necessitates the design of forward-looking algorithms.

1.1 Main Contribution

In this work, we develop a simple forward-looking algorithm dubbed SOAR, which automatically preserves supply diversity and achieves near-optimal performance within our model across a variety of settings. We also develop an analytical framework for our algorithm SOAR which in turn enables us to prove near-optimal performance guarantees and may be of broader interest. As a corollary of our techniques, we also resolve some open problems posed in Kanoria [2022]. We now elaborate on our contributions.

(*i*) *A simple forward-looking algorithm.* We propose a principled approach dubbed SOAR (short for Simulate-Optimize-Assign-Repeat), that combines real-world applicability and a theoretical performance guarantee. SOAR is inspired by model predictive control (MPC), a well-known heuristic in dynamic control theory (see e.g. Bertsekas [2012], Camacho and Alba [2013] for detailed discussion). The key idea is to utilize a simulated scenario of future demands to facilitate efficient online decision-making. More precisely, at each decision epoch (namely, when a new customer arrives), the algorithm forms a projected demand pool, which includes both the currently arriving customer and a simulated stream of future customers. It then matches the arriving customer as per their assignment under the optimal offline matching between the projected demand pool and the remaining supply pool. The two main steps of SOAR at each decision epoch are (*i*) simulating a future demand stream and (*ii*) solving an offline assignment problem, both easy to implement. Regarding (*i*), we note that many companies collect and store large amounts of demand data. This rich data can be leveraged to build high-fidelity simulators for SOAR. Furthermore, our algorithm only requires simulating one sample path of future customers in each decision epoch, setting it apart from most simulation-based algorithms which fundamentally rely on sample average approximation (SAA) and require simulating a large number of independent streams of future customers. Regarding (*ii*), the bipartite assignment problem is a classical computational problem. Theoretically, a worst-case guarantee of $\mathcal{O}(n^{2+o(1)})$ has been shown for this problem [Chen et al., 2022], ensuring its tractability. In practice, several fast and highly scalable solvers have been developed for this problem that can be directly plugged into SOAR.

(*ii*) *Novel analytical framework for* SOAR*.* We note that MPC is generally considered to be a sub-optimal heuristic (see Section 6.5 of Bertsekas [2012]). Our main technical contribution lies in developing a novel framework that allows us to systematically analyze and establish the near-optimal performance of SOAR for stochastic feature-based dynamic matching. Serving as the technical backbone of this work, our framework relies on a key observation (Theorem 1), which expresses the performance of SOAR as the average of a sequence of hindsight optimum values. In particular, when there are n supply providers and customers, let $U_n(\mathsf{SOAR})$ denote the expected average matching quality under SOAR and U_n^{H} denote the expected average matching quality achievable in hindsight. Then Theorem 1 asserts that $U_n(\mathsf{SOAR}) = \frac{1}{n}\sum_{k=1}^{n} U_k^{\mathsf{H}}$. Equivalently in terms of regret, we have $\textsc{Reg}_n(\mathsf{SOAR}) = U_\infty - U_n(\mathsf{SOAR}) = \frac{1}{n}\sum_{k=1}^{n}(U_\infty - U_k^{\mathsf{H}})$, where U_∞ denotes the thick market limit (refer to Section 2). This result leverages the symmetry induced by *i.i.d.* random variables, and is otherwise completely general; it holds for arbitrary demand and supply distributions P, Q, and for arbitrary matching quality function $\varphi(X, Y)$ that is bounded over the support of X and Y (boundedness is required only to ensure these performance metrics are well-defined). It reduces bounding the regret of the *online* algorithm SOAR to characterizing a sequence of bounds on the regret under the hindsight relaxation for $k = 1, 2, \ldots, n$, which are purely *offline*. We leverage this framework to prove near-optimal regret guarantees for different quality functions and under general structural assumptions on the demand and supply distributions.

(iii) *Near-optimal performance of* SOAR. SOAR enjoys a guarantee of near-optimal performance across broad classes of demand and supply distributions and quality functions. Our general performance guarantee, Corollary 2, states that SOAR is near optimal for any matching problem satisfying a "regular scaling" property (Definition 1), which we believe essentially incorporates all non-pathological matching instances. In particular, we demonstrate that under a specific performance measure, namely, the average matching regret relative to the hindsight limit matching value, the performance of SOAR is within a $\log n$ factor of that of the hindsight optimum for any matching instance which scales regularly. We then explicitly characterize the scaling of average matching regret incurred by SOAR in various interesting classes of matching problems. In Section 4, we investigate matching for a general class of quality functions $\varphi_p(X, Y) = -\|X - Y\|^p$ ($\|\cdot\|$ denotes Euclidean norm) for $p \geq 1$ under two sets of assumptions on the demand and supply distributions, one restricting P and Q to

be the uniform distribution over $[0,1]^d$, following the literature on dynamic spatial matching (Akbarpour et al. [2021], Kanoria [2022], Balkanksi et al. [2022]), and the other allowing for arbitrary distributions.

We characterize the regret scaling attained by SOAR under the two sets of assumptions respectively, which are both optimal (up to a factor of at most $\log n$) for each possible dimension d and exponent p (see Theorem 2). In particular, sharper regret scaling is achievable under the more restrictive assumption that both distributions P and Q are uniform over $[0,1]^d$. In the same section, we also study the practically relevant dot-product matching quality function $\langle X, Y\rangle$. In particular, we establish an equivalence between the dot-product quality function and φ_p with $p = 2$, and prove in this special case that SOAR attains the sharper regret scaling for a general class of smooth, regular, "uniform-like" distributions P and Q that can be non-uniform and unequal (see Corollary 3, Theorem 3, and Table 1). The matching upper and lower bounds in Table 1 together characterize the complete landscape of the regret scaling with dot product quality.

As a corollary of our analysis, we also solve an open problem in Kanoria [2022] and generalize some of the results in Kanoria [2022] to a setting with quality function $\varphi(X,Y) = -\|X-Y\|^p$ for $p > 1$. Our results significantly advance the previous understanding of dynamic spatial matching, which is mostly restricted to the special case with $P = Q = \text{Uniform}([0,1]^d)$ and $p = 1$, [Kanoria, 2022, Balkanksi et al., 2022, Akbarpour et al., 2021].

Our results reveal several interesting insights as we summarize below. First, the "cost of matching" dominates the "cost of uncertainty about the future" in our setting. In particular, SOAR attains the same regret scaling (up to logarithmic factors) as the hindsight optimal matching in all cases, indicating that knowing the future demand in advance does not allow us to substantially reduce regret. Second, the regret of SOAR increases as the dimension of vectors, d, increases. Since d corresponds to the level of heterogeneity on both sides of the market in our model, this observation can be interpreted as matching is harder in a market with more dimensions of heterogeneity. Finally, smoothness helps reduce regret, but only in low dimensions. For $d \geq 4$ the regret scaling achievable is the same under smooth versus arbitrary distributions. Section 5 describes a range of numerical studies, which provide evidence that SOAR is near optimal and exhibits regret which vanishes rapidly with n in various settings.

Table 1: The average regret scaling for $\varphi(X,Y) = \langle X, Y \rangle$. Here $a \wedge b := \min\{a, b\}$.

	$P = Q = \text{Uniform}([0,1]^d)$	P, Q Smooth	P, Q Arbitrary
Lower Bound	$\tilde{\Omega}(n^{-(\frac{2}{d}\wedge 1)})$	$\tilde{\Omega}(n^{-(\frac{2}{d}\wedge 1)})$	$\tilde{\Omega}(n^{-(\frac{2}{d}\wedge \frac{1}{2})})$
Algorithm of Kanoria [2022]	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d}\wedge \frac{1}{2})})$	-	-
SOAR (this work)	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d}\wedge 1)})$	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d}\wedge 1)})$	$\tilde{\mathcal{O}}(n^{-(\frac{2}{d}\wedge \frac{1}{2})})$

1.2 Related Literature

Matching Markets. Devising algorithms and guarantees for online matching is a central topic for researchers at the intersection of CS, Economics, and OR/OM; see, e.g., Aouad and Ma [2022], Delong et al. [2022], Braverman et al. [2022], Udwani [2021], Papadimitriou et al. [2021], Ezra et al. [2020], Ma and Simchi-Levi [2020], Ashlagi et al. [2019] among many others. In recent years amid the rise of the platform economy, matching markets, with a rich set of newly emerging economical/operational/computational challenges, have attracted significant attention in the academic literature. [Aouad and Saban, 2022, Ashlagi et al., 2022, Shi, 2022, Manshadi et al., 2022, Derakhshan et al., 2022, Feng and Niazadeh, 2022, Immorlica et al., 2021, Aouad and Saritaç, 2020]. In this paper, we study a centralized online platform that aims at maximizing social welfare (expected total matching quality). Like us, several works consider a reward-maximizing platform, where the reward can be total revenue, total match number, etc. Aouad and Saritaç [2020] studies a centralized matching problem in a stochastic environment with departures and formulate it as an MDP. Manshadi et al. [2022] and Immorlica et al. [2021] investigate a decentralized platform, and how to choose a good market equilibrium through appropriately designed match recommendations. Shi [2022] conducts an insightful study revealing *why* platforms in the home services industries tend to implement centralized matching policies, complementary to our work.

Dynamic Resource Allocation. We formulate and solve an online matching problem with *i.i.d.* demand and supply, which is categorized as a dynamic resource allocation problem in the OR/OM literature. The $d = 1$ special case of the model has appeared in earlier works in the

OR/OM community under the name of (stochastic) sequential assignment problem (Derman et al. [1972], Albright Jr [1972], with applications in the kidney exchange markets Su and Zenios [2005]). The workhorse policy commonly used in several classical dynamic resource allocation instances is the Certainty Equivalent (CE) policy (Balseiro et al. [2023b]). It is worth pointing out that directly applying CE in our setting requires solving a fluid problem equivalent to an optimal transport (OT) problem, which can be challenging when P and Q are supported on infinite or even uncountable sets (Taşkesen et al. [2022]). Indeed, the finite-ness of the support of demand/supply distributions is a key requirement not only to the design of CE but also to the analysis of several new algorithms recently proposed in the field (Vera and Banerjee [2021], Bumpensanti and Wang [2020]). By comparison, SOAR is a simulation-based computationally efficient proxy for CE with performance guarantees, that naturally copes with infinite types of demand and supply in the matching context. We note here that simulation-based policies have previously been proposed and studied in the context of dynamic resource allocation problems [Talluri and Van Ryzin, 1999, Kunnumkal et al., 2012, Freund and Banerjee, 2019, Sinclair et al., 2023, Ahani et al., 2023, Besbes et al., 2023]. The most relevant to our work is Besbes et al. [2023]. Subsequent to our work, in a recently revised version of Besbes et al. [2023], the authors study a multi-simulation variant of our SOAR-like policy. They use the compensated coupling framework of Vera and Banerjee [2021] to analyze hindsight-based regret, and their theoretical guarantees on the dynamic matching problem are only applicable under the setting with a few demand and supply types. Different from previous work, our performance guarantee does not rely on sample-average-approximation (SAA) type analysis, which depends on the number of simulated sample paths used. Instead, the SOAR algorithm developed in this paper mimics MPC and only uses one simulated sample path in each decision epoch. We develop a formula that characterizes the expected performance of SOAR (cf. Theorem 1), enabling us to prove optimal regret scalings in various settings for very general demand and supply type distributions.

Stochastic Online Matching. The problem studied in this paper is relevant to the literature on stochastic online matching. This literature examines the stochastic online bipartite matching problem in the known *i.i.d.* input model, initiated by Feldman et al. [2009]. Many subsequent papers have generalized the base model of Feldman et al. [2009], which assumed an unweighted underlying graph and integral arrival rates, and proposed new algorithms. In particular, Manshadi et al. [2012], Jaillet and Lu [2014], and Brubach et al. [2016] derived algorithms with improved competitive ratio guarantees and studied models that relaxed the integrality restriction on the arrival rates, allowing the arrival rates to be arbitrarily close to zero. Additionally, Haeupler et al. [2011] and Brubach et al. [2016] extended the base model to the edge-weighted setting. A key difference between our setting and the existing literature is the power of the adversary. In the stochastic online matching literature, such as Feldman et al. [2009] and its successors, the adversary can choose the supply units arbitrarily, leading to hard instances with non-vanishing regret. In our setting, the supply units are *i.i.d.* draws from a distribution, limiting the adversary's power as n grows (to "selecting" the "worst-case" distributions and the matching quality function, in particular). Consequently, our setting inherently enjoys vanishing regret, regardless of regularity assumptions like the continuity or smoothness of P, Q, φ (see Corollary 1). Subsequent to our work, Saberi et al. [2024] studies a stochastic online metric matching problem where the adversary chooses supply unit locations in a metric space.They leverage the algorithmic analysis framework developed in this paper and achieve improved competitive ratio and provably near optimal regret.

Dynamic Spatial Matching. An intimately related literature to this work is dynamic spatial matching. There the matching cost is typically chosen as the distance between two units. The stochastic version of dynamic spatial matching is a specific instance in our model, with $P = Q = \text{Uniform}([0,1]^d)$ and $\varphi(x, y) = -\|x - y\|$. Kanoria [2022] studies this problem (under the so-called "semi-dynamic setting") and gives a complete characterization of the regret scaling. Their Hierarchical Greedy algorithm achieves optimal scaling for all $d \geq 1$. However, the results and techniques do not extend to general quality $-\|x-y\|^p$ and unequal demand and supply distributions as we previously alluded to. SOAR resolves an open problem posed in Kanoria [2022] by achieving a provably tight regret scaling under $-\|x - y\|^p$ and $P = Q$; see Proposition 2. Other related works include Holden et al. [2021] which proposes a gravitational allocation method for spatial matching in both dynamic and offline settings between two sets of uniform points on a $3D$ sphere that achieves tight guarantee. Akbarpour et al. [2021], Balkanksi et al. [2022], and Kanoria [2022] all study the case of excess supply, whereas we focus on the case of scarce supply in this work. In general, randomized algorithms that are in spirit similar to MPC have been proposed to solve online matching problems (e.g., Gupta et al. [2019]), that come with different performance guarantees (e.g. competitive ratio) under various modeling assumptions. We believe that an attractive feature of our work relative to prior work is the broad generality of the conditions under which we establish near optimality of SOAR.

1.3 Notation

We denote $f(n) = \Theta(g(n))$ if there exists constant $C > 0$ independent of n such that $C^{-1}g(n) \leq f(n) \leq Cg(n)$. Similarly, we say $f(n) = \mathcal{O}(g(n))$ if there exists constant $C < \infty$ independent of n such that $f(n) \leq Cg(n)$ and we say that $f(n) = \Omega(g(n))$ if exists a constant $C > 0$ independent of n such that $f(n) \geq Cg(n)$. The notation $\tilde{\mathcal{O}}, \tilde{\Theta}, \tilde{\Omega}$ will ignore the polylogarithmic factors in n. For example $\mathcal{O}(n \log n) = \tilde{\mathcal{O}}(n)$. We denote $a \wedge b := \min\{a, b\}, a \vee b := \max\{a, b\}$.

1.4 Organization

The rest of the paper is organized as follows. In Section 2 we formally introduce the model. In Section 3.1, we discuss how myopic policies like Greedy perform badly even in simple cases. In Section 3.2, we formally define the Simulate-Optimize-Assign-Repeat (SOAR) policy, establish a meta analysis framework and present the algorithm's universal near optimality. In section 4, we characterize the near-optimal regret scaling achieved by SOAR under several different sets of assumptions. In Section 5, we present the results of numerical simulations for various settings and compare the performances of several algorithms. In Section 6, we conclude with a discussion of key takeaways and some challenging open problems.

2 Model

Let $Y_1, Y_2, \ldots, Y_n$ denote an initial endowment of n supply units, where each $Y_k \in \mathcal{Y} \subseteq \mathbb{R}^d$ is sampled *i.i.d.* from a supply feature distribution Q. A sequence of n demand units arrive sequentially over time, each seeking one supply unit. The t^{th} demand unit for $t = 1, 2, \ldots, n$ is a vector $X_t \in \mathcal{X} \subseteq \mathbb{R}^d$, which is i.i.d. drawn from a demand weight distribution P. Upon the arrival of the demand unit X_t, the platform must immediately and irrevocably assign an available supply unit Y to the current demand unit X_t. Such an assignment decision generates a feature and weight-dependent match value of $\varphi(X_t, Y)$, after which the matched pair leaves the system, depleting the available supply units by one. We refer to the function $\varphi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ as the (match) *quality* function. The above decision-making process is repeated n times, until there is no supply unit left.

The platform seeks a dynamic matching policy that maximizes the average expected match value. Let $\mathcal{H}_t$ denote the history of the system up to decision epoch t (namely, when the t^{th} customer arrives) that incorporates the first $t-1$ demand units and their corresponding matched supply units. Let $\mathcal{S}_t$ denote the set of indices for the remaining supply units when the t^{th} demand unit arrives, and $\Delta(\mathcal{S}_t)$ be the set of probability distributions over $\mathcal{S}_t$. We formally define a dynamic matching policy as a collection of mappings $\pi := (\pi_t)_{1\le t\le n}$, where $\pi_t(X_t, \mathcal{H}_t) \in \Delta(\mathcal{S}_t)$ is the possibly randomized assignment of supply unit given the t^{th} demand unit X_t and the current history $\mathcal{H}_t$. We denote the average expected match value under π by

$$U_n(\pi; P, Q, \varphi) := \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} \varphi(X_t, Y_{\pi_t})\right], \tag{1}$$

where the expectation is taken over π, P, Q, and we slightly abuse notation π_t to represent both the random variable and its associated distribution.

We do not assume the platform knows P and Q. Instead, we assume that the platform has access to m *i.i.d.* samples from P, collected from its historical matching activities. Our algorithm and the corresponding theoretical results only require a moderate m relative to the scale of the problem. In particular, $m = \Omega(n^2)$.

Benchmark. We use the limiting hindsight optimum of the problem as a benchmark to measure and compare the performance of matching policies. To define this benchmark, we first introduce the following hindsight relaxation

$$U_n^{\mathsf{H}}(P, Q, \varphi) := \frac{1}{n}\mathbb{E}\left[\sup_{\tau\in S_n} \sum_{t=1}^{n} \varphi(X_t, Y_{\tau_t})\right], \tag{2}$$

where S_n is the symmetric group on $\{1, \dots, n\}$. We refer to $U_n^{\mathsf{H}}(P, Q, \varphi)$ as the *hindsight optimum value* of the problem. In words, the hindsight optimum is achieved by relaxing the non-anticipative constraint on π and allowing access to the actual values of all arriving demand units $X_1, \dots, X_n$ at time 0. The *limiting hindsight optimum value* of the problem, denoted as $U_\infty(P, Q, \varphi)$, represents the thick market limit of the hindsight optimum: $U_\infty(P, Q, \varphi) := \lim_{n\to\infty} U_n^{\mathsf{H}}(P, Q, \varphi)$, whose existence is guaranteed by the boundedness and monotonicity of U_n^{H} (see Appendix A for details.) For any dynamic matching policy π and $n \ge 1$, we have $U_\infty(P, Q, \varphi) \ge U_n^{\mathsf{H}}(P, Q, \varphi) \ge U_n(\pi; P, Q, \varphi)$ (due to the monotonicity of U_n^{H}; as formalized in Lemma A.1 in Appendix A). We take $U_\infty(P, Q, \varphi)$ to be our performance benchmark, and define *regret* (of a policy π) as

$$\textsc{Reg}_n(\pi; P, Q, \varphi) := U_\infty(P, Q, \varphi) - U_n(\pi; P, Q, \varphi). \tag{3}$$

For $\textsc{Reg}_n(\pi; P, Q, \varphi)$ to be well-defined, we assume the quality function φ is bounded. Note that $U_\infty(P, Q, \varphi), U_n^{\mathsf{H}}(P, Q, \varphi)$ and $U_n(\pi; P, Q, \varphi)$ are all average (per match) quantities. We further note that under proper regularity conditions, U_∞ coincides with the optimal transport value between the distributions P and Q with respect to the function φ (For a counterexample, see Remark A.3). We shall hereafter drop the notation for dependence on distributions P and Q and the quality function φ, denoting by $U_\infty, U_n^{\mathsf{H}}, U_n(\pi)$ and $\textsc{Reg}_n(\pi)$ the corresponding objectives, since the distributions are always clear from the context.

Remark 1. *So far in our discussion, we have focused on a balanced market setting with an equal number of supply and demand units. However, our model is flexible enough to incorporate the setting with scarce supply and the corresponding rejection cost due to not being able to serve all demand units. To this end, we introduce the notion of a* virtual *or* dummy *supply unit denoted as* **dum**. *Matching a demand unit to a virtual supply unit* **dum** *is akin to rejecting that particular demand unit. To capture scarce supply within our model, we assume each supply unit is drawn* i.i.d. *from a specific distribution Q', which is a mixture of Q, the distribution of* real *supply units, and $Q^{\boldsymbol{dum}}$, a Dirac measure on the atom* **dum**. *Such a distribution ensures each supply unit is virtual with a certain probability p. Consequently, the number of Q-distributed* real *supply units corresponding to n demand units follows a $Bin(n, 1-p)$ distribution and is always (weakly) less than n, thus capturing scarce supply. Furthermore, our model can incorporate rejection cost. Let $c(x)$ denote the cost of rejecting a particular demand unit $x \in \mathcal{X}$. Then the matching quality function $\varphi'(x, y) := \varphi(x, y)\mathbb{1}\{y \neq \boldsymbol{dum}\} - c(x)\mathbb{1}\{y = \boldsymbol{dum}\}$ effectively captures the value generated from matching the demand unit to a real supply unit and the cost incurred from rejecting the demand unit.*

3 SOAR: Algorithmic Principle and Performance Analysis

The two primary properties we seek from our matching technology are: (i) computational efficiency, and (ii) provably near optimal performance. In the following discussion, we will first assess the practically popular myopic algorithm against our desiderata and argue that it can lead to highly sub-optimal matching outcomes. We then propose our algorithmic approach, SOAR, that possesses both properties, and present a meta guarantee on the performance of SOAR, which will drive the guarantees on regret scaling presented in the next section.

3.1 Insufficiency of Myopic Policies

We first consider the Greedy policy where each arriving demand unit is matched to a myopically optimal supply unit. In terms of computational efficiency, each matching decision can be computed in linear time. Motivated in part by ride hailing platforms, Kanoria [2022] and Akbarpour et al. [2021] show that Greedy has near optimal performance if the demand and supply distributions are identical (and uniform). However, this (near) optimality of Greedy is quite fragile. Beyond the very special cases studied in Akbarpour et al. [2021], Kanoria [2022], Greedy can suffer from significant performance degradation, resulting in non-vanishing regret. We formalize this using the following instance.

Proposition 1 (Failure of Greedy). *Suppose the supply distribution Q is supported on the atoms $\{0,1\}$, i.e., $\mathbb{P}(Y=1)=1-\mathbb{P}(Y=0)=\rho$ and the demand distribution P has a continuous distribution over the interval $[0,1]$ with density bounded below and above, i.e., there exists $\gamma>0$ such that $\gamma^{-1}\leq f_P\leq\gamma$ and has a CDF F_P. Fix $\rho\in(0,1)$ and assume that $F_P(1/2)\neq 1-\rho$. Fix $p\geq 1$ and consider the quality function $\varphi(X,Y)=-|X-Y|^p$. Then there exists a universal constant $c=c(\rho,F_p,p)>0$ and $n_0\in\mathbb{N}$ such that for all $n\geq n_0$, we have that $\textsc{Reg}_n(\mathsf{Greedy})\geq c$.*

Note that Proposition 1 holds for a class of quality functions including the standard euclidean distance with $p=1$. To obtain some intuition for this result, consider the case of $P=\text{Uniform}([0,1])$ and let $\rho=1/4$. Consider the fluid limit of the problem. Whenever there are supply units available, the Greedy algorithm matches demand unit arrivals located in $[1/2,1]$ to the supply units at location 1 and demand unit arrivals in $[0,1/2]$ to supply units at location 0. However, as $\mathbb{P}(Y=1)=1/4$, the Greedy allocation would prematurely exhaust all the supply units at 1 by the end of the first half of the time horizon, and then for the remaining half of the time horizon, all the demand units will be matched to the supply units at 0. In contrast, the optimal fluid policy would match all the demand units located in $[3/4,1]$ to the supply units located at 1 and the demand units located in $[0,3/4]$ to the supply units located at 0. Due to the myopic nature of Greedy, the demand units with location in $[3/4, 1]$ that arrive in the second half are forced to be matched to supply units at 0 resulting in large matching costs and hence the non-vanishing regret. This intuition is formalized in our proof deferred to Appendix B.

Remark 2. *In Appendix B, we further illustrate the fragility of* Greedy *by analyzing its performance on a broader class of instances (cf. Proposition B.3 and Conjecture 1). Specifically, for continuous distributions P and Q with densities bounded above and below on two intervals, and general quality function $\varphi(X, Y) = -|X - Y|^p$ for some $p > 1$, we show that the matching induced by* Greedy *is sufficiently different from the hindsight optimal matching unless P and Q are identical (or sufficiently similar), as measured by the Lévy–Prokhorov distance. This result strongly underscores the inherent suboptimality of* Greedy*.*

3.2 Simulate, Optimize, Assign, Repeat (SOAR) Principle

We now present SOAR, a principled simulation-based approach to general stochastic dynamic two-sided matching, and establish a meta-performance analysis framework. Our key result is a weighted sum representation of the expected average matching quality of SOAR, where the terms in the summation are precisely the hindsight optimum values. Such a direct connection between the policy performance and the hindsight optimum leads to several interesting structural results, as well as concrete regret analysis in specific settings. The latter is summarized in the next section.

3.2.1 Policy Description

Our policy does not require the precise distributional knowledge of P and Q. Instead, we utilize independent demand samples of P to facilitate matching. For ease of mathematical exposition, we conceptualize this sampling access requirement by assuming access to a demand unit simulator `SIM`. More precisely, upon calling `SIM`, we get an independent demand sample drawn from P. A pool of m historical *i.i.d.* demand units can thus be viewed as being generated from m repeated calls to `SIM`.

We formally state SOAR in Algorithm 1. The algorithmic principle is simple and can be summarized as follows. Upon each arrival of a new demand, the algorithm simulates a future demand scenario and solves a hindsight assignment problem, based on which the new demand is assigned to a supply unit. It is worth mentioning that to operationalize SOAR, we only need $\Theta(n^2)$ independent samples of demand, since only one simulated sample path and no more than n samples are needed for each demand arrival. This sets SOAR apart from many other simulation-based algorithms employing sample average approximation (SAA), which significantly relies on the volume of independently simulated samples to enhance performance. Also, there are standard algorithms, e.g., the Hungarian method and its variants, for solving the perfect assignment problem (4) and the state-of-the-art runtime has been reduced to almost linear time in the number of edges [Chen et al., 2022]. Hence, for each demand arrival, our problem of computing a perfect assignment requires a runtime of $\mathcal{O}(n^{2+o(1)})$. Combining the above, SOAR is both easy to implement and computationally feasible. We remark that in our base model each individual demand is matched upon arrival, but in practice it is often possible to batch and match. SOAR can be easily modified to suit such a scenario, in which the computational and sampling burden is further reduced.

Algorithm 1: Simulate-Optimize-Assign-Repeat (SOAR)

Input: supply units $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_n\}$, simulator `SIM`, quality function φ

$\mathcal{Y}_{\text{matched}} \leftarrow \emptyset$

for $t \leftarrow 1$ **to** n **do**

- Observe demand X_t and denote it as $\hat{X}_0 := X_t$ `// re-label demand unit`
- $\mathcal{Y}_{\text{un-matched}} := \{\tilde{Y}_0, \tilde{Y}_1, \ldots, \tilde{Y}_{n-t}\} \leftarrow \mathcal{Y} \backslash \mathcal{Y}_{\text{matched}}$ `// re-label supply units`
- Call the simulator `SIM` with random seed $U \sim \text{Uniform}([0,1])$ and denote the simulated demand scenario as $\mathcal{X}_t^{\text{sim}} := \{\hat{X}_1, \ldots, \hat{X}_{n-t}\}$ `// Simulate a future demand scenario`
- Randomly permute the demand pool $\hat{X}_0 \cup \mathcal{X}_t^{\text{sim}}$ and denote the random permutation as σ
- Solve the maximum quality matching optimization problem in (4) `// Optimize based on a simulated demand scenario`

$$\eta^\star \leftarrow \underset{\eta \text{ is a permutation}}{\arg\max} \sum_{k=0}^{n-t} \varphi(\hat{X}_{\sigma(k)}, \tilde{Y}_{\eta(k)}), \tag{4}$$

- Allocate $\tilde{Y}_{\eta^\star(\sigma(0))}$ to X_t `// Allocate the supply unit according to the optimal permutation`
- $\mathcal{Y}_{\text{matched}} \leftarrow \mathcal{Y}_{\text{matched}} \cup \left\{\tilde{Y}_{\eta^\star(\sigma(0))}\right\}$

end

3.2.2 Meta Performance Analysis

A simple formula connects the performance of SOAR with the sequence of hindsight optimum values. Recall that for the setting with n supplies and demands, $U_n(\pi)$ denotes the expected average matching quality achieved by policy π for the dynamic matching problem, and U_n^{H} denotes the hindsight optimum value.

Theorem 1 (Meta Performance of SOAR)**.** *Assuming that n supply units are drawn i.i.d. from a distribution Q and the demand units are drawn i.i.d. from a distribution P. For any quality function φ that is bounded on the support of P and Q, we have that the expected average match value of* SOAR *is given as*

$$U_n(\mathsf{SOAR}) = \frac{1}{n}\sum_{k=1}^{n} U_k^{\mathsf{H}}. \tag{5}$$

Theorem 1 holds regardless of the choice of P, Q and the quality function φ. The boundedness of φ is imposed only to ensure U_k^{H} is well defined. The key observation is that in each decision period, all remaining supply units, irrespective of their feature vectors, are equally likely to be assigned by SOAR to the current arriving demand unit. Such a property induces strong symmetry, and in particular, the remaining supply units continue to be distributed *i.i.d.* as per Q throughout the decision-making process, from which the theorem follows immediately.

Proof. The proof relies on the following key observations:

(i) *Remaining supply units are i.i.d*: The remaining supply units at each decision epoch $t \in \{1, 2, \ldots, n\}$ are located i.i.d. according to Q.

(ii) *Expected quality is the average hindsight optimum value.* In particular,

$$\mathbb{E}\left[\varphi(X_t, Y_{\pi_t^{\mathsf{SOAR}}})\right] = \frac{1}{n-t+1}\mathbb{E}_{\hat{X}\sim P, \hat{Y}\sim Q}\left[\max_{\sigma}\sum_{k=1}^{n-t+1}\varphi\left(\hat{X}_k, \hat{Y}_{\sigma(k)}\right)\right] = U_{n-t+1}^{\mathrm{H}}.$$

Note that (ii) is a corollary of (i). Indeed, upon the arrival of the t-th demand, SOAR calls an offline matching solver while outputs a perfect assignment with respect to quality function φ between $\{\hat{X}_0, \ldots, \hat{X}_{n-t}\}$ and $\{\tilde{Y}_0, \ldots, \tilde{Y}_{n-t}\}$, where $\hat{X}_0, \ldots, \hat{X}_{n-t}$ is an *i.i.d.* P sequence by algorithm design, and $\tilde{Y}_0, \ldots, \tilde{Y}_{n-t}$ is an *i.i.d.* Q sequence by property (i). Hence, the expected cumulative matching quality achieved by the offline optimizer equals the hindsight optimum (cumulative) quality, $(n - t+1)U^H_{n-t+1}$. Due to symmetry, the expected matching quality of the matched pair $(\hat{X}_0, \tilde{Y}_{\eta^*(0)})$ (or equivalently $(X_t, Y_{\pi_t^{\mathsf{SOAR}}})$) equals that of any pair $(\hat{X}_i, \tilde{Y}_{\eta^*(i)})$, from which (ii) follows.

We prove (i) inductively in t. Suppose (i) holds up to $t-1$. We show that upon the arrival of the t^{th} demand, each remaining supply unit is equally likely to be matched and leave. Indeed, let A_i denote the event that $\pi_t^{\mathsf{SOAR}} = i$, namely $\tilde{Y}_i$ is matched to X_t. Let $B(x_0, \ldots, x_{n-t}, \eta)$ denote the event that $\{x_0, \ldots, x_{n-t}\}$ is the demand pool and η is the permutation the offline matching output by the optimizer, i.e., x_i is matched to $\tilde{Y}_{\eta(i)}$. By design, one of $x_0, \ldots, x_{n-t}$ is the true realized demand of t, and the rest are the simulated units (note the random permutation of indices of the demand units in line 5 of Algorithm 1). Crucially, event B does not specify which unit is the true realized demand, and η is independent of which unit is the true realized demand. We thus have

$$
\begin{aligned}
&\mathbb{P}\Big(A_i | B(x_0, \ldots, x_{n-t}, \eta)\Big) = \mathbb{P}\Big(x_{\eta^{-1}(i)} \text{ is the true demand } \mid B(x_0, \ldots, x_{n-t}, \eta)\Big) \\
&= \mathbb{P}\Big(x_{\eta^{-1}(i)} \text{ is the true demand } \mid \{x_0, \ldots, x_{n-t}\} \text{ is the demand pool}\Big) \\
&= \frac{1}{n-t+1},
\end{aligned}
$$

where in the second equality we use the independence of η with $\{x_{\eta^{-1}(i)}$ is the true demand$\}$, and the third equality follows from the fact that the $n-t+1$ units in the demand pool are *i.i.d.* The above implies $\mathbb{P}(A_i) = \frac{1}{n-t+1}$, which then further implies property (i). Finally, Theorem 1 is an immediate corollary of (ii). □

Note that the proof of Property (i) does not require the offline matching solver to follow a particular tie-breaking rule (such as uniform-at-random tie-breaking) in facing multiple optimal solutions. Thus, Theorem 1 holds true for SOAR when implementing *any* tie-breaking rule under multiple optimal solutions to the matching problem in (4).

Remark 3 (Extension to the setting of Bayesian learning with unknown demand distributions)**.** *Performance guarantee similar to Theorem 1 continues to hold when the demand units form an* i.i.d. *sequence with unknown distribution, where the platform is assumed to have a prior belief about the true demand distribution from a class of candidate distributions and keep updating its knowledge as demands arrive (cf. Theorem F.1). In such cases, we characterize the regret of a Bayesian learning* *version of* SOAR*, which consists of the regret of* SOAR *in the setting where the distribution is known, and an additional regret term representing the learning cost due to not knowing the distribution in advance (cf. Corollary F.2). We provide a detailed description and proof in Appendix F.*

Remark 4 (Regret Decomposition)**.** *Observe that Theorem 1 implies that the regret of the SOAR algorithm can be written as the average of the regret of a sequence of hindsight problems, i.e.,* $\textsc{Reg}_n(\textit{SOAR}) = U_\infty - U_n(\textit{SOAR}) = \frac{1}{n}\sum_{k=1}^{n}(U_\infty - U_k^{\mathsf{H}}) \triangleq \frac{1}{n}\sum_{k=1}^{n}\textsc{Reg}_k\,(\textit{H-OPT})$*, where H-OPT stands for the hindsight optimal algorithm.*

Theorem 1 reveals the connection between the performance of SOAR and the convergence behavior of the sequence of hindsight optimum values. Rather surprisingly, Theorem 1 implies that SOAR achieves vanishing regret under no assumption other than the boundedness of φ, as we summarize in the next Corollary.

Corollary 1. *Under the same assumption as in Theorem 1,* SOAR *achieves vanishing regret:* $\lim_{n\to\infty}\textsc{Reg}_n(\textit{SOAR}) = 0$.

The proof follows immediately from Remark 4 and the monotonicity of U_k^{H}, which we delegate to Appendix C. When the matching instance satisfies additional conditions, Theorem 1 further implies the regret scaling of SOAR. In particular, we expect the regret of SOAR to converge at the same rate as the regret of H-OPT as n grows in most spatial matching settings, as the former is the Cesàro sum of the later via Remark 4. To formalize this intuition, we first introduce the following regular-scaling property of an offline bipartite matching instance specified by primitives P, Q, and φ.

Definition 1 (Regular and polynomial regret scaling)**.** *A matching instance* (P, Q, φ) *is said to* scale regularly *with parameter* $\beta > 0$ *if, for any* $0 < \epsilon < \beta$*, we have* $\limsup_{n\to\infty} n^{\beta-\epsilon}\cdot(U_\infty - U_n^{\mathsf{H}}) = 0$ *and* $\liminf_{n\to\infty} n^{\beta+\epsilon}\cdot(U_\infty - U_n^{\mathsf{H}}) = \infty$*. If in addition* $\lim_{n\to\infty} n^{\beta}\cdot(U_\infty - U_n^{\mathsf{H}}) = l_0$ *for some constant* l_0*, then we say the matching instance* (P, Q, φ) scales polynomially *with parameter* β*.*

No policy can do better than the hindsight optimal policy and, in particular, we have $\textsc{Reg}_n(\pi) \geq \textsc{Reg}_n(\text{H-OPT})$ for any π. Under regular or polynomial scaling, the following corollary of Theorem 1 tells us that $\textsc{Reg}_n(\text{SOAR})$ scales like $\textsc{Reg}_n(\text{H-OPT})$.

Corollary 2. *For a matching instance that scales regularly with parameter $\beta > 0$, we have that, for any $\epsilon > 0$, $\limsup_{n\to\infty} n^{\beta-\epsilon}\text{Reg}_n(\mathsf{SOAR}) = 0$ and $\liminf_{n\to\infty} n^{\beta+\epsilon}\text{Reg}_n(\mathsf{SOAR}) = \infty$, and*

$$\text{Reg}_n(\mathsf{SOAR}) \leq n^{\epsilon} \cdot \text{Reg}_n(\text{H-OPT}) \text{ for } n \text{ sufficiently large.}$$

If, in addition, the matching instance scales polynomially with parameter β, then there exists a constant $l_1 = l_1(P, Q, \varphi)$ independent of n, such that for all $n \in \mathbb{N}$

$$\text{Reg}_n(\mathsf{SOAR}) \leq \begin{cases} l_1 \text{Reg}_n(\text{H-OPT}), & \text{if } \beta \in (0,1), \\ l_1 \log n \cdot \text{Reg}_n(\text{H-OPT}), & \text{if } \beta = 1. \end{cases}$$

Furthermore, in this case, the regret scaling of SOAR*, which is also the optimal regret scaling (up to at most a logarithmic factor), can be tightly characterized:*

$$\text{Reg}_n(\mathsf{SOAR}) = \begin{cases} \Theta(n^{-\beta}), & \text{if } \beta \in (0,1), \\ \Theta(n^{-1}\log n), & \text{if } \beta = 1. \end{cases}$$

This corollary shows that SOAR enjoys a universal guarantee of near-optimal regret scaling when the matching instance scales regularly. We note that regular (or polynomial) scaling seems to be a relatively mild requirement, which is satisfied by a number of matching instances of practical interest. For example, when both demand and supply units are uniformly distributed on $[0,1]^d$ and the cost function is $-\|X - Y\|^p$, $U_\infty = 0$, the hindsight optimum scales regularly for all $d, p \geq 2$, with $\beta = p/d$ [Caracciolo et al., 2014]. Then it follows immediately from Corollary 2 that in these examples, SOAR provides a tight regret scaling for all $d, p \geq 2$. Refer to Appendix E for details. Indeed, we expect that matching instances satisfying certain basic conditions (e.g., continuity or boundedness of quality function φ) should all scale regularly or even polynomially. Determining primitive conditions on P, Q, φ under which the associated matching instances scale regularly or polynomially is in general a mathematically intriguing question, and we defer further investigation to future research. We defer the proof of the corollary to Appendix D.

To determine the precise value of β for a particular matching instance (P, Q, φ) can be quite challenging (see Caracciolo et al. [2014]). In many cases, we are often only able to establish bounds on the hindsight regret $\textsc{Reg}_n(\textsf{H-OPT})$. Indeed, $\textsc{Reg}_n(\textsf{H-OPT})$ emerges as the objective of an extensively studied problem: *Empirical optimal transport* (see Manole and Niles-Weed [2021], Appendix G.1). Its rich literature provides upper bounds on $\textsc{Reg}_n(\textsf{H-OPT})$ for matching instances belonging to various particular classes, e.g. P, Q satisfying certain smoothness conditions and/or $\varphi(X,Y) = -\|X-Y\|^p$. Leveraging the analytical framework set up in this section, we immediately get upper bounds on the regret scaling of SOAR for these classes of matching instances. Such upper bounds are tight if they are matched by an instance in these classes. In the next section, we take the above approach to establish the near-optimal regret scaling of SOAR for some interesting classes of matching instances.

4 Near-optimal Regret Scaling of SOAR for the $-\|X-Y\|^p$ and $\langle X, Y \rangle$ Quality Functions

In this section, we leverage our analytical framework (Theorem 1) to establish the near-optimal regret scaling of SOAR for a general quality function $\varphi_p(X,Y) = -\|X-Y\|^p$ for $p \geq 1$. Two important cases of this general quality function are $p = 1$ and $p = 2$. Our matching problem for the case of $p = 1$ (euclidean distance cost) corresponds to the setting studied in the semi-dynamic model of Kanoria [2022] assuming uniform supply and demand distributions. The Euclidean distance cost has been widely studied in the context of ride hailing platforms [Akbarpour et al., 2021, Kanoria, 2022, Besbes et al., 2022]. In terms of regret scaling, the quality function $\varphi_p(X,Y) = -\|X-Y\|^p$ for the case of $p = 2$ is equivalent to the dot product quality function $\varphi_{\textsf{dot}}(X,Y) = \langle X, Y \rangle$ (see Lemma G.7), which is a practically motivated quality function inspired by recommender systems [Koren et al., 2009, Aggarwal et al., 2016], a close cousin of our matching problem. To obtain our results, we leverage the fact that the offline version of our matching problem for the quality function $\varphi_p(X,Y) = -\|X-Y\|^p$ is intimately related to the problem of quantifying the empirical Wasserstein$-p$ distance between two measures, which has been extensively studied in the literature (Niles-Weed and Rigollet [2019], Ledoux [2019], Weed and Bach [2019], Manole et al. [2021]). The meta performance analysis (Theorem 1) allows us to directly infer from these results the regret scaling of SOAR in the corresponding online setting. We obtain two sets of regret scalings achieved by SOAR: one set allows for general distributions P and Q (Theorem 2 and Corollary 3), and the other set requires P and Q to be sufficiently smooth (Proposition 2 and Theorem 3). For the latter, we observe sharper regret scaling in low dimensions ($d \leq 3$). We construct hard instances to demonstrate the tightness of these regret scalings.

4.1 Performance Guarantees for $-\|X-Y\|^p$ Quality Functions

We begin by providing near-optimal regret guarantees for SOAR for the $\varphi_p(X,Y) = -\|X-Y\|^p$ quality functions under very general assumptions on the demand and supply distributions. The proof of Theorem 2 is deferred to Appendix H.

Theorem 2. *Suppose P and Q are supported on bounded sets. Under the quality function $\varphi_p(X,Y) = -\|X-Y\|^p$ for $p \geq 1$, there exists a universal constant $C := C(P,Q,d,p) < \infty$ such that the regret of* SOAR *is bounded above as*

$$\textsc{Reg}_n\,(\mathsf{SOAR}) \leq \begin{cases} Cn^{-\frac{1}{2}}, & d < 2(p \wedge 2)\,, \\ Cn^{-\frac{1}{2}}\log n, & d = 2(p \wedge 2)\,, \\ Cn^{-\frac{p\wedge 2}{d}}, & d > 2(p \wedge 2)\,. \end{cases}$$

Furthermore, the aforementioned regret scaling is nearly the best possible, formalized as follows. For each $d \geq 1$, there exists a pair of distributions P and Q (supported on bounded sets) such that there exists a constant $c := c(P,Q,d,p) > 0$ and the optimal regret scaling is bounded below as

$$\inf_{\pi \in \Pi} \textsc{Reg}_n\,(\pi) \geq \begin{cases} cn^{-\frac{1}{2}}, & d < 2(p \wedge 2)\,, \\ cn^{-\frac{1}{2}}, & d = 2(p \wedge 2)\,, \\ cn^{-\frac{p\wedge 2}{d}}, & d > 2(p \wedge 2)\,. \end{cases}$$

In contrast to myopic policies like Greedy that incurs non-vanishing regret (cf. Proposition 1), our forward-looking simulation-based policy SOAR achieves vanishing regret. Moreover, the upper bounds when viewed in conjunction with their corresponding lower bounds establish the near-optimality of SOAR. We observe that the regret increases with dimension d. This is due to an increase in the intrinsic hardness of matching, cf. Kanoria [2022]. Indeed, the available supply units become sparser in higher dimensional spaces. Thus finding supply units that are compatible with the demand units along all dimensions becomes harder, resulting in larger regret.

Remark 5 (Dependence of constants)**.** *The universal constant C in the upper bound in Theorem 2 is exponential in the dimension d. In contrast, the constant c in the lower bound scales polynomially in the dimension d. It is unclear if the exponential dependence on d of the constant C is tight or merely an artifact of the analysis and we leave the investigation of this question and possibly improving the dependence of the constant C on dimension d for future research.*

Under more stringent conditions on P and Q, sharper regret scalings are achievable. We state such a result for the special case of $P = Q = \text{Uniform}([0,1]^d)$.

Proposition 2. *Suppose $P = Q = \text{Uniform}([0,1]^d)$. Under the quality functions $\varphi_p(X,Y) = -\|X-Y\|^p$ for $p \geq 1$, there exists a universal constant $C := C(P,Q,d,p) < \infty$ such that*

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \begin{cases} Cn^{-(\frac{p}{2}\wedge 1)}\mathbb{1}\{p \neq 2\} + C\left(n^{-1}\log n\right)\mathbb{1}\{p=2\}, & d=1, \\ C\left(n^{-1}\log n\right)^{\frac{p}{2}}\mathbb{1}\{p<2\} + C\left(n^{-1}(\log n)^2\right)\mathbb{1}\{p=2\} + Cn^{-1}\mathbb{1}\{p>2\}, & d=2, \\ Cn^{-(\frac{p}{d}\wedge 1)}\mathbb{1}\{p \neq d\} + C\left(n^{-1}\log n\right)\mathbb{1}\{p=d\}, & d\geq 3. \end{cases}$$

Furthermore, the aforementioned regret scaling can not be improved in general. For each $d \geq 1$, there exists a constant $c := c(d) > 0$ and the optimal regret scaling is bounded below as

$$\inf_{\pi\in\Pi}\textsc{Reg}_n(\pi) \geq \begin{cases} cn^{-(\frac{p}{2}\wedge 1)}\mathbb{1}\{p \neq 2\} + cn^{-1}\mathbb{1}\{p=2\}, & d=1, \\ c(n^{-1}\log n)^{\frac{p}{2}}\mathbb{1}\{p<2\} + c(n^{-1}\log n)\mathbb{1}\{p=2\} + cn^{-1}\mathbb{1}\{p>2\}, & d=2, \\ cn^{-(\frac{p}{d}\wedge 1)}\mathbb{1}\{p \neq d\} + c(n^{-1}\log n)\mathbb{1}\{p=d\}, & d\geq 3. \end{cases}$$

We defer the proof of Proposition 2 to Appendix I. We note that in comparison with Theorem 2, the performance of SOAR improves for $P = Q = \text{Uniform}([0,1]^d)$ in various parameter regimes. For example, in Theorem 2 where P and Q can be general distributions supported on bounded sets, for $p \in (2, d]$ and $d \geq 4$, the regret scales as $\Theta(n^{-2/d})$. However, in Proposition 2, the regret scaling improves to $\tilde{\Theta}(n^{-p/d})$. There are two main drivers for the sharper regret scaling: (i) smoothness of the demand and supply distribution and (ii) both the demand and supply distributions being identical. In light of the hard instance for Theorem 2 that consists of discrete distributions (cf. proof of Theorem 2 in Appendix H), the smoothness of distribution $\text{Uniform}([0,1]^d)$ in Proposition

2 seems necessary for the sharper regret scalings. Later in Subsection 4.2, we present a similar regret guarantee in the case of $p = 2$ (equivalently, the case of dot-product quality function) (cf. Theorem 3), for a more general class of P and Q that allow for possibly non-uniform and unequal distributions, assuming some proper smoothness conditions. However, such an extension is not immediate for general $p \neq 2$ due to lack of regularity of the optimal transport map (Ma et al. [2005]), and we defer a more in-depth exploration of the assumptions that may yield sharper regret scaling for general quality function to future research.

Open Problem in Kanoria [2022]. The case of $p = 1$ (euclidean distance cost) has been extensively studied for all $d \geq 1$ in Kanoria [2022]. Kanoria [2022] develops a greedy-like algorithm called Hierarchical Greedy and shows that Hierarchical Greedy achieves near-optimal regret scaling for all $d \geq 1$. However, this near optimality does not hold for general $p > 1$: for a fixed $d \geq 2$ and $p \in (d/2, d]$, the regret of Hierarchical Greedy scales as $\Theta(n^{-1/2})$ whereas a sharper regret scaling of $\tilde{\Theta}(n^{-p/d})$ is achievable via SOAR and up to logarithmic factors, this regret scaling is tight. In fact, Kanoria [2022] leaves it as an open problem to close the gap between the regret scaling achieved via Hierarchical Greedy and optimal regret scaling of $\tilde{\Omega}(n^{-p/d})$ (assuming $d \geq 2, p \leq d$) for uniform demand and supply distributions and $\varphi_p(X, Y) = -\|X - Y\|^p$. As a consequence of Proposition 2, we resolve this open problem.

4.2 Performance Guarantees for the $\langle X, Y \rangle$ Quality Function

We first observe that in terms of regret, the dot-product quality function $\langle X, Y \rangle$ is equivalent to $\varphi_2(X, Y) = -\|X - Y\|^2$ (cf. Lemma G.7). Therefore, the regret scaling of SOAR for the dot-product quality function under the assumptions of the demand and supply distributions being supported on bounded sets follows immediately from Theorem 2.

Corollary 3. *Suppose P and Q are supported on bounded sets. The regret scalings for the dot product function $\varphi(X, Y) = \langle X, Y \rangle$ correspond to the regret scalings demonstrated in Theorem 2 with $p = 2$.*

Remark 6 (Near-optimal regret scaling with scarce supply and rejection cost)**.** *Recall that we modelled scarce supply and rejection cost in Section 2 by considering an augmented demand and supply distribution P' and Q' and a modified quality function. For the particular quality function $\varphi(X, Y) = \langle X, Y \rangle$, its corresponding quality function that incorporates the rejection cost can be* *reformulated so as to retain the dot-product form, albeit in a different, augmented feature space. The key idea is to view the rejection cost as an additional dimension for the demand unit's weight vector. As before, let $c(x)$ denote the cost of rejection for a demand unit $x \in \mathcal{X}$ that is measurable and bounded. Let $\boldsymbol{dum} = (\mathbf{0}_{d\times 1}, 1)$ and define the following $d+1$ dimensional spaces*

$$\mathcal{X}' = \{(x, -c(x)) : x \in \mathcal{X}\} \subseteq \mathbb{R}^{d+1}, \qquad \mathcal{Y}' = \{(y, 0) : y \in \mathcal{Y}\} \cup \boldsymbol{dum} \subseteq \mathbb{R}^{d+1}.$$

Then for any $x' \in \mathcal{X}'$ and $y' \in \mathcal{Y}'$, we have that $\varphi'(x', y') = \langle x, y\rangle \mathbb{1}\{y \neq \boldsymbol{dum}\} - c(x)\mathbb{1}\{y = \boldsymbol{dum}\} = \langle x', y'\rangle$. Let P' denote a distribution supported on $\mathcal{X}'$ and Q' denote a distribution supported on $\mathcal{Y}'$. As previously discussed, Q' is a mixture distribution of Q and $Q^{\boldsymbol{dum}}$. By modelling the virtual supply unit $\boldsymbol{dum}$ in the distribution Q' and the rejection cost in distribution P', we are effectively back to the balanced setting with n supply units $Y_1', Y_2', \dots, Y_n'$ independently sampled from Q' and n demand units $X_1', X_2', \dots, X_n'$ independently sampled from P', associated with the dot-product quality function. Furthermore, we note that P' and Q' will be supported on bounded sets as long as P and Q are supported on bounded sets and c is a measurable bounded function. Hence from Corollary 3, even under scarce supply, we are able to achieve near-optimal regret scaling for the quality function $\varphi(X, Y) = \langle X, Y\rangle$. Note that the benchmark in the case of scarce supply is $U_\infty(P', Q')$.

Remark 7 (Extension to polynomial kernel quality functions)**.** *Our regret bound on* SOAR *for the dot product quality function implies regret bounds for a broader class of quality functions which we refer to as the polynomial kernel quality functions. For some $q \in \mathbb{N}$, we refer to $\varphi^q_{\mathsf{ker}}(X, Y) = \langle X, Y\rangle^q$ as the polynomial kernel quality function. Using Corollary 3, one can easily establish vanishing regret for* SOAR *under this more general class of polynomial kernel quality functions $\varphi^q_{\mathsf{ker}}(X, Y)$ (see Corollary J.1). The key idea is the well-known result that for each choice of $q \in \mathbb{N}$, there exists a mapping $\phi_q : \mathbb{R}^d \to \mathbb{R}^{d_q}$ where $d_q = \binom{d+q}{q}$ such that $\varphi^q_{\mathsf{ker}}(X, Y) = \langle X, Y\rangle^q = \langle \phi_q(X), \phi_q(Y)\rangle$ [Murphy, 2022]. This dot product representation allows us to invoke Corollary 3 and establish vanishing regret. Furthermore, the vanishing regret guarantee of* SOAR *can be extended to a conic combination (weighted sum with non-negative coefficients) of polynomial kernel quality functions corresponding to different values of $q \in \mathbb{N}$, i.e., $\varphi_{\mathsf{ker}}(X, Y) = \sum_{q=0}^m a_q \varphi^q_{\mathsf{ker}}(X, Y) = \sum_{q=0}^m a_q \langle X, Y\rangle^q$ (see Corollary J.1). For further details, refer to Appendix J.*

In Proposition 2, we showed that sharper regret scalings are achievable by SOAR when the demand and supply distributions are uniform. Under the dot-product quality function, we can further generalize those results to settings where demand and supply distributions are smooth (uniform-like) and possibly distinct and under a key curvature condition on the *Brenier potential*. The Brenier potential (Brenier [1991]) is a convex function ψ_0 whose gradient uniquely defines the optimal transport map $\mathcal{T}_{P\to Q}$ from P to Q under the quadratic cost (equivalently, the dot-product quality function), for P and Q satisfying some smoothness assumptions. We refer the interested readers to Villani [2009] for a detailed introduction of the Brenier potential and the related background knowledge of optimal transport (a brief discussion is provided in Appendix G).

Assumption 1 (Curvature condition). *The Brenier potential ψ_0 is a closed, convex, function such that $\psi_0 \in \mathcal{C}^2([0,1]^d)$ and $(1/\lambda)\boldsymbol{I}_{d\times d} \preceq \nabla^2\psi_0(x) \preceq \lambda \boldsymbol{I}_{d\times d}$ for all $x \in [0,1]^d$ and for some $\lambda \geq 1$.*

The curvature condition on the the Brenier potential in Assumption 1 has been borrowed from Manole et al. [2021] and it enables us to prove sharper regret scaling under smoothness assumptions of the demand and supply distributions. This sharper regret scaling is formalized in Theorem 3 presented below.

Theorem 3. *Suppose P and Q are absolutely continuous distributions on $[0,1]^d$ with densities p and q. Assume that there exists $\gamma_p \geq 1$ and $\gamma_q \geq 1$ such that $\gamma_p^{-1} \leq p \leq \gamma_p$ and $\gamma_q^{-1} \leq q \leq \gamma_q$ over $[0,1]^d$ and further assume the curvature condition in Assumption 1. Under the quality function $\varphi(X,Y) = \langle X, Y\rangle$ there exists a universal constant $C := C(P,Q,d) < \infty$ such that the regret of* SOAR *is bounded above as*

$$\textsc{Reg}_n\,(\mathsf{SOAR}) \leq \begin{cases} Cn^{-1}\log n, & d = 1\,, \\ Cn^{-1}\,(\log n)^3\,, & d = 2\,, \\ Cn^{-\frac{2}{d}}, & d \geq 3\,. \end{cases}$$

Furthermore, the aforementioned regret scaling can not be improved in general. The lower bound on the optimal regret scaling with quality function $\varphi(X,Y) = \langle X, Y\rangle$ follows from Proposition 2 for the case of $p = 2$.

We defer the proof of Theorem 3 to Appendix K. Comparing Corollary 3 with Theorem 3 for $\varphi(X,Y) = \langle X, Y\rangle$, we observe that for $d \leq 3$, Theorem 3 provides a sharper regret scaling:

contrast the regret scaling of $\Theta(n^{-1/2})$ for $d \leq 3$ and $p = 2$ in Corollary 3 with $\tilde{\Theta}(n^{-1})$ for $d = 1, 2$ and $\Theta(n^{-2/3})$ for $d = 3$ in Theorem 3. The special structure of the dot-product quality function $\varphi(X, Y) = \langle X, Y \rangle$ (a special case of $\varphi_p(X, Y) = -\|X - Y\|^p$ for $p = 2$) enables us to derive sharper regret scalings under some smoothness assumptions of the demand P and supply Q distributions and the curvature condition in Assumption 1.

5 Numerics

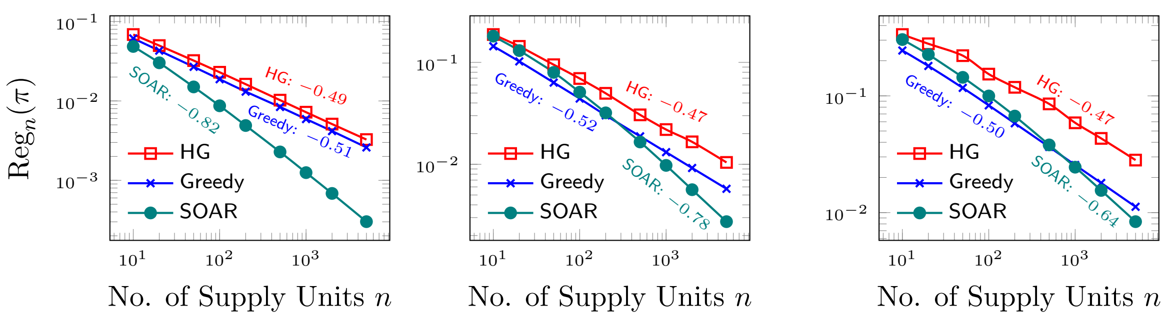
In this section, we describe our numerical simulations for different dimensions d and different sets of demand and supply distributions. We focus on the quality function $\varphi(X, Y) = -\|X - Y\|^2$ unless mentioned otherwise, where we recall the equivalence between this quality function and $\varphi(X, Y) = \langle X, Y \rangle$ in terms of regret. We consider different demand and supply distributions and evaluate the performance of different algorithms as listed in Table 2. In this section, we focus on settings that satisfy our theoretical assumptions and as such our numerical experiments exemplify our theoretical guarantees that our proposed algorithm SOAR achieves vanishing regret at near-optimal rate in all the settings we consider whereas myopic policies like Greedy or OT + Greedy (a smarter variant of Greedy) either suffer from non-vanishing regret or the rate is sub-optimal. In the following subsections, we describe the settings considered in detail and present the regret scaling for different algorithms. Note that all the plots are presented in the $\log - \log$ scale. In Appendix L, we compliment our numerical investigations in this section by broadening the scope of the experiments beyond the confines of our theoretical assumptions. In particular, we study natural variants of SOAR for different environments: known i.i.d setting, unknown i.i.d setting and time-varying demand distribution setting. Our numerical investigations demonstrate strong numerical performance of natural variants of SOAR.

Table 2: Summary of algorithms considered for numerical simulations

Algorithm	Requires OT Map	Description
SOAR	No	Algorithm 1
Greedy	No	Match X_t to nearest supply
OT + Greedy	Yes	Transport X_t to Q and match to nearest supply
Hierarchical Greedy	N/A	Algorithm 1 in Kanoria [2022]

5.1 Setting (I) $P = Q = \textbf{Uniform}\left([0,1]^d\right)$

We consider the case of demand and supply distributions being equal and in particular, being Uniform($[0,1]^d$). For this setting, we present the results for $d \in \{1,2,3\}$ in Figure 1. We compare the performance of the Hierarchical Greedy (Kanoria [2022]), Greedy, and SOAR algorithms. In Figure 1, we observe that as the number of supply units n increases, SOAR performs better than Greedy and Hierarchical Greedy. Observe the slopes corresponding to the different algorithms. Both Hierarchical Greedy and Greedy algorithms have a slope close to -0.5 which corresponds to the regret of these algorithms scaling as $\mathcal{O}(n^{-1/2})$. For SOAR, the slope for different values of d closely matches the theoretical regret guarantees of $\mathcal{O}(\log n/n)$ for $d = 1$ (slope is -0.82), $\mathcal{O}(\log^3 n/n)$ for $d = 2$ (slope is -0.78) and $\mathcal{O}(n^{-2/3})$ for $d = 3$ (slope is -0.64).

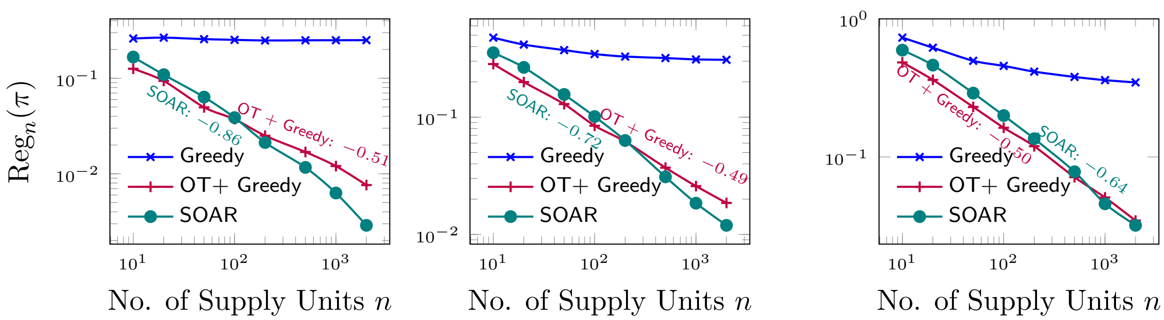


(a) $d = 1$ (b) $d = 2$ (c) $d = 3$

Figure 1: Comparing the performance of Hierarchical Greedy (HG), Greedy and SOAR for $P = Q =$ Uniform $\left([0,1]^d\right)$.

5.2 Setting (II) $P = \textbf{Uniform}\left([0,1]^d\right), Q = \textbf{Uniform}\left([0,2]^d\right)$

We consider an example where the demand and supply distributions are unequal and the optimal transport map is fairly easy to compute i.e $\mathcal{T}_{P\to Q}(X) = 2X$. This enables us to implement the OT + Greedy algorithm, which is a smarter variant of Greedy. The OT + Greedy algorithm is as follows: upon observing a demand request X_t, we first transport the demand request X_t using the optimal transport map $\mathcal{T}_{P\to Q}$ and then *greedily* match the transport demand unit $\mathcal{T}_{P\to Q}(X_t)$ to its nearest existing supply unit. By computing the transport map of the demand units, the OT + Greedy algorithm reduces to a dynamic greedy matching between random points of the same distribution. Moreover, instead of directly implementing a greedy matching between the demand unit and existing supply units, by utilizing the optimal transport map $\mathcal{T}_{P\to Q}$, we are able to make the Greedy algorithm forward looking. For this setting, we present the results for the case of $d \in \{1,2,3\}$ in Figure 2. Note that this setting is equivalent to the previous setting upto the optimal transport. In this setting, we observe that the Greedy algorithm suffers from non-vanishing regret and this is in line with Proposition 1. Both the smarter variant of Greedy, which we dub as OT + Greedy, and SOAR are able to achieve vanishing regret. However, we note that OT + Greedy requires nontrivial knowledge of the underlying model, namely the optimal transport map between P and Q, which quickly becomes a computation burden in more complex settings. Moreover, observe the slopes of the curve corresponding to OT + Greedy and SOAR algorithm. The slope of the curves corresponding to SOAR are steeper compared to the ones corresponding to OT + Greedy for $d \in \{1,2,3\}$ which implies that SOAR achieves a sharper regret scaling in comparison to the OT + Greedy policy.

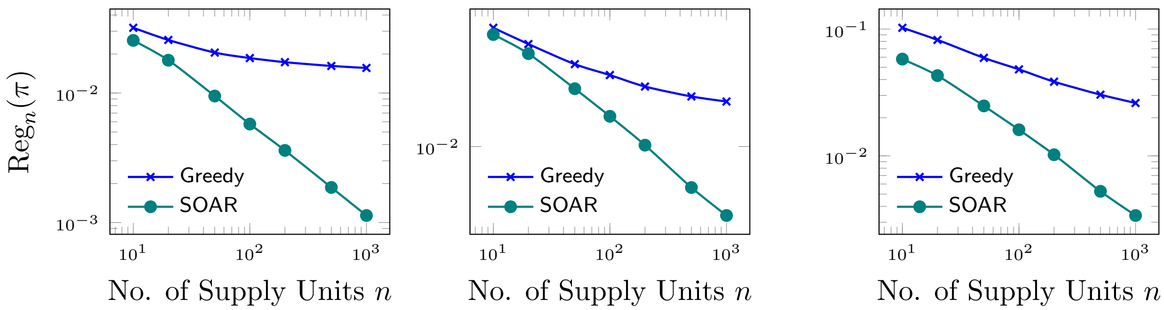


(a) $d = 1$ (b) $d = 2$ (c) $d = 3$

Figure 2: Comparing the performance of Greedy, OT + Greedy and SOAR for $P = \text{Uniform}([0,1/2]^d)$ and $Q = \text{Uniform}([0,1]^d)$.

5.3 Setting (III) $P = \textbf{TruncNorm}(\mu, \Sigma), Q = \textbf{Uniform}\left([0,1]^d\right)$

We consider an example where the demand and supply distributions are unequal however it is non-trivial to compute the optimal transport map. In particular, we assume that the demand distribution is a truncated normal with mean $\mu = (1/2) \times \mathbb{1}_{d\times 1}$ and covariance $\Sigma = 0.1 \times \mathbb{I}_{d\times d}$ and the supply distribution is Uniform $\left([0,1]^d\right)$. Since the optimal transport map is non-trivial to compute, we approximate the value of the fluid optimum for each d via simulation. For this setting, we present the results for the case of $d \in \{1,2,3\}$ in Figure 3 and compare the Greedy algorithm and the SOAR algorithm. As before, we observe that the SOAR algorithm outperforms the Greedy algorithm.



(a) $d = 1$ (b) $d = 2$ (c) $d = 3$

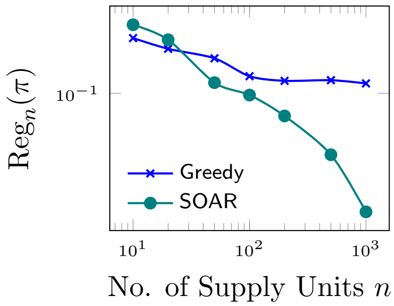
Figure 3: Comparing the performance of Greedy and SOAR for $P = \text{TruncNorm}(\mu, \Sigma)$ and $Q = \text{Uniform}([0,1]^d)$.

5.4 Setting (IV) $P = \textbf{Uniform}\left([0,1]^d\right), Q = \textbf{TruncNorm}\left(\mu_{d-2\times 1}, \Sigma_{d-2\times d-2}\right)\times\textbf{Ber}(0.7)\times\textbf{Ber}\left(0.2\right)$

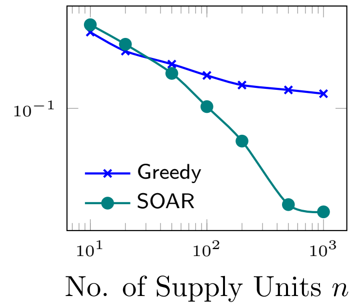
We consider the case where demand and supply distributions are unequal, the optimal transport is non-trivial to compute and moreover, the supply distribution is not smooth unlike the previously consider settings. We focus on the case of $d \geq 3$. The supply distribution is $Q = \text{TruncNorm}\left(\frac{1}{2} \times \mathbb{1}_{d-2\times 1}, 0.1 \times \mathbb{I}_{d-2\times d-2}\right) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$. Since the optimal transport map is non-trivial to compute, we approximate the fluid optimum value via simulation. For this setting, we present the results for the case of $d \in \{3, 4, 5\}$ in Figure 4 and compare the Greedy and SOAR algorithms. We observe that as n increases, the performance of SOAR dominates the performance of Greedy algorithm.

6 Conclusion and Future Research

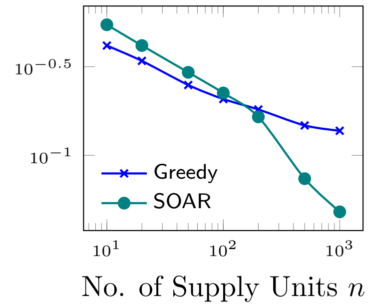
In this work, we study a dynamic two-sided matching problem as the market thickness n scales, and characterize the optimal regret scaling for all dimensions d. We develop a principled simulation-based approach dubbed Simulate-Optimize-Assign-Repeat (SOAR) and demonstrate that this forward-looking policy is vastly superior to myopic policies like Greedy. En-route we develop a novel framework for regret analysis where we provide a simple formula connecting the performance of SOAR



(a) $d = 3$



(b) $d = 4$



(c) $d = 5$

Figure 4: Comparing the performance of Greedy and SOAR for $P = \text{Uniform}([0,1]^d)$ and $Q = \text{TruncNorm}(\mu, \Sigma) \times \text{Ber}(0.7) \times \text{Ber}(0.2)$.

with a sequence of hindsight optimum values. As a corollary of our techniques, we also resolve one of the open problems in Kanoria [2022].

Our algorithm SOAR and results crucially rely on knowledge of the horizon n. Given the recent burgeoning interest in the study of online resource allocation problems in the presence of horizon uncertainty [Besbes and Sauré, 2014, Aouad and Ma, 2022, Balseiro et al., 2023a, Bai et al., 2023], an interesting follow-up would be to design and analyze near-optimal algorithms for such uncertain scenarios. One approach to modelling horizon uncertainty is to assume distributional knowledge of the horizon [as done in Bai et al., 2023, Aouad and Ma, 2022], where the horizon length (total demand units) is modelled as a random variable N and this is known to the platform. Assume that $n = \mathbb{E}[N]$, then in assigning n supply units to the N arriving demand units, SOAR can be implemented by fixing the number of demand units as if it was n. If in addition, N is well concentrated, i.e., $\text{var}(N) = o(n^2)$ (e.g., N is a Binomial or Poisson random variable), SOAR achieves vanishing regret where the rate may depend on the variance of N. On the other hand, if the variance is large, i.e., $\text{var}(N) = \Omega(n^2)$, then SOAR suffers from non-vanishing regret and this observation is in line with other resource allocation works studying horizon uncertainty [Aouad and Ma, 2022, Bai et al., 2023]. The large variance case is quite challenging and developing near-optimal algorithms would require novel approaches and we leave this line of research for future work.

References

Charu C Aggarwal et al. *Recommender systems*, volume 1. Springer, 2016.

Narges Ahani, Paul Gölz, Ariel D Procaccia, Alexander Teytelboym, and Andrew C Trapp. Dynamic placement in refugee resettlement. *Operations Research*, 2023.

Mohammad Akbarpour, Yeganeh Alimohammadi, Shengwu Li, and Amin Saberi. The value of excess supply in spatial matching markets. *arXiv preprint arXiv:2104.03219*, 2021.

Samuel Christian Albright Jr. Stochastic sequential assignment problems. Technical report, STANFORD UNIV CALIF DEPT OF OPERATIONS RESEARCH, 1972.

Ali Aouad and Will Ma. A nonparametric framework for online stochastic matching with correlated arrivals. *arXiv preprint arXiv:2208.02229*, 2022.

Ali Aouad and Daniela Saban. Online assortment optimization for two-sided matching platforms. *Management Science*, 2022.

Ali Aouad and Ömer Saritaç. Dynamic stochastic matching under limited time. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 789–790, 2020.

Itai Ashlagi, Maximilien Burq, Chinmoy Dutta, Patrick Jaillet, Amin Saberi, and Chris Sholley. Edge weighted online windowed matching. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 729–742, 2019.

Itai Ashlagi, Anilesh K Krishnaswamy, Rahul Makhijani, Daniela Saban, and Kirankumar Shiragur. Assortment planning for two-sided sequential matching markets. *Operations Research*, 70(5): 2784–2803, 2022.

Yicheng Bai, Omar El Housni, Billy Jin, Paat Rusmevichientong, Huseyin Topaloglu, and David P Williamson. Fluid approximations for revenue management under high-variance demand. *Management Science*, 2023.

Eric Balkanksi, Yuri Faenza, and Noemie Perivier. The power of greedy for online minimum cost matching on the line. *arXiv preprint arXiv:2210.03166*, 2022.

Santiago Balseiro, Christian Kroer, and Rachitesh Kumar. Online resource allocation under horizon uncertainty. In *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 63–64, 2023a.

Santiago R Balseiro, Omar Besbes, and Dana Pizarro. Survey of dynamic resource-constrained reward collection problems: Unified model and analysis. *Operations Research*, 2023b.

Siddhartha Banerjee, Daniel Freund, and Thodoris Lykouris. Pricing and optimization in shared vehicle systems: An approximation framework. *Operations Research*, 70(3):1783–1805, 2022.

Dimitri Bertsekas. *Dynamic programming and optimal control: Volume I*, volume 4. Athena scientific, 2012.

Omar Besbes and Denis Sauré. Dynamic pricing strategies in the presence of demand shifts. *Manufacturing & Service Operations Management*, 16(4):513–528, 2014.

Omar Besbes, Francisco Castro, and Ilan Lobel. Spatial capacity planning. *Operations Research*, 70(2):1271–1291, 2022.

Omar Besbes, Yash Kanoria, and Akshit Kumar. Dynamic resource allocation: Algorithmic design principles and spectrum of achievable performances. *arXiv preprint arXiv:2205.09078*, 2023.

Gabriel Bitran and René Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

Mark Braverman, Mahsa Derakhshan, and Antonio Molina Lovett. Max-weight online stochastic matching: Improved approximations against the online benchmark. *arXiv preprint arXiv:2206.01270*, 2022.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

Brian Brubach, Karthik Abinav Sankararaman, Aravind Srinivasan, and Pan Xu. New algorithms, better bounds, and a novel model for online stochastic matching. In *24th Annual European Symposium on Algorithms (ESA 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

Pornpawee Bumpensanti and He Wang. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science*, 66(7):2993–3009, 2020.

Eduardo F Camacho and Carlos Bordons Alba. *Model predictive control*. Springer science & business media, 2013.

Sergio Caracciolo, Carlo Lucibello, Giorgio Parisi, and Gabriele Sicuro. Scaling hypothesis for the euclidean bipartite matching problem. *Physical Review E*, 90(1):012118, 2014.

Li Chen, Rasmus Kyng, Yang P Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. *arXiv preprint arXiv:2203.00671*, 2022.

Steven Delong, Alireza Farhadi, Rad Niazadeh, and Balasubramanian Sivan. Online bipartite matching with reusable resources. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 962–963, 2022.

Mahsa Derakhshan, Negin Golrezaei, Vahideh Manshadi, and Vahab Mirrokni. Product ranking on online platforms. *Management Science*, 68(6):4024–4041, 2022.

Cyrus Derman, Gerald J Lieberman, and Sheldon M Ross. A sequential stochastic assignment problem. *Management Science*, 18(7):349–355, 1972.

Tomer Ezra, Michal Feldman, Nick Gravin, and Zhihao Gavin Tang. Online stochastic max-weight matching: Prophet inequality for vertex and edge arrival models. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 769–787, 2020.

Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and Shan Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 117–126. IEEE, 2009.

Yiding Feng and Rad Niazadeh. Batching and optimal multi-stage bipartite allocations. *arXiv preprint arXiv:2211.16581*, 2022.

Daniel Freund and Siddhartha Banerjee. Good prophets know when the end is near. *Available at SSRN 3479189*, 2019.

Alfred Galichon. *Optimal transport methods in economics.* Princeton University Press, 2018.

Anupam Gupta, Guru Guruganesh, Binghui Peng, and David Wajc. Stochastic online metric matching. *arXiv preprint arXiv:1904.09284*, 2019.

Bernhard Haeupler, Vahab S Mirrokni, and Morteza Zadimoghaddam. Online stochastic weighted matching: Improved approximation algorithms. In *Internet and Network Economics: 7th International Workshop, WINE 2011, Singapore, December 11-14, 2011. Proceedings 7*, pages 170–181. Springer, 2011.

Nina Holden, Yuval Peres, and Alex Zhai. Gravitational allocation for uniform points on the sphere. *The Annals of Probability*, 49(1):287–321, 2021.

Shayan Hundrieser, Thomas Staudt, and Axel Munk. Empirical optimal transport between different measures adapts to lower complexity. *arXiv preprint arXiv:2202.10434*, 2022.

Nicole Immorlica, Brendan Lucier, Vahideh Manshadi, and Alexander Wei. Designing approximately optimal search on matching platforms. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 632–633, 2021.

Patrick Jaillet and Xin Lu. Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research*, 39(3):624–646, 2014.

Yash Kanoria. Dynamic spatial matching. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 63–64, 2022.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

Sumit Kunnumkal, Kalyan Talluri, and Huseyin Topaloglu. A randomized linear programming method for network revenue management with product-specific no-shows. *Transportation Science*, 46(1):90–108, 2012.

Michel Ledoux. On optimal matching of gaussian samples. *Journal of Mathematical Sciences*, 238 (4):495–522, 2019.

Will Ma and David Simchi-Levi. Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios. *Operations Research*, 68(6):1787–1803, 2020.

Xi-Nan Ma, Neil S Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Archive for rational mechanics and analysis*, 177:151–183, 2005.

Tudor Manole and Jonathan Niles-Weed. Sharp convergence rates for empirical optimal transport with smooth costs. *arXiv preprint arXiv:2106.13181*, 2021.

Tudor Manole, Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman. Plugin estimation of smooth optimal transport maps. *arXiv preprint arXiv:2107.12364*, 2021.

Vahideh Manshadi, Scott Rodilitz, Daniela Saban, and Akshaya Suresh. Online algorithms for matching platforms with multi-channel traffic. *arXiv preprint arXiv:2203.15037*, 2022.

Vahideh H Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012.

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

Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL probml.ai.

Jonathan Niles-Weed and Philippe Rigollet. Estimation of wasserstein distances in the spiked transport model. *arXiv preprint arXiv:1909.07513*, 2019.

Christos Papadimitriou, Tristan Pollner, Amin Saberi, and David Wajc. Online stochastic max-weight bipartite matching: Beyond prophet inequalities. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 763–764, 2021.

Amin Saberi, Mingwei Yang, and Sophie H Yu. Stochastic online metric matching: Adversarial is no harder than stochastic. *arXiv preprint arXiv:2407.14785*, 2024.

Peng Shi. Optimal matchmaking strategy in two-sided marketplaces. *Management Science*, 2022.

Sean R Sinclair, Felipe Vieira Frujeri, Ching-An Cheng, Luke Marshall, Hugo De Oliveira Barbalho, Jingling Li, Jennifer Neville, Ishai Menache, and Adith Swaminathan. Hindsight learning for mdps with exogenous inputs. In *International Conference on Machine Learning*, pages 31877–31914. PMLR, 2023.

Xuanming Su and Stefanos A Zenios. Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations research*, 53(3):443–455, 2005.

Kalyan Talluri and Garrett Van Ryzin. A randomized linear programming method for computing network bid prices. *Transportation science*, 33(2):207–216, 1999.

Kalyan T Talluri, Garrett Van Ryzin, and Garrett Van Ryzin. *The theory and practice of revenue management*, volume 1. Springer, 2004.

Bahar Taşkesen, Soroosh Shafieezadeh-Abadeh, and Daniel Kuhn. Semi-discrete optimal transport: Hardness, regularization and numerical solution. *Mathematical Programming*, pages 1–74, 2022.

Rajan Udwani. Periodic reranking for online matching of reusable resources. *arXiv preprint arXiv:2110.02400*, 2021.

Alberto Vera and Siddhartha Banerjee. The bayesian prophet: A low-regret framework for online decision making. *Management Science*, 67(3):1368–1391, 2021.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.

Electronic Companion:
Feature-Based Dynamic Matching

Appendix

Table of Contents

A Proof of $U_\infty(P,Q,\varphi) \geq U_n^{\mathsf{H}}(P,Q,\varphi) \geq U_n(\pi;P,Q,\varphi)$

That $U_n^{\mathsf{H}}(P,Q,\varphi) \geq U_n(\pi;P,Q,\varphi)$ is straightforward. To prove the result it suffices to show the first inequality, i.e. $U_\infty(P,Q,\varphi) \geq U_n^{\mathsf{H}}(P,Q,\varphi)$, which follows from the next Lemma. With bounded quality function φ, U_n^{H} are trivially bounded, then by monotone convergence theorem $U_\infty(P,Q,\varphi)$ is also bounded, making it a valid performance benchmark.

Lemma A.1. *$U_n^H, n \geq 1$ is a monotone increasing sequence.*

Proof of Lemma A.1. For each realized $(X_t, Y_{(t)})_{1\leq t\leq n}$, we consider a specific randomized permutation. Simulate $n-1$ demand units $X'_1, \dots X'_{n-1}$ *i.i.d.* from P. Let λ denote the optimal assignment between $\{X'_1, \dots X'_{n-1}, X_n\}$ and $\{Y_1, \dots, Y_n\}$. We then optimally match $\{X_1, \dots, X_{n-1}\}$ and $\{Y_1, \dots Y_n\}/\{Y_{\lambda_n}\}$. The above constitutes a permutation (possibly randomized) between $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_n\}$, which we call σ'. Notice that

$$\begin{aligned} U_n^{\mathsf{H}} &= \frac{1}{n}\mathbb{E}\left[\sup_{\sigma\in S_n}\sum_{t=1}^{n}\varphi(X_t, Y_{\sigma_t})\right] \geq \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n}\varphi(X_t, Y_{\sigma'_t})\right] \\ &= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n-1}\varphi(X_t, Y_{\sigma'_t})\right] + \frac{1}{n}\mathbb{E}\left[\varphi(X_n, Y_{\sigma'_n})\right] \\ &= \frac{n-1}{n}U_{n-1}^{H} + \frac{1}{n}U_n^{\mathsf{H}}, \end{aligned}$$

where the last step follows from (a) $X_1, \dots X_{n-1}$ is *i.i.d.* P, (b) $\{Y_1, \dots Y_n\}/\{Y_{\lambda_n}\}$ is *i.i.d.* Q and (c) (X_n, Y_{λ_n}) is a pair taken from the optimal offline assignment of size n. The desired result thus follows. □

Remark A.3. *We provide an example to highlight the distinction between the limiting hindsight optimum and the fluid optimum, which in our setting is the optimal transport value between the distributions P and Q w.r.t. function φ. Suppose the supply and demand distributions $P = Q =$ Uniform$([0,1])$, and the quality function is $\varphi(X,Y) = \mathbb{1}\{X = Y\}$, then for any finite number of supply and demand units n, the hindsight optimum value $U_n(P,Q,\varphi) = 0$. Consequently, $U_\infty(P,Q,\varphi) = 0$ whereas the fluid optimum is 1.*

B Proof and Additional Discussion on the Failure of Greedy in Section 3.1

In this section, we will provide a proof for Proposition 1 and state and prove Proposition B.3.

B.1 Proof of Proposition 1

Proof of Proposition 1. We assume that the quality function is $\varphi(X,Y) = -|X-Y|^p$. Instead of taking the quality function maximizing perspective, we will take the cost minimization perspective and consider the cost function $d(X,Y) = |X-Y|^p$. Note that the we will use $U_n(\mathsf{Greedy}; P, Q), U_n^{\mathsf{H}}(P,Q)$ and $U_\infty(P,Q)$ to denote the average matching cost under Greedy, the hindsight optimal cost and the limiting hindsight cost. Note that $U_n^{\mathsf{H}}(P,Q) \geq U_\infty(P,Q)$ and hence we have that

$$\textsc{Reg}_n(\mathsf{Greedy}; P, Q) = U_n(\mathsf{Greedy}; P, Q) - U_\infty(P,Q) \geq U_n(\mathsf{Greedy}; P, Q) - U_n^{\mathsf{H}}(P,Q) \tag{B.1}$$

Note that since $f_p(x) > 0$ for all $x \in [0,1]$, we have that F_P and F_P^{-1} are strictly increasing functions over the interval $[0,1]$, where $F_P^{-1}(y) = \{x : F_P(x) = y\}$. Fix $\rho \in (0,1)$ and let $F_P(0.5) := \alpha \neq 1-\rho$ and let $\bar{\alpha} := \bar{F}_P(0.5) = 1 - F_P(0.5)$. Let us assume that $\alpha < 1-\rho$. Note that this assumption is without any loss because if $\alpha > 1-\rho$, then the entire analysis can be symmetrically made for the supply units located at 0.

Define $r \triangleq \rho/(1-\alpha)$. Note that since $\alpha < 1-\rho$ (by assumption), we have that $r < 1$. Define $\delta \triangleq 1-\alpha-\rho > 0$. Note that $(1-r)\rho = r\delta$. Define $\underline{x}^\star \triangleq F_P^{-1}(\alpha+\delta/2)$, $x^\star \triangleq F_P^{-1}(1-\rho)$ and $\overline{x}^\star \triangleq F_P^{-1}(1-\rho/2)$. Since $\alpha < \alpha+\delta/2 < 1-\rho < 1-\rho/2$ and F_P^{-1} is strictly increasing, we have that $0.5 < \underline{x}^\star < x^\star < \overline{x}^\star$.

For $x_2 > x_1$ and $t_2 > t_1$, define $N_{t_1:t_2}^X(x_1,x_2) \triangleq \sum_{k=t_1}^{t_2} \mathbb{1}\{X_k \in (a,b)\}$ to be the number of demand units that arrive in the time interval $\{t_1, t_1+1, \ldots, t_2\}$ and are located in the interval (x_1,x_2). Let $N_1^Y = \sum_{k=1}^n \mathbb{1}\{Y_k = 1\}$ denote the number of supply units located at 1.

We will now define some events

$$
\begin{aligned}
E^Y &= \{\rho n - \sqrt{n}/16 \le N_1^Y \le \rho n - \sqrt{n}/8\} \\
L_1^X &= \{\delta r n/2 - \sqrt{n}/32 \le N_{1:\lfloor rn \rfloor}^X(0.5, \underline{x}^\star) \le \delta r n/2 + \sqrt{n}/32\} \\
L_2^X &= \{\delta r n/2 - \sqrt{n}/32 \le N_{1:\lfloor rn \rfloor}^X(\underline{x}^\star, x^\star) \le \delta r n/2 + \sqrt{n}/32\} \\
L_3^X &= \{\rho r n/2 - \sqrt{n}/32 \le N_{1:\lfloor rn \rfloor}^X(x^\star, \overline{x}^\star) \le \rho r n/2 + \sqrt{n}/32\} \\
L_4^X &= \{\rho r n/2 - \sqrt{n}/32 \le N_{1:\lfloor rn \rfloor}^X(\overline{x}^\star, 1) \le \rho r n/2 + \sqrt{n}/32\} \\
H_1^X &= \{\delta(1-r)n/2 - \sqrt{n}/32 \le N_{\lfloor rn \rfloor+1:n}^X(0.5, \underline{x}^\star) \le \delta(1-r)n/2 + \sqrt{n}/32\} \\
H_2^X &= \{\delta(1-r)n/2 - \sqrt{n}/32 \le N_{\lfloor rn \rfloor+1:n}^X(\underline{x}^\star, x^\star) \le \delta(1-r)n/2 + \sqrt{n}/32\} \\
H_3^X &= \{\rho(1-r)n/2 - \sqrt{n}/32 \le N_{\lfloor rn \rfloor+1:n}^X(x^\star, \overline{x}^\star) \le \rho(1-r)n/2 + \sqrt{n}/32\} \\
H_4^X &= \{\rho(1-r)n/2 - \sqrt{n}/32 \le N_{\lfloor rn \rfloor+1:n}^X(\overline{x}^\star, 1) \le \rho(1-r)n/2 + \sqrt{n}/32\}
\end{aligned}
$$

We have that $N_1^Y \sim \text{Binomial}(n, \rho)$ and hence using CLT there exists a constant $c > 0$ such that $\mathbb{P}(E^Y) \ge c$. We have that $\mathbb{P}(X \in (0.5, \underline{x}^\star)) = F_P(\underline{x}^\star) - F_P(0.5) = \delta/2$. Similarly, we have that $\mathbb{P}(X \in (\underline{x}^\star, x^\star)) = \delta/2$, $\mathbb{P}(X \in (x^\star, \overline{x}^\star)) = \rho/2$ and $\mathbb{P}(X \in (\overline{x}^\star, 1)) = \rho/2$. Therefore, using standard CLT arguments, we have that there exists a constant $c > 0$ such that $\mathbb{P}(L_i^X) \ge c$ for all $i \in \{1, 2, 3, 4\}$ and $\mathbb{P}(H_i^X) \ge c$ for all $i \in \{1, 2, 3, 4\}$. Define $G = E^Y \cap \left(\cap_{i=1}^4 L_i^X\right) \cap \left(\cap_{i=1}^4 H_i^X\right)$, then using standard conditioning arguments (refer to Besbes et al. [2023]), we have that for some constant $\beta > 0$, $\mathbb{P}(G) \ge \beta$.

Hindsight Optimum. Under the event G, we have that $N_{1:n}^X(x^\star, 1) \ge \rho n - \sqrt{n}/8$ and $N_1^Y \le \rho n - \sqrt{n}/8$. Therefore, under the event G, the hindsight optimal will match all the supply units located to 1 to the demand units that arrive in the interval $(x^\star, 1)$. Note that since $N_{1:n}^X(x^\star, 1) \ge \rho n - \sqrt{n}/8$, not all demand units in the interval $(x^\star, 1)$ will be matched to the supply units located at 1. The demand units not matched to the supply units located 1 will be matched to the supply units located at 0.

Greedy Algorithm. Under the event G, we have that $N_{1:\lfloor rn \rfloor}^X(0.5, 1) \ge \rho n - \sqrt{n}/8$ and $N_1^Y \le \rho n - \sqrt{n}/8$. Therefore under the event G, the Greedy algorithm will match all the supply units located at 1 to the demand units that arrive in the interval $(0.5, 1)$ up till the time $t = \lfloor rn \rfloor$. For

$t \geq \lfloor rn \rfloor + 1$, Greedy will match all the demand units to the supply units located at 0. This includes the demand units that will arrive in the interval $(x^\star, 1)$. The two places that Greedy differs from the hindsight optimal matching is:

(*i*) at least $\delta rn - \sqrt{n}/8$ of the demand arrivals in the interval $(0.5, x^\star)$ during the first $\lfloor rn \rfloor$ time steps are matched to the supply units at 1 under Greedy whereas they are matched to the supply units at 0 under the hindsight optimal matching. This is because all the arrivals in $(0.5, x^\star)$ are matched under the hindsight optimal to the supply units at 0. Under Greedy, we have that at most $\rho rn + \sqrt{n}/16$ arrivals in the interval $(x^\star, 1)$ get matched to supply units at 1 and since there are at least $\rho n - \sqrt{n}/16$ supply units at 1, we have the at least $\delta rn - \sqrt{n}/8$ of the arrivals in $(0.5, x^\star)$ get matched to supply units at 1.

(*ii*) at least $\rho(1-r)n - \sqrt{n}/8$ of the demand arrivals in the interval $(x^\star, 1)$ during the last $n - \lfloor rn \rfloor$ time steps are matched to the supply units at 0 under Greedy whereas under the hindsight optimal matches these demand units are matched the supply units at 1. This is because all the demand arrivals in the last $n - \lfloor rn \rfloor$ time steps are matched the supply units at 0 under Greedy since Greedy prematurely exhausts all the supply units at 1. Since there are $\rho n - \sqrt{n}/16$ supply units at location 1 and at most $\rho n + \sqrt{n}/16$ of the demand arrivals in the interval $(x^\star, 1)$ get matched in first $\lfloor rn \rfloor$ time steps, we have that at least $\rho(1-r)n - \sqrt{n}/8$ demand units are available to be matched in the last $n - \lfloor rn \rfloor$ time steps.

Let π^{g} denote the allocation under the Greedy algorithm. Then we have that

$$
\begin{aligned}
\text{REG}_n(\mathsf{Greedy}; P, Q) &\overset{(a)}{\geq} U_n(\mathsf{Greedy}; P, Q) - U_n^{\mathsf{H}}(P, Q), \\
&\overset{(b)}{=} n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_{\sigma} \sum_{t=1}^{n} c(X_t, Y_{\sigma_t})\right], \\
&\overset{(c)}{=} n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_{\sigma} \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) \middle| G\right] \mathbb{P}(G) \\
&\quad + n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_{\sigma} \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) \middle| G^c\right] \mathbb{P}(G^c), \\
&\overset{(d)}{\geq} \beta n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_{\sigma} \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) \middle| G\right],
\end{aligned}
$$

where (a) follows from (B.1), (b) from the definition of $U_n(\mathsf{Greedy}; P, Q)$ and $U_n^{\mathsf{H}}(P, Q)$, (c) follows from law of total expectation and (d) follows from the fact that $\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_\sigma \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) \geq 0$, $\mathbb{P}(G^c) \geq 0$ and $\mathbb{P}(G) \geq \beta$. Now it suffices to show that there exists a constant $\kappa > 0$ such that

$$n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_\sigma \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) \middle| G\right] \geq \kappa$$

Given the supply and demand units $\{Y_1, Y_2, \ldots, Y_n\}$ and $\{X_1, X_2, \ldots, X_n\}$, let $\sigma(X)$ and $\pi^{\mathsf{g}}(X)$ denote the hindsight optimal assignment and the Greedy assignment of demand unit X respectively. Furthermore, let $\mathcal{A}_{t_1:t_2}(a, b) = \{X_k : k \in \{t_1, t_1 + 1, \ldots, t_2\} \text{ and } X_k \in (a, b)\}$. Now we have can re-write the summation $\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}})$ and $\sum_{t=1}^{n} c(X_t, Y_{\sigma_t})$ as follows

$$\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) = \sum_{X \in \mathcal{A}_{1:n}(0,0.5)} c(X, \pi^{\mathsf{g}}(X)) + \sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \pi^{\mathsf{g}}(X))$$
$$\sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) = \sum_{X \in \mathcal{A}_{1:n}(0,0.5)} c(X, \sigma(X)) + \sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \sigma(X))$$

Note that under the event G, we have that $c(X, \pi^{\mathsf{g}}(X)) = c(X, \sigma(X))$ for all $X \in (0, 0.5)$ and therefore we have that under the event G,

$$\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_\sigma \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) = \sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \pi^{\mathsf{g}}(X)) - \sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \sigma(X))$$

For the Greedy algorithm we have that,

$$\begin{aligned}\sum_{X \in \mathcal{A}_{1:n}(0.5,1)} c(X, \pi^{\mathsf{g}}(X)) = &\sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(0.5,x^\star)} c(X, \pi^{\mathsf{g}}(X)) + \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(0.5,x^\star)} c(X, \pi^{\mathsf{g}}(X)) \\ &+ \sum_{X \in \mathcal{A}_{1:\lfloor rn \rfloor}(x^\star,1)} c(X, \pi^{\mathsf{g}}(X)) + \sum_{X \in \mathcal{A}_{\lfloor rn \rfloor+1:n}(x^\star,1)} c(X, \pi^{\mathsf{g}}(X))\end{aligned}$$

For the hindsight optimal we have that,

$$\sum_{X\in\mathcal{A}_{1:n}(0.5,1)} c(X,\sigma(X)) = \sum_{X\in\mathcal{A}_{1:\lfloor rn\rfloor}(0.5,x^\star)} c(X,\sigma(X)) + \sum_{X\in\mathcal{A}_{\lfloor rn\rfloor+1:n}(0.5,x^\star)} c(X,\sigma(X)) + \sum_{X\in\mathcal{A}_{1:\lfloor rn\rfloor}(x^\star,1)} c(X,\sigma(X)) + \sum_{X\in\mathcal{A}_{\lfloor rn\rfloor+1:n}(x^\star,1)} c(X,\sigma(X))$$

Under the event G, we have that for all $X \in \mathcal{A}_{\lfloor rn\rfloor+1:n}(0.5,x^\star)$, we have that $c(X,\pi^{\mathrm{g}}(X)) = c(X,\sigma(X))$ and we have that $|\sum_{X\in\mathcal{A}_{1:\lfloor rn\rfloor}(x^\star,1)} c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X))| \le \sqrt{n}$.
Therefore we have that

$$\sum_{X\in\mathcal{A}_{1:n}(0.5,1)} (c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X))) \ge \sum_{X\in\mathcal{A}_{1:\lfloor rn\rfloor}(0.5,x^\star)} c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X)) + \sum_{X\in\mathcal{A}_{\lfloor rn\rfloor+1:n}(x^\star,1)} c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X)) - \sqrt{n}$$

Next we will provide a lower bound for the sum $\sum_{X\in\mathcal{A}_{1:\lfloor rn\rfloor}(0.5,x^\star)}(c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X)))$. Consider the demand arrivals in $\mathcal{A}_{1:\lfloor rn\rfloor}(0.5,\underline{x}^\star)$, we have that $c(X,\pi^{\mathrm{g}}(X)) \ge (1-\underline{x}^\star)^p$ and $c(X,\sigma(X)) \le (\underline{x}^\star)^p$ and this is true for at least $\delta rn/2 - \sqrt{n}$ arrivals. Similarly, consider the demand arrivals in $\mathcal{A}_{1:\lfloor rn\rfloor}(\underline{x}^\star,x^\star)$, we have that $c(X,\pi^{\mathrm{g}}(X)) \ge (1-x^\star)^p$ and $c(X,\sigma(X)) \le (x^\star)^p$ and this is also true for at least $\delta rn/2 - \sqrt{n}$. Therefore we have that

$$\sum_{X\in\mathcal{A}_{1:\lfloor rn\rfloor}(0.5,x^\star)} c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X)) \ge [(1-\underline{x}^\star)^p + (1-x^\star)^p - (\underline{x}^\star)^p - (x^\star)^p](\delta rn/2) - 2\sqrt{n}$$

Next we will provide a lower bound for the sum $\sum_{X\in\mathcal{A}_{\lfloor rn\rfloor+1:n}}(x^\star,1)(c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X)))$. Consider the demand arrivals in $\mathcal{A}_{\lfloor rn\rfloor+1:n}(x^\star,\overline{x}^\star)$, we have that $c(X,\pi^{\mathrm{g}}(X)) \ge (x^\star)^p$ and $c(X,\sigma(X)) \le (1-x^\star)^p$ and this is true for at least $\rho(1-r)n/2 - \sqrt{n}$ arrivals. Similarly, consider the demand arrivals in $\mathcal{A}_{\lfloor rn\rfloor+1:n}(\overline{x}^\star,1)$, we have that $c(X,\pi^{\mathrm{g}}(X)) \ge (\overline{x}^\star)^p$ and $c(X,\sigma(X)) \le (1-\overline{x}^\star)^p$ and this is true for at least $\rho(1-r)n/2 - \sqrt{n}$ arrivals. Therefore we have that

$$\sum_{X\in\mathcal{A}_{\lfloor rn\rfloor+1:n}(x^\star,1)} (c(X,\pi^{\mathrm{g}}(X)) - c(X,\sigma(X))) \ge [-(1-\overline{x}^\star)^p - (1-x^\star)^p + (\overline{x}^\star)^p + (x^\star)^p](\rho(1-r)n/2) - 2\sqrt{n}$$

Define $\beta' = (1-\underline{x}^\star)^p - (1-\overline{x}^\star)^p + (\overline{x}^\star)^p - (\underline{x}^\star)^p > 0$. Since $\rho(1-r) = \delta r$, we have that

$$\sum_{X \in \mathcal{A}_{1:n}(0.5,1)} (c(X, \pi^{\mathsf{g}}(X)) - c(X, \sigma(X))) \geq \beta' \cdot \frac{\rho(1-\rho-\alpha)}{1-\alpha} \cdot n - 5\sqrt{n}$$

Therefore we have that

$$n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} c(X_t, Y_{\pi_t^{\mathsf{g}}}) - \min_{\sigma} \sum_{t=1}^{n} c(X_t, Y_{\sigma_t}) \middle| G\right] \geq \beta' \frac{\rho}{1-\alpha}(1-\rho-\alpha) - 5/\sqrt{n}$$

This concludes the proof as we can choose a large enough $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, the lower bound in the Proposition 1 holds. □

B.2 Additional Discussion on the Sub-optimality of Greedy

In this section, we provide additional evidence regarding the sub-optimality of Greedy. The formal statement of the results require the following mathematical notion.

Definition 2 (Lévy–Prokhorov (LP) Distance)**.** *For two probability measures P and Q defined on $\mathbb{R}$ with CDF F_P and F_Q, we recall that the LP distance $\Delta_{\mathsf{LP}}(P,Q)$ is defined as*

$$\Delta_{\mathsf{LP}}(P,Q) = \inf\left\{\epsilon > 0 : F_P(x-\epsilon) - \epsilon < F_Q(x) < F_P(x+\epsilon) + \epsilon, \forall x \in \mathbb{R}\right\}.$$

More generally, the LP distance $\Delta_{\mathsf{LP}}(\lambda_1, \lambda_2)$ for a pair of probability measures on a metric space (X,d) is defined as:

$$\Delta_{\mathsf{LP}}(\lambda_1,\lambda_2) = \inf\left\{\epsilon > 0 : \lambda_1(A) \leq \lambda_2(A^\epsilon) + \epsilon \text{ and } \lambda_2(A) \leq \lambda_1(A^\epsilon) + \epsilon \text{ for all measurable } A \subseteq X\right\},$$

where A^ϵ is the ϵ-enlargement of the set A, defined as: $A^\epsilon = \{x \in X : \exists a \in A \text{ such that } d(x,a) < \epsilon\}$.

Proposition B.3. *Suppose the demand distribution P and supply distribution Q are one dimensional continuous distributions supported over bounded open intervals $\mathcal{I}_P$ and $\mathcal{I}_Q$, with densities bounded above and below, i.e., there exists $\gamma \in [1,\infty)$ such that $\gamma^{-1} \leq f_P(x) \leq \gamma$ for all $x \in \mathcal{I}_P$ and $\gamma^{-1} \leq f_Q(x) \leq \gamma$ for all $x \in \mathcal{I}_Q$. Consider the quality function $\varphi(X,Y) = -|X-Y|^p$ for some $p > 1$. Suppose $P \neq Q$ and the LP distance between them is strictly positive, $\Delta_{\mathsf{LP}}(P,Q) > 0$. For any matching policy π, let μ_n^π denote the induced empirical (joint) measure on the set of demand* *nodes $X_1 \dots, X_n$ and supply nodes $Y_1, \dots, Y_n$. More precisely, $\mu_n^\pi = \frac{1}{n}\sum_{i=1}^n \delta_{\left(X_i, Y_{\pi_i}\right)}$ with $\delta_{(\cdot)}$ the Dirac measure. In particularly, μ_n^{g} and μ_n^{hind} denote the empirical (joint) distributions induced by the greedy algorithm and the hindsight optimal algorithm, respectively. Then there exists a universal constant $c = c(\gamma, p, \Delta_{\mathsf{LP}}(P, Q)) > 0$ such that*

$$\liminf_{n\to\infty} \Delta_{\mathsf{LP}}(\mu_n^{\mathsf{g}}, \mu_n^{\mathsf{hind}}) \geq c\,,$$

where $\Delta_{\mathsf{LP}}(\cdot,\cdot)$ is the LP distance between two probability measures.

Conjecture 1. *Suppose $\varphi(X, Y) = -|X - Y|^p$ with $p > 1$. If $\liminf_{n\to\infty} \Delta_{\mathsf{LP}}(\mu_n^{\mathsf{g}}, \mu_n^{\mathsf{hind}}) > 0$, then* Greedy *incurs non-vanishing regret as n scales, namely*

$$\liminf_{n\to\infty} \textsc{Reg}_n(\mathsf{Greedy}) > 0.$$

The proof of Proposition B.3 relies on the following well-known structural result of the hindsight optimal matching under matching quality functions $-|X - Y|^p$ for $p > 1$.

Lemma B.1. *Under the matching quality function $\varphi(X, Y) = -|X - Y|^p$ for $p > 1$, the hindsight-optimal matching is unique and assortative. Specifically, the optimal matching first sorts both the demand units and the supply units in descending order, then matches the i-th demand unit to the i-th supply unit for all $1 \leq i \leq n$. More generally, the optimal transport between probability measures P and Q with respect to $\varphi(X, Y)$ is given by $T(X) = F_Q^{-1}(F_P(X))$, where F_P and F_Q are the CDFs of P and Q, respectively.*

Lemma B.1 is a well-known result in optimal transport theory. In fact, the optimal matching is assortative as long as $\frac{\partial^2 \varphi(x,y)}{\partial x \partial y} > 0$, which is satisfied by the class of quality functions $-|X - Y|^p$. We omit the proof of Lemma B.1 and refer the interested reader to classic textbooks, e.g. Galichon [2018] for details.

Proof of Proposition B.3. Suppose $\Delta_{\mathsf{LP}}(P, Q) > \delta > 0$. Then there must (WLOG) exist $x \in \mathbb{R}$ such that $F_P(x - \delta) \geq F_Q(x) + \delta$. Note that we can WLOG assume that $x \in \mathcal{I}_Q$, because on the one hand $F_Q(x) + \delta \leq F_P(x - \delta) \leq 1$ implies $F_Q(x) \leq 1 - \delta$, on the other hand if $F_Q(x) = 0$, it is WLOG to set x to be the left end point of $\mathcal{I}_Q$, as $F_P(x - \delta)$ only increases and the inequality $F_P(x - \delta) \geq F_Q(x) - \delta$ remains valid. Similarly, we may assume WLOG that $x - \delta \in \mathcal{I}_P$. Let

$\eta = \frac{\delta}{8\gamma^3}$. Denote by $A = A_1 \times A_2 = [x-\delta-\eta, x-\delta+\eta] \times (-\infty, x+2\eta]$. Let $\xi = \frac{\delta}{16\gamma^4}$. For notational simplicity, we assume $\xi n \in \mathbb{Z}$. We proceed with the proof by stating a sequence of claims. We provide proof sketches of the claims to simplify exposition.

Claim 1. $\liminf_{n\to\infty} \mathbb{P}\left(|i \in [1,n] : \ Y_i \in [x, x+2\eta]| > \xi n\right) = 1.$

Recall that $Y_1, \ldots, Y_n$ are *i.i.d.* Q random variables. We have

$$|i \in [1,n] : \ Y_i \in [x, x+2\eta]| = \sum_{i=1}^{n} \mathbb{1}\{Y_i \in [x, x+2\eta]\}.$$

Since $x \in \mathcal{I}_Q$ and $F_Q(x) \leq 1-\delta$, we thus have $[x, x+2\eta] \subseteq [x, x+\delta/\gamma] \subseteq \mathcal{I}_Q$ (where we used $f_Q(\cdot) \leq \gamma$ on $\mathcal{I}_Q$). We thus have $\mathbb{E}[\mathbb{1}\{x \leq Y_i \leq x+2\eta\}] = \mathbb{P}(x \leq Y_i \leq x+2\eta) \geq 2\gamma^{-1}\eta$. Since $2\gamma^{-1}\eta > \xi$, Claim 1 follows from the weak law of large numbers.

Claim 2. $\liminf_{n\to\infty} \mathbb{P}\left(|i \in [1,\xi n] : \ X_i \in A_1| > \xi^2 n\right) = 1.$

Similar to Claim 1, here we use the fact that $x-\delta \in \mathcal{I}_P$. $\mathcal{I}_P$ is an interval with $|\mathcal{I}_P| > \eta$, hence either $[x-\delta-\eta, x-\delta] \subseteq \mathcal{I}_P$ or $[x-\delta, x-\delta+\eta] \subseteq \mathcal{I}_P$. Therefore $|i \in [1,\xi n] : \ X_i \in A_1| = \sum_{i=1}^{\xi n} \mathbb{1}\{X_i \in A_1\}$ with $\mathbb{E}[\mathbb{1}\{X_i \in A_1\}] = \mathbb{P}(X_i \in A_1) \geq \gamma^{-1}\eta > \xi$. Claim 2 again follows from the weak law of large numbers.

Claim 3. $\liminf_{n\to\infty} \mathbb{P}\left(|i \in [1,n] : \ X_i \in A_1, Y_{\mathsf{g}_i} \in A_2| > \xi^2 n\right) = 1.$

Essentially, we argue that those among the first ξn demand units $X_1 \ldots X_{\xi n}$ that belong to A_1 will all be matched to supply units in A_2 under Greedy. Observe that by Claim 1, with probability 1 there are more than ξn supply units that are initially available in $[x, x+2\eta]$. Hence with probability 1 upon the arrival of the (ξn)-th demand unit, there is at least one supply unit in $[x, x+2\eta]$ that remains available. We make the following observation.

Observation 1. *If an arrival demand unit belongs to A_1, and there is at least one available supply unit in $[x, x+2\eta]$, then* Greedy *matches the demand unit to a supply unit in A_2.*

Combining Observation 1 with Claim 1, we thus have $|i \in [1,n] : \ X_i \in A_1, Y_{\mathsf{g}_i} \in A_2| \geq |i \in [1,\xi n] : \ X_i \in A_1, Y_{\mathsf{g}_i} \in A_2| = |i \in [1,\xi n] : \ X_i \in A_1|$. Now by Claim 2, we conclude the proof of Claim 3. Claim 3 immediately implies that $\mu_n^{\mathsf{g}}(A) = \mu_n^{\mathsf{g}}(A_1 \times A_2) = \frac{1}{n}\sum_{i=1}^{n} 1\left((X_i, Y_{\mathsf{g}_i}) \in A\right)) > \xi^2$ with probability 1 in the n limit. Now on the other hand, we treat μ_n^{hind} in the next claim.

Claim 4. *Let μ^{OT} denote the optimal transport (joint measure) between P and Q. Then $\mu^{\mathsf{OT}}(A^\epsilon) = \mu^{\mathsf{OT}}(A_1^\epsilon \times A_2^\epsilon) = 0$ for any $\epsilon \leq \frac{\delta}{8\gamma^3}$. Here for any interval $B = [l_B, r_B]$, the ϵ-enlargement $A^\epsilon = [l_B - \epsilon, r_B + \epsilon]$.*

By definition $A^\epsilon = A_1^\epsilon \times A_2^\epsilon = [x-\delta-\eta-\epsilon, x-\delta+\eta+\epsilon] \times (-\infty, x+2\eta+\epsilon]$. Since $F_P(x-\delta) \geq F_Q(x)+\delta$ and our assumption on the boundedness of f_Q, we know that $F_Q^{-1}(F_P(x-\delta)) \geq x+\gamma^{-1}\delta$. Furthermore, $F_Q^{-1}(F_P(x-\delta-\eta-\epsilon)) \geq x+\gamma^{-1}\delta-\gamma^2(\eta+\epsilon)$. To see this:

$$F_P(x-\delta-\eta-\epsilon) \overset{(a)}{\geq} F_P(x-\delta)-\gamma(\eta+\epsilon) \geq F_Q(x+\gamma^{-1}\delta)-\gamma(\eta+\epsilon) \overset{(b)}{\geq} F_Q(x+\gamma^{-1}\delta-\gamma^2(\eta+\epsilon)),$$

where (a) follows as $f_P(x) \leq \gamma$; (b) follows as $f_Q(x) \geq 1/\gamma$. By Lemma B.1, the optimal transport between P and Q with respect to φ matches any demand unit in A_1^ϵ to a supply unit that is greater than $x+\gamma^{-1}\delta-\gamma^2(\eta+\epsilon)$. Notice that the right end point of interval A_2^ϵ is $x+2\eta+\epsilon$, which is smaller than $x+\gamma^{-1}\delta-\gamma^2(\eta+\epsilon)$ for any $\epsilon \leq \frac{\delta}{8\gamma^3}$. As a result $\mu^{\mathsf{OT}}(A_1^\epsilon \times A_2^\epsilon) = 0$.

Claim 5. *For any measurable set A' it holds true that $\lim_{n\to\infty}(\mu_n^{\mathsf{hind}}(A')) = \mu^{\mathsf{OT}}(A')$.*

We omit the proof. Combining Claim 4 and 5, we conclude that $\lim_{n\to\infty} \mu_n^{\mathsf{hind}}(A^\epsilon) = 0$ for all $\epsilon \leq \frac{\delta}{8\gamma^3}$. Now combining the above with the definition of LP distance (Definition 2), we conclude that $\liminf_{n\to\infty} \Delta_{\mathsf{LP}}\left(\mu_n^{\mathsf{g}}, \mu_n^{\mathsf{hind}}\right) \geq \min\left(\epsilon, \xi^2\right) \geq \frac{\delta^2}{256\gamma^8}$. Combining the above completes the proof of Proposition B.3 with the specific constant choice of $c = \frac{(\Delta_{\mathsf{LP}}(P,Q))^2}{256\gamma^8}$ □

C Proof of Corollary 1

Proof of Corollary 1. From the proof of Lemma A.1, $U_\infty(P,Q,\varphi) = \lim_{n\to\infty} U_n^{\mathsf{H}}(P,Q,\varphi)$ exists. Equivalently, $\{\textsc{Reg}_k(\mathsf{H\text{-}OPT})\}_{k\geq 1}$ is a non-negative monotone decreasing sequence and

$$\lim_{k\to\infty} \textsc{Reg}_k(\mathsf{H\text{-}OPT}) = 0.$$

By Remark 4, we have $\textsc{Reg}_n(\mathsf{SOAR}) = \frac{1}{n}\sum_{k=1}^n \textsc{Reg}_k(\mathsf{H\text{-}OPT})$, which also converges to 0 as $n \to \infty$. □

D Proof of Corollary 2

Proof of Corollary 2. First, we show $\beta \le 1$. By definition, $\textsc{Reg}_n(\text{H-OPT}) = U_\infty - U_n^{\mathsf{H}}$. As the cumulative regret is at least a constant, $U_\infty - U_n^{\mathsf{H}} \ge \frac{C}{n}$. Take lim sup on both sides, $\limsup_{n\to\infty} n^{\beta-\epsilon} \cdot (U_\infty - U_n^{\mathsf{H}}) \ge \limsup_{n\to\infty} n^{\beta-1-\epsilon} = 0$, thus $\beta < 1+\epsilon$ for all $\epsilon > 0$.

1) $\limsup_{n\to\infty} n^{\beta-\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) = 0$.

As $\limsup_{n\to\infty} n^{\beta-\epsilon} \cdot (U_\infty - U_n^{\mathsf{H}}) = 0$, for any $\delta > 0$, there exists $N_\delta \in \mathbb{N}$ such that $n^{\beta-\epsilon} \cdot (U_\infty - U_n^{\mathsf{H}}) < \delta$ for $n \ge N_\delta$. Then we have for all $n \in \mathbb{N}$,

$$\textsc{Reg}_n(\mathsf{SOAR}) \overset{(a)}{=} U_\infty - U_n(\mathsf{SOAR}) \overset{(b)}{=} \frac{1}{n}\sum_{k=1}^{n}(U_\infty - U_k^{\mathsf{H}}) < \frac{N_\delta U_\infty}{n} + \frac{\delta}{n}\sum_{k=N_\delta+1}^{n} k^{-\beta+\epsilon} \overset{(c)}{<} \frac{N_\delta U_\infty}{n} + \frac{\delta/(-\beta+\epsilon+1)}{n^{\beta-\epsilon}},$$

where (a) follows by definition; (b) follows Theorem 1; and (c) follows as $\epsilon < \beta$, and $\sum_{k=N_\delta+1}^{n} k^{-\beta+\epsilon} \le \int_{N_\delta}^{n} k^{-\beta+\epsilon}\, dk$. Therefore, for any $\delta' > 0$, let $N_{\delta'}$ be the minimum n such that $N_\delta U_\infty n^{\beta-\epsilon-1} + \frac{\delta}{-\beta+\epsilon+1} < \delta'$, where the existence of $N_{\delta'}$ follows from $\beta < 1+\epsilon$. Then we have for all $n \ge N_{\delta'}$,

$$0 < n^{\beta-\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) < N_\delta U_\infty n^{\beta-\epsilon-1} + \frac{\delta}{-\beta+\epsilon+1} < \delta'.$$

Therefore $\limsup_{n\to\infty} n^{\beta-\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) = 0$.

2) $\limsup_{n\to\infty} n^{\beta+\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) = \infty$.

As $\liminf_{n\to\infty} n^{\beta+\epsilon} \cdot (U_\infty - U_n^{\mathsf{H}}) = \infty$, for any $M > 0$, there exists $N_M \in \mathbb{N}$ such that $n^{\beta+\epsilon} \cdot \textsc{Reg}_n(\text{H-OPT}) > M$ for $n \ge N_M$. Then we have for all $n > N_M$,

$$\begin{aligned}\textsc{Reg}_n(\mathsf{SOAR}) &> \frac{1}{n}\sum_{k=1}^{N_M}\frac{C}{k} + \frac{1}{n}\sum_{k=N_M+1}^{n}\frac{M}{k^{\beta+\epsilon}} > \frac{1}{n}\int_1^{N_M}\frac{C}{k}\,dk + \frac{1}{n}\int_{N_M+1}^{n}\frac{M}{k^{\beta+\epsilon}}\,dk \\ &= \frac{C\log(N_M)}{n} + \frac{M/(1-\beta-\epsilon)}{n}(n^{1-\beta-\epsilon} - (N_M+1)^{1-\beta-\epsilon}).\end{aligned}$$

- If $\beta + \epsilon > 1$, then

$$n^{\beta+\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) > C\log(N_M) n^{\beta+\epsilon-1},$$

thus $\liminf_{n\to\infty} n^{\beta+\epsilon}\textsc{Reg}_n(\mathsf{SOAR}) = \infty$.

- If $\beta + \epsilon < 1$, then

$$n^{\beta+\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) > \frac{M}{1-\beta-\epsilon}(1-(N_M+1)/n)^{1-\beta-\epsilon}).$$

Therefore, for any $M' > 0$, let $N_{M'}$ be the minimum n such that $\frac{M}{1-\beta-\epsilon}(1-(N_M+1)/n)^{1-\beta-\epsilon}) > M'$, where the existence of $N_{M'}$ follows from $\beta + \epsilon < 1$. Then for all $n \geq N_{M'}$,

$$n^{\beta+\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) > M'.$$

Therefore $\limsup_{n\to\infty} n^{\beta+\epsilon} \cdot \textsc{Reg}_n(\mathsf{SOAR}) = \infty$.

3) $\textsc{Reg}_n(\mathsf{SOAR}) \leq n^{\epsilon} \cdot \textsc{Reg}_n(\text{H-OPT})$ for n sufficiently large.

We have shown that

$$\textsc{Reg}_n(\mathsf{SOAR}) < \frac{N_\delta U_\infty}{n} + \frac{\delta/(-\beta+\epsilon+1)}{n^{\beta-\epsilon}} = \Theta(n^{-\beta+\epsilon}) \text{ for } n \in \mathbb{N},$$

and for any $M > 0$, there exists $N_M \in \mathbb{N}$ such that

$$\textsc{Reg}_n(\text{H-OPT}) > Mn^{-\beta-\epsilon} \text{ for } n \geq N_M.$$

Combining them gives $\textsc{Reg}_n(\mathsf{SOAR}) \leq n^{\epsilon} \cdot \textsc{Reg}_n(\text{H-OPT})$ for n sufficiently large.

Polynomial Regret Scaling Case.

If in addition, $\lim_{n\to\infty} n^{\beta} \cdot (U_\infty - U_n^{\mathsf{H}}) = l_0$, i.e. for any $\delta > 0$, there exists N such that for any $n > N$, $|n^{\beta} \cdot \textsc{Reg}_n(\text{H-OPT}) - l_0| < \delta$. Applying Theorem 1, we have for all $n \in \mathbb{N}$,

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{NU_\infty}{n} + \frac{1}{n}\sum_{k=N}^{n} \frac{l_0+\delta}{k^{\beta}} = \begin{cases} \Theta(n^{-\beta}), & \beta \neq 1, \\ \Theta(n^{-1}\log n), & \beta = 1. \end{cases}$$

Therefore we have

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \begin{cases} l_1 \textsc{Reg}_n(\text{H-OPT}), & \beta \neq 1, \\ l_1 \log n \cdot \textsc{Reg}_n(\text{H-OPT}), & \beta = 1. \end{cases}$$

Similarly we have for n sufficiently large,

$$\textsc{Reg}_n(\mathsf{SOAR}) \geq \frac{1}{n}\sum_{k=1}^{N}\frac{C}{k} + \frac{1}{n}\sum_{k=N+1}^{n}\frac{l_0-\delta}{k^\beta}\begin{cases}\Theta(n^{-\beta}), & \beta \neq 1,\\ \Theta(n^{-1}\log n), & \beta = 1.\end{cases}$$

Combining the above gives the tight characterization of the scaling of $\textsc{Reg}_n(\mathsf{SOAR})$. □

E Examples of Matching Instances Scale Regularly

In this section, we list some problem settings when the offline optima scale regularly. First, we introduce two useful results.

Theorem E.1 (Eq. (6) and (25) of Caracciolo et al. [2014])**.** *Suppose a set of n demand points and n supply points are generated independently and uniformly at random in the hypercube $[0,1]^d$, with matching cost $\|X-Y\|^p$. Then the average cost of the optimal assignment, denoted by $U_n^{\mathsf{H}}(p,d)$, is given by*

$$U_n^{\mathsf{H}}(2,d) \approx \begin{cases}\frac{1}{6n} + \frac{e_1^{(2)}}{n^2}, & d=1,\\ \frac{1}{2\pi}\frac{\ln n}{n} + \frac{e_2^{(2)}}{n}, & d=2,\\ e_d^{(2)}n^{-\frac{2}{d}} + \frac{\zeta_d(1)}{2\pi^2}n^{-1}, & d>2.\end{cases}$$

Here and in the following the symbol $\approx$ means that the term on the l.h.s. is asymptotically equal to the r.h.s. except for some additional terms decaying faster than each term in the r.h.s. (e.g. $U_n^{\mathsf{H}}(2,1) = \frac{1}{6n} + \frac{e_1^{(2)}}{n^2} + o\left(\frac{1}{n^2}\right)$). Conjectured based on numerical simulations:

$$U_n^{\mathsf{H}}(p,d) \approx e_d^{(p)}n^{-\frac{p}{d}} + \alpha_d^{(p)}n^{\frac{2-p-d}{d}} \quad \textit{for } d>2, p>0,$$

where coefficients $e_d^{(p)}, \alpha_d^{(p)}$ are constants and $\zeta_d(x)$ is the Epstein zeta function.

By Definition 1, as $U_\infty = 0$, and $U_n^{\mathsf{H}}(p,d) = \tilde{\Theta}(n^{-p/d})$, it is easy to see this matching instance scale regularly with $\beta = p/d$. For example, $\lim_{n\to\infty} n \cdot U_n^{\mathsf{H}}(2,1) = \Theta(1)$, thus SOAR gives a tight regret scaling.

F Extension to the setting of *i.i.d.* demand units with unknown distributions

The various regret scaling results of SOAR can be generalized to the setting where the demand feature distribution is unknown a priori to the platform but can be learned on-the-fly in a Bayesian fashion. More specifically, suppose the supply units are drawn *i.i.d* from distribution Q, while the demand units are drawn *i.i.d.* from distribution P that is *unknown a priori* to the platform, who only has a prior belief, characterized by a prior distribution ρ_0 over a candidate distribution set $\mathcal{P}$. Upon seeing the sequentially arriving demand units, the platform updates its posterior belief in a Bayesian fashion $\rho_t \leftarrow \rho_{t-1}|X_t$. The platform implements SOAR, which simulates *i.i.d.* future demand sequence from a distribution in $\mathcal{P}$ sampled according to the current posterior belief ρ_t at each time t to facilitate matching. We provide a detailed description of Bayesian-SOAR. At each time $t \geq 1$,

1. Observes the real demand X_t and updates posterior distribution $\rho_{t-1} \to \rho_t$ accordingly.
2. Draw a demand distribution from ρ_t, then draw *i.i.d.* virtual demands $\hat{X}_{t+1}, \ldots \hat{X}_n$ from that distribution and, together with the real demand X_t, form the input to the offline matching subroutine to determine the current-step matching decision.

For clarity, we denote SOAR in this Bayesian setting by Bayesian-SOAR. See Algorithm 2 for a detailed description. We establish an analogue of Theorem 1 for Bayesian-SOAR. We then derive a decomposition of the regret of Bayesian-SOAR into the regret of SOAR with a known demand distribution, and an extra learning cost. This result can be utilized to derive various regret scalings under additional distributional/structural assumptions.

Theorem F.1. *The average matching quality under* Bayesian-SOAR *is*

$$\begin{aligned} &U_n(\textit{Bayesian-SOAR}) \triangleq \mathbb{E}_{P\sim\rho_0} U_n(\textit{Bayesian-SOAR}; P, Q, \varphi), \\ &= \frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=1}^{n} \mathbb{E}_{\rho_{t-1}, P_{t-1}\sim\rho_{t-1}, \mathcal{Y}^{t-1}_{\textit{un-matched}}}\left[U^{\text{H-OPT}}_{n-t+1}(P_{t-1}, \mathcal{Y}^{t-1}_{\textit{un-matched}})\middle| P\right]\right], \end{aligned} \tag{F.2}$$

where P_{t-1} is the time $t-1$ "guess" of the true demand distribution, sampled from the current posterior belief ρ_{t-1}. Here $\mathcal{Y}^{t-1}_{\text{un-matched}} = \{Y_{j_1}, \ldots Y_{j_{n-t+1}}\}$ denotes the set of unmatched supply

Algorithm 2: Bayesian-SOAR

```
Input: supply units 𝒴 = {Y_1, Y_2, ..., Y_n}, prior distribution ρ_0 over a distribution set 𝒫,
       quality function φ
𝒴_matched ← ∅
for t ← 1 to n do
    Observe demand X_t and denote it as X̂_0 := X_t                       // re-label demand unit
    𝒴_un-matched := {Ỹ_0, Ỹ_1, ..., Ỹ_{n−t}} ← 𝒴 \ 𝒴_matched           // re-label supply units
    Draw a distribution P_{t−1} ~ ρ_{t−1}, then draw i.i.d. P_{t−1} simulated demand scenario
      𝒳_t^sim := {X̂_1, ..., X̂_{n−t}}                          // Simulate a future demand scenario
    Randomly permute the demand pool X̂_0 ∪ 𝒳_t^sim and denote the random permutation as
      σ
    Solve the maximum quality matching optimization problem in (F.1)   // Optimize based
      on a simulated demand scenario
```

$$\eta^\star \leftarrow \underset{\eta \text{ is a permutation}}{\arg\max} \sum_{k=0}^{n-t} \varphi(\hat{X}_{\sigma(k)}, \tilde{Y}_{\eta(k)}), \tag{F.1}$$

```
    Allocate Ỹ_{η*(σ(0))} to X_t          // Allocate the supply unit according to the optimal
      permutation
    𝒴_matched ← 𝒴_matched ∪ {Ỹ_{η*(σ(0))}}
    Update ρ_t ← ρ_{t−1} | X_t              // Bayesian update given observed real demand X_t
end
```

units before X_t arrives.

$$U^{\text{H-OPT}}_{n-t+1}(\tilde{P}, \mathcal{Y}^{t-1}_{\textit{un-matched}}) \triangleq \frac{1}{n-t+1}\mathbb{E}_{X_t, \hat{X}_{t+1}, \ldots, \hat{X}_n \overset{\text{i.i.d.}}{\sim} \tilde{P}} \left[\max_{\tau \in \mathcal{S}_{n-t+1}} \varphi\left(X_t, Y_{j_{\tau_1}}\right) + \sum_{i=1}^{n-t} \varphi\left(\hat{X}_{t+i}, Y_{j_{\tau_{i+1}}}\right)\right].$$

denote the hindsight optimum average matching quality between $n-t+1$ i.i.d. $\tilde{P}$ demand units and supply units in $\mathcal{Y}^{t-1}_{\textit{un-matched}}$.

Corollary F.2. *The regret of* Bayesian-SOAR *can be decomposed as follows*

$$\begin{aligned}\text{REG}_n(\textit{Bayesian-SOAR}) &= \mathbb{E}_{P\sim\rho_0}\left[\text{REG}_n(\mathsf{SOAR}; P, Q, \varphi)\right] \\ &+ \frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=1}^{n} \mathbb{E}_{\rho_{t-1}, P_{t-1}\sim\rho_{t-1}, \mathcal{Y}^{t-1}_{\textit{un-matched}}}\left[U^{\text{H-OPT}}_{n-t+1}(P, \mathcal{Y}^{t-1}_{\textit{un-matched}}) - U^{\text{H-OPT}}_{n-t+1}(P_{t-1}, \mathcal{Y}^{t-1}_{\textit{un-matched}}) \,\middle|\, P\right]\right],\end{aligned}$$

where $\text{REG}_n(\mathsf{SOAR}; P, Q, \varphi)$ *denotes the regret of* SOAR *in the setting with known demand and supply distributions* P *and* Q.

Remark F.3. *The expectation in the second term of eq. (F.2) in Theorem F.1 involves nested expectations. The outer expectation is taken over the randomness in the true underlying demand distribution* P. *Conditional on* P *and for each fixed* $t \geq 0$, *the inner expectation is taken over the randomness of (i) the posterior distribution* ρ_{t-1} *on the demand distribution, whose randomness comes from the* i.i.d. *(real) demand sequence* $X_1, \ldots X_{t-1}$ *drawn from* P, *(ii) the "guess" of the demand distribution at time* t, *which is a distribution sampled from the posterior distribution* ρ_{t-1}, *and (iii) the set of un-matched supply units,* $\mathcal{Y}^{t-1}_{\textit{un-matched}}$, *which is determined by the initial set of* i.i.d. Q *supply units* $\mathcal{Y}^{0}_{\textit{un-matched}}$, *and the supply units that were matched to the previous demand units* $X_1, \ldots X_{t-1}$, *selected by the* Bayesian-SOAR *algorithm. Later in Lemma F.6, we prove that* $\mathcal{Y}^{t}_{\textit{un-matched}}$ *has a simple probabilistic characterization thanks to symmetry.*

Remark F.4. *Corollary F.2 decomposes the regret incurred by* SOAR *into two terms. The first term captures the familiar regret term* $\text{REG}_n(\mathsf{SOAR}, P, Q, \varphi)$ *as in Remark 4, quantifying the quality loss even when the platform knows the underlying demand distribution* P. *In addition to this term,* Bayesian-SOAR *pays an extra regret term due to not knowing the demand distribution a priori. Here* ρ_{t-1} *(for* $t \geq 2$*) is random — its randomness comes from* $X_1, \ldots X_{t-1}$. *As* t *grows, the posterior* ρ_{t-1} *becomes more and more concentrated around the true underlying distribution* P, *and ultimately converges to the delta measure* δ_P. *The rate at which* ρ_{t-1} *converges to* δ_P *determines the scaling of this extra regret term, and hence* $\text{REG}_n(\textit{Bayesian-SOAR})$.

We now make a number of observations leading to the proof of Theorem F.1 and Corollary F.2.

Lemma F.5. *For any* $t \geq 1$, *conditional on* $X_1, \ldots, X_{t-1}$, *the input of* Bayesian-SOAR *to the offline optimal matching solver* $(X_t, \hat{X}_{t+1}, \ldots, \hat{X}_n)$ *is a set of* i.i.d. *demand units drawn from a distribution sampled according to* ρ_{t-1}, *identically distributed to the sequence of future real demand units* $(X_t, \ldots, X_n)$. *In particular,* $(X_t, \hat{X}_{t+1}, \ldots, \hat{X}_n)$ *are exchangeable conditional on* $X_1, \ldots, X_{t-1}$.

Proof of Lemma F.5. Recall that for each $t \geq 1$, the input of the Bayesian-SOAR algorithm to the offline optimal matching solver consists of the real demand X_t, and the virtual demands $\hat{X}_{t+1}, \ldots, \hat{X}_n$, that are drawn *i.i.d.* from some distribution sampled from ρ_t. Recall that X_t can be viewed as an independent demand unit drawn from a distribution sampled from ρ_{t-1}, and ρ_t is the posterior distribution updated from ρ_{t-1} with observation X_t. $(X_t, \hat{X}_{t+1}, \ldots, \hat{X}_n)$ constructed that way is thus equivalent in distribution to an *i.i.d.* sequence of $n-t+1$ demand units drawn from a distribution sampled from ρ_{t-1}. Conditional on any $X_1, \ldots, X_{t-1}$, the future real demand units $X_t, \ldots X_n$ are drawn *i.i.d.* from a distribution sampled from the posterior ρ_{t-1}. Thereby it is identically distributed to what Bayesian-SOAR inputs to the offline matching solver. Furthermore, as a mixture of (conditionally) *i.i.d.* sequences, $X_t, \ldots, X_n$ (and thus $X_t, \hat{X}_{t+1}, \ldots, \hat{X}_n$) are exchangeable. □

Lemma F.6. *For any fixed realized supply units* $\{Y_1, \ldots, Y_n\}$*, under* Bayesian-SOAR*,* $\mathcal{Y}^{t-1}_{\textit{un-matched}}$ *is a subset of* $\{Y_1, \ldots, Y_n\}$ *with* $n-t+1$ *elements, drawn uniformly at random.*

Proof. By the design of Bayesian-SOAR, in the first period we observe the true demand X_1, update the posterior distribution $\rho_1 \leftarrow \rho_0 | X_1$, and simulate from $\hat{P}_1 \sim \rho_1$ the virtual future demand $\hat{X}_2, \ldots, \hat{X}_n$. By Lemma F.5, the input demand sequence $X_1, \hat{X}_2, \ldots, \hat{X}_n$ to the offline matching solver is a realization of *i.i.d.* sequence drawn from some $\hat{P}_0 \sim \rho_0$. From the perspective of the offline matching solver, it takes as input a set of vectors $\{x_1, x_2, \ldots, x_n\}$ **without knowing their identity, i.e. which vector is the true demand of the first period** (Recall in the description of SOAR we randomly permute the sequence $X_1, \hat{X}_2, \ldots, \hat{X}_n$). An important implication is that, the output from the offline matching solver, given the set of input vectors $\{x_1, x_2, \ldots, x_n\}$, regardless of being randomized or deterministic, is oblivious, thus independent of the index of the actual (non-simulated) demand vector. Now by Lemma F.5, the sequence $X_1, \ldots, \hat{X}_n$ are exchangeable. Thus, conditional on the set of values of $\{X_1, \hat{X}_2, \ldots, \hat{X}_n\} = \{x_1, \ldots, x_n\}$, the actual order is uniform at random. Namely,

$$\mathbb{P}(X_1 = x_i | \{X_1, \hat{X}_2, \ldots, \hat{X}_n\} = \{x_1, \ldots, x_n\}) = \frac{1}{n}$$

for any $i = 1, \ldots, n$. This property of exchangeable random variables, along with the previously established independence between the output of the offline solver and the index of the true demand vector together implies that

$$\mathbb{P}\left(Y_i \text{ was matched in the first period }\right) = \frac{1}{n} \text{ for all } i = 1, \ldots, n.$$

In other words, with respect to the randomness in the real demand X_1 and the simulated future demand $\hat{X}_2, \dots \hat{X}_n$, in effect SOAR uniformly discards a supply unit in $Y_1, \dots, Y_n$. Recall the resulting set is denoted $\mathcal{Y}^1_{\text{un-matched}}$. Since with conditioning the statement is stronger, the unconditional version of the statement directly follows. Specifically, $\mathcal{Y}^2_{\text{un-matched}}$ is in distribution equivalent to repeating the following twice to $\mathcal{Y}^0_{\text{un-matched}}$: discarding a remaining supply unit uniformly-at-random, which is further equivalent to picking a size $n-2$ random subset of $\mathcal{Y}^0_{\text{un-matched}}$. The above argument can be immediately generalized to any $t > 2$, and we complete the proof. □

Remark F.7. *Lemma F.6 is crucially about the unconditional distribution of $\mathcal{Y}^{t-1}_{\text{un-matched}}$. The conditional distribution given the historical demand sequence $X_1, \dots, X_{t-1}$ for $t \geq 2$ is no longer uniform.*

With Lemma F.5 and Lemma F.6, we now prove Theorem F.1 and Corollary F.2.

Proof of Theorem F.1. By definition

$$U_n(\mathsf{Bayesian\text{-}SOAR}) = \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} \varphi\left(X_t, Y_t^{\mathsf{Bayesian\text{-}SOAR}}\right)\right],$$

where $Y_t^{\mathsf{Bayesian\text{-}SOAR}}$ denote the time t match to the incoming real demand X_t. Hence by the tower rule, to show eq. (F.2), it suffices to prove

$$\mathbb{E}_{P_{t-1}\sim\rho_{t-1}, \mathcal{Y}^{t-1}_{\text{un-matched}}}\left[U^{\text{H-OPT}}_{n-t+1}(P_{t-1}, \mathcal{Y}^{t-1}_{\text{un-matched}})\middle| P\right] = \mathbb{E}_{P_{t-1}\sim\rho_{t-1}, \mathcal{Y}^{t-1}_{\text{un-matched}}}\left[\varphi\left(X_t, Y_t^{\mathsf{Bayesian\text{-}SOAR}}\right)\middle| P\right].$$

We prove a stronger version of the above equation, that is, the two conditional expectations are equal even with ρ_{t-1} and $\mathcal{Y}^{t-1}_{\text{un-matched}}$ further fixed and conditioned on. Recall that

$$U^{\text{H-OPT}}_{n-t+1}(P_{t-1}, \mathcal{Y}^{t-1}_{\text{un-matched}}) = \frac{1}{n-t+1}\mathbb{E}_{X'_t, X'_{t+1}\dots, X'_n \overset{i.i.d.}{\sim} P_{t-1}}\left[\max_{\tau\in\mathcal{S}_{n-t+1}} \sum_{i=1}^{n-t+1} \varphi\left(X'_{t+i-1}, Y_{j_{\tau_i}}\right)\right].$$

On the other hand, the Bayesian-SOAR algorithm inputs $\{X_t, \hat{X}_{t+1}, \dots \hat{X}_n\}$ to the offline optimal matching solver, where X_t is the real demand unit, and $\hat{X}_{t+1} \dots \hat{X}_n$ are the virtual ones. The total conditional expected matching quality that the offline solver achieves is

$$\mathbb{E}_{X_t, \hat{X}_{t+1}\dots, \hat{X}_n \overset{i.i.d.}{\sim} P_{t-1}}\left[\max_{\tau\in\mathcal{S}_{n-t+1}} \varphi\left(X_t, Y_{j_{\tau_1}}\right) + \sum_{i=1}^{n-t} \varphi\left(\hat{X}_{t+i}, Y_{j_{\tau_i}}\right)\middle| \rho_{t-1}, \mathcal{Y}^{t-1}_{\text{un-matched}}\right],$$

where $X_t, \hat{X}_{t+1} \ldots, \hat{X}_n \overset{i.i.d.}{\sim} P_{t-1}$ follows from Lemma F.5. Since the so-distributed $X_t, \hat{X}_{t+1} \ldots, \hat{X}_n$ are exchangeable, thus by symmetry we have the expected matching quality contributed by any of them is equal, and in particular, the expected matching quality of the real demand X_t is equal to

$$\frac{1}{n-t+1}\mathbb{E}_{X_t,\hat{X}_{t+1}\ldots,\hat{X}_n \overset{i.i.d.}{\sim} P_{t-1}}\left[\max_{\tau\in\mathcal{S}_{n-t+1}} \varphi\left(X_t, Y_{j_{\tau_1}}\right) + \sum_{i=1}^{n-t}\varphi\left(\hat{X}_{t+i}, Y_{j_{\tau_i}}\right)\middle|\rho_{t-1}, \mathcal{Y}^{t-1}_{\text{un-matched}}\right],$$

where we again use the fact that the offline matching solver is oblivious of the true identity of each input demand unit, just as in the proof of Lemma F.6. Combining the above with the definition of Bayesian-SOAR thus yields eq. (F.2). □

Proof of Corollary F.2. With Theorem F.1, and in particular, eq.(F.2), we have

$$\begin{aligned}
&\textsc{Reg}_n(\text{Bayesian-SOAR})\\
&= \mathbb{E}_{P\sim\rho_0} U_\infty(P,Q,\varphi) - U_n(\text{Bayesian-SOAR}),\\
&= \mathbb{E}_{P\sim\rho_0} U_\infty(P,Q,\varphi) - \frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=1}^{n}\mathbb{E}_{P_{t-1}\sim\rho_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}}}\left[U^{\text{H-OPT}}_{n-t+1}(P_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}})\middle|P\right]\right],\\
&= \frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=1}^{n}\mathbb{E}_{P_{t-1}\sim\rho_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}}}\left[U_\infty(P,Q,\varphi) - U^{\text{H-OPT}}_{n-t+1}(P_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}})\middle|P\right]\right],\\
&= \frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=1}^{n}\left(U_\infty(P,Q,\varphi) - U^{\text{H-OPT}}_{n-t+1}(P,\mathcal{Y}^{t-1}_{\text{un-matched}})\right)\right]\\
&\quad + \frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=1}^{n}\mathbb{E}_{P_{t-1}\sim\rho_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}}}\left[U^{\text{H-OPT}}_{n-t+1}(P,\mathcal{Y}^{t-1}_{\text{un-matched}}) - U^{\text{H-OPT}}_{n-t+1}(P_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}})\middle|P\right]\right].
\end{aligned}$$

By Lemma F.6, $\mathcal{Y}^{t-1}_{\text{un-matched}}$ under Bayesian-SOAR is a uniform-at-random subset of size $n-t+1$ of $\{Y_1,\ldots,Y_n\}$. Since the supply units are initially *i.i.d.* Q, we thus conclude that $\mathcal{Y}^{t-1}_{\text{un-matched}}$ is a set of *i.i.d.* Q supply units. Namely, $U^{\text{H-OPT}}_{n-t+1}(P,\mathcal{Y}^{t-1}_{\text{un-matched}})$ is the average hindsight optimum matching quality between $n-t+1$ *i.i.d.* P demand units and $n-t+1$ *i.i.d.* Q supply units. By definition of $\textsc{Reg}_n(\text{SOAR}; P, Q, \varphi)$ and Remark 4,

$$\frac{1}{n}\mathbb{E}_{P\sim\rho_0}\left[\sum_{t=0}^{n-1}\left(U_\infty(P,Q,\varphi) - U^{\text{H-OPT}}_{n-t+1}(P,\mathcal{Y}^{t-1}_{\text{un-matched}})\right)\right] = \mathbb{E}_{P\sim\rho_0}\left[\textsc{Reg}_n(\text{SOAR}; P,Q,\varphi)\right],$$

which is the first term in the regret decomposition in Corollary F.2. Combining the above thus conclude the proof of Corollary F.2. □

Next we discuss a concrete example, for which, utilizing Corollary F.2, we derive the regret scaling of Bayesian-SOAR.

Example. Consider a discrete setting with $\mathcal{P} = \{P^1, P^2, \ldots, P^m\}$ where P^k is associated with density function f^k, and $\rho_0 = \mathsf{Unif}(m)$. Let $P^* \in \mathcal{P}$ be the true underlying distribution. The posterior distribution ρ_t has a closed-form expression

$$\rho_t(k) = \frac{\Pi_{i=1}^t f^k(X_i)}{\sum_{j=1}^m \Pi_{i=1}^t f^j(X_i)}, \quad k = 1, \ldots, m.$$

Observe that $\Pi_{i=1}^t f^k(X_i) = \exp\left(\sum_{i=1}^t \log f^k(X_i)\right)$, and therefore

$$\frac{\Pi_{i=1}^t f^*(X_i)}{\Pi_{i=1}^t f^k(X_i)} = \exp\left(\sum_{i=1}^t \log \frac{f^*(X_i)}{f^k(X_i)}\right).$$

The above RHS can be rewritten as $\exp\left(t \cdot \frac{1}{t}\sum_{i=1}^t \log \frac{f^*(X_i)}{f^k(X_i)}\right)$, which, for large t, is approximately $\exp\left(t\mathbb{E}_{X\sim P^*}\left[\log \frac{f^*(X)}{f^k(X)}\right]\right)$, or, equivalently, $\exp\left(tD_{\mathrm{KL}}(P^*\|P^k)\right)$, where $D_{\mathrm{KL}}(P^*\|P^k)$ is the KL-divergence between the two distributions. For technical convenience, we assume the distributions in $\mathcal{P}$ are well-separated and regular. In particular, suppose $\delta \equiv \min_{j\neq k\in[m]} D_{\mathrm{KL}}(P^j\|P^k) > 0$. Furthermore, suppose $0 < l \leq f^k(x) \leq u$ for a pair of constants l, u, for any $k \in [m]$ and for any $x \in \mathbb{R}^d$. In addition, we shall assume the matching quality $0 \leq \varphi(\cdot,\cdot) \leq C$ for some universal constant $C > 0$. Formalizing the above intuition, we derive a bound on the convergence rate of the posterior distribution, stated in the following lemma.

Lemma F.8. *There exists a constant c_2 that depends only on the problem primitives δ, u, l, such that for any $P^k \neq P^*$, $\rho_t(k) \leq e^{-\frac{\delta}{2}t}$ with probability at least $1 - \exp(-tc_2)$*

Proof. Observe that

$$\begin{aligned}\rho_t(k) &= \frac{\Pi_{i=1}^t f^k(X_i)}{\sum_{j=1}^m \Pi_{i=1}^t f^j(X_i)}, \\ &\leq \frac{\Pi_{i=1}^t f^k(X_i)}{\Pi_{i=1}^t f^*(X_i)}, \\ &= \exp\left(-\sum_{i=1}^t \log\frac{f^*(X_i)}{f^k(X_i)}\right), \\ &\leq \exp\left(-\sum_{i=1}^t \log\frac{f^*(X_i)}{f^k(X_i)}\right)\end{aligned}$$

Denote by $Y_i \equiv \log\frac{f^*(X_i)}{f^k(X_i)}$ and $Z_t \equiv \sum_{i=1}^t Y_i$. Then $\rho_t(k) = \exp(-Z_t)$. By our assumption, $\mathbb{E}[Y_i] = D_{\mathrm{KL}}\left(P^*\|P^k\right) \geq \delta > 0$. By the regularity condition, we have $|Y_i| \leq \log(\frac{u}{l})$. The Hoeffding's bound then yields for any $\epsilon > 0$,

$$\mathbb{P}\left(\frac{Z_t}{t} \leq \delta - \epsilon\right) \leq \exp\left(-\frac{t\epsilon^2}{2\log(u/l)}\right).$$

Consequently (by setting $\epsilon = \delta/2$), with probability at least $1 - \exp(-tc_2)$, $\rho_t(k) \leq \exp\left(-\frac{\delta t}{2}\right)$, where $c_2 = \frac{\delta^2}{8\log(u/l)}$ is an absolute constant. □

We thus have

$$\begin{aligned}&\mathbb{E}_{\rho_{t-1},P_{t-1}\sim\rho_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}}}\left[U^{\text{H-OPT}}_{n-t+1}(P^*,\mathcal{Y}^{t-1}_{\text{un-matched}}) - U^{\text{H-OPT}}_{n-t+1}(P_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}}) \,\middle|\, P^*\right] \\ &\leq\ C\mathbb{E}_{\rho_{t-1},P_{t-1}\sim\rho_{t-1},\mathcal{Y}^{t-1}_{\text{un-matched}}}\left[\mathbb{1}\{P_{t-1}\neq P^*\}\middle|\, P^*\right] \\ &= C\sum_{k:P^k\neq P^*}\mathbb{E}_{\rho_{t-1}}[\rho_{t-1}(k)|P^*] \\ &\leq m\left(C\mathbb{P}\left(\rho_{t-1}(k) > \exp\left(-\frac{\delta}{2}t\right)\right) + C\exp\left(-\frac{\delta}{2}t\right)\right) \\ &\leq C\left(m^2\exp(-(t-1)c_2) + m\exp\left(-(t-1)\frac{\delta}{2}\right)\right) \leq a_1\exp(-(t-1)b_1),\end{aligned}$$

where $a_1 = 2m^2C$ and $b_1 = \max\left(\frac{\delta}{2}, \frac{\delta^2}{8\log(u/l)}\right)$. Combining the above with Corollary F.2, we immediately derive the following result.

Corollary F.4. *In the setting of Bayesian learning with discrete candidate distribution set* $\{P^1,\ldots,P^m\}$, *suppose the following conditions are satisfied:*

1. $\delta \equiv \min_{j \neq k \in [m]} D_{\mathrm{KL}}(P^j \| P^k) > 0$.

2. $0 < l \leq f^k(x) \leq u$ *for a pair of constants* l, u, *for any* $k \in [m]$ *and for any* $x \in \mathbb{R}^d$.

3. *The matching quality* $0 \leq \varphi(\cdot, \cdot) \leq C$ *for some universal constant* $C > 0$.

Then we have

$$\mathrm{REG}_n(\textsf{\textit{Bayesian-SOAR}}) \leq \mathbb{E}_{P \sim \rho_0}\left[\mathrm{REG}_n(\textsf{SOAR}; P, Q, \varphi)\right] + \frac{C_1}{n},$$

where $C_1 = 2m^2 C \left(\exp\left(\max\left(\frac{\delta}{2}, \frac{\delta^2}{8\log(u/l)}\right)\right) - 1\right)^{-1}$.

In this example, the extra cumulative regret term due to not knowing the true underlying distribution is $\mathcal{O}(1)$, thus dominated by $\mathrm{REG}_n(\textsf{SOAR})$, thus the regret scaling of $\mathrm{REG}_n(\textsf{Bayesian-SOAR})$ remains what we have derived in the corresponding known *i.i.d.* setting (see Theorem 2, Proposition 2 and Theorem 3).

G Details Related to Optimal Transport and Useful Known Results

In this section, we will provide some additional notation and some existing results in the empirical optimal transport literature which will leverage to prove the theorems in Section 4.

G.1 Background on Optimal Transport and Wasserstein-p distance

In this section, we provide some background on optimal transport and Wasserstein-p distance. Some of the notations are adapted from [Manole et al., 2021, Section 2]. For a fixed $d \geq 1$, let $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ and let $\mathcal{P}(\mathcal{X})$ denote the set of Borel probability measures with support contained in $\mathcal{X}$. Let $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$. An optimal transport map $\mathcal{T}_{P \to Q}$ from distribution P to Q (with support in the set Ω) is any solution to the *Monge problem* Monge [1781] defined as

$$\underset{\mathcal{T}_{P \to Q} \in \mathscr{T}(P,Q)}{\arg\min} \int_\Omega \|x - \mathcal{T}_{P \to Q}(x)\|^2 dP(x), \tag{G.1}$$

where $\mathscr{T}(P,Q)$ is the set of all transport maps between P and Q, i.e., the set of Borel-measurable functions $\mathcal{T}:\Omega\to\Omega$ such that $\mathcal{T}_{\#}P=Q$. Here, $\mathcal{T}_{\#}Q$ denotes the pushforward measure of Q induced by $\mathcal{T}$. The convex relaxation of the *Monge problem* is the *Kantorovich problem*,

$$\underset{\pi\in\Pi(P,Q)}{\arg\min}\int_{\Omega}\|x-y\|^2 d\pi(x,y), \tag{G.2}$$

where $\Pi(P,Q)$ is the set of couplings of probability distributions P and Q, i.e, the set of probability distributions over $\mathcal{X}\times\mathcal{Y}$ with first and second margins being P and Q. Therefore, a probability measure π over $\mathcal{X}\times\mathcal{Y}$ belongs to $\Pi(P,Q)$ if and only if $\pi(A\times\mathcal{Y})=P(A)$ and $\pi(\mathcal{X}\times B)=Q(B)$ holds for every subset A of $\mathcal{X}$ and B of $\mathcal{Y}$. It can be shown that a minimizer π always exists for (G.2) [Villani, 2009, Theorem 4.1] and is called the optimal coupling. The corresponding optimal value of (G.2) is referred to as the Wasserstein-2 distance

$$W_2(P,Q)=\left(\inf_{\pi\in\Pi(P,Q)}\int\|x-y\|^2 d\pi(x,y)\right)^{1/2}.$$

The above optimization problem is an (infinite-dimensional) convex program with linear constraints, and it admits a dual maximization problem, known as the Kantorovich dual problem given below.

$$W_2^2(P,Q)=\sup_{(\phi,\nu)\in\mathcal{K}}\int\phi dP+\int\nu dQ, \tag{G.3}$$

where $\mathcal{K}$ is the set of pairs $(\phi,\nu)\in L^1(\Omega)\times L^1(\Omega)$ such that $\phi(x)+\nu(y)\leq\|x-y\|^2$ for all $x,y\in\Omega$. Let ϕ_0,ν_0 be the pair for which the supremum is achieved. The Kantorovich dual problem can be reparameterized and it is equivalent to the following semi-dual problem

$$W_2^2(P,Q)=\sup_{\psi\in L^1(\Omega)}\int\psi dP+\int\psi^c dQ. \tag{G.4}$$

The Brenier potential ψ is the solution of the semi-dual problem in (G.4) and is related to ϕ_0 as $\psi=\|\cdot\|^2-2\phi_0$. So far, we focus on the cost function $c(x,y)=\|x-y\|^2$, but we can define the optimal transport cost more generally as well. Given a non-negative cost function $c:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$, the *optimal transport cost* based on c is defined by

$$\mathsf{OT}_c(P,Q) := \inf_{\pi\in\Pi(P,Q)} \int c(x,y)d\pi(x,y). \tag{G.5}$$

Consider i.i.d. random variables $X_1, X_2, \ldots, X_n \sim P$ and $Y_1, Y_2, \ldots, Y_n \sim Q$ and let P_n and Q_n denote their corresponding empirical measures, i.e. $P_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ and $Q_n = \frac{1}{n}\sum_{k=1}^n \delta_{Y_k}$. The *empirical optimal transport cost* based on c is defined as

$$\mathsf{OT}_c(P_n,Q_n) := \inf_{\sigma} \frac{1}{n}\sum_{k=1}^{n} c(X_k, Y_{\sigma(k)}), \quad \sigma \text{ is a permutation of } \{1,2,\ldots,n\}.$$

For the special case of $c(x,y) = c_p(x,y) := \|x-y\|^p$ (note that the cost function $c_p(x,y)$ is negative of the quality function $\varphi_p(x,y)$) for $p \geq 1$. For $p \in [1,\infty)$, the Wasserstein p-distance between two probability distributions P and Q on Borel sets of $\mathbb{R}^d$ with finite p-moments is defined as

$$W_p(P,Q) := \left(\inf_{\pi\in\mathcal{M}(P,Q)} \int \|x-y\|^p \mathrm{d}\pi(x,y)\right)^{1/p} = \left(\inf_{\pi\in\mathcal{M}(P,Q)} \mathbb{E}_{(x,y)\sim\pi}\left[\|x-y\|^p\right]\right)^{1/p}. \tag{G.6}$$

Note that $W_p^p(P,Q) = \mathsf{OT}_{c_p}(P,Q) = -U_\infty(P,Q)$, where $U_\infty(P,Q)$ is the fluid optimum for the quality function $\varphi_p(x,y) = -\|x-y\|^p$. Moreover, we have that

$$\mathsf{OT}_{c_p}(P_n,Q_n) = W_p^p(P_n,Q_n) = \inf_{\sigma} \frac{1}{n}\sum_{k=1}^{n} \|X_k - Y_{\sigma(k)}\|^p, \quad \sigma \text{ is a permutation over } \{1,2,\ldots,n\}.$$

Note that $\mathbb{E}\left[W_p^p(P_n,Q_n)\right] = -U_n^{\mathsf{H}}(P,Q)$ where U_n^{H} is the hindsight optimum matching value for the quality function $\varphi_p(X,Y) = -\|X-Y\|^p$.

G.2 Existing Results on convergence of Empirical Optimal Transport value

In this section, we will present some results on the convergence of the empirical optimal transport value to its limit for different quality functions and different assumptions on the distributions P and Q. A function $c : U \to \mathbb{R}$ on a convex domain $U \subseteq \mathbb{R}^d$ is (α,Λ)-Holder smooth for $0 < \alpha \leq 1$ and $\Lambda > 0$ if $\|c\|_\infty < \Lambda$ and $|c(x) - c(y)| \leq \Lambda\|x-y\|^\alpha$ and c is (α,Λ)-Hölder smooth for $1 < \alpha \leq 2$ if $\|c\|_\infty \leq \Lambda$ and c is differentiable with $(\alpha-1,\Lambda)$-Hölder smooth partial derivatives.

Lemma G.1 (Theorem 3.11 of Hundrieser et al. [2022])**.** *Fix $d \geq 1$ and a constant $M < \infty$. Let*

$\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces and $c : \mathcal{X} \times \mathcal{Y} \to [0, M]$ be continuous. If c is $(\alpha, \Lambda)-$Hölder smooth for some $\alpha \in [1, 2]$, then for any $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$, there exists a constant $C \equiv C(P, Q, d, \alpha)$ such that

$$\mathbb{E}\left[|\mathsf{OT}_c(P_n, Q_n) - \mathsf{OT}(P, Q)|\right] \leq \begin{cases} Cn^{-1/2}, & \text{if } d < 2\alpha, \\ Cn^{-1/2}\log n, & \text{if } d = 2\alpha, \\ Cn^{-\alpha/d}, & \text{if } d > 2\alpha. \end{cases}$$

Corollary G.2. *Fix $d \geq 1$. Let $\mathcal{X}$ and $\mathcal{Y}$ be bounded subsets of $\mathbb{R}^d$ with a non-empty interior. Consider $c_p(x, y) = \|x-y\|^p$ and assume that the demand P and supply Q distributions are supported on $\mathcal{X}$ and $\mathcal{Y}$ respectively, then there exists a constant $C \equiv C(P, Q, d, p) < \infty$ such that*

$$\mathbb{E}\left[W_p^p(P_n, Q_n) - W_p^p(P, Q)\right] \leq \begin{cases} Cn^{-1/2}, & \text{if } d < 2(p \wedge 2), \\ Cn^{-1/2}\log n, & \text{if } d = 2(p \wedge 2), \\ Cn^{-(p\wedge 2)/d}, & \text{if } d > 2(p \wedge 2). \end{cases}$$

Proof of Corollary G.2. Since both $\mathcal{X}$ and $\mathcal{Y}$ are closed, bounded and convex, we have that $\mathcal{X}$ and $\mathcal{Y}$ are Polish spaces. Since $\mathcal{X}$ and $\mathcal{Y}$ are compact and c_p is continuous, we have that $\|c_p\|_\infty \leq M$ for some constant $M < \infty$. From [Manole and Niles-Weed, 2021, Corollary 3], we know that c_p is $(p \wedge 2, \Lambda)-$Hölder smooth, i.e. $\alpha = p \wedge 2$. Hence all the conditions in Lemma G.1 are verified. From the convexity of $f(x) = |x|$, we have that

$$\mathbb{E}\left[W_p^p(P_n, Q_n) - W_p^p(P, Q)\right] \leq \left|\mathbb{E}\left[W_p^p(P_n, Q_n) - W_p^p(P, Q)\right]\right| \leq \mathbb{E}\left[|W_p^p(P_n, Q_n) - W_p^p(P, Q)|\right].$$

Finally, using the fact that $\mathsf{OT}_{c_p}(P_n, Q_n) = W_p^p(P_n, Q_n)$ and $\mathsf{OT}_{c_p}(P, Q) = W_p^p(P, Q)$, the results follows from Lemma G.1. □

Lemma G.3 (Proposition 21 of Manole and Niles-Weed [2021])**.** *Fix $d \geq 1$. $\mathcal{X}$ and $\mathcal{Y}$ are convex subsets of $\mathbb{R}^d$ with non-empty interior. The cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$ and takes the form $c(x, y) = h(x - y)$ where $h : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ is convex, even, and lower semi-continuous. Further assume that h is differentiable over $\mathcal{Z} = \mathcal{X} - \mathcal{Y}$. Furthermore, there exists $\lambda > 0, \alpha \in (0, 2]$ and*

$z_0 = x_0 - y_0 \in \mathcal{Z}$ *such that* $x_0 \in int(\mathcal{X}), y_0 \in int(\mathcal{Y})$ *and for all* $z \in \mathcal{Z}$,

$$h(z) - h(z_0) \geq \begin{cases} \lambda \|z - z_0\|^\alpha, & \alpha \leq 1, \\ \langle \nabla h(z_0), z - z_0 \rangle + \lambda \|z - z_0\|^\alpha, & \alpha > 1. \end{cases}$$

Then there exists a constant $c > 0$ *such that*

$$\sup_{P \in \mathcal{P}(\mathcal{X}), Q \in \mathcal{P}(\mathcal{Y})} \mathbb{E}\left[\mathsf{OT}_c(P_n, Q_n) - \mathsf{OT}_c(P, Q)\right] \geq c n^{-\alpha/d}.$$

Corollary G.4. *Fix* $d \geq 1$*. Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be closed, bounded and convex subsets of* $\mathbb{R}^d$ *with a non-empty interior. Consider* $c_p(x, y) = \|x - y\|^p$*. There exists demand* P *and supply* Q *distributions supported on* $\mathcal{X}$ *and* $\mathcal{Y}$ *respectively and a constant* $c \equiv c(P, Q, d, p) > 0$ *such that*

$$\mathbb{E}\left[W_p^p(P_n, Q_n) - W_p^p(P, Q)\right] \geq c n^{-(p \wedge 2)/d}.$$

Proof of Corollary G.4. From Manole and Niles-Weed [2021], it is straightforward to see that $h(x) = \|x\|^p$ satisfies the conditions in Lemma G.3 with $\alpha = p \wedge 2$. The conditions on $\mathcal{X}$ and $\mathcal{Y}$ are satisfied by assumption. The result follows from Lemma G.3. □

Lemma G.5 (Ledoux [2019])**.** *Fix* $d \geq 1$*. Let* $P = Q = \text{Uniform}([0, 1]^d)$*. Consider the cost function* $c_p(x, y) = \|x - y\|^p$ *for some* $p \in [1, d]$*. Then we have that*

$$\mathbb{E}\left[W_p^p(P_n, Q_n)\right] = \begin{cases} \Theta\left(n^{-p/2}\right), & d = 1, \\ \Theta\left((\log n)^{p/2} n^{-p/2}\right), & d = 2, \\ \Theta\left(n^{-p/d}\right), & d \geq 3. \end{cases}$$

Lemma G.6 (Proposition 13 of Manole et al. [2021])**.** *Let* P *and* Q *be absolutely continuous distributions with the support in* $[0, 1]^d$ *and with densities being bounded above and below over* $[0, 1]^d$*. Furthermore assume that Assumption 1 is satisfied. Then for the cost function* $c(x, y) = \|x - y\|^2$, *we have that there exists a constant $C \equiv C(P,Q,d) < \infty$ such that*

$$\mathbb{E}\left[W_2^2(P_n,Q_n) - W_2^2(P,Q)\right] \le \begin{cases} Cn^{-1}, & d=1, \\ Cn^{-1}\log n, & d=2, \\ Cn^{-2/d}, & d\ge 3. \end{cases}$$

G.3 Equivalence of $\varphi_{\mathsf{dot}}(X,Y) = \langle X,Y\rangle$ and $\varphi_2(X,Y) = -\|X-Y\|^2$

We formally establish that the regret corresponding to the dot-product quality function $\varphi_{\mathsf{dot}}(X,Y) = \langle X,Y\rangle$ scales exactly as the regret corresponding to the quality $\varphi(X,Y) = -\|X-Y\|^2$.

Lemma G.7. *The dot-product quality function $\varphi_{\mathsf{dot}}(X,Y) = \langle X,Y\rangle$ and the quality function $\varphi_2(X,Y) = -\|X-Y\|^2$ are equivalent in that the regret incurred by any policy π under $\varphi(X,Y) = \langle X,Y\rangle$ is exactly half the regret under $\varphi_2(X,Y) = -\|X-Y\|^2$, incurred by the same π.*

Proof of Lemma G.7. Consider the quality function $\varphi_2(X,Y) = -\|X-Y\|^2$, then we have that $U_\infty(P,Q,\varphi_2) = -\lim_{n\to\infty}\mathbb{E}\left[\inf_{\sigma\in S_n}\sum_{t=1}^n \|X_t - Y_{\sigma(t)}\|^2\right]/n$. For any policy $\pi\in\Pi$, the expected average quality under the policy π with quality function $\varphi_2(X,Y) = -\|X-Y\|^2$ is given as $U_n(\pi;P,Q,\varphi_2) = -\mathbb{E}\left[\sum_{t=1}^n \|X_t - Y_{\pi(t)}\|^2\right]/n$.

Consider the hindsight optimum value for the dot product utility $\varphi(X,Y) = \langle X,Y\rangle$ for $n\ge 1$,

$$\begin{aligned} U_n^{\mathsf{H}}(P,Q,\varphi_{\mathsf{dot}}) &\overset{(a)}{=} \frac{1}{n}\mathbb{E}\left[\sup_{\sigma\in S_n}\sum_{t=1}^n \langle X_t, Y_{\sigma(t)}\rangle\right] \overset{(b)}{=} \frac{1}{2n}\mathbb{E}\left[\sup_{\sigma\in S_n}\sum_{t=1}^n \langle X_t, X_t\rangle + \langle Y_{\sigma(t)}, Y_{\sigma(t)}\rangle - \|X_t - Y_{\sigma(t)}\|^2\right] \\ &\overset{(c)}{=} \frac{1}{2}\mathbb{E}\left[\langle X,X\rangle\right] + \frac{1}{2}\mathbb{E}\left[\langle Y,Y\rangle\right] - \frac{1}{2n}\mathbb{E}\left[\inf_{\sigma\in S_n}\sum_{t=1}^n \|X_t - Y_{\sigma(t)}\|^2\right], \end{aligned}$$

where (a) follows from the definition of $U_n^{\mathsf{H}}(P,Q,\varphi_{\mathsf{dot}})$ for $\varphi_{\mathsf{dot}}(x,y) = \langle x,y\rangle$, (b) follows from the fact that $\varphi_{\mathsf{dot}}(X,Y) = \langle X,Y\rangle = \frac{1}{2}\langle X,X\rangle + \frac{1}{2}\langle Y,Y\rangle - \frac{1}{2}\|X-Y\|^2$, (c) from the fact that $X_1,X_2,\ldots,X_n$ and $Y_1,Y_2,\ldots,Y_n$ are i.i.d. Therefore we have that $U_\infty(P,Q,\varphi_{\mathsf{dot}}) = \mathbb{E}\left[\langle X,X\rangle\right]/2 + \mathbb{E}\left[\langle Y,Y\rangle\right]/2 - \lim_{n\to\infty}\mathbb{E}\left[\inf_{\sigma\in S_n}\sum_{t=1}^n \|X_t - Y_{\sigma(t)}\|^2\right]/2n$. Hence we have that

$$U_\infty(P,Q,\varphi_{\mathsf{dot}}) = \mathbb{E}\left[\langle X,X\rangle\right]/2 + \mathbb{E}\left[\langle Y,Y\rangle\right]/2 + U_\infty(P,Q,\varphi_2)/2.$$

For any policy π and the quality function $\varphi_{\mathsf{dot}}(x,y) = \langle x,y\rangle$, the expected average quality under the policy π is given as

$$U_n(\pi; P, Q, \varphi_{\mathsf{dot}}) = \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} \langle X_t, Y_{\pi(t)}\rangle\right] = \frac{1}{2n}\mathbb{E}\left[\sum_{t=1}^{n} \langle X_t, X_t\rangle + \langle Y_{\pi(t)}, Y_{\pi(t)}\rangle - \|X_t - Y_{\pi(t)}\|^2\right]$$
$$= \frac{1}{2}\mathbb{E}\left[\langle X, X\rangle\right] + \frac{1}{2}\mathbb{E}\left[\langle Y, Y\rangle\right] - \frac{1}{2n}\mathbb{E}\left[\sum_{t=1}^{n} \|X_t - Y_{\pi(t)}\|^2\right]$$
$$= \frac{1}{2}\mathbb{E}\left[\langle X, X\rangle\right] + \frac{1}{2}\mathbb{E}\left[\langle Y, Y\rangle\right] + \frac{1}{2}U_n(\pi; P, Q, \varphi_2).$$

Hence we have that for any policy π,

$$\begin{aligned}\textsc{Reg}_n(\pi; P, Q, \varphi_{\mathsf{dot}}) &= U_\infty(P, Q, \varphi_{\mathsf{dot}}) - U_n(\pi; P, Q, \varphi_{\mathsf{dot}}) \\ &= U_\infty(P, Q, \varphi_2)/2 - U_n(\pi; P, Q, \varphi_2)/2 \\ &= \textsc{Reg}_n(\pi; P, Q, \varphi_2)/2.\end{aligned}$$

Hence we have that the regret for the quality functions $\varphi_2(x, y)$ and $\varphi_{\mathsf{dot}}(x, y)$ are equal up to a constant factor and this completes the proof. □

H Proof of Theorem 2

Proof of Theorem 2. We begin by proving the upper bound on the performance of the SOAR algorithm. We have that $U_n(\mathsf{SOAR}) = -\frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} \|X_t - Y_{\pi_t^{\mathsf{SOAR}}}\|^p\right]$ and $U_\infty = -W_p^p(P, Q)$. Using the definition of regret (per match) and Theorem 1, we have that

$$\begin{aligned}\textsc{Reg}_n(\mathsf{SOAR}) &= U_\infty - U_n(\mathsf{SOAR}) \\ &= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} \|X_t - \pi_t^{\mathsf{SOAR}}(X_t)\|^p\right] - W_p^p(P, Q), \\ &\overset{(a)}{=} \frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{k}\mathbb{E}\left[\min_\sigma \sum_{j=1}^{k} \|X_j - Y_{\sigma(j)}\|^p\right] - W_p^p(P, Q)\right) \\ &\overset{(b)}{=} \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[W_p^p(P_k, Q_k) - W_p^p(P, Q)\right],\end{aligned}$$

where (a) follows from Theorem 1 and (b) follows from the definition of $W_p^p(P_k, Q_k)$. Next we will consider three cases: *(i)* $d < 2(p \wedge 2)$, *(ii)* $d = 2(p \wedge 2)$ and *(iii)* $d > 2(p \wedge 2)$.

(i) $\underline{d < 2(p \wedge 2).}$ Using Corollary G.2 for $d < 2(p \wedge 2)$, there is a constant $C \equiv C(P, Q, d, p) < \infty$ such that $\mathbb{E}[W_p^p(P_k, Q_k) - W_p^p(P, Q)] \leq Ck^{-\frac{1}{2}}$ for any $k \geq 1$. Hence we have that,

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{1}{n} \sum_{k=1}^{n} Ck^{-\frac{1}{2}} \leq \frac{C}{n} \int_0^n x^{-\frac{1}{2}} dx = C' n^{-\frac{1}{2}}.$$

(ii) $\underline{d = 2(p \wedge 2).}$ Using Corollary G.2 for $d = 2(p \wedge 2)$, there is a constant $C \equiv C(P, Q, d, p) < \infty$ such that $\mathbb{E}[W_p^p(P_k, Q_k) - W_p^p(P, Q)] \leq Ck^{-\frac{1}{2}} \log k$ for any $k \geq 2$. Hence we have that,

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{1}{n} \left(C + \sum_{k=2}^{n} Ck^{-\frac{1}{2}} \log k \right) \leq \frac{C' \log n}{n} \int_0^n x^{-\frac{1}{2}} dx = 2C' n^{-\frac{1}{2}} \log n.$$

(iii) $\underline{d > 2(p \wedge 2).}$ Using Corollary G.2 for $d = 2(p \wedge 2)$, there is a constant $C \equiv C(P, Q, d, p) < \infty$ such that $\mathbb{E}[W_p^p(P_k, Q_k) - W_p^p(P, Q)] \leq Ck^{-\frac{p \wedge 2}{d}}$ for any $k \geq 1$. Hence we have that

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{1}{n} \sum_{k=1}^{n} Ck^{-\frac{p \wedge 2}{d}} \leq \frac{C}{n} \int_0^n x^{-\frac{p \wedge 2}{d}} dx = C' n^{-\frac{p \wedge 2}{d}}.$$

This completes the proof of the upper bound in Theorem 2.

Next we will prove the lower bound on the regret of any online optimal policy. For any feasible, non-anticipative online policy π, recall that $U_n^{\mathsf{H}}(P, Q, \varphi) \geq U_n(\pi; P, Q, \varphi)$ for any pair of distributions P and Q and hence we have that

$$\inf_{\pi \in \Pi} \textsc{Reg}_n(\pi; P, Q, \varphi) = \inf_{\pi \in \Pi} U_\infty(P, Q, \varphi) - U_n(\pi; P, Q, \varphi) \geq U_\infty(P, Q, \varphi) - U_n^{\mathsf{H}}(P, Q, \varphi).$$

Recall that for $\varphi(X, Y) = -\|X - Y\|^p$, we have that $U_n^{\mathsf{H}}(P, Q, \varphi) = -\frac{1}{n}\mathbb{E}\left[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p\right] = -\mathbb{E}\left[W_p^p(P_n, Q_n)\right]$ and $U_\infty(P, Q, \varphi) = -W_p^p(P, Q)$ and hence we have that

$$\inf_{\pi \in \Pi} \textsc{Reg}_n(\pi; P, Q, \varphi) \geq \mathbb{E}\left[W_p^p(P_n, Q_n) - W_p^p(P, Q)\right]. \tag{H.1}$$

As before we will consider different cases for d depending on the value of p.

(i) $\underline{d > 2(p \wedge 2).}$ From Corollary G.4, there exists a positive constant $c := c(P, Q, d, p) > 0$ such that $\inf_{\pi \in \Pi} \textsc{Reg}_n(\pi; P, Q, \varphi) \geq cn^{-\frac{p \wedge 2}{d}}$ for $d > 2(p \wedge 2)$.

(*ii*) $\underline{d \leq 2(p \wedge 2)}$. We will consider the case where $P = Q$ and distribution P is supported on the vertices of the hypercube $[0,1]^d$ namely on the points $\mathcal{V} = \{\boldsymbol{v} : \boldsymbol{v}_i \in \{0,1\} \forall i \in \{1,2,\ldots,d\}\}$ and moreover, the distribution P is uniform over the points in $\mathcal{V}$. Since $P = Q$, we have that $U_\infty = -\inf_{\mu \in \mathcal{M}(P,Q)} \mathbb{E}[\|X - \mu(Y)\|^p] = 0$. We have n i.i.d. samples of the demand $X_1, X_2, \ldots, X_n$ and supply $Y_1, Y_2, \ldots, Y_n$. Let $N_{\boldsymbol{v}}^X = \sum_{k=1}^n \mathbb{1}\{X_i = \mathbf{v}\}$ denote the number of demand random variables (i.e. $X_1, X_2, \ldots, X_n$) that are equal to the point $\boldsymbol{v}$. Similarly define the quantity $N_{\boldsymbol{v}}^Y$ for the supply random variables $Y_1, Y_2, \ldots, Y_n$. For a fixed $\boldsymbol{v} \in \mathcal{V}$, consider the following events $\mathcal{E}_X = \{N_{\boldsymbol{v}}^X \leq n/2^d - \sqrt{n/2^d}\}$ and $\mathcal{E}_Y = \{N_{\boldsymbol{v}}^Y \geq n/2^d + \sqrt{n/2^d}\}$. Since P and Q are uniformly distributed over the points in $\mathcal{V}$, for any $\boldsymbol{v} \in \mathcal{V}$, we have that $N_{\boldsymbol{v}}^X \sim \mathrm{Bin}(n, 1/2^d)$ and $N_{\boldsymbol{v}}^Y \sim \mathrm{Bin}(n, 1/2^d)$ and hence $\mathbb{E}\left[N_{\boldsymbol{v}}^X\right] = \mathbb{E}\left[N_{\boldsymbol{v}}^Y\right] = n/2^d$ and $\mathrm{var}(N_{\boldsymbol{v}}^X) = \mathrm{var}(N_{\boldsymbol{v}}^Y) = n(2^d - 1)/2^{2d}$. Using CLT, it is easy to observe that there exists a positive constant $\alpha > 0$ such that $\mathbb{P}(\mathcal{E}_X) \geq \alpha$ and $\mathbb{P}(\mathcal{E}_Y) \geq \alpha$ and since the events $\mathcal{E}_X$ and $\mathcal{E}_Y$ are independent, we have that $\mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \geq \alpha^2 > 0$. Now we have that

$$\begin{aligned}
\inf_{\pi \in \Pi} \textsc{Reg}_n(\pi; P, Q, \varphi) &\overset{(a)}{\geq} \frac{1}{n}\mathbb{E}\left[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p\right] \\
&\overset{(b)}{=} \frac{1}{n}\mathbb{E}\left[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| \mathcal{E}_X \cap \mathcal{E}_Y\right] \mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \\
&\quad + \frac{1}{n}\mathbb{E}\left[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| (\mathcal{E}_X \cap \mathcal{E}_Y)^c\right] \mathbb{P}((\mathcal{E}_X \cap \mathcal{E}_Y)^c) \\
&\overset{(c)}{\geq} \frac{1}{n}\mathbb{E}\left[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \middle| \mathcal{E}_X \cap \mathcal{E}_Y\right] \mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \\
&\overset{(d)}{\geq} 2\alpha^2 \sqrt{n/2^d}/n = cn^{-\frac{1}{2}},
\end{aligned}$$

where (a) follows from that $U_\infty = 0$ (since $P = Q$) and (H.1), (b) follows from law of total expectations, (c) follows from the fact that $\mathbb{E}[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p \,|\, (\mathcal{E}_X \cap \mathcal{E}_Y)^c] \geq 0$, (d) follows from fact that under the event $\mathcal{E}_X \cap \mathcal{E}_Y$, the number of supply units equal to $\boldsymbol{v}$ (recall that we fixed $\boldsymbol{v}$ while defining $\mathcal{E}_X$ and $\mathcal{E}_Y$) are at least $2\sqrt{n/2^d}$ times more than the demand units equal to $\boldsymbol{v}$ and hence these excess supply units must be matched to some $\boldsymbol{v}' \neq \boldsymbol{v}$ and since $\|\boldsymbol{v}' - \boldsymbol{v}\|^p \geq 1$ for any $\boldsymbol{v}' \neq \boldsymbol{v}$, we have that $\mathbb{E}[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p | \mathcal{E}_X \cap \mathcal{E}_Y] \geq 2\sqrt{n/2^d}$ and also the fact that $\mathbb{P}(\mathcal{E}_X \cap \mathcal{E}_Y) \geq \alpha^2 > 0$.

Together this completes the proof of Theorem 2. $\square$

I Proof of Proposition 2

Proof of Proposition 2. We will begin by proving the upper bound on the performance of the SOAR algorithm. We assume that $P = Q = \text{Uniform}([0,1]^d)$ and hence we have that $U_\infty = 0$ for $\varphi_p(X,Y) = -\|X-Y\|^p$ for any $p \geq 1$ and we have that $U_n(\mathsf{SOAR}) = -\frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n \|X_t - Y_{\pi_t^{\mathsf{SOAR}}}\|^p\right]$. Using the definition of regret (per match) and Theorem 1, we have that

$$\begin{aligned}
\textsc{Reg}_n(\mathsf{SOAR}) &\overset{(a)}{=} U_\infty - U_n(\mathsf{SOAR}), \\
&\overset{(b)}{=} \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n \|X_t - Y_{\pi_t^{\mathsf{SOAR}}}\|^p\right], \\
&\overset{(c)}{=} \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left[\frac{1}{k}\min_\sigma \sum_{j=1}^k \|X_j - Y_{\sigma(j)}\|^p\right], \\
&\overset{(d)}{=} \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left[W_p^p(P_k, Q_k)\right].
\end{aligned}$$

where (a) follows from definition of regret, (b) follows from $U_n(\mathsf{SOAR})$, (c) follows from Theorem 1, (d) follows from the definition of $W_p^p(P_k, Q_k)$. Next we will consider three cases : (i) $d=1$, (ii) $d=2$ and (iii) $d \geq 3$.

(i) $d=1$. Using Lemma G.5 for $d=1$, there is a constant $C \equiv C(P,Q,p) < \infty$ such that $\mathbb{E}\left[W_p^p(P_k,Q_k)\right] \leq Ck^{-p/2}$ for $k \geq 1$ and $p \geq 1$. Hence we have that,

$$\begin{aligned}
\textsc{Reg}_n(\mathsf{SOAR}) &\leq \frac{1}{n}\sum_{k=1}^n Ck^{-p/2} \\
&\leq \frac{C}{n}\left(1 + \int_1^n x^{-p/2}dx\right) \\
&\leq C'\left(n^{-p/2}\mathbb{1}\{p<2\} + n^{-1}\log n\mathbb{1}\{p=2\} + n^{-1}\mathbb{1}\{p>2\}\right).
\end{aligned}$$

(ii) $d=2$. Using Lemma G.5 for $d=2$, there is a constant $C \equiv C(P,Q,p) < \infty$ such that

$\mathbb{E}\left[W_p^p(P_k, Q_k)\right] \le C(\log k)^{p/2} k^{-p/2}$ for $k \ge 2$ and $p \ge 1$. Hence we have that,

$$
\begin{aligned}
\text{Reg}_n(\mathsf{SOAR}) &\le \frac{1}{n}\left(C + \sum_{k=2}^{n} C(\log k)^{p/2} k^{-p/2}\right), \\
&\overset{(a)}{\le} \frac{C'}{n}\left((\log n)^{p/2}\int_0^n x^{-p/2}dx\right)\mathbb{1}\{p<2\} + \left(\log n \int_1^n x^{-1}dx\right)\mathbb{1}\{p=2\} \\
&\quad + \frac{C'}{n}\left(\int_1^n x^{-1-\epsilon(p)}dx\right)\mathbb{1}\{p>2\}, \\
&\le C'\left((\log n)^{p/2} n^{-p/2}\mathbb{1}\{p<2\} + n^{-1}\log^2 n \mathbb{1}\{p=2\} + n^{-1}\mathbb{1}\{p>2\}\right),
\end{aligned}
$$

where (a) for $p \le 2$ follows from the fact that $(\log k)^{p/2} \le (\log n)^{p/2}$ for all $k \le n$ and for $p > 2$, we can write $p = 2+2\epsilon$ for some $\epsilon > 0$ and there exists $n_0 \in \mathbb{N}$ such that $(\log n)^{p/2} \le n^{\epsilon}$ for all $n \ge n_0$ and hence we have that $(\log x/x)^{p/2} \le x^{-1-\epsilon(p)}$ for sufficiently large x.

(*iii*) $\underline{d \ge 3.}$ Using Lemma G.5 for $d = 3$, there is a constant $C \equiv C(P, Q, p) < \infty$ such that $\mathbb{E}\left[W_p^p(P_k, Q_k)\right] \le Ck^{-p/d}$ for $k \ge 1$ and $p \ge 1$. Hence we have that,

$$
\begin{aligned}
\text{Reg}_n(\mathsf{SOAR}) &\le \frac{1}{n}\sum_{k=1}^{n} Ck^{-p/d} m \\
&\le \frac{C}{n}\left(1 + \int_1^n x^{-p/d}dx\right) \\
&\le C'\left(n^{-p/d}\mathbb{1}\{p<d\} + n^{-1}\log n \mathbb{1}\{p=d\} + n^{-1}\mathbb{1}\{p>d\}\right).
\end{aligned}
$$

This completes the upper bound proof. Next we will prove the lower bound on the regret of any online optimal policy. For any feasible, non-anticipative online policy π, recall that $U_n^{\mathsf{H}}(P, Q, \varphi) \ge U_n(\pi; P, Q, \varphi)$ for any pair of distributions P and Q and hence we have that

$$
\inf_{\pi \in \Pi} \text{Reg}_n(\pi; P, Q, \varphi) = \inf_{\pi \in \Pi} U_\infty(P, Q, \varphi) - U_n(\pi; P, Q, \varphi) \ge U_\infty(P, Q, \varphi) - U_n^{\mathsf{H}}(P, Q, \varphi).
$$

Recall that for $\varphi(X, Y) = -\|X - Y\|^p$, we have that $U_n^{\mathsf{H}}(P, Q, \varphi) = -\frac{1}{n}\mathbb{E}\left[\min_\sigma \sum_{k=1}^n \|X_k - Y_{\sigma(k)}\|^p\right] = -\mathbb{E}\left[W_p^p(P_n, Q_n)\right]$ and $U_\infty(P, Q, \varphi) = -W_p^p(P, Q) = 0$ since $P = Q$ and hence we have that

$$
\inf_{\pi \in \Pi} \text{Reg}_n(\pi; P, Q, \varphi) \ge \mathbb{E}\left[W_p^p(P_n, Q_n)\right]. \tag{I.1}
$$

We will consider the following three cases: (i) $d = 1$, (ii) $d = 2$ and (iii) $d \geq 3$.

(i) $\underline{d = 1.}$ For the case of $p \leq 2$, from Lemma G.5 and (I.1) it follows that $\inf_{\pi \in \Pi} \text{Reg}_n(\pi; P, Q, \varphi) \geq cn^{-p/2}$ for some constant $c \equiv c(P, Q, p) > 0$.

For the case of $p > 2$, observe that $n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} |X_t - Y_{\pi_t}|^p\right] \geq n^{-1}\mathbb{E}\left[|X_n - Y_{\pi_n}|^p\right] \geq c/n$ for some constant $c > 0$, where the last inequality follows from the fact that $\mathbb{E}\left[|X_n - Y_{\pi_n}|^p\right] \geq c$. This is because of the following reason. At time n, we have only one supply unit remaining. Let B be a ball of radius $r = \Gamma(d/2+1)^{1/d}/\sqrt{\pi} \cdot 2^{-1/d}$ around the location of the last supply unit. We have that $\text{Vol}(B) \leq 1/2$. Since $\text{Vol}([0,1]^d) = 1$, with probability at least 1/2, a demand unit arrives in $[0,1]^d \cap B^c$ and the distance of demand unit is at least r. Note that this is irrespective of the location of the supply unit and hence the expected matching cost at $t = n$ is at least $(r/2)^p$ using Jensen's inequality.

(ii) $\underline{d = 2.}$ For the case of $p \leq 2$, from Lemma G.5 and (I.1) it follows that $\inf_{\pi \in \Pi} \text{Reg}_n(\pi; P, Q, \varphi) \geq c(\log n)^{p/2} n^{-p/2}$ for some constant $c \equiv c(P, Q, p) > 0$.

For the case of $p > 2$, observe that $n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} \|X_t - Y_{\pi_t}\|^p\right] \geq n^{-1}\mathbb{E}\left[\|X_n - Y_{\pi_n}\|^p\right] \geq c/n$ for some constant $c > 0$. The reason for this lower bound follows the exact same reason as in the case of $d = 1$.

(iii) $\underline{d \geq 3.}$ For the case of $p < d$, from Lemma G.5 and (I.1) it follows that $\inf_{\pi \in \Pi} \text{Reg}_n(\pi; P, Q, \varphi) \geq cn^{-p/d}$ for some constant $c \equiv c(P, Q, p) > 0$.

For the case of $p > d$, observe that $n^{-1}\mathbb{E}\left[\sum_{t=1}^{n} \|X_t - Y_{\pi_t}\|^p\right] \geq n^{-1}\mathbb{E}\left[\|X_n - Y_{\pi_n}\|^p\right] \geq c/n$ for some constant $c > 0$. The reason for this lower bound follows the exact same reason as in the case of $d = 1$.

Finally for the case of $p = d$, we have that

$$\sum_{t=1}^{n} \mathbb{E}\left[\|X_t - Y_{\pi_t}\|^p\right] \overset{(a)}{\geq} \sum_{t=1}^{n} \left(\mathbb{E}\left[\|X_t - Y_{\pi_t}\|\right]\right)^p \overset{(b)}{\geq} c\sum_{t=1}^{n} (n-t+1)^{-1} \geq c\int_1^n x^{-1} dx \geq c' \log n,$$

where (a) follows from Jensen's inequality, (b) follows from the following argument: at decision epoch t, we have $n - t + 1$ supply units in the $[0,1]^d$. For any arbitrary location of supply units, we have that the expected distance between the incoming demand unit and the closest supply unit is at least $c(n-t+1)^{-\frac{1}{d}}$ and hence the expected matching cost is of the order

$c(n-t+1)^{-1}$. Let B_i be a ball of radius $r = 2c(n-t+1)^{-\frac{1}{d}}$ around the i-th supply unit for $1 \le i \le n-t+1$. The volume of the ball B_i is proportional to $r^d = 2^d c^d (n-t+1)^{-1}$. For c chosen small enough, we have that $\text{Vol}(B_i) \le \frac{1}{2(n-t+1)}$. Define $B = \cup_{i=1}^{n-t+1} B_i$. Now we have that $\text{Vol}(B) \le \sum_{i=1}^{n-t+1} \text{Vol}(B_i) \le \frac{1}{2}$. Since $\text{Vol}([0,1]^d) = 1$, with probability at least $1/2$, a demand unit arrives in $[0,1]^d \cap B^c$ and the distance of the demand unit is at least $c(n-t+1)^{-\frac{1}{d}}$. This implies that the expected matching cost is at least $\Omega((n-t+1)^{-1})$.

This completes the lower bound proof. □

J Vanishing Regret for polynomial kernel quality function

In this section, we discuss how the performance guarantees for the dot-product quality function $\varphi_{\mathsf{dot}}(X,Y) = \langle X, Y\rangle$ can be leveraged to establish vanishing regret guarantees for the broad class of quality functions which we refer to as the polynomial kernel quality functions $\varphi_{\mathsf{ker}}(X,Y) = \sum_{q=0}^{m} a_q \langle X, Y\rangle^q$.

Corollary J.1. *Suppose P and Q are supported on bounded sets with dimension d. Fix $m \in \mathbb{N}$ and consider the quality function $\varphi_{\mathsf{ker}}(X,Y) = \sum_{q=0}^{m} a_q \langle X, Y\rangle^q$ where $a_q \ge 0$ for all $q \le m$. Define $d' \triangleq \sum_{q=0}^{m} \binom{d+q-1}{q} \mathbb{1}\{a_q > 0\}$. There exists a universal constant $C := C(P, Q, d, \{a_q\}_{q=0}^m) < \infty$ such that*

$$\textsc{Reg}_n(\mathsf{SOAR}) \le C\left(n^{-\frac{1}{2}} \mathbb{1}\{d' \le 3\} + n^{-\frac{1}{2}} \log n \mathbb{1}\{d' = 4\} + n^{-\frac{2}{d'}} \mathbb{1}\{d' \ge 5\}\right)$$

Proof of Corollary J.1. We first consider the quality function $\varphi_{\mathsf{ker}}^q(X,Y) = \langle X, Y\rangle^q$ for some $q \in \mathbb{N}$. From Murphy [2022], it follows that there exists a continuous mapping $\phi_q : \mathbb{R}^d \to \mathbb{R}^{d_q}$ where $d_q \triangleq \binom{d+q-1}{q}$ such that

$$\varphi_{\mathsf{ker}}^q(X,Y) = \langle X, Y\rangle^q = \langle \phi_q(X), \phi_q(Y)\rangle \tag{J.1}$$

In the case that $a_q > 0, q = 1, \ldots, m$, we define ϕ_q in the following form:

$$\phi(z) \triangleq \begin{bmatrix} \sqrt{a_0} \\ \sqrt{a_1}\phi_1(z) \\ \sqrt{a_2}\phi_2(z) \\ \vdots \\ \sqrt{a_m}\phi_m(z) \end{bmatrix}_{d' \times 1},$$

In general, when there exists q s.t. $a_q = 0$, we simply remove the corresponding terms $\sqrt{a_q}\phi_q(z)$ from the above RHS. Hence the dimension of ϕ is $d' = \sum_{q=0}^{m} \binom{d+q-1}{q} \mathbb{1}\{a_q > 0\}$. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the support of the distributions P and Q respectively. Define $\mathcal{X}' = \{\phi(x) : x \in \mathcal{X}\} \subseteq \mathbb{R}^{d'}$ and $\mathcal{Y}' = \{\phi(y) : y \in \mathcal{Y}\} \subseteq \mathbb{R}^{d'}$. Let P' and Q' denote the resulting distributions on the set $\mathcal{X}'$ and $\mathcal{Y}'$ respectively. Using (J.1), we have that

$$\varphi_{\mathsf{ker}}(X, Y) = \sum_{q=0}^{m} a_q \langle X, Y \rangle^q = \langle \phi(X), \phi(Y) \rangle = \langle X', Y' \rangle$$

Note that P' and Q' satisfy the assumptions of Corollary 3 since P and Q are supported on bounded sets and ϕ is a continuous mapping. Therefore, the result in Corollary J.1 follows by invoking Corollary 3. □

K Proof of Theorem 3

Proof of Theorem 3. We begin by proving the upper bound on the regret for the SOAR algorithm. From Lemma G.7, we know that regret of SOAR for the dot-product quality function $\varphi_{\mathsf{dot}} = \langle X, Y \rangle$ is given as

$$\begin{aligned} \textsc{Reg}_n(\mathsf{SOAR}) &= \frac{1}{2}\left[\frac{1}{n}\mathbb{E}\left[\sum_{t=1}^{n} \|X_t - Y_{\pi_t^{\mathsf{SOAR}}}\|^2\right] - W_2^2(P, Q)\right] \\ &= \frac{1}{2n}\sum_{k=1}^{n} \mathbb{E}\left[W_2^2(P_k, Q_k) - W_2^2(P, Q)\right], \end{aligned}$$

where P_k and Q_k denote the empirical measure corresponding to k i.i.d samples from the distributions P and Q respectively. Next we consider the following three cases: (i) $d = 1$, (ii) $d = 2$ and (*iii*) $d \geq 3$.

(*i*) $\underline{d = 1.}$ Using Lemma G.6 for $d = 1$, there is a constant $C \equiv C(P, Q) < \infty$ such that $\mathbb{E}\left[W_2^2(P_k, Q_k) - W_2^2(P, Q)\right] \leq Ck^{-1}$ for all $k \geq 1$. Hence we have that

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{1}{n}\sum_{k=1}^{n} Ck^{-1} \leq \frac{C}{n}\left(1 + \int_1^n x^{-1}dx\right) \leq C'n^{-1}\log n.$$

(*ii*) $\underline{d = 2.}$ Using Lemma G.6 for $d = 2$, there is a constant $C \equiv C(P, Q) < \infty$ such that $\mathbb{E}[W_2^2(P_k, Q_k) - W_2^2(P, Q)] \leq Ck^{-1}(\log k)^2$ for any $k \geq 2$. Hence we have that,

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{1}{n}\left(C + \sum_{k=2}^{n} Ck^{-1}(\log k)^2\right) \leq \frac{C'(\log n)^2}{n}\int_1^n x^{-1}dx = 2C'n^{-1}(\log n)^3.$$

(*iii*) $\underline{d \geq 3.}$ Using Lemma G.6 for $d \geq 3$, there is a constant $C \equiv C(P, Q) < \infty$ such that $\mathbb{E}[W_2^2(P_k, Q_k) - W_2^2(P, Q)] \leq Ck^{-2/d}$ for any $k \geq 1$. Hence we have that

$$\textsc{Reg}_n(\mathsf{SOAR}) \leq \frac{1}{n}\sum_{k=1}^{n} Ck^{-\frac{2}{d}} \leq \frac{C}{n}\int_0^n x^{-\frac{2}{d}}dx = C'n^{-\frac{2}{d}}.$$

This completes the proof of the upper bound in Theorem 2. The lower bound follows from the lower bound in Proposition 2 for $p = 2$. □

L Additional Numerics

In this section, we conduct extensive numerical studies to assess the performance of three interesting variants of SOAR across different input models. For i.i.d. demand with a known distribution, we examine the multi-scenario variant of SOAR, where at each time step, the algorithm simulates multiple future demand sequences to inform decision-making. For i.i.d. demand with an unknown distribution, we test the bootstrap variant of SOAR, where the future demand sequence is drawn from the demand units observed so far. For demand with time-varying distributions, we explore a variant where the future demands are simulated from the corresponding time-dependent distribution. Across all three settings, variants of SOAR demonstrated strong numerical performances, providing encouraging evidence of the versatility of our algorithmic prescription. We hope these numerical findings will spur further research into simulation-based algorithms like SOAR for general resource allocation problems. In all subsequent figures, the confidence intervals are 95%.

L.1 Multi-Scenario Variant of SOAR

The SOAR algorithm can be extended to incorporate multiple future scenarios. Specifically, we draw multiple future demand sequences and choose the supply unit which is most often matched to the true realized demand unit across the scenarios. Let K denote the number of scenarios drawn at each decision time step, and refer to the multi-scenario variant as mSOAR(K) ($K = 1$ is the vanilla SOAR stated in 1). In Figure 5 below, we numerically compare the performance of vanilla SOAR with mSOAR(K for $K = 1, 10, 50$ (refer to the caption for details on the quality function and demand and supply distributions). The results show a clear benefit from simulating the future multiple times: the expected regret (empirical average over 100 instances) for each value of n (number of supply units) decreases as K increases.

For both instances, the numerics suggest that the regret scales at the same rate for different K values, which is expected since SOAR (i.e., $K = 1$) is provably near-optimal scaling, and therefore mSOAR with more future scenarios cannot achieve superior regret scaling. Moreover, our numerical investigation also reveals *diminishing returns* from increasing the number of simulations K; as shown in Figure 5, there is a substantial reduction in regret when moving from $K = 1$ to $k = 10$, but only a minor further reduction when increasing K from 10 to 50.

L.2 Bootstrap Variant of SOAR

If the demand distribution is i.i.d. but unknown to the decision-maker, we can employ a bootstrap approach using observed demands. The bootstrap variant of SOAR operates as follows:

1. At time $t = 1$, we match X_1 with a uniformly random supply unit.
2. For $t > 1$, let $X_1, \ldots, X_{t-1}$ denote the demands that arrive before period $t > 1$, we then sample, with replacement, $n - t - 1$ demands from these as future arrivals.

We conduct experiments on the settings 1) both demand and supply distributions are Uniform($[0, 1]^2$); 2) both demand and supply distributions are discrete with different supports. The matching quality we consider is $\varphi(X, Y) = -\|X - Y\|$ and the results are empirical average over 100 sample paths.

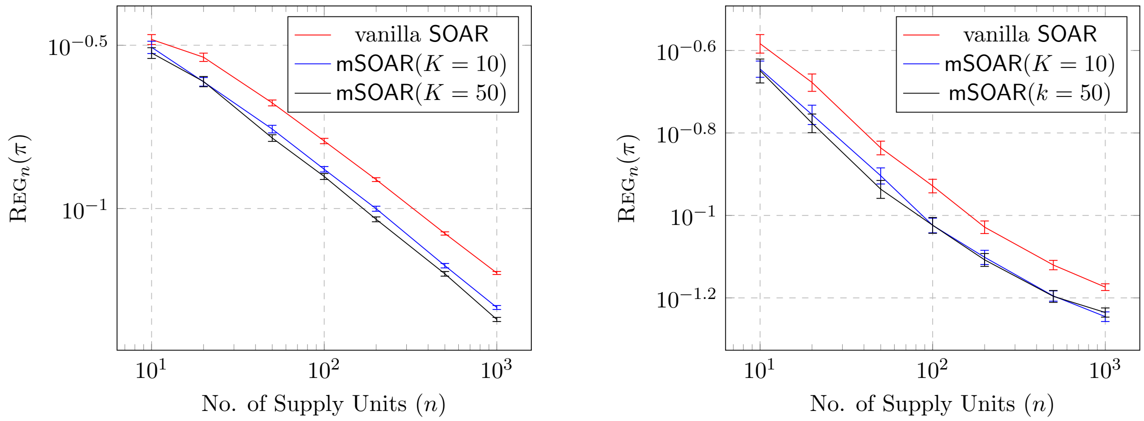
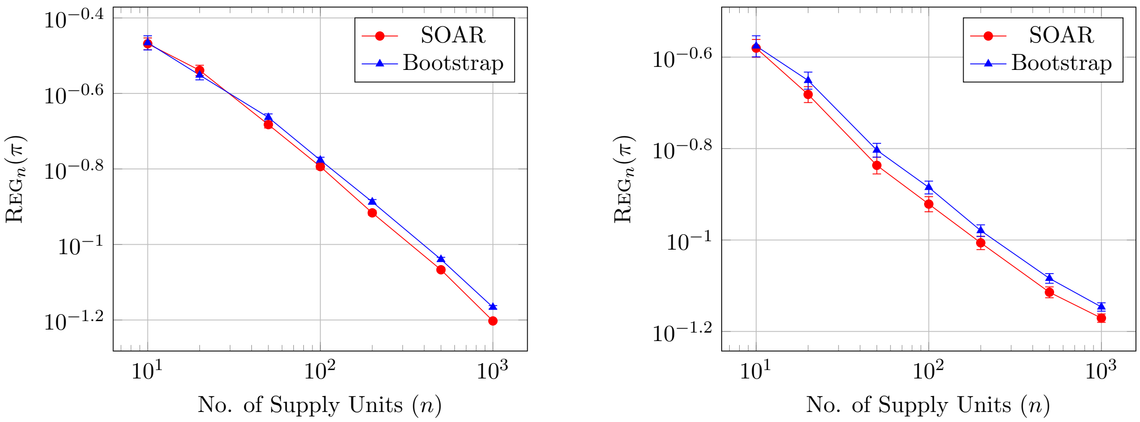


Figure 5: Comparison of the regret for $\varphi(X, Y) = -\|X - Y\|$ for different demand P and supply Q distributions (a) $P = Q = \text{Uniform}([0, 1]^2)$ (b) $P((x_1, x_2)) = Q((y_1, y_2)) = \frac{1}{25}, \forall(x_1, x_2) \in [0.1, 0.3, 0.5, 0.7, 0.9]^2$ and $\forall(y_1, y_2) \in [0, 0.2, 0.4, 0.6, 0.8]^2$

Our numerical experiments (Figure 6) shows that bootstrap-SOAR–initiated without knowledge of the demand distribution–performs comparably to the vanilla SOAR (which **knows** the underlying distribution in advance). In fact, the bootstrap-SOAR and the vanilla SOAR can be viewed as being designed for cold-start (with 0 demand data samples at time 0), and warm-start (with $n(n-1)/2$ demand data samples at time 0) situations, respectively. Our simulation results suggest that the performance gap between the cold-start and warm-start versions of SOAR is minimal, supporting its robustness and applicability.

L.3 Time-Varying Variant of SOAR

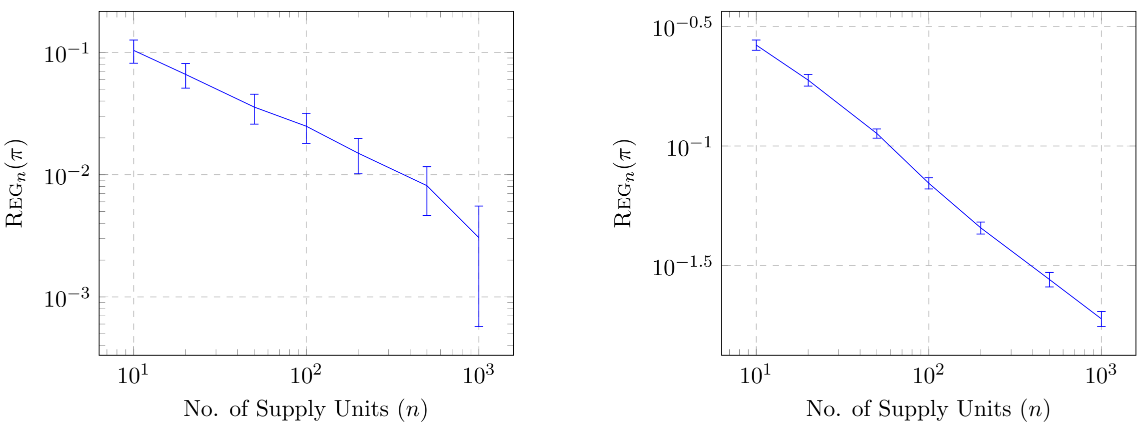
In practice, demand patterns can be time-varying. Let P_t denote the demand distribution for period t, then we can apply a natural generalization of SOAR, the time-varying variant, which simulates future demands using the corresponding distributions $P_{t+1}, P_{t+2}, \ldots, P_n$ at decision period t. We test this variant on the following instances with seasonal demand shift causing non-stationarity: $P^1 := P_1 = P_2 = \ldots = P_{n/2} = \text{Uniform}([0, 1]^2)$ and $P^2 := P_{n/2+1} = \ldots = P_n = \text{Uniform}([1, 2]^2)$ (assume n is even). We also consider another instance with discrete supply and demand distributions. The matching quality functions considered were $\varphi(X, Y) = -\|X - Y\|$, and the results are empirical average over 100 sample paths. We observe that the time-varying variant of SOAR achieves vanishing regret (see Figure 7). Intuitively, given time-varying demand distribu-



(a) (b)

Figure 6: Comparing the performance of (modified) SOAR with 0 demand samples (bootstrap) and $n(n-1)/2$ demand samples for $\varphi(X,Y) = -\|X - Y\|$, (a) $P = Q = \text{Uniform}([0,1]^2)$ (b) $P((x_1,x_2)) = Q((y_1,y_2)) = \frac{1}{25}, \forall (x_1,x_2) \in [0.1, 0.3, 0.5, 0.7, 0.9]^2$ and $\forall (y_1,y_2) \in [0, 0.2, 0.4, 0.6, 0.8]^2$.

tions $\text{Uniform}([0,1]^2)$ and $\text{Uniform}([1,2]^2)$, the variant matches demand X to supply near $X/2$ by simulating future scenarios according to the time-dependent demand distributions.



(a) (b)

Figure 7: Performance of SOAR with time-varying demand distribution. Matching quality function $\varphi(X,Y) = -\|X - Y\|$, (a) $P^1 = \text{Uniform}([0,1]^2), P^2 = \text{Uniform}([1,2]^2), Q = \text{Uniform}([0,1]^2)$, (b) $P^1((x_1,x_2)) = Q((y_1,y_2)) = \frac{1}{25}, \forall (x_1,x_2),(y_1,y_2) \in [0.1, 0.3, 0.5, 0.7, 0.9]^2$ and $P^2((x_1,x_2)) = \frac{1}{25}, \forall (x_1,x_2) \in [0, 0.2, 0.4, 0.6, 0.8]^2$. In (a), note that the confidence intervals appear larger due to the log-log scale, which unevenly stretches the intervals. The true empirical standard deviations are decreasing with the number of supply units.