Online Vector Balancing and Geometric Discrepancy

Nikhil Bansal[*] Haotian Jiang[†] Sahil Singla[‡] Makrand Sinha[§]

Abstract

We consider an online vector balancing question where T vectors, chosen from an arbitrary distribution over $[-1, 1]^n$, arrive one-by-one and must be immediately given a $\pm$ sign. The goal is to keep the discrepancy—the ℓ_∞-norm of any signed prefix-sum—as small as possible. A concrete example of this question is the online interval discrepancy problem where T points are sampled one-by-one uniformly in the unit interval $[0, 1]$, and the goal is to immediately color them $\pm$ such that every sub-interval remains always nearly balanced. As random coloring incurs $\Omega(T^{1/2})$ discrepancy, while the offline bounds are $\Theta((n \log T)^{1/2})$ for vector balancing and 1 for interval balancing, a natural question is whether one can (nearly) match the offline bounds in the online setting for these problems. One must utilize the stochasticity as in the worst-case scenario it is known that discrepancy is $\Omega(T^{1/2})$ for any online algorithm.

In a special case of online vector balancing, Bansal and Spencer [BS19] recently show an $O(\sqrt{n} \log T)$ bound when each coordinate is *independently* chosen. When there are *dependencies* among the coordinates, as in the interval discrepancy problem, the problem becomes much more challenging, as evidenced by a recent work of Jiang, Kulkarni, and Singla [JKS19] that gives a non-trivial $O(T^{1/\log\log T})$ bound for online interval discrepancy. Although this beats random coloring, it is still far from the offline bound.

In this work, we introduce a new framework that allows us to handle online vector balancing even when the input distribution has dependencies across coordinates. In particular, this lets us obtain a $\mathrm{poly}(n, \log T)$ bound for online vector balancing under arbitrary input distributions, and a $\mathrm{polylog}(T)$ bound for online interval discrepancy. Our framework is powerful enough to capture other well-studied geometric discrepancy problems; e.g., we obtain a $\mathrm{poly}(\log^d(T))$ bound for the online d-dimensional Tusńady's problem. All our bounds are tight up to polynomial factors.

A key new technical ingredient in our work is an *anti-concentration* inequality for sums of pairwise uncorrelated random variables, which might also be of independent interest.

[*]CWI Amsterdam and TU Eindhoven, `N.Bansal@cwi.nl`. Supported by the ERC Consolidator Grant 617951 and the NWO VICI grant 639.023.812.

[†]Paul G. Allen School of CSE, University of Washington, `jhtdavid@cs.washington.edu`. Supported in part by the National Science Fundation, Grant Number CCF-1749609, CCF-1740551, DMS-1839116.

[‡]Institute for Advanced Study and Princeton University, `singla@cs.princeton.edu`. Supported in part by the Schmidt Foundation.

[§]CWI Amsterdam, `makrand@cwi.nl`. Supported by the Netherlands Organization for Scientific Research, Grant Number 617.001.351 and the NWO VICI grant 639.023.812.

Contents

1 Introduction

Consider the following online vector balancing question, originally proposed by Spencer [Spe77]: vectors $v_1, v_2, \ldots, v_T \in [-1,1]^n$ arrive online, and upon the arrival of v_t, a sign $\varepsilon_t \in \{\pm 1\}$ must be chosen irrevocably, so that the ℓ_∞-norm of the signed sum $d_t = \varepsilon_1 v_1 + \ldots + \varepsilon_t v_t$ remains as small as possible. That is, find the smallest B such that $\max_{t\in[T]} \|d_t\|_\infty \leq B$. As we shall see later, the problem arises naturally in various contexts where one wants to divide an incoming stream of objects, so that the split is as even as possible along each of the various dimensions that one might care about.

A naïve algorithm is to pick each sign ε_t randomly and independently, which by standard tail bounds gives $B = \Theta((T \log n)^{1/2})$ with high probability. In most of the interesting settings, $T \gg n$, and a natural question is whether the dependence on T can be improved from $T^{1/2}$ to say, $\log T$, or removed altogether (possibly with a worse dependence on n).

Offline setting. The offline version of the problem, where the vectors $v_1, \ldots, v_T$ are given in advance and the goal is to minimize $\max_{t\in[T]} \|d_t\|_\infty$, is known as the signed-series problem. It was first studied by Spencer [Spe77], who obtained a bound independent of T, but exponential in n. This was later improved by Bárány and Grinberg [BG81] to $B \leq 2n$. Chobanyan [Cho94] showed a beautiful connection between the signed-series problem and the classic Steinitz problem on rearrangement of vector sequences—any upper bound on B also holds for the latter problem. Steinitz problem has a much longer history, originating from a question of Riemann and Lévy in the 19th century (c.f. the survey [Bár08] for some fascinating history). A long-standing conjecture for both the problems, still open, is that $B = O(n^{1/2})$. Another notable bound is due to Banaszczyk [Ban12], who showed that $B = O((n \log T)^{1/2})$. While the original argument in [Ban12] was non-constructive, a polynomial time algorithm to find such a signing was recently given in [BG17].

In general, there has been extensive work on various offline discrepancy problems over last several decades, and several powerful techniques such as the partial coloring method [Spe85] and convex geometric methods [Gia97, Ban98, Ban12, MNT14] have been developed, which significantly improve upon the bounds given by random coloring. While these initial methods were mostly non-algorithmic, several new algorithmic techniques and insights have been developed in recent years [Ban10, LM15, Rot14, ES18, BDG16, LRR17, BDGL18, DNTT18].

Online setting. The online setting was first studied in the 70's and 80's, but it did not receive much interest later as it was realized that the best guarantees are already achieved by trivial algorithms. In particular, the $T^{1/2}$ dependence on T achieved by random coloring cannot be improved [Spe77]. See [Spe87, Bár79] for even more specific lower bounds. The difficulty is that the all-powerful adversary, upon seeing the signs chosen by the algorithm until time $t-1$, can choose the next input vector v_t to be *orthogonal* to d_{t-1}. Now, irrespective of the choice of the sign ε_t, the resulting signed sum d_t satisfies

$$\|d_t\|_2^2 \;=\; \|d_{t-1} + \varepsilon_t v_t\|_2^2 \;=\; \|d_{t-1}\|_2^2 + 2\varepsilon_t \langle d_{t-1}, v_t\rangle + \|v_t\|_2^2 \;=\; \|d_{t-1}\|_2^2 + \|v_t\|_2^2. \tag{1}$$

For any d_{t-1}, one can always pick v_t with[1] $\|v_t\|_\infty \leq 1$ and $\|v_t\|_2^2 \geq n-1$, resulting in $\|d_t\|_2^2 \geq (n-1)t$, and hence $\|d_t\|_\infty = \Omega(t^{1/2})$ for all $t \in [T]$ (as long as $n > 1$).

It is therefore natural to ask if relaxing the power of the adversary, or making additional assumptions on the input sequence, can lead to interesting new ideas and to algorithms that perform much better, and in particular give bounds that only mildly depend on T.

A natural assumption is that of *stochasticity*: if the arriving vectors are chosen in an i.i.d. manner from some distribution p, can we maintain that the ℓ_∞ norm of the current signed-sum d_t—henceforth, referred to as discrepancy—is $\mathrm{poly}(n)$ or $\mathrm{poly}(n, \log T)$?

Previous work and challenges. Recently, this stochastic setting was studied by Bansal and Spencer [BS19], where they considered the case where p is the uniform distribution on all $\{-1,1\}^n$ vectors.

[1]For any $d \in \mathbb{R}^n$, any basic feasible solution to $\langle d, x\rangle = 0$ with $x \in [-1,1]^n$ has at least $n-1$ coordinates ± 1.

They give an online algorithm achieving a bound of $O(\sqrt{n})$ on the *expected* discrepancy, matching the best possible offline bound, and an $O(\sqrt{n}\log T)$ discrepancy bound at all times $t \in [T]$, with high probability.

In general, the algorithmic discrepancy approaches developed in the last decade do not seem to help in the online setting. This is because in the offline setting, the algorithms can ensure that the discrepancy stays low by *simultaneously* updating the colors of various elements in a correlated way. In the online setting, however, the discrepancy must necessarily rise (in the ℓ_2 sense) whenever the incoming vector v_t is almost orthogonal to d_{t-1}, which can happen quite often. The only thing that the online algorithm can do is to *actively* try to *cancel* this increase, whenever possible, by choosing the sign ε_t cleverly.

The algorithm of [BS19] crucially uses that if the coordinates of v_t are independently distributed and mean-zero[2], then for any d_{t-1} the incoming vector v_t will typically be far from being orthogonal to d_{t-1}. More quantitatively, the *anti-concentration* property for independent random variables gives that for any $d_{t-1} = (d_1, \ldots, d_n)$, the random vector $v_t = (X_1, \ldots, X_n)$ with $X_1, \ldots, X_n$ being independent and mean-zero satisfies

$$\mathbb{E}_v\Big[|\langle d_{t-1}, v_t\rangle|\Big] = \Omega\left(\Big(\sum_{i=1}^n d_i^2 \cdot \mathbb{E}[X_i]^2\Big)^{1/2}\right).$$

Whenever $|\langle d_{t-1}, v_t\rangle|$ is large, the algorithm can choose ε_t appropriately to create a *negative drift* in (1), to offset the increase due to the $\|v_t\|^2$ term. We give a more detailed description below in §2.1.

In many interesting settings, however, the X_i's can be *dependent*. For example, motivated by an envy minimization problem, Jiang, Kulkarni, and Singla [JKS19] considered the following natural online interval discrepancy problem: points $x_1, \ldots, x_T$ arrive uniformly in the interval $[0,1]$, and the goal is to assign them signs online to minimize the discrepancy of every sub-interval of $[0,1]$. Viewing the sub-intervals as coordinates, this becomes a stochastic online vector balancing problem, but where the random variables X_i corresponding to the various sub-intervals are dependent (details in §2.2). They give a non-trivial algorithm that achieves $T^{1/\log\log T}$ discrepancy, which is much better than the $T^{1/2}$ bound obtained by random coloring, but still substantially worse than $\mathrm{polylog}(T)$.

In general, the difficulty with dependent coordinates X_i is that even a small correlation can destroy anti-concentration, which makes it difficult to create a negative drift. For example, suppose the distribution p is mostly supported on vectors with equal number of $+1$ and -1 coordinates. Now if d has the form $d = c(1, \ldots, 1)$, then the incoming vector v_t is almost always orthogonal to it, and $\|d_T\|_2$ can potentially increase as fast as $\Omega(T^{1/2})$.

In this paper, we focus on the stochastic setting where the coordinates have dependencies, and give several results both for specific geometric problems and for general vector balancing under arbitrary distributions. In general, there are various other ways in which one can relax the power of the adversary, and in §8 we describe several interesting open questions and directions in this area.

1.1 Our Discrepancy Bounds

We first consider the following interval discrepancy problem. Let $x = x_1, \ldots, x_T$ be a sequence of points drawn uniformly in $[0,1]$ and let $\varepsilon_1, \ldots, \varepsilon_T \in \{\pm 1\}$ be a signing. For an interval $I \subseteq [0,1]$, let $\mathbf{1}_I$ denote the indicator function of the interval I. For any time $t \in [T]$, we define the discrepancy of interval I to be

$$\mathsf{disc}_t(I) := \Big|\varepsilon_1 \mathbf{1}_I(x_1) + \cdots + \varepsilon_t \mathbf{1}_I(x_t)\Big|.$$

We show the following bounds on discrepancy.

Theorem 1.1 (Interval Discrepancy)**.** *There is an online algorithm which selects signs $\varepsilon_t \in \{\pm 1\}$ such that, with high probability,* for every interval $I \subseteq [0,1]$ *we have* $\max_{t\in[T]} \mathsf{disc}_t(I) = O(\log^3 T)$. *Moreover, with constant probability, for any online algorithm,* $\max_{I\subseteq[0,1]} \max_{t\in[T]} \mathsf{disc}_t(I) = \Omega\left(\sqrt{\log T}\right)$.

[2] Note that this holds in the case of uniform distribution over $\{-1, 1\}^n$.

This gives an exponential improvement over the $T^{1/\log\log T}$ bound of [JKS19], and is tight up to polynomial factors. The lower bound also improves a previous bound of $\Omega(\log^{1/4} T)$ of [JKS19].

There are two natural d-dimensional generalizations of the interval discrepancy problem, and our framework, which we will describe in §1.2, can handle both of them.

d-dimensional Online Interval Discrepancy: Consider a sequence of points $x_1, \ldots, x_T$ drawn uniformly from the unit cube $[0,1]^d$. The goal is to simultaneously minimize the discrepancy of every interval for all the d-coordinates. In other words, to minimize the following for every interval I and every coordinate $i \in [d]$:

$$\mathsf{disc}_t^i(I) := \Big|\varepsilon_1 \mathbf{1}_I(x_1(i)) + \ldots + \varepsilon_t \mathbf{1}_I(x_t(i))\Big|.$$

The offline version of this problem for $d \geq 2$ is equivalent to the classic d-permutations problem, where an upper bound of $O(\sqrt{d}\log T)$ [SST97] and a breakthrough lower bound of $\Omega(\log T)$ [NNN12, Fra18] for $d \geq 3$, and $\Omega(\sqrt{d})$ in general is known for the worst-case placement of points.

We show the following result that matches the best offline bounds, up to polynomial factors.

Theorem 1.2 (d-dimensional Interval Discrepancy)**.** *There is an online algorithm which selects signs $\varepsilon_t \in \{\pm 1\}$ such that, with high probability, for each $i \in [d]$ and $I \subseteq [0,1]$, we have $\max_{t\in[T]} \mathsf{disc}_t^i(I) = O(d \log^3 T)$. Moreover, with constant probability, for any online algorithm there exists an interval I and a coordinate $i \in [d]$, such that $\max_{t\in[T]} \mathsf{disc}_t^i(I) = \Omega\big(\sqrt{d\log{(T/d)}}\big)$.*

Previously, Jiang et al. [JKS19] could extend their analysis for online interval discrepancy to the $d = 2$ case and prove the same $T^{1/\log\log T}$ bound. However, their proof is rather ad-hoc and does not seem to generalize to higher d. In contrast, our bound holds for any d, and is tight up to polynomial factors.

The second natural generalization of interval discrepancy is to d-dimensional axis-parallel boxes, which gives the following online version of the extensively studied Tusnády's Problem.

d-dimensional Online Tusnády's Problem: Consider a sequence of points $x_1, \ldots, x_T$ drawn uniformly from the unit cube $[0,1]^d$. The goal is to simultaneously minimize the discrepancy of all axis-parallel boxes. In other words, to minimize the following for every box B:

$$\mathsf{disc}_t(B) := \Big|\varepsilon_1 \mathbf{1}_B(x_1) + \ldots + \varepsilon_t \mathbf{1}_B(x_t)\Big|.$$

The (offline) Tusnády's problem has a fascinating history (see [Mat09] and references there in), and after a long line of work, it is known that for the worst-case placement of points, the offline discrepancy is at most $O_d(\log^{d-\frac{1}{2}} T)$ [Nik17] and at least $\Omega_d(\log^{d-1} T)$ [MN15]. We show the following result in the online setting, which is tight to within polynomial factors.

Theorem 1.3 (Tusnády's problem)**.** *There is an online algorithm which selects signs $\varepsilon_t \in \{\pm 1\}$ such that, with high probability, for every axis-parallel box B, we have $\max_{t\in[T]} \mathsf{disc}_t(B) = O_d(\log^{2d+1} T)$. Moreover, for any online algorithm, there exists a box B, such that $\max_{t\in[T]} \mathsf{disc}_t(B) = \Omega_d(\log^{d/2} T)$.*

In contrast, the proof approach of [JKS19] completely breaks down for the Tusnády's problem even in two dimensions, and does not give any better lower bounds in terms of d.

Remark: Although all the problems above are stated for uniform distributions, one can use the probability integral transformation to reduce any product distribution to the uniform distribution without increasing the discrepancy, so our results in Theorems 1.2 and 1.3 also apply to any product distribution over $[0,1]^d$.

Finally, note that Theorem 1.1 follows as a direct corollary of either of the above theorems.

General distributions. We now consider the setting of *arbitrary* distributions for the online vector balancing problem. Here we need to tackle the orthogonality issue which gave $\Omega(T^{1/2})$ lower bounds discussed in (1). As discussed earlier, for the uniform distribution over $\{-1,+1\}^n$, Bansal and Spencer [BS19] get around this issue since this does not happen for the uniform distribution reasonably often, and hence, $\mathbb{E}[\langle d_{t-1}, v_t\rangle]$ is large for any vector d_{t-1}. Using this, they obtain the bound $O(n^{1/2}\log T)$. Our next result shows that such a $\mathrm{poly}(n, \log T)$ upper bound is possible even for arbitrary distributions.

Theorem 1.4. (Vector balancing under dependencies) *For any sequence of vectors $v_1, \ldots, v_T \in [-1,1]^n$ sampled i.i.d. from some arbitrary distribution p, there is an online algorithm which selects signs $\varepsilon_t \in \{\pm 1\}$ such that, with high probability, we have*

$$\max_{t\in[T]} \left\|\varepsilon_1 v_1 + \ldots + \varepsilon_t v_t\right\|_\infty = O(n^2(\log T + \log n)).$$

We remark that the dependencies on n and $\log T$ in this theorem are tight up to polynomial factors.

All of the above results follow from a general framework that we discuss next. In addition to the framework below, the key new technical ingredient is an anti-concentration inequality for dependent random variables, which we describe below in Theorem 1.5. This may be of independent interest.

1.2 Our Framework

To tackle the orthogonality issue, one of our key idea is to work with a *different basis* for the discrepancy vectors. More specifically, instead of maintaining bounds on the individual coordinate discrepancies $d_t(i)$, we maintain bounds on suitable linear combinations of them. This basis ensures that the (new) coordinates of the incoming vector are *uncorrelated*, i.e., $\mathbb{E}[X(i)\cdot X(j)] = \mathbb{E}[X(i)]\cdot\mathbb{E}[X(j)]$ for distinct coordinates i, j. Note that this condition is only on the expected values, and is much weaker, e.g., even pairwise independence. Once one finds a suitable new basis, which turns out to be an eigenbasis of the covariance matrix, the anti-concentration bound for such random variables (proved below in Theorem 1.5), together with the standard exponential penalty based framework used in previous works [BS19, JKS19], gives Theorem 1.4.

For our results on geometric discrepancy problems, there is an additional challenge, we cannot afford to lose a $\mathrm{poly}(n)$ factor, as in Theorem 1.4 above, since the dimension $n = \Theta(T)$. In this case, however, the update vectors are $(\log T)$-sparse in the original basis (see §2) and one could hope to utilize this sparsity. Yet another challenge in this case is that bounding the discrepancy in a new basis preserves ℓ_2-discrepancy in the original basis, but could lead to a $\sqrt{n}$ loss in ℓ_∞-discrepancy. To get $\mathrm{polylog}(T)$ bounds, we use a natural basis from wavelet theory, called the *Haar system*, which simultaneously has sparsity, uncorrelation, and avoids the ℓ_2 to ℓ_∞ loss. This also easily extends to higher dimensions as these wavelets can be tensorized in a natural way to get a suitable basis for higher dimensional versions of the problems. A more detailed description of our framework is given in §2. Next we discuss our anti-concentration results.

1.3 Our Anti-Concentration Results for Non-Independent Random Variables

Suppose $X_1, \ldots, X_n$ are independent $\{-1,+1\}$ random variables with mean zero. Then, it is well-known that $|\sum_i X_i|$ has mean $\Theta(n^{1/2})$, and moreover, this value is at least $\Omega(n^{1/2})$ with constant probability.

Now, on the other hand, consider the following distribution. Let H_n be $n\times n$ Hadamard matrix and let $H_n(i)$ denote its i-th row for $i \in [n]$. Consider the random vector $X = (X_1, \ldots, X_n)$, where $X = \xi \cdot H_n(i)$ for a Rademacher random variable $\xi \in \{-1,+1\}$ and a uniformly chosen $i \in [n]$. Then the X_i's are still mean-zero and $\{-1,+1\}$, and in fact, pairwise independent. However, the magnitude of the sum $|\sum_i X_i|$ behaves very differently from the i.i.d. setting above. It takes value n with probability only $1/n$ (if $X = \xi \cdot H_n(1)$, the row of all 1's) and is 0 otherwise. In particular the mean is $\mathbb{E}[|\sum_i X_i|] = 1$ (instead of $n^{1/2}$ above), and moreover the entire contribution to the mean comes from an event with probability only $1/n$.

Nevertheless, we can say interesting things about the anti-concentration of sums of such random variables. In particular, we show the following results for uncorrelated or pairwise independent random variables.

Theorem 1.5. (Uncorrelated anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be uncorrelated random variables that are bounded $|X_i| \le c$, satisfy $\mathbb{E}[X_i X_j] = 0$ for all $i \ne j$, and have sparsity s (the number of non-zero X_i's in any outcome). Then*

$$\mathbb{E}\Big|\sum_i a_i X_i\Big| \;\ge\; \mathbb{E}\Big[\sum_i |a_i| X_i^2\Big] \cdot \frac{1}{cs}. \tag{2}$$

Moreover, this bound is tight, even for pairwise independent random variables.

The tightness holds for the Hadamard example above, where $\mathbb{E}|\sum_i X_i| = 1$, $s = n$, $c = 1$, and $\mathbb{E}[\sum_i X_i^2] = n$.

Theorem 1.6. (Pairwise independent anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be mean-zero pairwise independent random variables with sparsity $s \le n$. Then*

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big|\Big] \;\ge\; \mathbb{E}\Big[\sum_i |a_i X_i|\Big] \cdot \frac{1}{s}. \tag{3}$$

Note that this bound is also tight for the Hadamard example. In general, the bound (3) is stronger than in (2); and a simple example in §3.2 shows that (3) cannot hold for uncorrelated random variables.

Although the anti-concentration properties and the small-ball probabilities for independent variables have been extensively studied (c.f. [NV13]), the uncorrelated and pairwise independent setting does not seem to have been studied before, and Theorems 1.5 and 1.6 do not seem to be known, to the best of our knowledge.

1.4 Applications to Envy Minimization

A classic measure of fairness in the field of fair division is envy [Fol67, TV85, LMMS04, Bud11]. In a recent work, Benade et al. [BKPP18] defined the *online envy minimization* where there are two players and T items arrive one by one. On arrival of an item $t \in \{1, \ldots, T\}$, we get to see the valuations $v_{it} \in [0, 1]$ for both the players $i \in \{1, 2\}$. The goal is to immediately and irrevocably allocate the item to one of the players while minimizing the maximum *envy*. There are two natural notions of envy: cardinal and ordinal (see §7 for definitions). Benade et al. [BKPP18] show an $\Omega(T^{1/2})$ lower bound for online envy minimization in the *adversarial* model—the reason is again similar to the lower bound for online discrepancy. Can we obtain better bounds when the player valuations are drawn from a distribution?

In the special case of product distributions (each player independently draws their value), Jiang et al. [JKS19] observed that the 2-dimensional interval discrepancy bounds also hold for online envy minimization. In particular, they obtained a $T^{1/\log\log T}$ bound on the ordinal envy. Our new interval discrepancy bound from Theorem 1.2 immediately improves this to an $O(\log^3 T)$ bound on ordinal envy. Moreover, we can also use our vector balancing result to obtain an $O(\log T)$ bound on the cardinal envy for general distributions.

Corollary 1.7. *Suppose valuations of two players are drawn i.i.d. from some distribution p over $[0, 1] \times [0, 1]$. Then, for an arbitrary distribution p (i.e., player valuations for the same item could be correlated), the online cardinal envy is $O(\log T)$. Moreover, if p is a product distribution (i.e., player valuations for the same item are independent) then the online ordinal envy is also $O(\log^3 T)$.*

Paper Organization

The rest of the paper is organized as follows: in §2, we give an overview of previous challenges and our main ideas. In §3, we prove our key anti-concentration theorems that are necessary for our upper bounds on discrepancy. In §4, we give upper and lower bounds for online discrepancy under certain "uncorrelation" assumptions on the distribution. Then, we apply these bounds in §5 to obtain our vector balancing result (Theorem 1.4). In §6, we again apply these bounds to obtain our geometric discrepancy results (Theorems 1.2 and 1.3). In §7, we show why our results immediately apply to online envy minimization. Finally, in §8 we end with some discussion of open problems and directions.

2 Proof Overview

Let us start by reviewing the approach considered by Bansal and Spencer [BS19] in the case of independent coordinates. We also discuss the challenges involved in extending it to the setting of dependent coordinates.

2.1 Independent Coordinates: Bansal and Spencer

Consider the online vector balancing problem, when each arriving vector is uniformly chosen from $\{\pm 1\}^n$, so that all the coordinates are independent. To design an online algorithm, it is natural to keep a potential function that keeps track of the discrepancy and chooses a sign ε_t for the current vector v_t that minimizes the increase in the potential. Formally, let $d_t = \varepsilon_1 v_1 + \ldots + \varepsilon_t v_t$ denote the discrepancy vector at time t. For a parameter $0 < \lambda < 1$, define the potential function

$$\Phi_t = \sum_{i \in [n]} \cosh(\lambda d_t(i)),$$

where $d_t(i)$ denotes the ith coordinate of d_t and $\cosh(x) = \frac{1}{2} \cdot (e^x + e^{-x})$ for all $x \in \mathbb{R}$. One should think of the above potential function as a proxy for the maximum discrepancy as Φ_t is dominated by the maximum discrepancy: $\Phi_t \approx e^{\lambda \|d_t\|_\infty}$.

On the arrival of vector v_t, the algorithm chooses a sign $\varepsilon_t \in \{\pm 1\}$, which updates the discrepancy vector to $d_t = d_{t-1} + \varepsilon_t v_t$ and changes the potential from Φ_{t-1} to Φ_t. If we can show that whenever $\Phi_t > 2n$, the *drift* $\Delta\Phi_t := \Phi_t - \Phi_{t-1}$ is negative in expectation for the sign ε_t chosen by the algorithm, then we can say that the potential after T arrivals, Φ_T, is bounded by $\mathrm{poly}(nT)$ with high probability. This implies $\cosh(\lambda\|d_T\|_\infty)$ is bounded by $\mathrm{poly}(nT)$, which means a bound of $O(\lambda^{-1} \log T)$ on the maximum discrepancy.

Let us try to compute the expected drift. Define $d = d_{t-1}$. By considering the second order Taylor expansion of $\cosh(x + \delta) \approx \cosh(x) + \sinh(x)\delta + \cosh(x)\delta^2$, where $\sinh(x) = \frac{1}{2} \cdot (e^x - e^{-x})$ for all $x \in \mathbb{R}$, we get that

$$\Delta\Phi_t \;\approx\; \sum_{i \in [n]} \Big(\lambda \sinh(\lambda d(i)) \cdot (\varepsilon_t v_t(i)) + \lambda^2 \cosh(\lambda d(i)) \cdot (\varepsilon_t v_t(i))^2\Big) \;=\; \varepsilon_t \lambda L + \lambda^2 Q,$$

where $L = \sum_{i \in [n]} \sinh(\lambda d(i)) \cdot v_t(i)$ is the *linear* term and $Q = \sum_{i \in [n]} \cosh(\lambda d(i))$ is the *quadratic* term from the Taylor expansion (note that $(\varepsilon_t v_t(i))^2 = 1$). Since the algorithm is free to choose the sign ε_t to minimize the drift, $\Delta\Phi_t \approx -\lambda|L| + \lambda^2 Q$. Now if one can show that $\mathbb{E}_{v_t}[|L|] \geq \frac{\mathbb{E}[Q]}{2\lambda}$, we would get that expected drift $\mathbb{E}[\Delta\Phi_t] < 0$, and if λ is large then this would translate to a good discrepancy bound of $O(\lambda^{-1} \log T)$ as described above.

Since $\cosh(x)$ and $|\sinh(x)|$ only differ by at most 1, we can make the approximation $Q \approx \sum_{i \in [n]} |\sinh(\lambda d(i))|$ up to some small error. So, denoting $\beta = 1/\lambda$ and $a_i = \sinh(\lambda d(i))$, our task reduces to proving the following anti-concentration statement:

Question. Let $X_1, \ldots, X_n$ be independent random variables with $|X_i| \leq 1$. What is the smallest β such that the following holds:

$$\mathbb{E}\Big[\Big|\sum_{i \in [n]} a_i X_i\Big|\Big] \geq \frac{1}{\beta} \cdot \mathbb{E}\Big[\sum_{i \in [n]} |a_i| X_i^2\Big]. \tag{4}$$

In the case where the X_i's are independent Rademacher (± 1) random variables, classical Khintchine's inequality and Cauchy-Schwarz tell us that

$$\mathbb{E}\Big[\Big|\sum_{i \in [n]} a_i X_i\Big|\Big] \;\geq\; \frac{1}{\sqrt{2}} \cdot \Big(\sum_{i \in [n]} a_i^2\Big)^{1/2} \;\geq\; \frac{1}{\sqrt{2n}}\Big(\sum_{i \in [n]} |a_i|\Big) \;=\; \frac{1}{\sqrt{2n}} \cdot \mathbb{E}\Big[\sum_{i \in [n]} |a_i| X_i^2\Big],$$

so $\beta = O(\sqrt{n})$, which suffices for the discrepancy application. In general, when X_i's are not Rademacher but are still bounded ($|X_i| \leq 1$), mean-zero, and *independent*, then following [BS19] one can still show that $\beta = O(\sqrt{n})$.

The above gives a bound of $O(\sqrt{n}\log T)$ on the maximum discrepancy at every time $t \in [T]$. However, when the input distribution has dependencies across coordinates, i.e. the X_i's are dependent, one can not take β to be small in general. For example, $\beta \to \infty$ when all a_i's are one and a random set of coordinates $S \subset [n]$ of size $n/2$ (say n is even) take value $+1$ and the remaining coordinates in $[n] \setminus S$ take value -1.

Next we discuss the simplest geometric discrepancy problem—the interval discrepancy problem in one dimension—where such a situation already arises if we use the same approach as above.

2.2 Interval Discrepancy: Previous Barriers

Recall, we have T points $x_1, \ldots, x_T$ chosen uniformly from $[0,1]$ which need to be given ± 1 signs online. Consider the *dyadic* intervals $I_{j,k} := [k2^{-j}, (k+1)2^{-j}]$ where $0 \le k < 2^j$ and $0 \le j \le \log T$. For intuition, imagine embedding the unit interval on a complete binary tree of height $\log T$; now sub-intervals corresponding to every node of the binary tree are dyadic intervals. Note that the smallest dyadic interval has size $2^{-\log T} = 1/T$. By a standard reduction, every sub-interval of $[0,1]$ is contained in a union of some $O(\log T)$ dyadic intervals, so it suffices to track the discrepancy of these dyadic intervals.

Denoting by $\mathbf{1}_I$ the indicator function for an interval I, define

$$d_t(I) := \varepsilon_1 \mathbf{1}_I(x_1) + \ldots + \varepsilon_t \mathbf{1}_I(x_t).$$

Note that $|d_t(I_{j,k})|$ is the discrepancy of the interval $I_{j,k}$ at time t. A natural choice of algorithm is to use the potential function

$$\Phi_t = \sum_{j,k} \cosh(\lambda d_t(I_{j,k})),$$

which is a proxy for the maximum discrepancy of any dyadic interval. Ideally, we want to set $0 < \lambda < 1$ as large as possible. Defining $d_{j,k} = d_{t-1}(I_{j,k})$, and doing a similar analysis as before, we derive

$$\Delta\Phi_t \approx \varepsilon_t \lambda L + \lambda^2 Q,$$

where $L = \sum_{j,k} \sinh(\lambda d_{j,k}) \cdot \mathbf{1}_{I_{j,k}}(x_t)$ and $Q = \sum_{j,k} \cosh(\lambda d_{j,k}) \cdot \mathbf{1}_{I_{j,k}}(x_t)^2$. The problem again reduces to showing an anti-concentration statement as in Eq. (4) with X_i's being the indicators $\mathbf{1}_{I_{j,k}}$ for all j,k. It turns out that the smallest β one can hope for this setting is exponential in the height of the tree (see Appendix A for an example), which for binary trees of height $\log T$ only yields a $\mathrm{poly}(T)$ bound on the discrepancy.

One can still leverage something out of this approach—letting $B = T^{1/\log\log T}$, it was shown by Jiang, Kulkarni, and Singla [JKS19] that by embedding B-adic intervals on a B-ary tree of height $\log\log T$, the above approach gives a sub-polynomial $T^{1/\log\log T}$ bound for the interval discrepancy problem. However, this cannot be pushed to give a $\mathrm{polylog}(T)$ bound because the above obstruction does not allow us to handle trees of height $\log T$.

2.3 Interval Discrepancy: A New Potential and the BDG Inequality

To get around the previous problem, we take a different approach and instead of directly using the discrepancies in the potential Φ_t, we work with linear combinations of discrepancies with the following desirable properties. First, if there is a bound on these linear combinations then it should imply a bound on the original discrepancies. Second, and more importantly, the term L in $\Delta\Phi_t$ can be viewed as a martingale, which leads to much better anti-concentration properties, i.e., smaller β in (4).

More specifically, consider the previous embedding of the dyadic intervals of length at least $1/T$ on the complete binary tree of depth $\log T$. For any interval $I_{j,k}$, let the left half interval be $I^l_{j,k}$ and the right half interval be $I^r_{j,k}$, and consider the difference (see Figure 1) of their discrepancies

$$d^-_t(I_{j,k}) := d_t(I^l_{j,k}) - d_t(I^r_{j,k}).$$

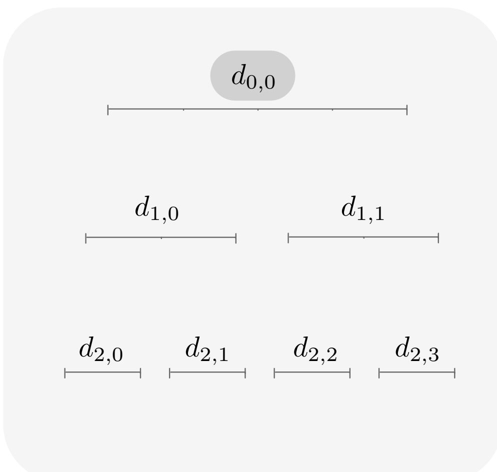
Note that if $|d_t(I_{j,k})| \le \alpha$ and also $|d^-_t(I_{j,k})| \le \alpha$, then both $|d_t(I^l_{j,k})| \le \alpha$ and $|d_t(I^r_{j,k})| \le \alpha$. A simple inductive argument now shows that if $|d_t([0,1])| \le \alpha$ and the differences of discrepancy for every dyadic interval $I_{j,k}$ satisfies $|d_t^-(I_{j,k})| \leq \alpha$, then every dyadic interval also has discrepancy at most α, thus satisfying the first property above. So let us consider a different potential function:

$$\Xi_t := \cosh(\lambda d_t(I_{0,0})) + \sum_{j,k} \cosh(\lambda d_t^-(I_{j,k}))$$

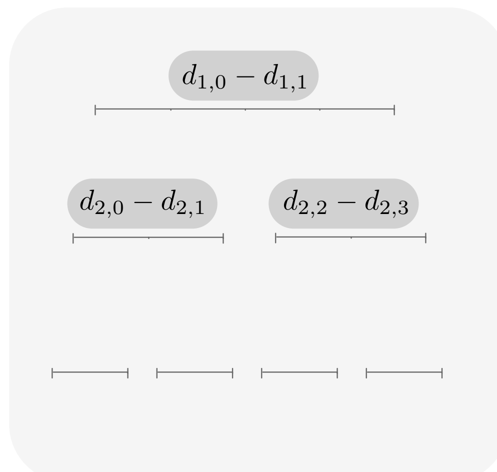
with j,k ranging over all the dyadic intervals (corresponding to internal nodes of the tree) and $0 < \lambda < 1$ is a parameter that we want to set as large as possible. Denoting $d_{j,k}^- = d_{t-1}^-(I_{j,k})$, as before, we can write $\Delta\Xi_t \approx \varepsilon_t \lambda L + \lambda^2 Q$, with

$$L = \sinh(\lambda d_t(I_{0,0})) + \sum_{j,k} \sinh(\lambda d_{j,k}^-) \cdot X_{j,k}(x_t) \quad \text{and}$$
$$Q = \cosh(\lambda d_t(I_{0,0})) + \sum_{j,k} \cosh(\lambda d_{j,k}^-) \cdot X_{j,k}(x_t)^2,$$

where $X_{j,k} = \mathbf{1}_{I_{j,k}^l} - \mathbf{1}_{I_{j,k}^r}$ for any interval $I_{j,k}$. Note that $X_{j,k}$ takes value 1 on the left half of $I_{j,k}$, and -1 on the right half of $I_{j,k}$, and is zero otherwise.



(a) The discrepancy $d_{j,k}$ terms for intervals $I_{j,k}$



(b) The difference of discrepancy $d_{j,k}^- := d_t^-(I_{j,k})$ terms for intervals $I_{j,k}$

Figure 1: Some terms appearing in the new potential function Ξ_t. Note that the hyperbolic cosine for the highlighted terms appears in Ξ_t.

Anti-concentration via Martingale analysis. Now we show how the random variable L can be viewed as a $(\log T)$-step martingale. Let us view a uniform point $x \in [0,1]$ as being sampled one *bit* at a time, starting with the most significant bit. At any point where j bits of x have been revealed, the interval $I_{j,k}$ on the j^{th} level of the dyadic tree is determined. Now, consider the process that starts with the value $Y_0 = \sinh(\lambda d_{0,0})$ at the root and at any time $0 \leq j \leq \log T$, the process is on some node of the j^{th} level. Conditioned on this node being $I_{j,k}$, the payoff $Y_j := a_j X_j$ where $a_j = \sinh(d_{j,k}^-)$ and X_j equals 1 if the process moves to the left child and equals -1 otherwise. Defining $L_j = Y_0 + Y_1 \ldots + Y_j$, it follows that the sequence $L_0, \ldots, L_{\log T}$ is a martingale and $L = L_{\log T}$.

Moreover, by the approximation $\cosh(x) \approx |\sinh(x)|$, we get that $Q = |Y_0| + |Y_1| + \ldots + |Y_{\log T}|$. Letting $a_0 = Y_0$ and $X_0 = 1$, the question then becomes—what is the smallest β such that the following holds:

$$\mathbb{E}\left|\sum_{i=0}^{\log T} a_i X_i\right| \;\geq\; \frac{1}{\beta} \cdot \mathbb{E}\left[\sum_{i=0}^{\log T} |a_i| X_i^2\right] \;=\; \frac{1}{\beta} \cdot \mathbb{E}\left[\sum_{i=0}^{\log T} |a_i|\right].$$

For martingales, a statement similar to Khintchine's inequality is implied by the well-known Burkholder-Davis-Gundy (BDG) inequality (see Theorem B.1 in Appendix B):

$$\mathbb{E}\left[\max_{t\leq \log T}\left|\sum_{i=0}^{t} a_i X_i\right|\right] \;\geq\; c\cdot\mathbb{E}\left[\Big(\sum_{i=0}^{\log T} a_i^2\Big)^{1/2}\right]$$

for a positive constant c. One can also prove (see Lemma B.2 in Appendix B) that

$$(1+\log T)\cdot\mathbb{E}\left|\sum_{i=0}^{\log T} a_i X_i\right| \;\geq\; \mathbb{E}\left[\max_{t\leq \log T}\left|\sum_{i=0}^{t} a_i X_i\right|\right].$$

Then, similar to the analysis for independent Rademacher random variables, using Cauchy-Schwarz,

$$(1+\log T)\cdot\mathbb{E}\left|\sum_{i=0}^{\log T} a_i X_i\right| \quad\geq\quad c\cdot\mathbb{E}\left[\Big(\sum_{i=0}^{\log T} a_i^2\Big)^{1/2}\right] \quad\geq\quad \frac{c}{\sqrt{\log T}}\cdot\mathbb{E}\left[\sum_{i=0}^{\log T}|a_i|\right].$$

So we can conclude that $\beta = \text{polylog}(T)$, which gives a $\text{polylog}(T)$ bound on interval discrepancy.

How to extend this analysis to d-dimensional Tusńady's problem? The martingale analysis above strongly relied on the interval structure of the problem, which is not clear even for the two-dimensional Tusńady's problem. To answer this question, we have to take a much more general view of our online discrepancy problem.[3]

2.4 A More General View of Changing Basis: Polynomial Discrepancy

One can also view the above analysis of the interval discrepancy problem as a more general underlying principle—that of working with a different *basis*. For example, let us take a linear algebraic approach to interval discrepancy and consider it as a vector balancing problem in $\mathbb{R}^{\mathcal{D}}$, where $\mathcal{D} = \{I_{j,k} \mid 0 \leq j \leq \log T, 0 \leq k < 2^j\}$ is the set of all dyadic intervals. When a new point $x \in [0,1]$ arrives, the coordinate $I \in \mathcal{D}$ of the update vector v_t is given by

$$v_t(I) = \mathbf{1}_I(x).$$

Note that the update v_t lives in a T-dimensional subspace $\mathcal{V}$ of the $(2T-1)$-dimensional space $\mathbb{R}^{\mathcal{D}}$ since the T-intervals, $I_{\log T,k}$, at the bottom layer determine the rest of the coordinates.

The original potential function Φ from §2.1 corresponded to working with the original basis, but with the potential function Ξ from §2.3, our approach consisted of bounding the ℓ_∞-discrepancy in a different basis of the subspace $\mathcal{V}$. In general, we may choose any basis and then define a potential function as the sum of hyperbolic-cosines of the coordinates. To choose the right basis, we need several properties from it, but most importantly we need uncorrelation.

Uncorrelation and anti-concentration via the Eigenbasis. Recall that we say random variables X, Y are *uncorrelated* if $\mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$, which is a condition only on the expected values of the random variables. Using Theorem 1.5, to show anti-concentration it suffices that the coordinates in the *new basis* are mean-zero and uncorrelated, i.e., $\mathbb{E}_v[v(i)v(j)] = 0$ for distinct coordinates i, j.

For our vector balancing results under arbitrary distributions in Theorem 1.4, we work in an *eigenbasis* of the covariance matrix. As will be shown in the proof later, standard results from linear algebra imply that the coordinates are uncorrelated in any eigenbasis. Our next lemma uses this anti-concentration (along with the hyperbolic cosine potential) to bound discrepancy in the *new basis* in terms of *sparsity*—number of non-zero coordinates—of the incoming vectors.

[3] The more general view in fact gives a (slightly) better bound for interval discrepancy than the martingale based argument above. However, we include this argument here, as it is insightful and could be useful for other problems.

Lemma 2.1. (Bounded discrepancy) *Let* p *be a distribution supported over* s*-sparse vectors in* $[-1,1]^n$ *satisfying* $\mathbb{E}_{v\sim\mathsf{p}}[v(i)v(j)] = 0$ *for all* $i \neq j \in [n]$*. Then for vectors* $v_1, \ldots, v_T$ *sampled i.i.d. from* p*, there is an online algorithm that maintains* $O(s(\log n + \log T))$ *discrepancy with high probability.*

Even though this lemma implies low discrepancy in the new basis, we need to be careful in bounding discrepancy in the original basis.

Sparsity and going back to the original basis. As discussed briefly in §1.2, although working in an eigenbasis allows us to obtain polynomial bounds for vector balancing, this is apriori not sufficient for our polylogarithmic geometric discrepancy bounds. There are two main challenges—firstly, working in a new basis might lose any sparsity that we might have in the original basis; e.g., in the one-dimensional interval discrepancy problem the arriving vectors are $(\log T)$-sparse (dyadic intervals) in the original basis, but could be $\Omega(T)$-sparse in the new basis; and secondly, even if one can find a new basis where the coordinates are uncorrelated and have low sparsity, Lemma 2.1 only implies low ℓ_∞-discrepancy in the new basis. So going back to the original basis might lose us a factor $\sqrt{n}$ more (we can only claim ℓ_2-discrepancy is the same). Recall, when we view interval discrepancy as vector balancing, $n = \Theta(T)$, so we cannot afford losing $\sqrt{n}$. Fortunately, there is a special basis consisting of *Haar wavelets* that allows us to prove polylog(T) geometric discrepancy bounds.

2.5 Haar Wavelets: Polylogarithmic Geometric Discrepancy

There is a natural orthogonal basis associated with the unit interval—the basis of Haar wavelet functions. These consist of the functions $\Psi_{j,k}$'s shown in Figure 2. Together these functions are known to form an orthogonal basis for functions on the unit interval with bounded L_2-norm.

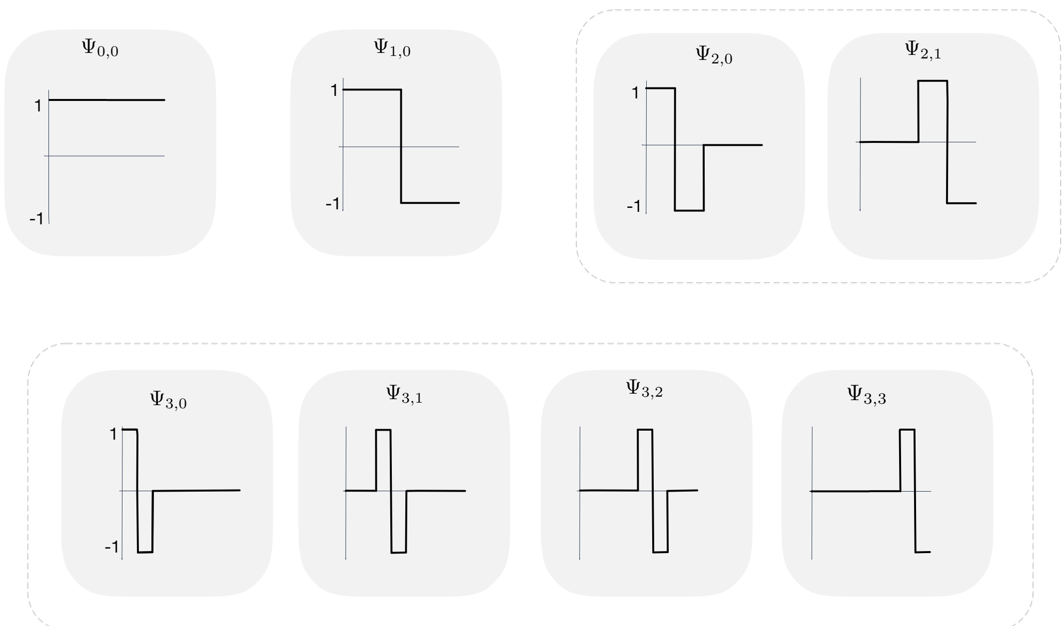


Figure 2: Haar wavelets in one dimension

Associated with the one-dimensional Haar wavelets is a natural martingale, which is the same martingale that our previous analysis in §2.3 relied on (e.g., $X_{j,k} = \Psi_{j+1,k}$ in the notation of §2.3.). It turns out that the Haar wavelets have nice orthogonality and sparsity properties that allow us to use Lemma 2.1—in particular, $\mathbb{E}_x[h(x)h'(x)] = 0$ for distinct Haar wavelet functions $h \neq h'$ and x sampled uniformly from $[0,1]$.

Moreover, moving from the basis of Haar wavelets to the original basis does not incur any additional loss in the discrepancy bound, since for any dyadic interval I, one can show that its discrepancy

$$|d_t(I)| \quad \leq \quad \alpha \|\widehat{\mathbf{1}}_I\|_1,$$

where α is a bound on the discrepancy in the Haar basis and $\|\widehat{\mathbf{1}}_I\|_1$ is the ℓ_1-norm of the function $\mathbf{1}_I$ in the Haar basis. We prove that this ℓ_1-norm is one, so $|d_t(I)| \leq \alpha$. This gives a more direct proof of the polylog(T) interval discrepancy bound and also extends easily to the d-dimensional interval discrepancy problem.

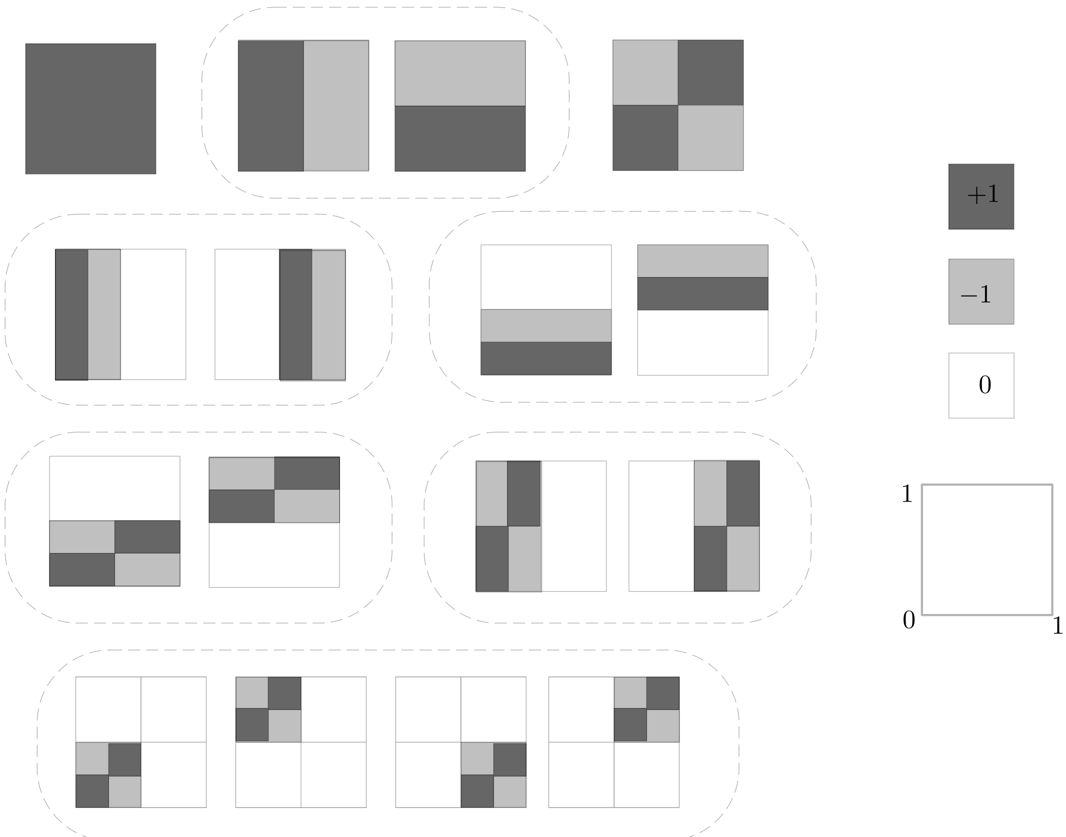


Figure 3: Haar wavelets in two dimensions

Tusńady's problem. Given the above framework of working in the Haar basis, our extension to the d-dimensional Tusńady's problem now naturally follows. For example, in two dimensions, we work with the basis of Haar wavelet functions which is formed by a taking tensor product $\Psi_{j,k} \times \Psi_{j',k'}$ of the one dimensional wavelets (see Figure 3). These functions form an orthogonal basis for all bounded *product* functions over $[0,1]^2$ and have nice sparsity properties. Moreover, we prove that for any axis-parallel box, the ℓ_1-norm of the Haar basis coefficients is one, so we do not lose any additional factor in the discrepancy bound while moving from the Haar basis to the original basis. This gives a polylogarithmic bound for two-dimensional Tusńady's problem, and also extends easily to higher dimensions.

Notation

All logarithms in this paper will be base two. For any integer k, throughout the paper $[k]$ will denote the set $\{1, \ldots, k\}$. For a vector $u \in \mathbb{R}^d$, we use $u(i)$ to denote the i^{th} coordinate of u for $i \in [d]$. Given another vector $v \in \mathbb{R}^d$, the notation $u \leq v$ denotes that $u(i) \leq v(i)$ for each $i \in [d]$. The all ones vector is denoted by $\mathbf{1}$. Given a distribution p, we use the notation $x \sim \mathsf{p}$ to denote an element x sampled from the distribution p. For a real function f, we will write $\mathbb{E}_{x\sim\mathsf{p}}[f(x)]$ to denote the expected value of $f(x)$ under x sampled from p. If the distribution is clear from the context, then we will abbreviate the above as $\mathbb{E}_x[f(x)]$.

3 Anti-Concentration Estimates

In this section we prove the anti-concentration results: we first prove it for uncorrelated random variables, and then give an improved bound for pairwise independent random variables. Although in the rest of this paper we only use the weaker bound for uncorrelated random variables, we think the improved anti-concentration for pairwise independent random variables is of independent interest and will find applications in future.

3.1 Pairwise Uncorrelated Random Variables

The following anti-concentration bound will be used in our discrepancy applications.

Theorem 1.5. (Uncorrelated anti-concentration) *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$, let $X_1, \ldots, X_n$ be uncorrelated random variables that are bounded $|X_i| \leq c$, satisfy $\mathbb{E}[X_i X_j] = 0$ for all $i \neq j$, and have sparsity s (the number of non-zero X_i's in any outcome). Then*

$$\mathbb{E}\Big|\sum_i a_i X_i\Big| \;\geq\; \mathbb{E}\Big[\sum_i |a_i| X_i^2\Big] \cdot \frac{1}{cs}. \tag{2}$$

Moreover, this bound is tight, even for pairwise independent random variables.

Note that if we have pairwise uncorrelated mean-zero random variables $X_1, \ldots, X_n$, then we get $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i] \cdot \mathbb{E}[X_j] = 0$, so the above lemma implies anti-concentration in this case. The bound in the above lemma is tight because of the Hadamard example described previously in §1.3.

We need the following main claim to prove (2) in Theorem 1.5.

Claim 3.1. *For any $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and random variables $X_1, \ldots, X_n$ satisfying $|X_i| \leq c$ and $\mathbb{E}[X_i X_j] = 0$ for distinct i, j, the following holds for any $k \in [n]$,*

$$\mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\right] \;\geq\; \frac{1}{c} \cdot \mathbb{E}[|a_k| X_k^2].$$

Proof. Using that $|X_k| \leq c$, we have

$$\begin{aligned} c \cdot \mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\right] &\geq \mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot |X_k|\right] \\ &= \mathbb{E}\left[\Big|a_k X_k^2 + \sum_{i \neq k} a_i X_i X_k\Big|\right] \;\geq\; \mathbb{E}\left[\mathsf{sign}(a_k)\Big(a_k X_k^2 + \sum_{i \neq k} a_i X_i X_k\Big)\right]. \end{aligned}$$

Since $\mathbb{E}[X_i X_k] = 0$ for $i \neq k$, it follows that

$$\begin{aligned} c \cdot \mathbb{E}\left[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\right] &\geq \mathbb{E}\left[|a_k| X_k^2\right] + \sum_{i \neq k} a_i \cdot \mathsf{sign}(a_k) \cdot \mathbb{E}\left[X_i X_k\right] \\ &= \mathbb{E}\left[|a_k| X_k^2\right]. \end{aligned}$$

∎

When combined with the following easy claim, this will prove Theorem 1.5.

Claim 3.2. *Let $Y_1, \dots, Y_n$ be correlated random variables such that for any outcome at most s of them are non-zero. Moreover, suppose there is a random variable L which satisfies*

$$\mathbb{E}\Big[|L| \cdot 1_{Y_k \neq 0}\Big] \geq \mathbb{E}\Big[|Y_k|\Big] \quad \textit{for all } k \in [n].$$

Then, $\mathbb{E}[|L|] \geq \frac{1}{s}\sum_k \mathbb{E}[|Y_k|]$.

Proof. Sum the given inequality for all $k \in [n]$ to get

$$\sum_k \mathbb{E}\Big[|Y_k|\Big] \quad \leq \quad \sum_k \mathbb{E}\Big[|L| \cdot 1_{Y_k \neq 0}\Big] \quad = \quad \mathbb{E}\Big[|L| \cdot \sum_k 1_{Y_k \neq 0}\Big] \quad \leq \quad \mathbb{E}\Big[|L| \cdot s\Big]. \qquad \square$$

Proof of Theorem 1.5. Applying Claim 3.1 and Claim 3.2 (with $L = \sum_i a_i X_i$ and $Y_i = |a_i| X_i^2$), we get that

$$\mathbb{E}\left[\Big|\sum_i a_i X_i\Big|\right] \quad \geq \quad \mathbb{E}\left[\sum_k |a_k| X_k^2\right] \cdot \frac{1}{cs}. \qquad \square$$

3.2 Pairwise Independent Random Variables

In the special case of pairwise independent random variables, it is possible to obtain an improved inequality over Theorem 1.5.

Theorem 1.6. (Pairwise independent anti-concentration) *For any $(a_1, \dots, a_n) \in \mathbb{R}^n$, let $X_1, \dots, X_n$ be mean-zero pairwise independent random variables with sparsity $s \leq n$. Then*

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big|\Big] \;\geq\; \mathbb{E}\Big[\sum_i |a_i X_i|\Big] \cdot \frac{1}{s}. \tag{3}$$

Notice, (3) immediately implies (2) for mean-zero pairwise independent random variables with $|X_i| \leq c$. One cannot hope to prove the stronger statement (3) for uncorrelated random variables due to the following example.

Example. Let $0 < \delta \ll 1$. Suppose X_1, X_2 are real random variables distributed over four outcomes:

$$(X_1, X_2) = \begin{cases} \left(\frac{1}{\delta}, \frac{1}{\delta}\right) \text{ or } \left(-\frac{1}{\delta}, -\frac{1}{\delta}\right) & \text{w.p. } \delta^2 \text{ each,} \\ (1, -1) \text{ or } (-1, 1) & \text{w.p. } \frac{1}{2} - \delta^2 \text{ each.} \end{cases}$$

Now it is easy to verify that X_1 and X_2 are mean zero, and

$$\mathbb{E}[|X_1 + X_2|] = 4\delta \quad \text{and} \quad \mathbb{E}[|X_1| + |X_2|] = 1 + 2\delta - 2\delta^2.$$

Therefore, the ratio between the two expectations can be made arbitrarily bad by making $\delta \to 0$.

Next, we prove Theorem 1.6. We start with the following claim.

Claim 3.3. *For any $(a_1, \dots, a_n) \in \mathbb{R}^n$ and mean-zero pairwise independent random variables $X_1, \dots, X_n$, the following holds for any $k \in [n]$,*

$$\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\Big] \;\geq\; \mathbb{E}[|a_k X_k|].$$

Proof. We have

$$\begin{aligned} \mathbb{E}\Big[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\Big] &= \mathbb{E}\Big[\Big|a_k X_k + \sum_{i \neq k} a_i X_i \cdot 1_{X_k \neq 0}\Big|\Big] \\ &\geq \mathbb{E}\Big[\mathsf{sign}(a_k X_k)\Big(a_k X_k + \sum_{i \neq k} a_i X_i \cdot 1_{X_k \neq 0}\Big)\Big] \\ &= \mathbb{E}\Big[|a_k X_k| + \mathsf{sign}(a_k X_k) \sum_{i \neq k} a_i X_i \cdot 1_{X_k \neq 0}\Big]. \end{aligned}$$

Since X_i and X_k are mean-zero and pairwise independent for $i \neq k$, we have $\mathbb{E}[X_i f(X_k)] = \mathbb{E}[X_i]\cdot\mathbb{E}[f(X_k)] = 0$ for any function f. Therefore,

$$\begin{aligned}\mathbb{E}\Big[\Big|\sum_i a_i X_i\Big| \cdot 1_{X_k \neq 0}\Big] &\geq \mathbb{E}[|a_k X_k|] + \sum_{i\neq k} \mathbb{E}\Big[\mathsf{sign}(a_k X_k)\cdot a_i X_i \cdot 1_{X_k\neq 0}\Big] \\ &= \mathbb{E}[|a_k X_k|]. \qquad \square\end{aligned}$$

Proof of Theorem 1.6. Combining Claim 3.3 with Claim 3.2 completes the proof of Theorem 1.6. $\square$

4 Online Discrepancy under Uncorrelated Arrivals

In this section we consider the vector balancing problem in the special case when the input distribution has uncorrelated coordinates. All our upper and lower bounds will then follow from choosing a suitable basis to reduce the original problem to a basis with uncorrelated coordinates.

4.1 Upper Bounds

We say a vector in $\mathbb{R}^d$ is *s-sparse* if it has at most s non-zero coordinates. The following lemma bounds the discrepancy for uncorrelated sparse distributions.

Lemma 2.1. (Bounded discrepancy) *Let* p *be a distribution supported over s-sparse vectors in* $[-1,1]^n$ *satisfying* $\mathbb{E}_{v\sim\mathsf{p}}[v(i)v(j)] = 0$ *for all* $i \neq j \in [n]$. *Then for vectors* $v_1, \ldots, v_T$ *sampled i.i.d. from* p, *there is an online algorithm that maintains* $O(s(\log n + \log T))$ *discrepancy with high probability.*

Proof of Lemma 2.1. Our algorithm will use the same potential function approach described in §2, and uses our anti-concentration lemma from §3 to argue that the potential always remains polynomially bounded.

Algorithm. At any time step t, let $d_t = \varepsilon_1 v_1 + \ldots + \varepsilon_t v_t$ denote the current discrepancy vector after the signs $\varepsilon_1, \ldots, \varepsilon_t \in \{\pm 1\}$ have been chosen. Set $\lambda = \frac{1}{2s}$ and define the potential function

$$\Phi_t := \sum_{i\in[n]} \cosh(\lambda d_t(i)).$$

When the vector v_t arrives, the algorithm chooses the sign ε_t that minimizes the increase $\Phi_t - \Phi_{t-1}$.

Bounded Positive Drift. Let us fix a time t. To simplify the notation, let $\Delta\Phi = \Phi_t - \Phi_{t-1}$, let $d = d_{t-1}$, and let $v = v_t$.

After choosing the sign ε_t, the discrepancy vector $d_t = d + \varepsilon_t v$. To bound the change $\Delta\Phi$, since $\cosh'(x) = \sinh(x)$ and $\sinh'(x) = \cosh(x)$, using Taylor expansion

$$\begin{aligned}\Delta\Phi &= \sum_i \Big(\lambda \sinh(\lambda d(i))\cdot(\varepsilon_t v(i)) + \frac{\lambda^2}{2!}\cosh(\lambda d(i))\cdot(\varepsilon_t v(i))^2 + \frac{\lambda^3}{3!}\sinh(\lambda d(i))\cdot(\varepsilon_t v(i))^3 + \cdots\Big), \\ &\leq \sum_i \Big(\lambda \sinh(\lambda d(i))\cdot(\varepsilon_t v(i)) + \lambda^2\cosh(\lambda d(i))\cdot(\varepsilon_t v(i))^2\Big),\end{aligned}$$

where the last inequality follows since $|\sinh(x)| \leq \cosh(x)$ for all $x \in \mathbb{R}$, and since $|\varepsilon_t v(i)| \leq 1$ and $\lambda < 1$, the higher order terms are dominated by the first and second order terms.

Set $L = \sum_i \sinh(\lambda d(i)) v(i)$, and $Q^* = \sum_i \cosh(\lambda d(i)) v(i)^2$, and $Q = \sum_i |\sinh(\lambda d(i))| v(i)^2$. Since $\cosh(x) \leq |\sinh(x)| + 1$ for $x \in \mathbb{R}$ and $|v(i)| \leq 1$, we have $Q^* \leq Q + n$. Therefore,

$$\Delta\Phi \leq \varepsilon_t \cdot \lambda \cdot L + \lambda^2 \cdot Q + \lambda^2 n.$$

Since, the algorithm chooses ε_t to minimize the increase in the potential:

$$\Delta\Phi \leq -\lambda \cdot |L| + \lambda^2 \cdot Q + \lambda^2 n.$$

Now, since $\mathbb{E}_v[v(i)v(j)] = 0$ for all $i, j \in [n]$, we can apply Theorem 1.5 with $X_i = v(i)$ and $a_i = \sinh(\lambda d(i))$ to get that $\mathbb{E}_v[|L|] \geq \frac{1}{s} \cdot \mathbb{E}[Q] = 2\lambda \cdot \mathbb{E}[Q]$, which yields that

$$\mathbb{E}_v[\Delta\Phi] \quad \leq \quad -\lambda \cdot \mathbb{E}_v[|L|] + \lambda^2 \cdot \mathbb{E}_v[Q] + \lambda^2 n \quad \leq \quad -\lambda^2 \cdot \mathbb{E}_v[Q] + \lambda^2 n \quad \leq \quad n.$$

Discrepancy Bound. The above implies that for any time $t \in [T]$, the expectation $\mathbb{E}[\Phi_t] \leq nT$. By Markov's inequality and union bound, with probability at least $1 - T^{-2}$, the potential $\Phi_t \leq nT^4$ for every time $t \in [T]$. Since at any time t, we have $\cosh(\lambda \|d_t\|_\infty) \leq \Phi_t$, this implies that with probability at least $1 - T^{-2}$, the discrepancy at every time is

$$O\left(\frac{\log(nT^4)}{\lambda}\right) = O(s(\log n + \log T)),$$

which finishes the proof of Lemma 2.1. □

4.2 Lower Bounds

We now show that the dependence on s and $\log T$ in Lemma 2.1, cannot be improved up to polynomial factors. In particular, a lower bound of $\Omega(s^{1/2})$, even when the time horizon is $T = n$, follows directly from the following more general statement for the vector balancing problem under distributions with uncorrelated coordinates. This general version will later also imply our lower bounds for geometric discrepancy.

Lemma 4.1. *Let* p *be a distribution supported over vectors in* $[-1, 1]^n$ *with* ℓ_2*-norm* k*, such that for every* $i \neq j \in [n]$ *we have* $\mathbb{E}_{v\sim\mathsf{p}}[v(i)v(j)] = 0$*. Then, for any online algorithm that receives as input vectors* $v_1, \ldots, v_n$ *sampled i.i.d. from* p*, with probability at least* $3/4$*, the discrepancy is* $\Omega(k)$ *at some time* $t \in [n]$*.*

We remark that the above lower bound may not hold if the algorithms are offline.

Proof of Lemma 4.1. Since the distribution p over inputs is fixed, we may assume that the algorithm is deterministic. Let $d_t = \varepsilon_1 v_1 + \ldots + \varepsilon_t v_t$ denote the discrepancy vector at any time $t \in [n]$. Consider the quadratic potential function:

$$\Phi_t \quad := \quad \|d_t\|_2^2 \quad = \quad \sum_{i \in [n]} |d_t(i)|^2.$$

We will need the following claim that shows Φ_t increases in expectation for any online algorithm. Let us define $\Delta\Phi_t = \Phi_t - \Phi_{t-1}$.

Claim 4.2. *Conditioned on any* $v_1, \ldots, v_{t-1}$ *and signs* $\varepsilon_1, \ldots, \varepsilon_{t-1}$ *such that* $\|d_{t-1}\|_\infty \leq \frac{k}{4}$*, we have*

$$\mathbb{E}_{v_t}[\Delta\Phi_t] \geq k^2/2 \tag{5}$$

where the expectation is taken only over the update $v_t \sim \mathsf{p}$*.*

Proof. Set $\Delta\Phi = \Delta\Phi_t$, vector $v = v_t$, and $d = d_{t-1}$. When the update v arrives, note that $d_t = d + \varepsilon_t v$. Therefore, the increase in the potential is given by

$$\Delta\Phi \quad = \quad \sum_{i=1}^{n} \Big(2d(i) \cdot \varepsilon_t v(i) + \big(\varepsilon_t v(i)\big)^2\Big) \quad = \quad 2\varepsilon_t \Big(\sum_{i=1}^{n} d(i)v(i)\Big) + \|v\|_2^2 \quad = \quad 2L + k^2, \tag{6}$$

where $L = \varepsilon_t \left(\sum_{i=1}^n d(i)v(i)\right)$.

To bound the expected value of L, we use Jensen's inequality and $\mathbb{E}_v[v(i)v(j)] = 0$ for $i \neq j$ to get:

$$(\mathbb{E}_v[L])^2 \;\leq\; \mathbb{E}_v[L^2] \;=\; \sum_{i=1}^{n} |d(i)|^2 \cdot \mathbb{E}_v[v(i)^2] + \sum_{i \neq j} d(i)d(j) \cdot \mathbb{E}_v[v(i)v(j)]$$

$$= \sum_{i=1}^{n} |d(i)|^2 \cdot \mathbb{E}_v[v(i)^2] \;\leq\; \|d\|_\infty^2 \cdot \sum_{i=1}^{n} \mathbb{E}_v[v(i)^2] \;=\; \|d\|_\infty^2 \, k^2 \;\leq\; \frac{k^4}{16}.$$

Therefore, plugging the above in (6), we get

$$\mathbb{E}_v[\Delta\Phi] \;\geq\; -2 \cdot |\mathbb{E}_v[L]| + k^2 \;\geq\; -2 \cdot \left(\frac{k^4}{16}\right)^{1/2} + k^2 \;\geq\; \frac{k^2}{2}. \qquad \square$$

To prove Lemma 4.1 using the last claim, we define τ to be the first time that $\|d_\tau\|_\infty > k/4$ if such a τ exists, or $\tau = n$ otherwise. Let us define a new potential Φ_t^* which remains the same as Φ_t for $t \leq \tau$ and increases by $k^2/2$ deterministically for every $t > \tau$.

Note that for all possible random choices,

$$\Phi_n^* \;\leq\; \Phi_{\tau-1} + \frac{nk^2}{2} \;\leq\; \frac{nk^2}{16} + \frac{nk^2}{2},$$

where the second inequality holds since $\|d_{\tau-1}\|_\infty \leq k/4$ and therefore, $\Phi_{\tau-1} \leq \frac{1}{16} \cdot nk^2$.

Moreover, let $\mathcal{E}$ be the event that $\|d_t\|_\infty \leq k/4$ for every $t \leq n$. Note that when $\mathcal{E}$ occurs then the final potential $\Phi_n^* \leq \frac{1}{16} \cdot nk^2$. Defining $p = \mathbb{P}[\mathcal{E}]$, we have

$$\mathbb{E}[\Phi_n^*] \;\leq\; p \cdot \frac{nk^2}{16} + (1-p)\left(\frac{nk^2}{16} + \frac{nk^2}{2}\right) \;=\; \frac{nk^2}{16} + (1-p)\frac{nk^2}{2}. \tag{7}$$

Moreover, from Claim 4.2 and the definition of Φ_t^*, it follows that $\mathbb{E}[\Phi_n^*] \geq \frac{1}{2} \cdot nk^2$. Comparing this with (7) yields that $p \leq 1/8$. Hence, with probability at least $7/8$, the discrepancy must be $k/4$ at some point. $\square$

Dependence on T. We next show that the discrepancy must be $\Omega((\log T/\log\log T)^{1/2})$ with high probability even when $n = O(1)$. We only sketch the proof here as the arguments are standard. The idea is that for large T, there is a high probability of getting a long enough run of consecutive vectors with each v_t almost orthogonal to d_{t-1}.

Let p be the uniform distribution over vectors on the unit sphere S^{n-1}. For any vector $u \in \mathbb{R}^n$, and v sampled from p, there is a universal constant c so that for all $\delta \leq 1$, we have $\Pr[|\langle u, v\rangle| \leq \delta\|u\|_2/n^{1/2}] \geq c\delta$.

Let $\beta \geq 1$ be some parameter that we optimize later. Setting $\delta = 1/(4\beta)$ gives that whenever $\|d_{t-1}\|_2 \leq \beta n^{1/2}$, there is at least $c/(4\beta)$ probability that $|\langle d_{t-1}, v_t\rangle| \leq 1/4$, and hence irrespective of the sign ε_t,

$$\|d_t\|_2^2 \;\geq\; \|d_{t-1}\|_2^2 - 2|\langle d_{t-1}, v_t\rangle| + \|v_t\|_2^2 \;\geq\; \|d_{t-1}\|_2^2 + 1/2.$$

So for any τ consecutive steps, with at least $(c/4\beta)^\tau$ probability, this happens at every step (or the ℓ_2-discrepancy already exceeds $\beta n^{1/2}$ at some step), and hence the discrepancy has ℓ_2-norm at least $\Omega(\tau^{1/2})$.

Partitioning the time horizon T into T/τ disjoint blocks, and setting $\beta = \log(T)$, and $\tau = \Omega(\log T/\log\log T)$, the probability such a run does not occur in any block is at most $(1 - (c/4\beta)^\tau)^{(T/\tau)} = T^{-\Omega(1)}$ by our choice of the parameters. This gives the claimed lower bound.

5 Online Vector Balancing: Polynomial Bounds

In this section, we prove our vector balancing result for arbitrary distributions.

Theorem 1.4. (Vector balancing under dependencies) *For any sequence of vectors $v_1, \ldots, v_T \in [-1,1]^n$ sampled i.i.d. from some arbitrary distribution p, there is an online algorithm which selects signs $\varepsilon_t \in \{\pm 1\}$ such that, with high probability, we have*

$$\max_{t \in [T]} \left\| \varepsilon_1 v_1 + \ldots + \varepsilon_t v_t \right\|_\infty = O(n^2(\log T + \log n)).$$

Proof of Theorem 1.4. Without loss of generality, we may assume that the distribution p is *symmetric*, i.e. both v and $-v$ have the same probability density, since we can always multiply the incoming vector v with a Rademacher ± 1 random variable without changing the problem. Let $P \in \mathbb{R}^{d \times d}$ denote the covariance matrix of our input distribution, and since p is symmetric, we get $P = \mathbb{E}_{v \sim \mathsf{p}}[vv^T]$. Let U denote the orthogonal matrix whose columns $u_1, \ldots, u_n$ form an eigenbasis for P. Note that in terms of its spectral decomposition, $P = \sum_{k=1}^n \lambda_k u_k u_k^T$ for $\lambda_k \in \mathbb{R}$.

To prove our discrepancy bound, instead of working in the original basis, we will view our problem as a vector balancing problem in the basis given by the columns of U. Now the update sequence is given by $w_1, \ldots, w_T$ where $w_t = \frac{1}{\sqrt{n}} \cdot U^T v$ is the normalized update vector in the basis U.

Since $\|v\|_2 \leq \sqrt{n}$ and orthogonal matrices preserve ℓ_2-norm, we have $\|U^T v\|_2 = \|v\|_2 \leq \sqrt{n}$. It follows that for any t, we have $\|w_t\|_\infty \leq \|w_t\|_2 = \frac{1}{\sqrt{n}} \cdot \|U^T v\|_2 \leq 1$. Furthermore, any two coordinates of the update vectors w_t's are uncorrelated, i.e., for any $i \neq j \in [n]$ we have

$$\mathbb{E}[w_t(i) \cdot w_t(j)] \;=\; \frac{1}{n}\mathbb{E}[\langle u_i, v\rangle \langle u_j, v\rangle] \;=\; \frac{1}{n}\mathbb{E}[u_i^T v v^T u_j] \;=\; \frac{1}{n} u_i^T P u_j \;=\; 0,$$

where the last equality holds since $P = \sum_{k=1}^n \lambda_k u_k^T u_k$.

Thus, we can use the online algorithm from Lemma 2.1 to select signs $\varepsilon_1, \ldots, \varepsilon_T \in \{\pm 1\}$. Let $d_t = \varepsilon_1 v_1 + \ldots + \varepsilon_t v_t$ denote the discrepancy in the original basis. Now using the trivial bound of $s \leq n$ on sparsity in Lemma 2.1, we get that with high probability,

$$\frac{1}{\sqrt{n}} \left\| U^T d_t \right\|_\infty = O(n(\log n + \log T)).$$

Again, using that orthogonal matrices preserve ℓ_2-norm,

$$\|d_t\|_\infty \;\leq\; \|d_t\|_2 \;=\; \|U^T d_t\|_2 \;\leq\; \sqrt{n} \cdot \left\|U^T d_t\right\|_\infty \;=\; O(n^2(\log n + \log T)). \qquad \square$$

6 Online Geometric Discrepancy: Polylogarithmic Bounds

In this section, we will prove our results on geometric discrepancy problems. For this, we will need a special basis of orthogonal functions on the unit interval called the Haar system. We briefly review its properties.

6.1 Haar System

Let $\Psi : \mathbb{R} \to \mathbb{R}$ denote the *mother wavelet* function

$$\Psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The *unnormalized* Haar wavelet functions (recall Figure 2) are defined as follows: let $\Psi_{0,0}(x) = 1$ for all $x \in \mathbb{R}$, and for any $j \in \mathbb{N}^*$ and $0 \leq k < 2^{j-1}$ define

$$\Psi_{j,k}(x) := \Psi(2^{j-1}x - k).$$

We call j as the *scale* and k as the *shift* of the wavelet.

The Haar wavelet functions have nice orthogonality properties. In particular, let x be drawn uniformly from the unit interval $[0,1]$. Then, one can easily check that

$$\begin{aligned}
&\mathbb{E}_x[\Psi_{j,k}(x)^2] = 2^{-(j-1)} && \text{for } j > 0, \\
&\mathbb{E}_x[\Psi_{j,k}(x)] = 0 && \text{for } j > 0, \\
&\mathbb{E}_x[\Psi_{j,k}(x)\Psi_{j',k'}(x)] = 0 && \text{unless } j = j' \text{ and } k = k'.
\end{aligned} \tag{8}$$

The Haar wavelet functions are not just orthogonal, but they form an orthogonal basis (not orthonormal), called the *Haar system*, for the class of functions on the unit interval with bounded L_2-norm. In particular, we have the following proposition where for $j \in \mathbb{Z}_{\geq 0}$ we denote $\mathcal{H}_j = \bigcup_{0 \leq k < 2^{j-1}} \{\Psi_{j,k}\}$ and let $\mathcal{H} = \bigcup_{j \geq 0} \mathcal{H}_j$.

Proposition 6.1 ([Wal04], Chapter 5). *For any $f : [0,1] \to \mathbb{R}$ such that $\mathbb{E}_x[f(x)^2] < \infty$, we have*

$$f = \sum_{h \in \mathcal{H}} \widehat{f}(h) \cdot h(x)$$

where $\widehat{f}(h) = \frac{\mathbb{E}_x[f(x)h(x)]}{\mathbb{E}_x[h(x)^2]}$ is the corresponding coefficient in the Haar system basis for $h \in \mathcal{H}$.

Indeed, since the Haar system forms an orthogonal basis, we also have that

$$\mathbb{E}_x[f(x)^2] = \sum_{h \in \mathcal{H}} \widehat{f}(h)^2 \cdot \mathbb{E}_x[h(x)^2].$$

A simple corollary of Proposition 6.1 is that $\mathcal{H}^{\otimes d}$ is an orthogonal basis for all functions over the unit cube $[0,1]^d$ that have a product structure and a bounded L_2-norm. In particular, let $\boldsymbol{h} = (h_1, \ldots, h_d)$ be an element of $\mathcal{H}^{\otimes d}$ which we will view as a function from $[0,1]^d \to \mathbb{R}$ by defining $\boldsymbol{h}(x) = \prod_{i=1}^d h_i(x(i))$ for $x \in [0,1]^d$. Note that distinct $\boldsymbol{h}$ and $\boldsymbol{h}'$ are orthogonal since for x drawn uniformly from $[0,1]^d$,

$$\mathbb{E}_x[\boldsymbol{h}(x)\boldsymbol{h}'(x)] \quad = \quad \prod_{i=1}^d \mathbb{E}_{x(i)}[h_i(x(i))h_i'(x(i))] \quad = \quad 0. \tag{9}$$

Moreover, they form a basis for product functions as given in the following proposition.

Proposition 6.2. *For any $f : [0,1]^d \to \mathbb{R}$ such that $f(x) = \prod_{i=1}^d f_i(x(i))$ for some $f_i : [0,1] \to \mathbb{R}$ satisfying $\mathbb{E}_{x(i)}[f_i(x(i))^2] < \infty$, we have that*

$$f = \sum_{\boldsymbol{h} \in \mathcal{H}^{\otimes d}} \widehat{f}(\boldsymbol{h})\boldsymbol{h},$$

where $\widehat{f}(\boldsymbol{h}) = \frac{\mathbb{E}_x[f(x)\boldsymbol{h}(x)]}{\mathbb{E}_x[\boldsymbol{h}(x)^2]}$.

Proof. Expressing each f_i in the Haar system basis using Proposition 6.1, we get the statement of the proposition by tensoring. □

Let $\mathcal{H}_{\leq j} = \bigcup_{j' \leq j} \mathcal{H}_{j'}$, and define $\mathcal{H}_{<j}, \mathcal{H}_{>j}, \mathcal{H}_{\geq j}$ analogously. Then, we have the following lemma about the Haar system decomposition of indicator functions of dyadic intervals.

Proposition 6.3. *Let $\mathbf{1}_{I_{\ell,m}}$ denote the indicator function for the interval $I_{\ell,m} = \left[m2^{-\ell}, (m+1)2^{-\ell}\right)$. Then,*

$$\sum_{h \in \mathcal{H}_0} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-\ell},$$

$$\sum_{h \in \mathcal{H}_j} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-(\ell+1-j)} \text{ for any } 1 \leq j \leq \ell \text{ and}$$

$$\widehat{\mathbf{1}}_{I_{\ell,m}}(h) = 0 \text{ for any } h \in \mathcal{H}_{>\ell}.$$

In particular, we have $\sum_{h \in \mathcal{H}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = \sum_{h \in \mathcal{H}_{\leq \ell}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 1$.

Proof. First, observe that for any $j > \ell$, either $\Psi_{j,k}(x) = 0$ identically on the interval $I_{\ell,m}$ or it takes $+1$ and -1 values on equal size sub-intervals of $I_{\ell,m}$, so that $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{j,k}(x)] = 0$.

For $\Psi_{0,0}$, notice that $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{0,0}(x)] = 2^{-\ell}$ and $\mathbb{E}_x[\Psi_{0,0}(x)^2] = 1$. Therefore, we have

$$\sum_{h \in \mathcal{H}_0} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-\ell}.$$

Now consider any $1 \le j \le \ell$. Then, there exists a unique $0 \le k^* < 2^{j-1}$ such that Ψ_{j,k^*} takes the constant value $+1$ or -1 identically on the interval $I_{\ell,m}$, and the function $\Psi_{j,k}$ is identically zero on the interval $I_{\ell,m}$ for any $k \neq k^*$. It follows that $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{j,k^*}(x)] = \pm 2^{-\ell}$, $\mathbb{E}_x[\Psi_{j,k^*}(x)^2] = 2^{-(j-1)}$ and $\mathbb{E}_x[1_{I_{\ell,m}}(x)\Psi_{j,k}(x)] = 0$ for any $k \neq k^*$. Therefore, for $1 \le j \le \ell$, we have

$$\sum_{h \in \mathcal{H}_j} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-(\ell+1-j)}.$$

From the above, it also follows that

$$\sum_{h \in \mathcal{H}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = \sum_{h \in \mathcal{H}_{\le \ell}} |\widehat{\mathbf{1}}_{I_{\ell,m}}(h)| = 2^{-\ell} + \sum_{j=1}^{\ell} 2^{-(\ell+1-j)} = 2^{-\ell} + (1 - 2^{-\ell}) = 1. \qquad \square$$

We also get a similar proposition about dyadic boxes. In particular, let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d)$ for non-negative integers ℓ_i's and let $\boldsymbol{m} = (m_1, \ldots, m_d)$ for integers $0 \le m_i < 2^{\ell_i}$. Let $\mathcal{H}^{\otimes d}_{\le \boldsymbol{\ell}} = \mathcal{H}_{\le \ell_1} \times \cdots \times \mathcal{H}_{\le \ell_d}$. Then, for the dyadic box

$$I_{\boldsymbol{\ell},\boldsymbol{m}} = I_{\ell_1,m_1} \times \cdots \times I_{\ell_d,m_d},$$

we have the following proposition. Below we write $\min\{\mathbf{e}, \mathbf{f}\}$ to denote the vector whose i^{th} coordinate is $\min\{\mathbf{e}(i), \mathbf{f}(i)\}$ for $\mathbf{e}, \mathbf{f} \in \mathbb{R}^d$.

Proposition 6.4. *Let $\mathbf{1}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}$ denote the indicator function for the dyadic box $I_{\boldsymbol{\ell},\boldsymbol{m}}$. Then,*

$$\sum_{\boldsymbol{h} \in \mathcal{H}^{\otimes d}_{\boldsymbol{j}}} |\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h})| = 2^{-\|\min\{\boldsymbol{\ell}, \boldsymbol{\ell}+\mathbf{1}-\boldsymbol{j}\}\|_1} \text{ for any } \boldsymbol{j} \le \boldsymbol{\ell} \text{ and}$$

$$\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h}) = 0 \text{ for any } \boldsymbol{h} \notin \mathcal{H}_{\le \boldsymbol{\ell}}.$$

In particular, we have $\displaystyle\sum_{\boldsymbol{h} \in \mathcal{H}^{\otimes d}} |\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h})| = \sum_{\boldsymbol{h} \in \mathcal{H}^{\otimes d}_{\le \boldsymbol{\ell}}} |\widehat{\mathbf{1}}_{I_{\boldsymbol{\ell},\boldsymbol{m}}}(\boldsymbol{h})| = 1.$

The proof of the above proposition follows from Proposition 6.3 by tensoring.

6.2 Online Interval Discrepancy Problem

Now we prove Theorem 1.2 for the d-dimensional interval discrepancy problem. Let $\mathbf{x} = (x_1, \ldots, x_T)$ be a sequence of points in $[0,1]^d$ and let $\varepsilon \in \{\pm 1\}^T$ be a signing. For any interval $I \subseteq [0,1]$ and time $t \in [T]$, recall that the discrepancy of interval I along coordinate direction i at time t is denoted

$$\mathsf{disc}^i_t(I, \mathbf{x}, \varepsilon) := \Big|\varepsilon_1 \mathbf{1}_I(x_1(i)) + \cdots + \varepsilon_t \mathbf{1}_I(x_t(i))\Big|.$$

We will just write $\mathsf{disc}^i_t(I)$ when the input sequence and signing is clear from the context.

6.2.1 Upper Bounds

To maintain the discrepancy of all intervals, it will suffice to bound the discrepancy of every dyadic interval $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ of length at least $1/T$ along every coordinate direction i. Let $\mathcal{D} = \{I_{j,k} \mid 0 \le j \le \log T, 0 \le k < 2^j\}$. Then, we prove the following.

Lemma 6.5. *Given any sequence $x_1, \ldots, x_T$ sampled independently and uniformly from $[0,1]^d$, there is an online algorithm that chooses a signing such that w.h.p. for every time $t \in [T]$, we have*

$$\max_{i \in [d]} \mathsf{disc}_t^i(I) = O(d \log^2 T) \quad \textit{for all } I \in \mathcal{D}.$$

Before proving Lemma 6.5, we first show why it implies the upper bound in Theorem 1.2.

Proof of the upper bound in Theorem 1.2. Without loss of generality, it suffices to consider half-open intervals. Every half-open interval $I \subseteq [0,1]$ can be decomposed as a union of at most $2 \log T$ disjoint dyadic intervals in $\mathcal{D}$ and two intervals $I_1 \subseteq I_{\log T, k}$ and $I_2 \subseteq I_{\log T, k'}$ for some $0 \le k, k' < T$. Note that the length of I_1 and I_2 is at most $2^{-\log T} = 1/T$. We can then write,

$$\mathsf{disc}_t^i(I) \le (2 \log T) \cdot \max_{I \in \mathcal{D}} \mathsf{disc}_t^i(I) + \mathsf{disc}_t^i(I_1) + \mathsf{disc}_t^i(I_2).$$

Applying the algorithm from Lemma 6.5, the discrepancy of every dyadic interval can be bounded w.h.p. by $O(d \log^2 T)$. The last two terms can be bounded by N_1 and N_2 respectively where N_1 (resp. N_2) is the number of points whose projections on any of the i coordinates is in I_1 (resp. I_2).

The probability that a random point z drawn uniformly from $[0,1]^d$ has some coordinate $z(i)$ for $i \in [d]$ in I_1 or I_2 is at most $2d/T$. It follows that $\mathbb{E}[N_1 + N_2] \le 2d$, so by Chernoff bounds, with probability at least $1 - T^{-4}$, the number $N_1 + N_2 \le 4d \log T$.

Overall, w.h.p. for any interval I, we have

$$\max_{i \in [d]} \mathsf{disc}_t^i(I) \quad \le \quad 2 \log T \cdot (d \log^2 T) + 4d \log T \quad = \quad O(d \log^3 T). \qquad \square$$

Next, we prove the missing Lemma 6.5.

Proof of Lemma 6.5. We will consider the d-dimensional interval discrepancy problem as a vector balancing problem in $[d] \times \mathcal{H}_{\le \log T}$ dimensions, where $\mathcal{H}_{\le \log T}$ are the Haar wavelet functions with scale parameter at most $\log T$. Note that $|\mathcal{H}_{\le \log T}| = T$, so the update vector in the vector balancing version will be Td-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\le \log T}$.

At any time when the point $x_t \in [0,1]$ arrives, then the (i,h) coordinate of the update vector $v_t \in [-1,1]^{d \times \mathcal{H}'}$ is given by

$$v_t(i,h) = h(x_t(i)).$$

Note that all the coordinates $(i, \Psi_{0,0})$ for $i \in [d]$ will always have the same value where $\Psi_{0,0}$ is constant Haar wavelet. So, to apply the online algorithm given by Lemma 2.1 we will only consider the subspace spanned by the coordinates (i,h) where $i \in [d]$ and $h \neq \Psi_{0,0}$ and the extra coordinate $(1, \Psi_{0,0})$.

Let us check first that we satisfy the conditions Lemma 2.1. First, note that the $\|v_t\|_\infty \le 1$ and the vector v_t has at most $d \log T + 1$ non-zero coordinates, since for any fixed scale $0 \le j \le \log T$ and any point $z \in [0,1]$, all but one of the values $\{h(z)\}_{h \in \mathcal{H}_j}$ are zero. The last condition to check is that the coordinates of the vector v_t are uncorrelated. This is a consequence of (8), since whenever coordinates (i,h) and (i',h') satisfy $i \neq i'$ or $h \neq h'$, we have

$$\mathbb{E}_{v_t}[v_t(i,h) \cdot v_t(i',h')] \quad = \quad \mathbb{E}_{x_t}[h(x_t(i)) \cdot h'(x_t(i'))] \quad = \quad 0.$$

To elaborate more, first note that we cannot have $h = h' = \Psi_{0,0}$ since we are working in the aforementioned subspace. Now, if $i \neq i'$ then the coordinates $x_t(i)$ and $x_t(i)$ are sampled independently from $[0,1]$, and

$\mathbb{E}_z[h(z)] = 0$ for $h \neq \Psi_{0,0}$ when z is drawn uniformly from $[0,1]$. Otherwise, for $i = i'$ but $h \neq h'$, it follows from the orthogonality of the Haar system that $\mathbb{E}_z[h(z)h'(z)] = 0$.

Next, applying the online algorithm from Lemma 2.1, we select signs $\varepsilon_1, \ldots, \varepsilon_T$ such that we get an ℓ_∞ bound on the vector $d_t = \sum_{l \leq t} \varepsilon_l v_l$. In particular, with high probability we have

$$|d_t(i,h)| \quad = \quad \Big| \sum_{l \leq t} \varepsilon_l h(x_l(i)) \Big| \quad = \quad O(d \log^2 T) \quad \text{for any } i \in [d], h \in \mathcal{H}'.$$

Note that the bound on $|d_t(i, \Psi_{0,0})|$ for $i \neq 1$ follows since $|d_t(i, \Psi_{0,0})| = |d_t(1, \Psi_{0,0})|$.

To finish the proof, we need to bound the discrepancy of every dyadic interval in terms of $\|d_t\|_\infty$. Note that for any dyadic interval $I \in \mathcal{D}$, its coefficients in the Haar system basis $\widehat{\mathbf{1}}_I(h) = 0$ for $h \in \mathcal{H}_{>\log T}$ using Proposition 6.3. Now, for any $i \in [d]$ and dyadic interval $I \in \mathcal{D}$, we can write

$$\begin{aligned}
\mathsf{disc}_t^i(I) = \Big| \sum_{l \leq t} \varepsilon_l \mathbf{1}_I(x_l(i)) \Big| \quad &= \quad \Big| \sum_{l \leq t} \varepsilon_l \sum_{h \in \mathcal{H}'} \widehat{\mathbf{1}}_I(h) h(x_l(i)) \Big| \\
= \Big| \sum_{h \in \mathcal{H}'} \widehat{\mathbf{1}}_I(h) \Big(\sum_{l \leq t} \varepsilon_l h(x_l(i)) \Big) \Big| \quad &= \quad \Big| \sum_{h \in \mathcal{H}'} \widehat{\mathbf{1}}_I(h) d_t(i,h) \Big| \\
\leq \|d_t\|_\infty \cdot \Big(\sum_{h \in \mathcal{H}'} |\widehat{\mathbf{1}}_I(h)| \Big) \quad \leq \quad \|d_t\|_\infty \quad &= \quad O(d \log^2 T),
\end{aligned}$$

where the second last inequality follows again from Proposition 6.3. □

6.2.2 Lower Bounds

Proof of the lower bound in Theorem 1.2. Set $A = T/d$. We will again consider the d-dimensional interval discrepancy problem as a vector balancing problem in $[d] \times \mathcal{H}_{\leq \log A}$ dimensions where $\mathcal{H}_{\leq \log A}$ are the Haar wavelet functions with scale parameter at most $\log A$. Note that $|\mathcal{H}_{\leq \log T}| = A$, so the update vector in the vector balancing version will be T-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\leq \log A}$.

At any time when the point $x_t \in [0,1]$ arrives, then the (i,h) coordinate of the update vector v_t is given by

$$v_t(i,h) = \begin{cases} 0 & \text{if } h = \Psi_{0,0} \\ h(x_t(i)) & \text{otherwise.} \end{cases}$$

Since for any fixed scale $0 < j \leq \log A$ and any point $z \in [0,1]$, all but one of the values $\{h(z)\}_{h \in \mathcal{H}_j}$ are zero, the vector v_t has $d \log A$ non-zero coordinates all of which take value ± 1. It follows that the Euclidean norm of any update vector v_t is $\sqrt{d \log A}$.

Furthermore, from the orthogonality of the Haar system, it follows that the coordinates of the vector v_t are uncorrelated:

$$\mathbb{E}_{v_t}[v_t(i,h) v_t(i',h')] = \mathbb{E}_{x_t}[h(x_t(i)) h'(x_t(i'))] = 0.$$

Then, applying Lemma 4.1, we get that with probability at least $3/4$, there is a $t \in [T]$ and a coordinate (i,h) with $h \neq \Psi_{0,0}$ such that $|d_t(i,h)| = \Omega(\sqrt{d \log A})$.

Let $h = \Psi_{j,k}$ for some j,k where $j > 0$. Then, by definition $h = \mathbf{1}_{I_1} - \mathbf{1}_{I_2}$ where I_1 and I_2 are the first and second halves of the interval $I_{j-1,k}$. In this case,

$$|d_t(i,h)| \quad = \quad \Big| \Big(\sum_{s \leq t} \varepsilon_t \mathbf{1}_{I_1}(x_s) \Big) - \Big(\sum_{s \leq t} \varepsilon_t \mathbf{1}_{I_2}(x_s) \Big) \Big| \quad \leq \quad 2 \max \Big\{ |\mathsf{disc}_t(I_1)|, |\mathsf{disc}_t(I_2)| \Big\}.$$

Therefore, substituting $A = T/d$, there exists an interval I such that $\mathsf{disc}_t^i(I) = \Omega\left(\sqrt{d \log\left(\frac{T}{d}\right)} \right)$. □

6.3 Online Tusnády's Problem

Let $\mathbf{x} = (x_1, \ldots, x_T)$ be a sequence of points in $[0,1]^d$ and let $\varepsilon \in \{\pm 1\}^T$ be a signing. For any axis-parallel box $B \subseteq [0,1]^d$ and any time $t \in [T]$, recall that the discrepancy of axis-parallel box B at time t is denoted

$$\mathsf{disc}_t(B, \mathbf{x}, \varepsilon) := \Big|\varepsilon(1) \cdot \mathbf{1}_B(x_1) + \ldots + \varepsilon(t) \cdot \mathbf{1}_B(x_t)\Big|.$$

We will just write $\mathsf{disc}_t(B)$ when the input sequence and signing is clear from the context.

6.3.1 Upper Bounds

As in the interval case, it will we sufficient to work with dyadic boxes. Recall that $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ for $j \in \mathbb{Z}_{\geq 0}$ and $0 \leq k < 2^j$. To maintain the discrepancy of all intervals, it will suffice to bound the discrepancy of every dyadic box

$$B_{\boldsymbol{j},\boldsymbol{k}} := I_{\boldsymbol{j}(1),\boldsymbol{k}(1)} \times \ldots \times I_{\boldsymbol{j}(d),\boldsymbol{k}(d)},$$

with $\boldsymbol{j}, \boldsymbol{k} \in \mathbb{Z}^d$ with $0 \leq \boldsymbol{j}$ and $0 \leq \boldsymbol{k} < 2^{\boldsymbol{j}}$ with each side length at least $1/T$. In particular, let $\mathcal{D} = \{B_{\boldsymbol{j},\boldsymbol{k}} \mid 0 \leq \boldsymbol{j} \leq (\log T)\mathbf{1}\ ,\ 0 \leq \boldsymbol{k} < 2^{\boldsymbol{j}}\}$ where $\mathbf{1} \in \mathbb{R}^d$ is the all ones vector. Then, we prove the following lemma to bound the discrepancy of every dyadic box.

Lemma 6.6. *Given any sequence $x_1, \ldots, x_T$ sampled independently and uniformly from $[0,1]^d$, there is an online algorithm that chooses a signing such that w.h.p. for every time $t \in [T]$,*

$$\mathsf{disc}_t(B) = O\left(\log^{d+1} T\right), \text{ for all } B \in \mathcal{D}.$$

Before proving Lemma 6.6, we first show why it implies Theorem 1.3.

Proof of the upper bound in Theorem 1.3. Without loss of generality, it suffices to consider axis-parallel boxes $B = I_1 \times \cdots \times I_d$ where I_j's are half-open sub-intervals of $[0,1]$. Recall that every half-open interval $I \subseteq [0,1]$ can be decomposed as a union of at most $2\log T$ disjoint dyadic intervals in $\mathcal{D}$ and two intervals $I' \subseteq I_{\log T,k}$ and $I'' \subseteq I_{\log T,k'}$ for some $0 \leq k, k' < T$ (note that the length of I' and I'' is at most $2^{-\log T} = 1/T$).

From this, it follows that for any axis-parallel box B, there exists a set of dyadic boxes $\mathcal{D}' \subseteq \mathcal{D}$ of size $|\mathcal{D}'| = (2\log T)^d$ and a set $\mathcal{I}$ of size $|\mathcal{I}| = 2d$ of disjoint intervals of length at most $1/T$, such that B can be decomposed as the union of boxes in $\mathcal{D}'$ and some other boxes of the form $I'_1 \times \cdots \times I'_d$, where $I'_i \in \mathcal{I}$ for at least one $i \in [d]$. We can therefore bound,

$$\mathsf{disc}_t(B) \leq (2\log T)^d \cdot \left(\max_{B \in \mathcal{D}} \mathsf{disc}_t(B)\right) + N,$$

where N is the number of points z in the input sequence such that $z(i) \in I$ for some $i \in [d]$ and $I \in \mathcal{I}$.

Applying the algorithm from Lemma 6.6, the discrepancy of every dyadic box can be bounded by $O(\log^{d+1} T)$ with high probability. Also, since the length of every interval in $\mathcal{I}$ is at most $1/T$, for z drawn uniformly from $[0,1]^d$, we have

$$\mathbb{P}_z\Big[\exists i \in [d], \exists I \in \mathcal{I} \text{ such that } z(i) \in I\Big] \leq \frac{2d^2}{T}.$$

Therefore, we have that $\mathbb{E}[N] \leq 2d^2$ and applying Chernoff bounds, it follows that with probability at least $1 - T^{-4}$, the number $N \leq 4d^2 \log T$.

Overall, with high probability for any axis-parallel box B, we have

$$\mathsf{disc}_t(B) \quad \leq \quad (2\log T)^d (\log^{d+1} T) + 4d^2 \log T \quad = \quad O_d(\log^{2d+1} T). \qquad \square$$

Next, we prove the missing Lemma 6.6.

Proof of Lemma 6.6. We will consider the d-dimensional interval discrepancy problem as a vector balancing problem in $\mathcal{H}_{\leq \log T}^{\otimes d}$ dimensions where $\mathcal{H}_{\leq \log T}$ are the Haar wavelet functions with scale parameter at most $\log T$. Note that $|\mathcal{H}_{\leq \log T}| = T$, so the update vector in the vector balancing version will be T^d-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\leq \log T}$ and also recall that for any $\boldsymbol{h} = (h_1, \ldots, h_d)$ in $\mathcal{H}'^{\otimes d}$, we view it as a function from the cube $[0,1]^d$ to $\mathbb{R}$ by defining $\boldsymbol{h}(x) = \prod_{i=1}^d h_i(x(i))$.

At any time when the point $x_t \in [0,1]^d$ arrives, then the $\boldsymbol{h} := (h_1, \ldots, h_d)$ coordinate of the update vector $v_t \in [-1,1]^{\mathcal{H}'^{\otimes d}}$ is given by

$$v_t(\boldsymbol{h}) \quad = \quad \boldsymbol{h}(x_t) \quad = \quad \prod_{i=1}^d h_i(x_t(i)).$$

We will apply the online algorithm given by Lemma 2.1. Let us check first that we satisfy the conditions of that lemma. First, note that the $\|v_t\|_\infty \leq 1$ and the vector v_t has at most $(\log T + 1)^d$ non-zero coordinates, since for any fixed scale $0 \leq j \leq \log T$ and any point $z \in [0,1]$, all but one of the values $\{h(z)\}_{h \in \mathcal{H}_j}$ are zero. The last condition to check is that the coordinates of the vector v_t are uncorrelated. This follows from the orthogonality of $\boldsymbol{h}$ and $\boldsymbol{h}'$. In particular, if $\boldsymbol{h} \neq \boldsymbol{h}'$, then

$$\mathbb{E}_{v_t}[v_t(\boldsymbol{h}) v_t(\boldsymbol{h}')] \quad = \quad \mathbb{E}_{x_t}[\boldsymbol{h}(x_t)\boldsymbol{h}'(x_t)] \quad = \quad 0.$$

Applying the online algorithm from Lemma 2.1, we select signs $\varepsilon_1, \ldots, \varepsilon_T$ such that we get an ℓ_∞ bound on the vector $d_t = \varepsilon_1 v_1 + \ldots \varepsilon_t v_t$. In particular, with high probability

$$|d_t(\boldsymbol{h})| \quad = \quad \Big| \sum_{l \leq t} \varepsilon_l \boldsymbol{h}(x_l) \Big| \quad = \quad O(\log^{d+1} T) \text{ for any } \boldsymbol{h} \in \mathcal{H}'^{\otimes d}.$$

To finish the proof, we next bound the discrepancy of every dyadic box in terms of $\|d_t\|_\infty$. For any dyadic box $B \in \mathcal{D}$, since each side consists of dyadic interval $I_{j,k}$ where $j \leq \log T$, Proposition 6.4 implies that $\widehat{\mathbf{1}}_I(\boldsymbol{h}) = 0$ for any $\boldsymbol{h} \notin \mathcal{H}'^{\otimes d}$. Therefore, we have

$$\begin{aligned}
\mathsf{disc}_t(B) = \Big| \sum_{l \leq t} \varepsilon_l \mathbf{1}_B(x_l) \Big| \quad &= \quad \Big| \sum_{l \leq t} \varepsilon_l \sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} \widehat{\mathbf{1}}_B(\boldsymbol{h}) \boldsymbol{h}(x_l) \Big| \\
= \Big| \sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} \widehat{\mathbf{1}}_B(\boldsymbol{h}) \Big(\sum_{l \leq t} \varepsilon_l \boldsymbol{h}(x_l) \Big) \Big| \quad &= \quad \Big| \sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} \widehat{\mathbf{1}}_B(\boldsymbol{h}) d_t(\boldsymbol{h}) \Big| \\
\leq \|d_t\|_\infty \cdot \Big(\sum_{\boldsymbol{h} \in \mathcal{H}'^{\otimes d}} |\widehat{\mathbf{1}}_B(\boldsymbol{h})| \Big) \quad &\leq \quad \|d_t\|_\infty \quad = \quad O(\log^{d+1} T),
\end{aligned}$$

where the second last inequality follows again from Proposition 6.4. $\square$

6.3.2 Lower Bounds

Proof of the lower bound in Theorem 1.3. Set $A = T^{1/d}$. We will consider the d-dimensional interval discrepancy problem as a vector balancing problem in $\mathcal{H}_{\leq \log A}^{\otimes d}$ dimensions where $\mathcal{H}_{\leq \log A}$ are the Haar wavelet functions with scale parameter at most $\log A$. Note that $|\mathcal{H}_{\leq \log A}| = A$, so the update vector in the vector balancing version will be A^d-dimensional. Let us abbreviate $\mathcal{H}' = \mathcal{H}_{\leq \log A}$ and also recall that for any $\boldsymbol{h} = (h_1, \ldots, h_d)$ in $\mathcal{H}'^{\otimes d}$, we view it as a function from the cube $[0,1]^d$ to $\mathbb{R}$ by defining $\boldsymbol{h}(x) = \prod_{i=1}^d h_i(x(i))$.

At any time when the point $x_t \in [0,1]^d$ arrives, then the $\boldsymbol{h} := (h_1, \ldots, h_d)$ coordinate of the update vector $v_t \in [-1,1]^{\mathcal{H}'^{\otimes d}}$ is given by

$$v_t(\boldsymbol{h}) \quad = \quad \boldsymbol{h}(x_t) \quad = \quad \prod_{i=1}^d h_i(x_t(i)).$$

We will apply Lemma 4.1. Let us check first that we satisfy the conditions of that lemma. Similar to the proof of Lemma 6.6, we note that the vector v_t has exactly $(\log A + 1)^d$ non-zero coordinates that take the value ± 1. This implies that that Euclidean norm of any update v_t is $(\log A + 1)^{d/2}$. Also from the orthogonality of $\boldsymbol{h}$ and $\boldsymbol{h}'$, the coordinates of the vector v_t are uncorrelated — if $\boldsymbol{h} \neq \boldsymbol{h}'$, then

$$\mathbb{E}_{v_t}[v_t(\boldsymbol{h})v_t(\boldsymbol{h}')] \quad = \quad \mathbb{E}_{x_t}[\boldsymbol{h}(x_t)\boldsymbol{h}'(x_t)] \quad = \quad 0.$$

Applying Lemma 4.1 tells us that with probability at least $3/4$, there exists a time $t \in [T]$ and a $\boldsymbol{h} \in \mathcal{H}'$ such that $|d_t(\boldsymbol{h})| = \Omega\left((\log A + 1)^{d/2}\right)$. Note that since $\boldsymbol{h}(x) = \prod_{i=1}^{d} h_i(x(i))$ and h_i can always be expressed as $\mathbf{1}_{I_i}$ or $\mathbf{1}_{I_i} - \mathbf{1}_{I_i'}$ for some intervals I_i and I_i', it follows that there exists a set $\mathcal{B}$ of at most 2^d axis-parallel boxes and some $\varepsilon \in \{\pm 1\}^{\mathcal{B}}$ such that

$$d_t(\boldsymbol{h}) = \sum_{B \in \mathcal{B}} \varepsilon_B \cdot \mathsf{disc}_t(B).$$

By averaging, it follows that there is an axis-parallel box $B \in \mathcal{B}$ such that $\mathsf{disc}_t(B) \geq \dfrac{|d_t(\boldsymbol{h})|}{2^d}$.

Substituting $A = T^{1/d}$, we get that for some box B,

$$\mathsf{disc}_t(B) \quad = \quad \Omega\Big(\frac{1}{2^d} \cdot \log^{d/2} A\Big) \quad = \quad \Omega_d(\log^{d/2} T). \qquad \square$$

7 Applications to Online Envy Minimization

In this section we use our vector balancing and two-dimensional interval discrepancy results to bound online envy. Let us first give the formal definition of envy.

Recall that there are two players and T items where for item $t \in \{1, \ldots, T\}$, the valuation of the player $i \in \{1, 2\}$ is $v_{it} \in [0, 1]$. The *cardinal envy* is the standard notion of envy studied in the fair division literature [LMMS04, Bud11]. If Player i is allocated set S_i by an algorithm, the cardinal envy is defined as

$$\mathsf{envy}_C(\mathbf{v}_1, \mathbf{v}_2, S_1, S_2) := \max\Big\{ \sum_{t \in S_2} v_{1t} - \sum_{t \in S_1} v_{1t} \ , \ \sum_{t \in S_1} v_{2t} - \sum_{t \in S_2} v_{2t} \Big\}.$$

The notion of ordinal envy is defined ignoring the precise item valuations, but only with respect to the relative ordering of the items. Thus, the ordinal envy is always at least the cardinal envy [JKS19]. For $i \in \{1, 2\}$, let π_i denote the decreasing order with respect to the valuations v_{it}. Denote π_i^t the first t items in the order π. If Player i is allocated set S_i by an algorithm, the ordinal envy is defined as

$$\mathsf{envy}_O(\pi_1, \pi_2, S_1, S_2) := \max_{t \geq 0} \Big\{ |S_2 \cap \pi_1^t| - |S_1 \cap \pi_1^t| \ , \ |S_1 \cap \pi_2^t| - |S_2 \cap \pi_2^t| \Big\}.$$

Next, we prove Corollary 1.7, which is restated below.

Corollary 1.7. *Suppose valuations of two players are drawn i.i.d. from some distribution* p *over* $[0, 1] \times [0, 1]$*. Then, for an arbitrary distribution* p *(i.e., player valuations for the same item could be correlated), the online cardinal envy is* $O(\log T)$*. Moreover, if* p *is a product distribution (i.e., player valuations for the same item are independent) then the online ordinal envy is also* $O(\log^3 T)$*.*

Proof. When the player valuations are drawn independently in $[0, 1]$, the "moreover" part is immediate from the following lemma of [JKS19] along with our Theorem 1.2 for 2-dimensional interval discrepancy.

Lemma 7.1 (Lemma 26 in [JKS19])**.** *For two players with independent valuations, any upper bound for 2-dimensional interval discrepancy problem also holds for 2-player online ordinal envy minimization.*

Next, we bound online cardinal envy under arbitrary distributions. In the following lemma we reduce this problem to 2-dimensional vector balancing.

Lemma 7.2. *For two players taking values from an arbitrary distribution* p *over* $[0,1]\times[0,1]$*, any upper bound for* 2*-dimensional vector balancing problem also holds for* 2*-player online cardinal envy minimization.*

Proof. For $i \in \{1,2\}$, let u_{it} denote the valuation of Player i for t^{th} item. We define the corresponding vector $v_t = (u_{1t}, -u_{2t})$. If our online vector balancing algorithm assigns the next vector v_t a $+$ sign, we give the item to Player 2, and otherwise we give it to Player 1. The crucial observation is that $d_t(1)$ and $d_t(2)$ capture precisely the cardinal envy of Players 1 and 2, respectively. Thus, any bound $\|d_t\|_\infty$ implies a bound on the maximum cardinal envy. ☐

The last lemma when combined with Theorem 1.4 finishes the proof of Corollary 1.7. ☐

8 Open Problems and Directions

We close this paper by mentioning some interesting open problems that seem to require fundamental new techniques, and new directions in online discrepancy that remain unexplored.

Improving the dependence on n for general distributions. Theorem 1.4 gives a bound of $O(n^2 \log T)$ for online vector balancing problem under inputs sampled from an arbitrary distribution. However, an optimal dependence of $O(n^{1/2})$ on n is achievable in the special case where the distribution has independent coordinates [BS19], and also in the offline setting with worst-case inputs [Ban12]. This motivates the following question.

Question 1. *Given an arbitrary distribution* p *supported over* $[-1,1]^n$*, is there an online algorithm that maintains discrepancy* $\sqrt{n}\cdot\mathrm{polylog}(T)$ *on a sequence of* T *inputs sampled i.i.d. from* p*?*

As the anti-concentration bound in Theorem 1.5 for uncorrelated variables is a $n^{1/2}$ factor worse than that for independent random variables, even getting a dependence of $n \cdot \mathrm{polylog}(T)$ is an interesting first step.

Bounds in terms of sparsity. Several natural problems such as the d-dimensional interval discrepancy and d-dimensional Tusnády's problem are best viewed as vector balancing problems where the input vectors are sparse. This motivates the following online version of the Beck-Fiala problem, where the online sequence $x_1, \ldots, x_T$ is chosen independently from some distribution p supported over s-sparse n-dimensional vectors over $[-1,1]^n$. In the offline setting with worst-case inputs (and where we care about the discrepancy of every prefix), the methods of Banaszczyk [Ban12] give a bound of $(s\log T)^{1/2}$.

Question 2. *Given an arbitrary distribution* p *supported over* s*-sparse vectors in* $[-1,1]^n$*, is there an online algorithm that maintains discrepancy* $\mathrm{poly}(s, \log T, \log n)$ *on a sequence of* T *inputs sampled i.i.d. from* p*?*

Resolving the above question would imply polylogarithmic bounds for Tusnády's problem in d-dimensions (similar to that in Theorem 1.3) in the much more general setting where the points x_T are sampled from an arbitrary distribution over points in $[0,1]^d$. Currently, Theorems 1.2 and 1.3 only hold when the points x_t are sampled from a product distribution on $[0,1]^d$.

Oblivious adversary model. A very interesting direction that is strictly harder than the stochastic setting is to understand online vector balancing in a setting where the adversary is oblivious or non-adaptive, i.e., the adversary chooses the entire input sequence (without any stochastic assumptions) beforehand and is *not allowed* to change the inputs later based on the execution of the algorithm.

Recall that if the adversary is fully adaptive, then one cannot hope to prove a bound better than $\Theta(T^{1/2})$, but this might be possible for oblivious adversaries.

Question 3. *Is there an online algorithm that maintains discrepancy* $\mathrm{poly}(n, \log T)$ *on any sequence of* T *vectors in* $[-1,1]^n$ *chosen by an oblivious adversary?*

One could also consider the same question in the Beck-Fiala setting, and ask if better bounds are possible when there is sparsity.

Question 4. *Is there an online algorithm that maintains discrepancy* $\mathrm{poly}(s, \log T, \log n)$ *on any sequence of* T *vectors in* $[-1,1]^n$ *that are* s*-sparse and chosen by an oblivious adversary?*

Resolving Questions 3 and 4 would also have implications for both online geometric discrepancy and online envy minimization problems in the oblivious adversary setting.

Acknowledgements

We are thankful to Janardhan Kulkarni for several discussions on this project.

A Tight example for Anti-Concentration in the Original Basis for Interval Discrepancy

Let us briefly recall the setting. Consider the complete binary tree of height $\log T$ where the nodes are the dyadic intervals $I_{j,k}$ for $0 \le j \le \log T$ and $0 \le k < 2^j$. Our objective was to find the smallest β such that

$$\mathbb{E}_x\Big[\Big|\sum_{j,k}\sinh(\lambda d_{j,k})\cdot \mathbf{1}_{I_{j,k}}(x)\Big|\Big] \;\ge\; \frac{1}{\beta}\cdot \mathbb{E}_x\Big[\sum_{j,k}\cosh(\lambda d_{j,k})\cdot \mathbf{1}_{I_{j,k}}(x)\Big], \tag{10}$$

where x is a uniform point on the unit interval $[0,1]$, the function $\mathbf{1}_{I_{j,k}}$ is the indicator for the dyadic interval $I_{j,k}$, and $\lambda > 0$ and $d_{j,k} \in \mathbb{R}$. For simplicity, we set $\lambda = 1$ henceforth.

Observe that when a uniform random point x arrives at a leaf dyadic interval $I_{\log T,k}$, then only the variables along that root-leaf path contribute to both sides. Moreover, since x is uniform, the chosen leaf interval is also uniform among the leaves. Therefore, denoting by ℓ the random leaf and P_ℓ the corresponding root-leaf path, we want to ask for the smallest β satisfying

$$\mathbb{E}_{P_\ell}\Big[\Big|\sum_{I_{j,k}\in P_\ell} a_{j,k}\Big|\Big] \;\ge\; \frac{1}{\beta}\cdot \mathbb{E}_{P_\ell}\Big[\sum_{I_{j,k}\in P_\ell} |a_{j,k}|\Big], \tag{11}$$

where $a_{j,k} = \sinh(d_{j,k})$ for a node $I_{j,k}$ in the dyadic tree. Note that to get (11) from (10), we made the standard approximation that $\cosh(x) \approx |\sinh(x)|$ for $x \in \mathbb{R}$.

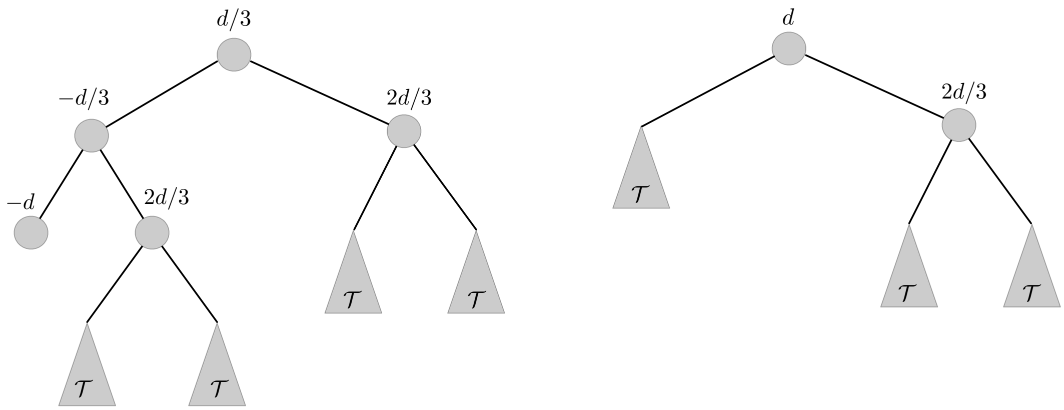
The following lemma shows that in general β could be exponentially large in the height of the tree, so in the above case since the height is $\log T$, the value of $\beta = \Omega(\mathrm{poly}(T))$. We remark that for non-binary trees, this was already shown by Jiang, Kulkarni, and Singla [JKS19].

Lemma A.1. *There exists* $d_{j,k}$ *for* $0 \le j \le h$ *and* $0 \le k < 2^j$*, such that* $\beta = \exp(\Omega(h))$ *in* (11).

Proof. Our construction has a fractal structure. Let $d > 0$ be a sufficiently large integer. Let $\mathcal{T}$ denote the tree structure shown in Figure 4(a) where the labels are the values that will be used for constructing $d_{j,k}$'s. We embed this structure in the complete binary tree of dyadic intervals and assign the $d_{j,k}$ values as follows: the root interval has value $d_{0,0} = d$ and its left children has the structure $\mathcal{T}$ with the values $d_{j,k}$ as assigned by the corresponding labels in $\mathcal{T}$, while the right child has value $d_{1,1} = 2d/3$ and has two child subtrees with structure $\mathcal{T}$ (see Figure 4(b)). The $d_{j,k}$ values for all the unassigned nodes (these lie in the subtree rooted at the nodes having values $d_{j,k} = -d$) are taken to be zero.

Note that $\mathcal{T}$ has the property that with probability $1/4$ it ends in a node $I_{j,k}$ with $a_{j,k} = \sinh(-d)$, and otherwise it enters another $\mathcal{T}$ (unless we already reached a leaf).

The proof now follows because if we take a random root-leaf path in our dyadic tree, with probability $1 - \exp(-\Omega(h))$ it will end in a leaf with $\sinh(-d)$, which will cancel with $\sinh(d)$ at the root. Since every



(a) Fractal structure $\mathcal{T}$ (b) Embedding of $\mathcal{T}$ in the dyadic tree

Figure 4: Construction of $d_{j,k}$'s satisfying (11)

other entry on a root leaf path has magnitude at most $\sinh(2d/3)$, the left hand side in (11) will be

$$\mathbb{E}_{P_\ell}\Big[\Big|\sum_{I_{j,k}\in P_\ell} a_{j,k}\Big|\Big] \;\leq\; \Big(1-\exp(-\Omega(h))\Big)\cdot h\cdot|\sinh(2d/3)| + \exp(-\Omega(h))\cdot h\cdot|\sinh(d)| \;\leq\; \frac{|\sinh(d)|}{\exp(\Omega(h))},$$

while the right hand side is

$$\mathbb{E}_{P_\ell}\Big[\sum_{I_{j,k}\in P_\ell} |a_{j,k}|\Big] \;\geq\; |\sinh(d)|.$$

Therefore, $\beta = \exp(\Omega(h))$ in (11). □

B Burkholder-Davis-Gundy Inequality

Let $Z_0, Z_1, \ldots, Z_t$ be a discrete martingale (with respect to $W_1, \ldots, W_t$) and let $\Delta Z_s = Z_s - Z_{s-1}$ denote the differences for all $s \in [t]$. Note that $Z_s = \Delta Z_1 + \Delta Z_2 + \ldots + \Delta Z_s$. Define $Z_t^* = \max_{0\leq s\leq t}|Z_s|$ to be the maximum value of the martingale process till time t. Then, the well-known Burkholder-Davis-Gundy inequality says the following.

Theorem B.1 ([BDG72])**.** *Let $1 \leq p < \infty$. Then, there exist positive constants c_p and C_p such that*

$$c_p\cdot\mathbb{E}\left[\left(\sum_{s=1}^{t}|\Delta Z_s|^2\right)^{p/2}\right] \;\leq\; \mathbb{E}[(Z_t^*)^p] \;\leq\; C_p\cdot\mathbb{E}\left[\left(\sum_{s=1}^{t}|\Delta Z_s|^2\right)^{p/2}\right].$$

Note that the inequality holds in much more general settings, but the above setting is sufficient for the purposes of this paper.

Furthermore, for $p = 1$, which is the case we need for the purposes of this paper, one can relate expected magnitude of Z_t^* and Z_t by the following inequality.

Lemma B.2. $\mathbb{E}[Z_t^*] \leq (t+1)\cdot\mathbb{E}[|Z_t|]$.

Proof. First note that $f(Z_0), \ldots, f(Z_t)$ is a sub-martingale with respect to $W_1, \ldots, W_t$ for any convex function f. Choosing $f(z) = |z|$, it follows that the absolute value of the above martingale is a sub-martingale.

Applying Doob's optional stopping theorem to this sub-martingale, one gets that $\mathbb{E}[|Z_t|] \geq \mathbb{E}[|Z_0|]$. Since, we could have started this sequence anywhere, it also follows for any $s < t$ that $\mathbb{E}[|Z_t|] \geq \mathbb{E}[|Z_s|]$.

Since $Z_t^* = \max_{s \leq t} |Z_s| \leq \sum_{s=0}^{t} |Z_s|$, using linearity of expectation, we get that

$$\mathbb{E}[Z_t^*] \quad \leq \quad \sum_{s=0}^{t} \mathbb{E}[|Z_s|] \quad \leq \quad (t+1)\mathbb{E}[|Z_t|]. \qquad \square$$

References

[Ban98] Wojciech Banaszczyk. Balancing vectors and Gaussian measures of n-dimensional convex bodies. *Random Struct. Algorithms*, 12(4):351–360, 1998.

[Ban10] Nikhil Bansal. Constructive Algorithms for Discrepancy Minimization. In *Proceedings of FOCS 2010*, pages 3–10, 2010.

[Ban12] Wojciech Banaszczyk. On series of signed vectors and their rearrangements. *Random Struct. Algorithms*, 40(3):301–316, 2012.

[Bár79] Imre Bárány. On a Class of Balancing Games. *J. Comb. Theory, Ser. A*, 26(2):115–126, 1979.

[Bár08] Imre Bárány. On the Power of Linear Dependencies. In *Building Bridges*, pages 31–45. Springer, 2008.

[BDG72] D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of BSMSP*, volume 2, pages 223–240, 1972.

[BDG16] Nikhil Bansal, Daniel Dadush, and Shashwat Garg. An Algorithm for Komlós Conjecture Matching Banaszczyk's Bound. In *Proceedings of FOCS 2016*, pages 788–799, 2016.

[BDGL18] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The Gram-Schmidt Walk: A Cure for the Banaszczyk Blues. In *Proceedings of STOC 2018*, pages 587–597, 2018.

[BG81] Imre Bárány and Victor S Grinberg. On some combinatorial questions in finite-dimensional spaces. *Linear Algebra and its Applications*, 41:1–9, 1981.

[BG17] Nikhil Bansal and Shashwat Garg. Algorithmic discrepancy beyond partial coloring. In *Proceedings of STOC 2017*, pages 914–926, 2017.

[BKPP18] Gerdus Benade, Aleksandr M. Kazachkov, Ariel D. Procaccia, and Christos-Alexandros Psomas. How to Make Envy Vanish Over Time. In *Proceedings of EC 2018*, pages 593–610, 2018.

[BS19] Nikhil Bansal and Joel H. Spencer. On-Line Balancing of Random Inputs. *CoRR*, abs/1903.06898, 2019.

[Bud11] Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *J. Polit. Econ.*, 119(6):1061–1103, 2011.

[Cho94] Sergej Chobanyan. Convergence as of rearranged random series in Banach space and associated inequalities. In *Probability in Banach Spaces, 9*, pages 3–29. Springer, 1994.

[DNTT18] Daniel Dadush, Aleksandar Nikolov, Kunal Talwar, and Nicole Tomczak-Jaegermann. Balancing Vectors in Any Norm. In *Proceedings of FOCS 2018*, pages 1–10, 2018.

[ES18] Ronen Eldan and Mohit Singh. Efficient algorithms for discrepancy minimization in convex sets. *Random Struct. Algorithms*, 53(2):289–307, 2018.

[Fol67] Duncan K Foley. Resource allocation and the public sector. *Yale Econ Essays*, 7:45–98, 1967.

[Fra18] Cole Franks. A simplified disproof of Beck's three permutations conjecture and an application to root-mean-squared discrepancy. *arXiv:1811.01102*, 2018.

[Gia97] Apostolos A Giannopoulos. On some vector balancing problems. *Studia Mathematica*, 122(3):225–234, 1997.

[JKS19] Haotian Jiang, Janardhan Kulkarni, and Sahil Singla. Online Geometric Discrepancy for Stochastic Arrivals with Applications to Envy Minimization. *arXiv:1910.01073*, 2019.

[LM15] Shachar Lovett and Raghu Meka. Constructive Discrepancy Minimization by Walking on the Edges. *SIAM J. Comput.*, 44(5):1573–1582, 2015.

[LMMS04] Richard J. Lipton, Evangelos Markakis, Elchanan Mossel, and Amin Saberi. On approximately fair allocations of indivisible goods. In *Proceedings of EC 2004*, pages 125–131, 2004.

[LRR17] Avi Levy, Harishchandra Ramadas, and Thomas Rothvoss. Deterministic Discrepancy Minimization via the Multiplicative Weight Update Method. In *Proceedings of IPCO 2017*, pages 380–391, 2017.

[Mat09] Jiri Matousek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 2009.

[MN15] Jiří Matoušek and Aleksandar Nikolov. Combinatorial discrepancy for boxes via the γ_2 norm. In *Proceedings of SoCG 2015*, pages 1–15, 2015.

[MNT14] Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy. *CoRR*, abs/1408.1376, 2014.

[Nik17] Aleksandar Nikolov. Tighter bounds for the discrepancy of boxes and polytopes. *CoRR*, abs/1701.05532, 2017.

[NNN12] Alantha Newman, Ofer Neiman, and Aleksandar Nikolov. Beck's three permutations conjecture: A counterexample and some consequences. In *Proceedings of FOCS 2012*, pages 253–262, 2012.

[NV13] Hoi H. Nguyen and Van H. Vu. *Small Ball Probability, Inverse Theorems, and Applications*, pages 409–463. Springer Berlin Heidelberg, 2013.

[Rot14] Thomas Rothvoß. Constructive Discrepancy Minimization for Convex Sets. In *Proceedings of FOCS 2014*, pages 140–145, 2014.

[Spe77] Joel Spencer. Balancing games. *J. Comb. Theory, Ser. B*, 23(1):68–74, 1977.

[Spe85] Joel Spencer. Six standard deviations suffice. *Trans. Am. Math. Soc.*, 289(2):679–706, 1985.

[Spe87] Joel H. Spencer. *Ten lectures on the probabilistic method*, volume 52. Society for Industrial and Applied Mathematics Philadelphia, 1987.

[SST97] Joel H. Spencer, Aravind Srinivasan, and Prasad Tetali. The discrepancy of permutation families. In *Proceedings of SODA*, 1997.

[TV85] William Thomson and Hal Varian. Theories of justice based on symmetry. *Social goals and social organizations: essays in memory of Elisha Pazner*, 126, 1985.

[Wal04] David Walnut. *An Introduction to Wavelet Analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser Basel, 1 edition, 1 2004.