

Colorful vector balancing

Gergely Ambrus^{1,2}  | Rainie Bozzai^{1,3}

¹Department of Geometry, Bolyai Institute, University of Szeged, Szeged, Hungary

²Alfréd Rényi Institute of Mathematics, Budapest, Hungary

³Department of Mathematics, University of Washington, Seattle, USA

Correspondence

Gergely Ambrus, Bolyai Institute, Aradi vertanuk tere 1, 6720 Szeged, Hungary.
Email: ambrus@renyi.hu

Funding information

ERC Advanced Grant, Grant/Award Number: Geoscape; Hungarian Scientific Research Fund, Grant/Award Number: NKFIH KKP-133819; National Science Foundation, Grant/Award Numbers: DGE-2140004, G20221001-4371; National Research, Development and Innovation Fund, Ministry of Innovation and Technology of Hungary, Grant/Award Number: TKP2021-NVA-09

Abstract

We extend classical estimates for the vector balancing constant of \mathbb{R}^d equipped with the Euclidean and the maximum norms proved in the 1980s by showing that for $p = 2$ and $p = \infty$, given vector families $V_1, \dots, V_n \subset B_p^d$ with $0 \in \sum_{i=1}^n \text{conv } V_i$, one may select vectors $v_i \in V_i$ with

$$\|v_1 + \dots + v_n\|_2 \leq \sqrt{d}$$

for $p = 2$, and

$$\|v_1 + \dots + v_n\|_\infty \leq O(\sqrt{d})$$

for $p = \infty$. These bounds are sharp and asymptotically sharp, respectively, for $n \geq d$. The proofs combine linear algebraic and probabilistic methods with a Gaussian random walk argument.

MSC 2020

52A40, 52C07, 05A99 (primary)

1 | HISTORY AND RESULTS

Vector balancing problems have been studied for over six decades. The prototype question was asked by Dvoretzky [15] in 1963:

Question 1.1 (Vector balancing in ℓ_p). Let $1 \leq p$ and $1 \leq d \leq n$. Find

$$\max \min \|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n\|_p,$$

© 2024 The Author(s). *Mathematika* is copyright © University College London and published by the London Mathematical Society on behalf of University College London. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

where the maximum is taken over $v_1, \dots, v_n \in B_p^d$, and the minimum is taken over sign sequences $\varepsilon_i \in \{\pm 1\}$ for $i = 1, \dots, n$.

In other words, given a collection of vectors $\{v_1, \dots, v_n\} \subset B_p^d$, the goal is to find a signed sum of them with small norm. (As usual, B_p^d stands for the unit ball of the ℓ_p -norm on \mathbb{R}^d .) The term “vector balancing” is readily motivated by the following interpretation: placing the vectors into the two plates of a scale according to their associated signs, the problem asks for achieving a nearly equal balance, that is, forcing the sum of the vectors in the plates to be as close as possible.

Even though the above formulation uses two parameters: d and n , surprisingly, the optimal bounds turn out to depend only on the dimension d [11, 28, 34, 35]. This was the first intriguing characteristic of the problem, with many more to follow in the subsequent decades.

In order to facilitate the forthcoming discussion, we introduce the notion of *vector balancing constants* by defining

$$\text{vb}(K, L, n) = \max_{v_1, \dots, v_n \in K} \min_{\varepsilon_1, \dots, \varepsilon_n \in \{\pm 1\}} \|\varepsilon_1 v_1 + \dots + \varepsilon_n v_n\|_L \quad (1)$$

for origin-symmetric convex bodies (i.e., compact, convex sets with nonempty interior) $K, L \subset \mathbb{R}^d$. Above, $\|\cdot\|_L$ denotes the Minkowski norm associated to L , which is defined by

$$\|x\|_L = \min\{r \geq 0 : x \in rL\}$$

for $x \in \mathbb{R}^d$ (note that L is the unit ball of $\|\cdot\|_L$). Then the vector balancing constant of K and L is given by

$$\text{vb}(K) = \sup_{n \geq d} \text{vb}(K, n).$$

When $K = L$, we simply write

$$\text{vb}(K, n) = \text{vb}(K, K, n)$$

and

$$\text{vb}(K) = \sup_{n \geq d} \text{vb}(K, n).$$

Thus, using the above notation, Question 1.1 asks to determine $\text{vb}(B_p^d, n)$ for $p \geq 1$ and $n \geq d$.

Sharp estimates for Dvoretzky’s question in the Euclidean norm were proven in the late 1970s by Sevast’yanov [30], independently by Bárány (unpublished at the time, for the proof, see [9]), and also, perhaps, by Grinberg. The proofs of Sevast’yanov and Bárány are of linear algebraic flavor. By the probabilistic method, Spencer [34] proved the same result, which reads as

$$\text{vb}(B_2^d, d) = \text{vb}(B_2^d) = \sqrt{d}. \quad (2)$$

In both approaches, the proof boils down to showing that any point of a parallelotope in \mathbb{R}^d may be approximated by a vertex with Euclidean error at most \sqrt{d} ; this is the direct predecessor of our Proposition 3.1.

The case of the ℓ_∞ -norm was solved by Spencer [35] by showing

$$\text{vb}(B_\infty^d, d) \leq C\sqrt{d} \tag{3}$$

and

$$\text{vb}(B_\infty^d) \leq 2C\sqrt{d} \tag{4}$$

for a universal constant $C < 6$. These estimates are asymptotically sharp, as demonstrated by vector systems consisting of vertices of B_∞^d . The weaker bound of $O(\sqrt{d \ln d})$ can be shown by applying the probabilistic method, but removing the $\sqrt{\ln d}$ factor is not possible using that approach. The upper bound (3) was also shown, independently, by Gluskin [18] who applied Minkowski’s theorem on lattice points and an argument of Kashin [22]. Indeed, these results rely on the parallelotope approximation in the maximum norm, which precedes our Proposition 4.1. These vertex approximation results are close relatives of the Beck-Fiala “integer-making” theorems [12].

Extending (2) and (4), Reis and Rothvoss [28] proved that there exists a universal constant C' for which $\text{vb}(B_p^d) \leq C'\sqrt{d}$ holds for each $2 \leq p \leq \infty$. This bound, up to constant factor, also matches the lower bound of Banaszczyk [3] for general norms.

Our article is mainly motivated by the work of Bárány and Grinberg [11], who generalized Question 1.1 to vector families which may also be interpreted as color classes. They proved the following general result.

Theorem 1.2 (Bárány, Grinberg [11]). *Assume that $B^d \subset \mathbb{R}^d$ is an origin-symmetric convex body, and $V_1, \dots, V_n \subseteq B^d$ are vector families so that $0 \in \sum_{i \in [n]} \text{conv } V_i$. Then there exists a selection of vectors $v_i \in V_i$ for $i \in [n]$ such that*

$$\left\| \sum_{i \in [n]} v_i \right\|_{B^d} \leq d. \tag{5}$$

The original vector balancing problem is retrieved by setting $V_i = \{\pm v_i\}$ for $i \in [n]$ (where, as usual, $[n] = \{1, \dots, n\}$).

Taking $B^d = B_1^d$, $n = d$, and $V_i = \{\pm e_i\}$ for $i \in [n]$ shows that Theorem 1.2 is sharp. Yet, for specific norms, asymptotically stronger estimates may hold. In light of (2) and (4), it is plausible to conjecture that for the Euclidean and the maximum norms, the sharp estimate is of order $O(\sqrt{d})$. For the case of the Euclidean norm, it is mentioned in [11] that V. V. Grinberg proved the sharp bound of \sqrt{d} , although this has never been published (or verified) — and 25 years later, the statement was again referred to as a conjecture [10]. Bárány and Grinberg [11] also note that “from the point of view of applications, it would be interesting to know more about” the case of the ℓ_∞ -norm. Based on a generalization of the Gram–Schmidt walk, Bansal et al. [7] recently proved the following theorem in the colorful setting.

Theorem 1.3 (Bansal, Dadush, Garg, Lovett [7]). *Let $V_1, \dots, V_n \subseteq B_2^d$ be vector families with $0 \in \text{conv } V_i$ for each $i \in [n]$. Then for any convex body K with $\gamma_d(K) \geq 1/2$, there exist vectors $v_i \in V_i$ such that $\sum_{i=1}^n v_i \in cK$, where $c > 0$ is an absolute constant.*

Applying this theorem to the Euclidean norm, one retrieves a sum of norm at most $C\sqrt{d}$ for some constant $C > 1$, and for the maximum norm one obtains a bound of $O(\sqrt{d \ln d})$. We note that their proof method, which is based on the techniques of Lovász, Spencer, and Vesztergombi [23], can be modified to show that the bound in the colorful setting is at most twice the original vector balancing constant, which implies $O(\sqrt{d})$ bounds for both the Euclidean and maximum norm. This asymptotically matches the estimates proved in the present paper up to constants; for details, see Section 7. Here, we provide a direct, constructive approach for proving an asymptotically matching, yet tighter estimate which also sheds more light on the geometry of the problem and its algorithmic aspects.

We set off to obtain sharp/asymptotically sharp bounds for the colorful vector balancing problem in the Euclidean and maximum norm. Our proof techniques yield estimates matching (2) and (4), the bounds for the original vector balancing problem.

Theorem 1.4 (Colorful balancing in Euclidean norm). *Given vector families $V_1, \dots, V_n \subseteq B_2^d$ with $0 \in \sum_{i \in [n]} \text{conv } V_i$, there exists a selection of vectors $v_i \in V_i$ for $i \in [n]$ such that*

$$\left\| \sum_{i \in [n]} v_i \right\|_2 \leq \sqrt{d}.$$

Note that the condition $0 \in \sum_{i \in [n]} \text{conv } V_i$ is weaker than requiring $0 \in \text{conv } V_i$ for each i — by applying a shift of each family, the more general estimate can be derived from the statement under this more restrictive condition, albeit with the loss of a factor 2 compared to the above bound.

For an estimate in the dual direction, the following result is well known: if V_1, \dots, V_n are sets of *unit* vectors with $0 \in \text{conv } V_i$ for each i , then one may select $v_i \in V_i$ for $i \in [n]$ so that $\left\| \sum_{i \in [n]} v_i \right\|_2 \geq \sqrt{n}$ (for a further generalization, see [1]).

Theorem 1.5 (Colorful balancing in maximum norm). *Given vector families $V_1, \dots, V_n \subseteq B_\infty^d$ with $0 \in \sum_{i \in [n]} \text{conv } V_i$, there exists $v_i \in V_i$ for $i \in [n]$ such that*

$$\left\| \sum_{i \in [n]} v_i \right\|_\infty \leq C\sqrt{d},$$

where $C = 22$ suffices.

As noted above, an $O(\sqrt{d \ln d})$ estimate is implied by Theorem 1.3 — we note that a matching bound can also be derived by applying the standard probabilistic method.

By Carathéodory's theorem we may assume that each V_i is finite (in fact, of cardinality at most $d + 1$). We are going to use this assumption without further mention wherever needed.

Let us conclude this section with a short selection of related results. Many classical vector balancing results are surveyed by Giannopoulos [17]. Vector balancing in the plane was studied by Swanepoel [39] and Lund, Magazinov [25]. Online versions of vector balancing and related combinatorial games were considered by Spencer in [24, 33]. Various anti-balancing questions were discussed by Banaszczyk [3] and Ambrus, González Merino [2]. Vector balancing for arbitrary convex bodies K was connected to the Gaussian measure of K by Banaszczyk [4]. Several recent

results have been given which provide algorithmic proofs for classical and online vector balancing problems, see, for example, Bansal [5], Bansal et al. [7, 8], Dadush et al. [14], and Lovett and Meka [24]. Additionally, vector balancing has recently been applied to various topics in machine learning, including sketches, coresets and randomized control trials, see Karnin, Liberty [21], Harshaw et al. [20], among others.

Vector balancing in the maximum norm is also closely related to discrepancy theory, since $\text{vb}(B_\infty^d, n)$ may be interpreted as an upper bound for the discrepancy of a set system of cardinality d on n elements. For that direction, see, for example, Matoušek [26].

One cannot miss to present a particularly attractive conjecture by Komlós (see [4, 35]), which states that

$$\text{vb}(B_2^d, B_\infty^d, n) \leq C$$

holds for each $n, d \geq 1$ with a universal constant C . The strongest result in this direction is due to Banaszczyk [4] who proved the upper bound of $O(\sqrt{\ln d})$. An algorithmic proof which yields a matching estimate was given recently by Bansal, Dadush, and Garg [6].

Due to space limitations, other classical vector summation problems, for example, the Steinitz lemma, questions about subset sums and partial sums, and the role of permutations are not addressed here. We note only that a colorful generalization of the Steinitz lemma has recently been given by Oertel, Paat, and Weismantel [27]. These directions are well portrayed in the book of Spencer [37] and the survey articles of Bárány [9] and Sevast’yanov [31, 32].

2 | REDUCING TO DIMENSION MANY FAMILIES

Using the method of linear dependencies, we first prove that the number of families can be reduced from n to at most d , moreover, the total number of vectors can also be bounded from above. This approach dates back to the classical work of Shapley and Folkman, and Starr in the 1960s [38]. Several applications of the method are well surveyed by Bárány [9].

We start with some terminology. Vectors in \mathbb{R}^d will be written as $x = (x^{(1)}, \dots, x^{(d)}) \in \mathbb{R}^d$ and will be understood as column vectors.

As usual, $\{e_1, \dots, e_d\}$ stands for the standard orthonormal basis of \mathbb{R}^d . We denote the standard $(m - 1)$ -dimensional simplex represented in \mathbb{R}^m by

$$\Delta^m := \text{conv} \{e_1, \dots, e_m\} = \left\{ \lambda \in \mathbb{R}^m : \lambda^{(i)} \geq 0 \forall i \in [m], \sum_{i \in [m]} \lambda^{(i)} = 1 \right\}.$$

In this paper, a convex polytope means a non-empty, bounded intersection of finitely many closed halfspaces (without any requirement on its interior).

We identify a set of vectors $U = \{u_1, \dots, u_m\} \subset \mathbb{R}^d$ with the $d \times m$ matrix

$$U = (u_1 \cdots u_m).$$

Definition 2.1. Given vector families $V_1, \dots, V_n \subset \mathbb{R}^d$ with $|V_i| = m_i$ and $\sum_{i \in [n]} m_i = m$, we define the associated *vector family matrix*

$$V := (V_1 | V_2 | \cdots | V_n) \in \mathbb{R}^{d \times m},$$

which is a partitioned matrix. We also introduce the associated *set of convex coefficients*

$$\Delta_V := \Delta^{m_1} \times \cdots \times \Delta^{m_n} \subset \mathbb{R}^m \quad (6)$$

which is a convex polytope arising as a direct product of simplices.

The relevance of Δ_V is shown by the fact that a vector $v \in \mathbb{R}^d$ is contained in $\text{conv } V_i$ if and only if $v = V_i \lambda$ for some $\lambda \in \Delta^{m_i}$. Accordingly,

$$\text{conv } V_1 + \cdots + \text{conv } V_n = \{V\lambda : \lambda \in \Delta_V\}.$$

In the above scenario, we will usually consider \mathbb{R}^m along with its orthogonal decomposition $\mathbb{R}^m = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_n}$. A collection of vector families $V = \{V_1, \dots, V_n\}$ will always be identified with its associated vector family matrix V — using the same notation for these two will cause no ambiguity and will be clarified by the context. From now on, U, V , and W will always stand for a collection of vector families or their associated vector family matrices.

Throughout the paper, Greek letters will be used to denote vectors in the coefficient space $\Delta_V \subset \mathbb{R}^m$, while letters of the Latin alphabet will stand for vectors in \mathbb{R}^d . To make the connection between these spaces explicit, coefficient vectors $\beta \in \Delta_V$ will also be indexed by members of V_i as follows:

$$\beta = (\beta(v_i))_{v_i \in V_i, i \in [n]} \in \mathbb{R}^m. \quad (7)$$

Given a vector family matrix $V \in \mathbb{R}^{m \times d}$ and a set of indices $J \subset [m]$, we naturally define $V|_J$, the restriction of V to the columns indexed by elements of J . This is again a vector family matrix which naturally induces a collection of vector families, the restrictions of the original ones to J . Naturally, $\Delta_{V|_J} \subset \mathbb{R}^{|J|}$ is the set of convex coefficients associated to $V|_J$. By virtue of the indexing (7), we may also define the restriction to a subcollection $W \subset V$. In particular, for $\beta \in \Delta_V$ and $V_i \in V$, $\beta|_{V_i} \in \Delta_{V_i}$ consists of the coefficients of vectors in V_i .

Given a partition $I \cup J = [m]$ and vectors $\lambda \in \Delta_{V|_I}$, $\mu \in \Delta_{V|_J}$, we introduce the natural concatenation of λ and μ by $\lambda \vee \mu \in \Delta_V$; that is, $(\lambda \vee \mu)|_I = \lambda$ and $(\lambda \vee \mu)|_J = \mu$.

Definition 2.2. A number $x \in [0, 1]$ is *fractional* if $x \notin \{0, 1\}$. Given a vector $\beta \in \Delta_V$, we say that family V_i is *locked* by β if none of the coordinates of $\beta|_{V_i}$ are fractional. Otherwise family V_i is *free* under β . A vector $\beta \in \Delta_V$ is a *selection vector* if every family is locked by β , equivalently, β is a vertex of Δ_V .

Note that for a selection vector $\beta \in \Delta_V$, $V_i \beta|_{V_i} = v_i$ for some $v_i \in V_i$ for each $i \in [n]$.

The main tool of the section is the following generalization of the Shapley–Folkman lemma [29, 38], a cornerstone result in econometric theory. Alternative versions were proved and used by Grinberg and Sevast'yanov [19] and Bárány and Grinberg [11].

Theorem 2.3. *Given a collection of vector families $V = \{V_1, \dots, V_n\}$ in \mathbb{R}^d with $0 \in \sum_{i \in [n]} \text{conv } V_i$, there exists a vector $\alpha \in \Delta_V$ such that*

- (i) $V\alpha = 0$;
- (ii) All but $k \leq d$ families V_i are locked by α ;
- (iii) α has at most $k + d$ fractional coordinates.

The proof is based on the Shapley–Folkman-style statement below which is related to the geometry of basic feasible solutions of linear programs.

Lemma 2.4 [19]. *Let K be a polyhedron in \mathbb{R}^m defined by a system*

$$\begin{cases} f_i(x) = a_i, & i = 1, \dots, p, \\ g_j(x) \leq b_j, & j = 1, \dots, q, \end{cases}$$

where f_i, g_j are linear functions. Let x_0 be a vertex of K and $A = \{j : g_j(x_0) = b_j\}$. Then $|A| \geq m - p$.

Proof of Theorem 2.3. Given vector families V_1, \dots, V_n in \mathbb{R}^d with $0 \in \sum_{i \in [n]} \text{conv } V_i$ and $m = \sum_{i \in [n]} |V_i|$, consider the set

$$P = \{\lambda \in \Delta_V : V\lambda = 0\} \\ = \left\{ \lambda \in \mathbb{R}^m : \sum_{i \in [n]} \sum_{v_i \in V_i} \lambda(v_i)v_i = 0, \sum_{v_i \in V_i} \lambda(v_i) = 1 \forall i \in [n], \lambda(v_i) \geq 0 \forall i \in [n], \forall v_i \in V_i \right\}. \tag{8}$$

By our assumption that $0 \in \sum_{i \in [n]} \text{conv } V_i$, P is a (non-empty) convex polytope in \mathbb{R}^m . Let $\alpha \in P$ be any extreme point of P . Define

$$S := \{i \in [n] : V_i \text{ is free under } \alpha\}$$

and let $k = |S|$. By Lemma 2.4, at most $n + d$ non-negativity inequalities in (8) are slack when substituting $\lambda = \alpha$. Each of the $n - k$ families locked by α contribute exactly one slack constraint, arising from the (unique) 1-coordinate. Let f denote the number of fractional coordinates of α ; then $f + (n - k)$ is the total number of slack constraints. Thus

$$f + (n - k) \leq n + d$$

which implies that $f \leq k + d$. Since, by definition, $f \geq 2k$, this also shows that $k \leq d$. □

By virtue of allowing to reduce consideration to at most d families, the following corollary is the main tool for proving upper bounds for the colorful vector balancing problem in arbitrary norms.

Corollary 2.5. *Let $\|\cdot\|$ be a norm on \mathbb{R}^d with unit ball B^d . Suppose there exists a constant $C(d)$ such that given any collection of $k \leq d$ families $U = \{U_1, \dots, U_k\}$ in B^d satisfying $|U_1| + \dots + |U_k| \leq k + d$, and any $\lambda \in \Delta_U$, there exists a selection vector $\mu \in \Delta_U$ such that*

$$\|V\lambda - V\mu\| \leq C(d).$$

Then given any collection of families $V_1, \dots, V_n \subseteq B^d$ with $0 \in \sum_{i \in [n]} \text{conv } V_i$, there exists a selection of vectors $v_i \in V_i$ for $i \in [n]$ such that

$$\left\| \sum_{i \in [n]} v_i \right\| \leq C(d).$$

Proof. Suppose that the hypothesis of the statement holds. Let $m := |V_1| + \dots + |V_n|$. Applying Theorem 2.3 to $V = \{V_1, \dots, V_n\}$, we find $\alpha \in \Delta_V$ such that $V\alpha = 0$, all but $k \leq d$ families V_i are locked by α , and α has at most $k + d$ fractional coordinates. Let $F \subset [m]$ be the set of indices of fractional coordinates, and set $L = [m] \setminus F$. Then $|F| \leq k + d$. By hypothesis, there exists a selection vector $\mu \in \Delta_{V|_F}$ such that $\|V|_F\alpha|_F - V|_F\mu\| \leq C(d)$, and so

$$\|V|_L\alpha|_L + V|_F\mu\| \leq \|V|_L\alpha|_L + V|_F\alpha|_F\| + \|-V|_F\alpha|_F + V|_F\mu\| \leq \|V\alpha\| + C(d) = C(d).$$

Taking the selection of vectors given by $\alpha|_L \vee \mu$ completes the proof. □

3 | THE CASE OF THE EUCLIDEAN NORM

To prove Theorem 1.4, it remains to prove the vertex approximation property for color classes, which generalizes the Lemma in [34], see also [9, Theorem 4.1].

Proposition 3.1 (Colorful vertex approximation in Euclidean norm). *Given a collection of k vector families $U = \{U_1, \dots, U_k\}$ in B_2^d and any point $\lambda \in \Delta_U$, there exists a selection vector $\mu \in \Delta_U$ such that $\|U\lambda - U\mu\|_2 \leq \sqrt{k}$.*

Our proof is inspired by Spencer’s argument for the vector balancing case [34], and it works in any finite dimensional Hilbert space.

Proof. Define $x := U\lambda \in \text{conv } U_1 + \dots + \text{conv } U_k$, so that

$$x = x_1 + \dots + x_k, \quad x_i = \sum_{u_i \in U_i} \lambda(u_i)u_i \quad \forall i \in [k],$$

where $\lambda|_{U_i} \in \Delta_{U_i}$ for each $i \in [k]$. We define a vector-valued random variable $w_i \in \mathbb{R}^d$ for each $i \in [k]$, which takes the value u_i with probability $\lambda(u_i)$ for each $u_i \in U_i$, independently of the other variables $w_j, j \in [k] \setminus \{i\}$. Then

$$\mathbb{E}[w_1 + \dots + w_k] = \sum_{i \in [k]} \mathbb{E}[w_i] = \sum_{i \in [k]} \sum_{u_i \in U_i} \lambda(u_i)U_i = \sum_{i \in [k]} x_i = x.$$

Component-wise this yields

$$\mathbb{E}\left[w_1^{(\ell)} + \dots + w_k^{(\ell)} - x^{(\ell)}\right] = 0, \quad \ell \in [d]. \tag{9}$$

For each $\ell \in [d]$, (9) and the independence of the variables w_i ’s imply

$$\begin{aligned} \mathbb{E}\left[(w_1^{(\ell)} + \dots + w_k^{(\ell)} - x^{(\ell)})^2\right] &= \mathbb{E}\left[(w_1^{(\ell)} + \dots + w_k^{(\ell)} - x^{(\ell)})^2\right] - \mathbb{E}\left[w_1^{(\ell)} + \dots + w_k^{(\ell)} - x^{(\ell)}\right]^2 \\ &= \text{Var}\left[w_1^{(\ell)} + \dots + w_k^{(\ell)} - x^{(\ell)}\right] \\ &= \sum_{i \in [k]} \text{Var}\left[w_i^{(\ell)}\right]. \end{aligned} \tag{10}$$

Since

$$\|w_1 + \dots + w_k - x\|_2^2 = \sum_{\ell=1}^d \left((w_1^{(\ell)} + \dots + w_k^{(\ell)}) - x^{(\ell)} \right)^2,$$

by linearity of expectation and (10) we conclude

$$\begin{aligned} \mathbb{E}[\|w_1 + \dots + w_k - x\|^2] &= \sum_{\ell \in [d]} \mathbb{E}[(w_1^{(\ell)} + \dots + w_k^{(\ell)} - x^{(\ell)})^2] \\ &= \sum_{\ell \in [d]} \sum_{i \in [k]} \text{Var}[w_i^{(\ell)}] \\ &= \sum_{\ell \in [d]} \sum_{i \in [k]} \mathbb{E}[(w_i^{(\ell)})^2] - \sum_{\ell \in [d]} \sum_{i \in [k]} \mathbb{E}[w_i^{(\ell)}]^2 \\ &= \sum_{i \in [k]} \mathbb{E}[\|w_i\|_2^2] - \sum_{\ell \in [d]} \sum_{i \in [k]} \mathbb{E}[w_i^{(\ell)}]^2. \end{aligned} \tag{11}$$

Finally, we note that

$$\mathbb{E}[\|w_i\|_2^2] = \sum_{u_i \in U_i} \lambda(u_i) \cdot \|u_i\|_2^2 \leq \sum_{u_i \in U_i} \lambda(u_i) = 1,$$

hence continuing calculation (11),

$$\mathbb{E}[\|w_1 + \dots + w_k - x\|^2] \leq k - \sum_{\ell \in [d]} \sum_{i \in [k]} \mathbb{E}[w_i^{(\ell)}]^2 \leq k.$$

It follows that for some specific choice of $u_i \in U_i, i \in [k]$, we have

$$\|u_1 + \dots + u_k - x\|_2^2 \leq k.$$

The corresponding selection vector $\mu \in \Delta_U$ satisfies the proposition. □

Theorem 1.4 now follows immediately from Corollary 2.5 and Proposition 3.1.

4 | THE CASE OF THE MAXIMUM NORM

To prove Theorem 1.5, we need to show that the vertex approximation property (the analogue of Proposition 3.1) holds for the maximum norm. This result for the original vector balancing problem is due to Spencer [35] and, independently, Gluskin [18]. As in the original vector balancing problem, the challenge is to remove the $\sqrt{\ln d}$ factor. Note that, unlike in the Euclidean case, we need to set an upper bound on the total cardinality of the vector systems.

Proposition 4.1 (Colorful vertex approximation in Maximum norm). *Given a collection of k vector families $U = \{U_1, \dots, U_k\}$ in B_∞^d satisfying $m := |U_1| + \dots + |U_k| \leq 2d$, and an arbitrary point $\lambda \in \Delta_U$, there exists a selection vector $\mu \in \Delta_U$ such that $\|U\lambda - U\mu\|_\infty \leq C\sqrt{d}$ for a universal constant $C > 0$.*

Note that applying the probabilistic method directly, by mimicking the second proof of Proposition 3.1 results in the weaker upper bound of $O(\sqrt{d}\sqrt{\ln d})$. Thus, in order to reach the bound of $O(\sqrt{d})$, one must apply an alternative argument, just as in the case of the original vector balancing problem in the maximum norm. In [35], Spencer utilized a partial coloring method in order to overcome this difficulty. This, and the argument of Lovett and Meka [24] are the predecessors of our approach described below.

We will prove Proposition 4.1 by iterating the following lemma, which is a close relative of the Partial Coloring Lemma in [24]. We call it the *skeleton approximation lemma*, as it approximates a point in the set of convex coefficients $\Delta_W \subset \mathbb{R}^m$ by a point on the $(m/2)$ -skeleton of Δ_W .

Lemma 4.2 (Skeleton approximation). *Let $W = \{W_1, \dots, W_k\}$ be a collection of vector families in B_∞^d with $|W_i| \geq 2$ for each i and $m := |W_1| + \dots + |W_k| \leq 2d$. Then for any point $\lambda \in \Delta_W$, there exists $\mu \in \Delta_W$ such that*

$$(i) \quad \|W\lambda - W\mu\|_\infty \leq \eta \sqrt{m \ln \frac{\xi d}{m}}$$

where η, ξ are constants specified as

$$\eta = \frac{7}{3} \text{ and } \xi = 18; \tag{12}$$

$$(ii) \quad \mu^{(i)} = 0 \text{ for at least } m/2 \text{ indices } i \in [m].$$

The proof of Lemma 4.2 is postponed to Section 6. Now we deduce Proposition 4.1 assuming Lemma 4.2.

Proof of Proposition 4.1. We may assume that $|U_i| \geq 2$ for each i , since any convex coefficient vector corresponding to a 1-element family is necessarily a selection vector.

By an inductive process, we are going to define points $\lambda(s) \in \Delta_U$, sets of indices $F(s), L(s) \subset [m]$, and cardinalities $m(s)$ for $s = 0, 1, \dots$ so that for a suitably large S , $\lambda(S)$ is a selection vector with the desired properties.

To initiate the recursive process, take $\lambda(0) = \lambda$, let $F(0) \subset [m]$ be the set of indices of fractional coordinates of $\lambda(0)$, and $L(0) = [m] \setminus F(0)$ be the set of indices of coordinates of $\lambda(0)$ equal to 0 or 1. Introducing $m(0) = |F(0)|$, we have $m(0) \leq m \leq 2d$.

Assuming that iterative step s has been taken, we define step number $s + 1$ as follows. Apply Lemma 4.2 to the vector family matrix $U(s) := U|_{F(s)}$ of total cardinality $m(s) \leq m$ and the point $\lambda(s)|_{F(s)} \in \Delta_{U(s)}$ to find $\mu(s + 1) \in \Delta_{U(s)}$ with the prescribed properties. Define $\lambda(s + 1) = \lambda(s)|_{L(s)} \vee \mu(s + 1)$ to be natural concatenation of these two vectors, obtained by replacing the fractional coordinates of $\lambda(s)$ by the approximating vector $\mu(s + 1)$. Let $F(s + 1) \subset [m]$ be the set of indices of fractional coordinates of $\lambda(s + 1)$, $L(s + 1) = [m] \setminus F(s + 1)$, and set $m(s + 1) = |F(s + 1)|$.

By Property (i) of Lemma 4.2 and the definition of $\lambda(s + 1)$, for each $s \geq 0$

$$\|U\lambda(s) - U\lambda(s + 1)\|_\infty \leq \eta \sqrt{m(s) \ln \frac{\xi d}{m(s)}}. \tag{13}$$

Also, by property (ii) of Lemma 4.2 we have that

$$m(s) \leq \frac{m}{2^s} \leq \frac{d}{2^{s-1}}. \tag{14}$$

Since $m(s) \in \mathbb{N}$, this also yields that after a finite number S of steps, $m(S) = 0$ will hold. Set $\mu = \lambda(S)$. We will show that μ fulfills the criteria of Proposition 4.1.

That $\mu \in \Delta_U$ is a selection vector is shown by $m(S) = 0$. To show the approximation property, note that the function $f(x) = x \ln(1/x)$ is increasing on the interval $[0, 1/4]$. Combined with (13),(14), and (12), this yields that

$$\begin{aligned} \|U\lambda - U\mu\|_\infty &\leq \sum_{s=0}^{S-1} \|U\lambda(s) - U\lambda(s + 1)\|_\infty \\ &\leq \sum_{s=0}^{S-1} \eta \sqrt{\ln \frac{\xi d}{m(s)}} \sqrt{m(s)} \\ &\leq \sum_{s=0}^{S-1} \eta \sqrt{\ln \frac{\xi d}{d/2^{s-1}}} \sqrt{\frac{d}{2^{s-1}}} \\ &\leq \eta \sqrt{d} \sum_{s=0}^{\infty} 2^{-(s-1)/2} \sqrt{\ln(\xi) + \ln 2 \cdot (s - 1)} \\ &< 22\sqrt{d}. \quad \square \end{aligned} \tag{15}$$

As in the Euclidean case, Theorem 1.5 now follows from Corollary 2.5 and Proposition 4.1. A simple modification of the proof yields the following version of Proposition 4.1, which provides a significant strengthening for $m \ll 2d$.

Proposition 4.3. *Given a collection of vector families $U = \{U_1, \dots, U_k\}$ in B_∞^d such that $m = |U_1| + \dots + |U_k| \leq 2d$ and any point $\lambda \in \Delta_U$, there exists a selection vector $\mu \in \Delta_U$ such that*

$$\|U\lambda - U\mu\|_\infty \leq K \sqrt{m} \sqrt{\ln \frac{18d}{m}}$$

for a universal constant $K > 0$.

Proof. Take $\mu \in \Delta_V$ as in the proof of Proposition 4.1. Then, substituting (14) in (15),

$$\begin{aligned} \|U\lambda - U\mu\|_\infty &\leq \sum_{s=0}^{S-1} \eta \sqrt{\ln \frac{\xi d}{m/2^{s-1}}} \sqrt{\frac{m}{2^{s-1}}} \\ &\leq \eta \sqrt{m} \sum_{s=0}^{\infty} \frac{\sqrt{\ln \left(\frac{\xi d}{m} \right) + \ln 2 \cdot s}}{\sqrt{2^{s-1}}} \end{aligned}$$

$$\begin{aligned} &\leq \eta \sqrt{m} \sqrt{\ln \frac{\xi d}{m}} \sum_{s=0}^{\infty} \frac{\sqrt{1+s}}{\sqrt{2^{s-1}}} \\ &< 9\eta \sqrt{m} \sqrt{\ln \frac{18d}{m}}. \end{aligned} \quad \square$$

5 | USEFUL PROPERTIES OF THE GAUSSIAN DISTRIBUTION

Proving Lemma 4.2, the skeleton approximation lemma, requires several standard facts about the behavior of Gaussian random variables. By $\mathcal{N}(\mu, \sigma^2)$ we denote the (one-dimensional) Gaussian distribution with mean μ and variance σ^2 . Given a linear subspace $A \subseteq \mathbb{R}^d$, $\mathcal{N}(A)$ denotes the standard multi-dimensional Gaussian distribution on A , that is, for $G \sim \mathcal{N}(A)$, $G = G_1 a_1 + \dots + G_m a_m$, where $\{a_1, \dots, a_m\}$ is any orthonormal basis of A and $G_1, \dots, G_m \sim \mathcal{N}(0, 1)$ are independent Gaussian random variables (for further details, see [13, 16]).

Lemma 5.1. *Let $A \subseteq \mathbb{R}^d$ be a linear subspace with $G \sim \mathcal{N}(A)$. Then given any $u \in \mathbb{R}^d$, $\langle G, u \rangle \sim \mathcal{N}(0, \sigma^2)$, with $\sigma^2 = \|P_A(u)\|^2 \leq \|u\|_2^2$, where $P_A(\cdot)$ denotes the orthogonal projection onto A .*

Corollary 5.2. *Let $A \subseteq \mathbb{R}^d$ be a linear subspace with $G \sim \mathcal{N}(A)$ and define σ_i by $\langle G, e_i \rangle \sim \mathcal{N}(0, \sigma_i^2)$. Then $\sum_{i \in [d]} \sigma_i^2 = \dim A$.*

A proof of Lemma 5.1 can be found in [16, Section III.6] (see also [24]). These results are particularly useful when combined with the following standard tail estimate.

Lemma 5.3. *Given a Gaussian random variable $G \sim \mathcal{N}(\mu, \sigma^2)$, for all $t > 0$,*

$$\mathbb{P}[|G - \mu| \geq t] \leq \exp(-t^2/2\sigma^2).$$

This result is a special case of the general version of Hoeffding's inequality (for a proof, see, e.g., [40]). We will also need a similar bound for *martingales* with Gaussian steps. Recall that a sequence $\{X_i\}_{i \in \mathbb{N}}$ of real-valued random variables is a martingale if $\mathbb{E}[X_{n+1} | X_1, \dots, X_n] = X_n$.

Lemma 5.4 [5]. *Let $0 = X_0, X_1, \dots, X_T$ be a martingale in \mathbb{R} with steps $Y_i = X_i - X_{i-1}$ for $i \geq 1$. Suppose that for all $i \in [T]$, $Y_i | X_0, \dots, X_{i-1}$ is a Gaussian random variable with mean zero and variance at most σ^2 . Then for any $c > 0$,*

$$\mathbb{P}[|X_T| \geq \sigma c \sqrt{T}] \leq 2 \exp(-c^2/2).$$

Finally, we will need the following result about sequences of Gaussian random variables. This is a well-known result that can be found for example in [40]; we provide the standard proof for the reader's convenience.

Lemma 5.5. *Let $X_i \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i \leq 1$ for $i = 1, 2, \dots$ be a sequence of not necessarily independent, jointly Gaussian random variables. Then for any $T \geq 2$,*

$$\mathbb{E} \max_{i \leq T} |X_i| \leq 6 \sqrt{\ln T}$$

and

$$\mathbb{E} \max_{i \leq T} |X_i|^2 \leq 10 \ln T.$$

Proof. We define the random variable $Y := \max_{i \in \mathbb{N}} \frac{|X_i|}{\sqrt{1+\ln i}}$. Then by Lemma 5.3, the union bound, and the fact that $\sigma_i \leq 1$ for all i ,

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^\infty \mathbb{P}[Y \geq y] dy \\ &= \int_0^2 \mathbb{P}[Y \geq y] dy + \int_2^\infty \mathbb{P}[Y \geq y] dy \\ &\leq 2 + \int_2^\infty \mathbb{P} \left[\max_{i \in \mathbb{N}} \frac{|X_i|}{\sqrt{1+\ln i}} \geq y \right] dy \\ &\leq 2 + \int_2^\infty \sum_{i=1}^\infty \mathbb{P} \left[|X_i| \geq y \sqrt{1+\ln i} \right] dy \\ &\leq 2 + \int_2^\infty \sum_{i=1}^\infty \exp(-y^2(1+\ln i)/2\sigma_i^2) dy \\ &\leq 2 + \int_2^\infty \left(\sum_{i=1}^\infty i^{-y^2/2} \right) \exp(-y^2/2) dy \\ &\leq 2 + \frac{\pi^2}{6} \cdot 0.06 < 3. \end{aligned} \tag{16}$$

Finally, note that $\sqrt{1+\ln i} \leq \sqrt{1+\ln T}$ for all $i \in [T]$, hence the calculation in (16) yields

$$\mathbb{E} \max_{i \leq T} \frac{|X_i|}{\sqrt{1+\ln T}} \leq \mathbb{E} \max_{i \leq T} \frac{|X_i|}{\sqrt{1+\ln i}} \leq \mathbb{E} \max_{i \in \mathbb{N}} \frac{|X_i|}{\sqrt{1+\ln i}} < 3.$$

Then for $T \geq 2$, $\mathbb{E} \max_{i \leq T} |X_i| < 3\sqrt{1+\ln T} \leq 6\sqrt{\ln T}$. The proof for $\max_{i \leq T} |X_i|^2$ follows from an analogous calculation. □

6 | PROOF OF THE SKELETON APPROXIMATION LEMMA

We complete the proof of Theorem 1.5 by proving the crux of the argument, Lemma 4.2. This will be done by means of providing an algorithm that proves the following slightly weaker statement.

Lemma 6.1. *Let $0.01 > \delta > 0$ be arbitrary, and let $W = \{W_1, \dots, W_k\}$ be a collection of vector families in B_∞^d which satisfies that $|W_i| \geq 2$ for each i , and $m := \sum_{i \in [k]} |W_i| \leq 2d$. Define*

$$\omega(m) := \eta \sqrt{m \ln \frac{\xi d}{m}}, \tag{17}$$

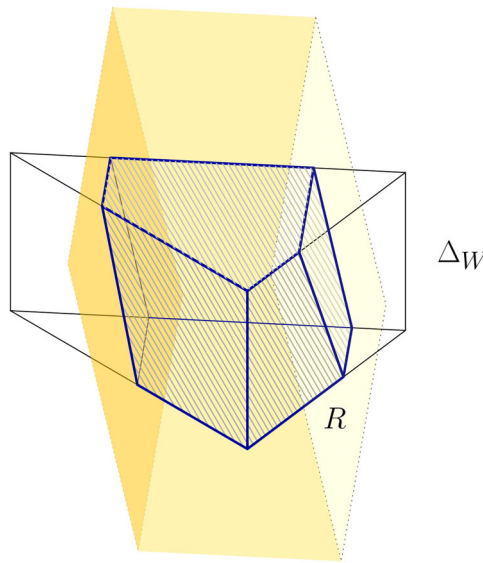


FIGURE 1 The polytope R .

where $\eta = \frac{7}{3}$ and $\xi = 18$ as in (12). Then for any $\gamma \in \Delta_W$ there exists $\hat{\gamma} \in \Delta_W$ such that

- (i) $\|W\gamma - W\hat{\gamma}\|_\infty \leq \omega(m)$;
- (ii) $\hat{\gamma}^{(i)} \leq \delta$ for at least $m/2$ indices $i \in [m]$.

Lemma 4.2 follows immediately from Lemma 6.1 by standard compactness arguments.

Proof of Lemma 6.1. For each $j \in [d]$, let $W^j \in \mathbb{R}^m$ denote the j th row of the vector family matrix W . The condition $W \subset B_\infty^d$ ensures that $\|W^j\|_\infty \leq 1$ for each $j \in [d]$. Accordingly,

$$\|W^j\|_2^2 \leq m \tag{18}$$

for each j .

Consider the polytope

$$R := \left\{ \alpha \in \mathbb{R}^m : \alpha \in \Delta_W, \|W\alpha - W\gamma\|_\infty \leq \omega(m) \right\},$$

which is the intersection of Δ_W with d slabs of width $\omega(m)/\|W^j\|_\infty$, $j \in [d]$ (see Figure 1). Equivalently, R is defined by the following set of linear equations and inequalities:

$$R = \left\{ \alpha \in \mathbb{R}^m : \sum_{w_\ell \in W_\ell} \alpha(w_\ell) = 1 \forall \ell \in [k], \alpha^{(i)} \geq 0 \forall i \in [m], |\langle \alpha - \gamma, W^j \rangle| \leq \omega(m) \forall j \in [d] \right\}. \tag{19}$$

We call the first and second sets of constraints *convexity constraints*, as they ensure that $W\alpha \in \text{conv}(W_1) + \dots + \text{conv}(W_k)$ for each $\alpha \in R$. The third set of constraints will be referred to as *maximum constraints*, as they imply that given $\alpha \in R$,

$$\|W\alpha - W\gamma\|_\infty = \max_{j \in [d]} |(W\alpha - W\gamma)_j| = \max_{j \in [d]} |\langle \alpha - \gamma, W^j \rangle| \leq \omega(m).$$

Let Z be the set of normal vectors of the inequality constraints in (19):

$$Z = \{e_1, \dots, e_m, W^1, \dots, W^d\}. \tag{20}$$

By the previous remarks, $\|Z\|_\infty = 1$.

The main tool of the argument is to introduce a suitable discrete time Gaussian random walk on \mathbb{R}^m , similar to that in [5, 24]. In order to help the reader navigate through the forthcoming technical details, we first give an intuitive description of the walk, whose position at time $t = 0, 1, \dots$ will be denoted by $\Gamma_t \in \mathbb{R}^m$.

The walk starts from $\Gamma_0 = \gamma$ and runs in $\text{aff } R$ with sufficiently small Gaussian steps as long as Γ_t is in the interior of R , far from its boundary. As Γ_t gets δ -close to crossing a facet of R , we confine the walk to an affine subspace parallel to that facet for the subsequent steps, by intersecting the current range with a hyperplane parallel to the facet. In particular, if any coordinate of Γ_t reaches a value less than δ , we freeze that coordinate for the remainder of the walk.

We show that running the walk long enough, until say time T , with high probability at least half of the coordinates of Γ_T become frozen, while $\Gamma_T \in R$ still holds. This will mean that $\hat{\gamma} = \Gamma_T$ satisfies the criteria of Lemma 6.1. For the proof it is essential that the value of $\omega(m)$ is carefully set (17), hence the slabs defining R are sufficiently wide so that the walk is unlikely to escape from them.

Let us turn to the formal definition of the random walk. Let $\varepsilon > 0$ and $T \in \mathbb{N}$ be parameters to be defined later. Define the sets

$$C_t^{conv} := \{i \in [m] : \Gamma_t^{(i)} \leq \delta\}, \quad C_t^{max} := \{j \in [d] : |\langle \Gamma_t - \Gamma_0, W^j \rangle| \geq \omega(m) - \delta\} \tag{21}$$

to be the convexity and maximum constraints, respectively, that are at most δ -close to being violated by Γ_t . We will say that coordinate i is *frozen if and only if* $i \in C_t^{conv}$. Recall that by (7), coordinates may be indexed by the vectors, that is, for each $i \in [m]$, $\Gamma_t^{(i)} = \Gamma_t(w_l)$ for some $l \in [k]$ and $w_l \in W_l$. In that case, coordinate w_l is frozen if and only if $i \in C_t^{conv}$.

Let A be the linear component of $\text{aff } \Delta_W$, that is, $A = \text{lin}(\Delta_W - \Delta_W)$. For each $t \leq 1$, step t is confined to occur in the linear subspace

$$S_t := \{\beta \in A : \beta^{(i)} = 0 \forall i \in C_{t-1}^{conv}, \langle \beta - \Gamma_0, W^j \rangle = 0 \forall j \in C_{t-1}^{max}\}$$

by taking a Gaussian step $\Lambda_t \sim \mathcal{N}(S_t)$ and defining

$$\Gamma_t = \Gamma_{t-1} + \varepsilon \Lambda_t.$$

The walk terminates after T steps: $\hat{\gamma} := \Gamma_T$, where T is to be determined later.

We will show that with certain restrictions on the parameters, Γ_T satisfies properties (i) and (ii) of Lemma 6.1 with probability at least 0.2.

Given $\varepsilon > 0$, we define

$$T := \left\lceil \frac{0.99^2 \eta^2}{2\varepsilon^2} \right\rceil. \tag{22}$$

Choose $\varepsilon > 0$ small enough so that the following inequalities hold simultaneously:

$$6Td \exp\left(-\frac{\delta^2}{2m\varepsilon^2}\right) < 0.01, \tag{23}$$

$$22\varepsilon m^2 \ln T \leq 0.01, \tag{24}$$

and

$$10\varepsilon^2 \ln T \leq 1. \tag{25}$$

This can indeed be guaranteed since the functions $\exp(-x)/x$, $x \ln \frac{1}{x^2}$ and $x \ln \frac{1}{x}$ converge to 0 as $x \searrow 0$.

We summarize a few useful properties of the random walk.

Lemma 6.2. *Let $\Gamma_0, \dots, \Gamma_T$ be the steps of the Gaussian random walk defined above and $i \in [m]$, $j \in [d]$. Then*

- (i) given Γ_{t-1} , $\mathbb{E}[\Lambda_t] = 0$;
- (ii) C_t^{conv}, C_t^{max} are nested increasing sets in t ;
- (iii) S_t is a nested decreasing sequence of linear subspaces in \mathbb{R}^m in t ;
- (iv) at any time $0 \leq t \leq T$ and for any $i \in [k]$, $\sum_{w_i \in W_i} \Gamma_t(w_i) = 1$;
- (v) if the walk leaves the polytope R at time $t \in [T]$, then $\Gamma_s \notin R$ for any $s \geq t$;
- (vi) if the walk leaves the polytope R at time $t \in [T]$, then $|\langle \Lambda_t, z \rangle| \geq \delta/\varepsilon$ for some $z \in Z$;
- (vii) if coordinate i is frozen at step t , that is $i \in C_t^{conv} \setminus C_{t-1}^{conv}$, then $\Gamma_T^{(i)} = \Gamma_t^{(i)} \geq \delta - \varepsilon|\Lambda_t^{(i)}|$.

Proof. Properties (i)–(v) are straightforward consequences of the definition of Γ_t .

To prove (vi), suppose that the walk leaves the polytope R at time t . Then an inequality constraint in (19) with normal vector $z \in Z$ is violated at time t . Suppose that $z = W^j$ for some $j \in [d]$. Since $j \notin C_{t-1}^{max}$,

$$|\langle \Gamma_{t-1} - \Gamma_0, W^j \rangle| < \omega(m) - \delta,$$

while on the other hand,

$$|\langle \Gamma_t - \Gamma_0, W^j \rangle| > \omega(m).$$

Combining these inequalities shows that

$$\varepsilon|\langle \Lambda_t, W^j \rangle| = |\langle \Gamma_t - \Gamma_{t-1}, W^j \rangle| \geq \delta.$$

The proof when $z = e_i$ for $i \in [m]$ is analogous.

To prove (vii), note that $i \notin C_{t-1}^{conv}$ implies that $\Gamma_{t-1}^{(i)} \geq \delta$. Therefore,

$$\Gamma_T^{(i)} = \Gamma_t^{(i)} = \Gamma_{t-1}^{(i)} + \varepsilon\Lambda_t^{(i)} \geq \delta + \varepsilon\Lambda_t^{(i)} \geq \delta - \varepsilon|\Lambda_t^{(i)}|. \quad \square$$

Equipped with these properties, we are ready to prove that Γ_T satisfies the required conditions of Lemma 6.1 with probability at least 0.01. To show (i), that is,

$$\|W\Gamma_0 - W\Gamma_T\|_\infty \leq \omega(m), \tag{26}$$

it is sufficient to argue that (with high probability) $\Gamma_T \in R$; that is, the walk does not leave the polytope R at any step.

Define the event $E_t := \{\Gamma_t \notin R \mid \Gamma_0, \dots, \Gamma_{t-1} \in R\}$ that the walk steps out of R at time t . If E_t occurs, then by Lemma 6.2(vi), $|\langle \Lambda_t, z \rangle| \geq \delta/\varepsilon$ for some $z \in Z$. By Lemma 5.1 and (18), for any $z \in Z$, $\langle \Lambda_t, z \rangle$ is a Gaussian random variable with mean 0 and variance $\sigma^2 \leq m$. Applying Lemma 5.3 to $\langle \Lambda_t, z \rangle$, we find

$$\mathbb{P}\left[|\langle \Lambda_t, z \rangle| \geq \frac{\delta}{\varepsilon}\right] \leq 2 \exp\left(-\left(\frac{\delta}{\varepsilon}\right)^2/2m\right). \tag{27}$$

Using the union bound, equations (20), (27), and the fact that $d \leq 2d$, we derive

$$\begin{aligned} \mathbb{P}[\exists t \in [T] : \Gamma_t \notin R] &= \sum_{t=1}^T \mathbb{P}[E_t] \\ &\leq \sum_{t=1}^T \sum_{z \in Z} \mathbb{P}\left[|\langle \Lambda_t, z \rangle| \geq \frac{\delta}{\varepsilon}\right] \\ &\leq 2T(d + m) \exp\left(-\left(\frac{\delta}{\varepsilon}\right)^2/2m\right) \\ &\leq 6Td \exp\left(-\frac{\delta^2}{2m\varepsilon^2}\right) \\ &< 0.01 \end{aligned} \tag{28}$$

by condition (23). This proves that (26) holds with probability at least 0.99.

It remains to address (ii) of Lemma 6.1, that (with positive probability) $\Gamma_T^{(i)} \leq \delta$ for at least $m/2$ indices $i \in [m]$. We will reach this by means of proving that

$$\mathbb{E}[|C_T^{conv}|] > 0.51m. \tag{29}$$

To this end we derive the following identity, using Lemma 6.2(i):

$$\begin{aligned} \mathbb{E}[\|\Gamma_t\|_2^2] &= \mathbb{E}[\|\Gamma_{t-1} + \varepsilon\Lambda_t\|_2^2] \\ &= \mathbb{E}[\|\Gamma_{t-1}\|_2^2] + \varepsilon^2\mathbb{E}[\|\Lambda_t\|_2^2] + 2\varepsilon\mathbb{E}[\langle \Gamma_{t-1}, \Lambda_t \rangle] \\ &= \mathbb{E}[\|\Gamma_{t-1}\|_2^2] + \varepsilon^2\mathbb{E}[\dim(S_t)], \end{aligned}$$

where in the last equation we use that, by Corollary 5.2,

$$\mathbb{E}[\|\Lambda_t\|_2^2] = \mathbb{E}\left[\sum_{i \in [m]} \langle \Lambda_t, e_i \rangle^2\right] = \sum_{i \in [m]} \mathbb{E}[\langle \Lambda_t, e_i \rangle^2] = \dim S_t.$$

Iterating this calculation and using Lemma 6.2(iii),

$$\mathbb{E}[\|\Gamma_t\|_2^2] \geq \varepsilon^2 \sum_{t \in [T]} \mathbb{E}[\dim(S_t)] \geq T\varepsilon^2\mathbb{E}[\dim S_T] = T\varepsilon^2\mathbb{E}[m - |C_T^{conv}| - |C_T^{max}|],$$

and rearranging yields

$$\mathbb{E}[|C_T^{conv}|] \geq m - \frac{\mathbb{E}[\|\Gamma_T\|_2^2]}{T\varepsilon^2} - \mathbb{E}[|C_T^{max}|]. \tag{30}$$

The above identity allows us to prove (29) by giving upper estimates on $\mathbb{E}[\|\Gamma_T\|_2^2]$ and $\mathbb{E}[|C_T^{max}|]$.

We start with the second of these and show that

$$\mathbb{E}[|C_T^{max}|] \leq \frac{2m}{\xi}. \tag{31}$$

To this end, we bound the probability that the walk gets close to escaping from a given slab. Note that for fixed $j \in [d]$, $\{\langle \Gamma_t - \Gamma_0, W^j \rangle\}_{t \in [T]}$ for $0 \leq t \leq T$ is a martingale satisfying the conditions of Lemma 5.4. As the step size is $\varepsilon \langle \Lambda_t, W^j \rangle$, by Lemma 5.1 the variance of any step is bounded by $\varepsilon^2 \|W^j\|_2^2 \leq \varepsilon^2 m$ (cf. (18)).

For any $j \in C_T^{max}$, by (21),

$$|\langle \Gamma_T - \Gamma_0, W^j \rangle| \geq \omega(m) - \delta \geq 0.99 \omega(m),$$

as we have $\delta \leq 0.01$ and $\omega(m) \geq 1$ by (17).

Therefore, by Lemma 5.4, (17), and (22),

$$\begin{aligned} \mathbb{P}[j \in C_T^{max}] &\leq \mathbb{P}[|\langle \Gamma_T - \Gamma_0, W^j \rangle| \geq 0.99 \omega(m)] \\ &\leq 2 \exp\left(\frac{-0.99^2 \cdot \eta^2 \ln(\xi d/m)}{2T\varepsilon^2}\right) \\ &< 2 \exp\left(\ln \frac{m}{\xi d}\right) = \frac{2m}{\xi d}. \end{aligned}$$

Thus

$$\mathbb{E}[|C_T^{max}|] = \sum_{j \in [d]} \mathbb{P}[j \in C_T^{max}] < \frac{2m}{\xi}$$

as desired.

To complete the proof of (29) we address the second term in (30) and show that

$$\mathbb{E}[\|\Gamma_T\|_2^2] \leq 1.01m. \tag{32}$$

By (7), we represent Γ_T in terms of the vector families as $\Gamma_T = (\Gamma_T(w_i))$, $w_i \in W_i$, $i \in [k]$. Then

$$\|\Gamma_T\|_2^2 = \sum_{i \in [k]} \sum_{w_i \in W_i} (\Gamma_T(w_i))^2. \tag{33}$$

As the above double sum has m terms in total, it suffices to show that the expectation of any of these terms is at most 1.01, that is, $\mathbb{E}[\Gamma_T(w_i)^2] \leq 1.01$ for any $i \in [k]$ and $w_i \in W_i$.

By Lemma 6.2(iv), $\sum_{w_i \in W_i} \Gamma_T(w_i) = 1$. Thus, for any fixed $w_i \in W_i$,

$$\Gamma_T(w_i) = 1 - \sum_{w \in W_i \setminus \{w_i\}} \Gamma_T(w). \tag{34}$$

Note that in the above sum, $\Gamma_T(w) \geq 0$ unless coordinate w is frozen. In this case, assuming that coordinate w is frozen at step t , by Lemma 6.2(vii) we have $\Gamma_T(w) \geq \delta - \varepsilon |\Lambda_t(w)|$. Accordingly,

$$\Gamma_T(w) \geq \delta - \max_{t \in [T]} \varepsilon |\Lambda_t(w)| > -\varepsilon \max_{t \in [T]} |\Lambda_t(w)|. \tag{35}$$

Thus, by (34),

$$\Gamma_T(w_i) \leq 1 + \varepsilon \sum_{w \in W_i \setminus \{w_i\}} \max_{t \in [T]} |\Lambda_t(w)|.$$

When $\Gamma_T(w_i) \geq 0$, this leads to

$$\begin{aligned} (\Gamma_T(w_i))^2 &\leq 1 + 2\varepsilon \sum_{w \in W_i \setminus \{w_i\}} \max_{t \in [T]} |\Lambda_t(w)| + \varepsilon^2 \sum_{w \in W_i \setminus \{w_i\}} \max_{t \in [T]} |\Lambda_t(w)|^2 \\ &\quad + \varepsilon^2 \sum_{w \neq u \in W_i \setminus \{w_i\}} \max_{t \in [T]} |\Lambda_t(w)| \max_{t \in [T]} |\Lambda_t(u)|. \end{aligned}$$

Note that by Lemma 5.1, for each vector w , $\Lambda_t(w)$ is a Gaussian random variable with variance at most 1. Also, for $a, b \geq 0$ we will use that $ab \leq a^2 + b^2$. Therefore, by taking expectations above, and applying Lemma 5.5,

$$\begin{aligned} &\mathbb{E}\left((\Gamma_T(w_i))^2 \mid \Gamma_T(w_i) \geq 0\right) \\ &= 1 + 2\varepsilon \sum_{w \in W_i \setminus \{w_i\}} \mathbb{E}\left[\max_{t \in [T]} |\Lambda_t(w)|\right] + \varepsilon^2 \sum_{w \in W_i \setminus \{w_i\}} \mathbb{E}\left[\max_{t \in [T]} |\Lambda_t(w)|^2\right] \\ &\quad + \varepsilon^2 \sum_{w \neq u \in W_i \setminus \{w_i\}} \mathbb{E}\left[\max_{t \in [T]} |\Lambda_t(w)| \max_{t \in [T]} |\Lambda_t(u)|\right] \\ &\leq 1 + 12\varepsilon m \sqrt{\ln T} + 10\varepsilon^2 m \ln T + \varepsilon^2 \sum_{w \neq u \in W_i \setminus \{w_i\}} \mathbb{E}\left[\max_{t \in [T]} |\Lambda_t(w)|^2 + \max_{t \in [T]} |\Lambda_t(u)|^2\right] \tag{36} \\ &\leq 1 + 12\varepsilon m \sqrt{\ln T} + 10\varepsilon^2 m \ln T + 2 \frac{m(m-1)}{2} 10\varepsilon^2 \ln T \\ &\leq 1 + 12\varepsilon m \sqrt{\ln T} + 10\varepsilon^2 m^2 \ln T \\ &\leq 1 + 22\varepsilon m^2 \ln T \\ &\leq 1.01 \end{aligned}$$

by (24) and that $m, \ln T \geq 1, \varepsilon < 1$.

When $\Gamma_T(w_i) < 0$, then coordinate w_i is frozen. Therefore, (35) and (25) imply that

$$\mathbb{E}\left((\Gamma_T(w_i))^2 \mid \Gamma_T(w_i) < 0\right) < \varepsilon^2 \mathbb{E} \max_{t \in [T]} |\Lambda_t(w_i)|^2 \leq 10\varepsilon^2 \ln T < 1.$$

Combining this with (36) shows that $\mathbb{E}\left((\Gamma_T(w_i))^2\right) \leq 1.01$ for each $i \in [k]$ and $w_i \in W_i$, and by invoking (33), we reach (32).

To prove (29), we may now combine (22) and (30)–(32) in order to derive that

$$\begin{aligned} \mathbb{E}[|C_T^{conv}|] &\geq m - \frac{1.01m}{T\varepsilon^2} - \frac{2m}{\xi} \\ &\geq \left(1 - \frac{2 \cdot 1.01}{0.992\eta^2} - \frac{2}{\xi}\right)m \\ &> 0.51m. \end{aligned}$$

Since $|C_T^{conv}| \leq m$, this leads to

$$\mathbb{P}[|C_T^{conv}| \geq m/2] \geq 0.02.$$

As the probability of the walk leaving R is less than 0.01 by (28), we conclude that the algorithm finds the desired vector Γ_T with probability greater than $0.02 - 0.01 = 0.01$, as claimed. \square

Finally, we illustrate how to transform the proof of Proposition 4.1 so as to provide a polynomial time algorithm.

Proposition 6.3. *There exists an algorithm of running time $O(d^7 \ln^2 d)$ which, in the setting of Proposition 4.1, yields the desired selection vector $\mu \in \Delta_U$.*

Proof. Along the course of the proof of Proposition 4.1, we replace the iteration of Lemma 4.2 by that of Lemma 6.1 so as to obtain a vector $\hat{\mu} \in \Delta_U$ such that, for each $i \in [k]$,

$$|\{w_i \in W_i : 0 \leq \hat{\mu}(w_i) \leq \delta\}| = |W_i| - 1. \tag{37}$$

The existence of such a vector is guaranteed as long as $\delta < 1/|W_i|$ for each $i \in [k]$. At the final step, we take μ to be the closest vertex of Δ_W to $\hat{\mu}$, that is, define

$$\mu(w_i) = \begin{cases} 0 & \text{if } \hat{\mu}(w_i) \leq \delta \\ 1 & \text{if } \hat{\mu}(w_i) > \delta. \end{cases}$$

In particular, taking $\chi := \mu - \hat{\mu}$, we have that by (15),

$$|\langle \mu - \lambda, W^j \rangle| = |\langle \hat{\mu} - \lambda, W^j \rangle + \langle \chi, W^j \rangle| \leq 22\sqrt{d} + |\langle \chi, W^j \rangle|$$

for each $j \in [d]$. We show that taking a sufficiently small value of δ ensures that $|\langle \chi, W^j \rangle| \leq O(\sqrt{d})$, accordingly, μ is an appropriate selection vector.

Let $i \in [k]$ be arbitrary, and let $w \in W_i$ be so that $\hat{\mu}(w) > \delta$. Then $\hat{\mu}(w) = 1 - \sum_{w_i \neq w \in W_i} \hat{\mu}(w_i)$. Accordingly,

$$\sum_{w_i \in W_i} |\chi(w_i)| = 2 \sum_{w \neq w_i \in W_i} |\hat{\mu}(w_i)| < 2|W_i| \delta.$$

Therefore, as $\|W^j\|_\infty \leq 1$,

$$|\langle \chi, W^j \rangle| \leq \sum_{i \in [k]} \sum_{w_i \in W_i} |\chi(w_i)| \leq \sum_{i \in [k]} (2|W_i| \delta).$$

Since $|W_i| \leq m \leq 2d$ for each $i \in [k]$ and $k \leq m \leq 2d$, we conclude that for each $j \in [d]$, $|\langle \chi, W^j \rangle| \leq 8d^2\delta$. Thus, fixing

$$\delta = 0.01d^{-3/2}, \tag{38}$$

we indeed obtain

$$\|W\mu - W\lambda\|_\infty = \max_{j \in [d]} |\langle \mu - \lambda, W^j \rangle| \leq 22\sqrt{d} + 8\sqrt{d} = O(\sqrt{d}).$$

Next, we estimate the running time of the algorithm at a given iteration s of Lemma 6.1. As before, let $m(s)$ be the number of active vectors. For a fixed step $t \in [T]$ of the Gaussian random walk, the calculation of the sets C_t^{var}, C_t^{max} takes time $O(d + m(s))$. An orthonormal basis of the subspace S_t may be determined in $O(d^3)$ time by applying a Gram–Schmidt orthogonalization process, and then the Gaussian vector Λ_{t+1} is sampled in $O(m(s))$ time. Hence, the time complexity of performing a given step of the Gaussian walk is dominated by the calculation of the orthonormal basis of S_t .

By (38), (14), and the condition that $m \leq 2d$, the maximal ϵ which satisfies the constraints (23), (24), and (25) simultaneously can be estimated by

$$\frac{1}{\epsilon^2} = O(m(s)d^3 \ln^2 d).$$

Since, by (22), the number of steps of the random walk is $T = O(1/\epsilon^2)$, the above estimate shows that the running time of the algorithm within a given iteration s is

$$O(m(s)d^6 \ln^2 d).$$

Let S be the total number of iterations until reaching the vector $\hat{\mu} \in \Delta_U$ satisfying (37). Note that by (14), we have $m(s) \leq m/2^s$. Therefore, the total running time of the algorithm is

$$\sum_{s=1}^S O\left(\frac{m}{2^s}d^6 \ln^2 d\right) = O(md^6 \ln^2 d) = O(d^7 \ln^2 d). \quad \square$$

7 | REDUCTION TO VECTOR BALANCING

In this section we describe an alternative approach for proving asymptotic estimates for colorful vector balancing constants matching Theorems 1.4 and 1.5, based on the proof techniques of Lovász, Spencer and Vesztergombi [23] and Bansal, Dadush, Garg, and Lovett [7]. For the following proof we denote the colorful vector balancing constant of two symmetric convex bodies $K, L \subseteq \mathbb{R}^d$ as

$$\text{colvb}(K, L) := \sup_{n \geq d} \max_{\substack{V_1, \dots, V_n \subseteq K \\ 0 \in \sum \text{conv } V_i}} \min_{v_i \in V_i, i \in [n]} \|v_1 + \dots + v_n\|_L.$$

Theorem 7.1. *Given any symmetric convex bodies $K, L \subseteq \mathbb{R}^d$,*

$$\text{colvb}(K, L) \leq 2\text{vb}(K, L).$$

Proof. We are given families $V_1, \dots, V_n \subseteq K$ and a vector $\lambda \in \Delta_V$ so that $V\lambda = 0$. Note that by Carathéodory's theorem we may assume that $|V_i| \leq d + 1$ for each $i \in [n]$. Let

$$\rho = \max_{u \in K} \|u\|_L.$$

Fix $\varepsilon > 0$ and take $\ell \in \mathbb{Z}$ so that $n(d + 1)2^{-(\ell-1)}\rho \leq \varepsilon$.

Each coordinate $\lambda(v)$ of λ , for $v \in V_i, i \in [n]$, has a binary expansion, which we truncate at the ℓ th digit after the radix point to obtain the vector μ with coordinates $\mu(v)$ so that $|\mu(v) - \lambda(v)| \leq 2^{-(\ell-1)}$ for each $v \in V_i, i \in [n]$. Then

$$\|V\lambda - V\mu\|_L = \left\| \sum_{i \in [n]} \sum_{v \in V_i} (\lambda(v) - \mu(v))v \right\|_L \leq \frac{1}{2^{\ell-1}} \sum_{i \in [n]} \sum_{v \in V_i} \|v\|_L \leq \frac{n(d + 1)\rho}{2^{\ell-1}} \leq \varepsilon. \tag{39}$$

Denote the j th digit of the binary expansion of $\mu(v)$ by $\mu(v)^{(j)}$. We define the set

$$S_\ell := \{v \in \cup_{i \in [n]} V_i : \mu(v)^{(\ell)} = 1\}$$

to be the set of vectors in our collection for which the ℓ th digit of the binary expansion of the corresponding coefficient is 1. Since $\sum_{v \in V_i} \lambda(v) = 1$ for each $i \in [n]$, it follows that $|S_\ell \cap V_i| = 2q_i$ for some $q_i \in \mathbb{Z}$, so we can write $S \cap V_i = \{v_1^i, \dots, v_{2q_i}^i\}$ for each $i \in [n]$. We define the auxiliary collections of vectors

$$W_i = \left\{ \frac{v_{2j}^i - v_{2j-1}^i}{2} \right\}_{j \in [q_i]} \subseteq K$$

and then balance the collection $W = \cup_{i \in [n]} W_i$, yielding signs $\chi_i(j) \in \{\pm 1\}$ so that

$$\left\| \sum_{i \in [n]} \sum_{j \in [q_i]} \chi_i(j) \frac{v_{2j}^i - v_{2j-1}^i}{2} \right\|_L \leq \text{vb}(K, L).$$

Color the elements of S_ℓ as follows: for each $i \in [n]$, for each $k \in [2q_i]$, we assign $\beta_i(k) = \chi_i(j)$ for k even and $\beta_i(k) = -\chi_i(j)$ for k odd, so that

$$\left\| \sum_{i \in [n]} \sum_{k \in [2q_i]} \beta_i(k)v_k^i \right\|_L = 2 \left\| \sum_{i \in [n]} \sum_{j \in [q_i]} \chi_i(j) \frac{v_{2j}^i - v_{2j-1}^i}{2} \right\|_L \leq 2\text{vb}(K, L). \tag{40}$$

We then update the vector μ as follows: for $v \notin S_\ell, \mu_1(v) = \mu(v)$. For $v \in S_\ell$, we know that $v = v_k^i$ for some $i \in [n], k \in [2q_i]$, and we update $\mu_1(v) := \mu(v) + 2^{-\ell}\beta_i(k)$. By construction, $\mu_1 \in \Delta_V \cap 2^{-(\ell-1)}$, and $\|V\mu_1\| \leq 2^{-(\ell-1)}\text{vb}(K, L)$. Iterating the argument for the successive digits leads to a selection vector μ_ℓ for which

$$\|V\lambda\|_L \leq \|V\lambda - V\mu\|_L + \|V\mu\|_L \leq \left\| \sum_{i=0}^{\ell-1} 2^i \text{vb}(K, L) \right\|_L \leq \varepsilon + 2\text{vb}(K, L).$$

As $\varepsilon > 0$ was arbitrary, the theorem follows. □

Combining Theorem 7.1 with (2) and (4) implies our Theorems 1.4 and 1.5 up to constants. We intended to give a direct proof that is more suited to algorithmic applications. Indeed, the

computational complexity of finding a solution by the above approach depends heavily on the number of vector families n , whereas our technique illuminates the geometric aspects of the problem and the independence of the number of vector families, including the reduction to $O(d)$ total vectors that is necessary in the maximum norm case. Moreover, it leads to the sharp bound of \sqrt{d} for the Euclidean case as opposed to the asymptotic bound above, and it improves on the constant for the maximum norm given by combining [24] with Theorem 7.

ACKNOWLEDGMENTS

The authors thank Imre Bárány and Thomas Rothvoss for their helpful suggestions and the anonymous referee for the helpful comments and pointing out the argument sketched in Section 7. Research of G. Ambrus was partially supported by ERC Advanced Grant “GeoScape,” by the Hungarian National Research grant no. NKFIH KKP-133819, and by project no. TKP2021-NVA-09. Project no. TKP2021-NVA-09 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme. Research of R. Bozzai was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2140004 and Grant No. G20221001-4371 in Aid of Research from Sigma Xi, The Scientific Research Honor Society.

JOURNAL INFORMATION

Mathematika is owned by University College London and published by the London Mathematical Society. All surplus income from the publication of *Mathematika* is returned to mathematicians and mathematics research via the Society’s research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

ORCID

Gergely Ambrus  <https://orcid.org/0000-0003-1246-6601>

REFERENCES

1. G. Ambrus, *A generalization of Bang’s lemma*, Proc. Amer. Math. Soc. **151** (2023), no. 3, 1277–1284.
2. G. Ambrus and B. González Merino, *Large signed subset sums*, Mathematika **67** (2021), no. 3, 579–595.
3. W. Banaszczyk, *Balancing vectors and convex bodies*, Studia Math. **106** (1993), no. 1, 93–100.
4. W. Banaszczyk, *Balancing vectors and Gaussian measures of n -dimensional convex bodies*, Random Structures Algorithms **12** (1998), no. 4, 351–360.
5. N. Bansal, *Constructive algorithms for discrepancy minimization*, 2010 IEEE 51st Annu. Sympos. Found. Comput. Sci., Las Vegas, NV, 2010, pp. 3–10.
6. N. Bansal, D. Dadush, and S. Garg, *An algorithm for Koplós conjecture matching Banaszczyk’s bound*, SIAM J. Comput. **48** (2019), no. 2, 534–553.
7. N. Bansal, D. Dadush, S. Garg, and S. Lovett, *The Gram–Schmidt walk: a cure for the Banaszczyk blues*, Theory Comput. **15** (2019), no. 21, 1–27.
8. N. Bansal, H. Jiang, S. Singla, and M. Sinha, *Online vector balancing and geometric discrepancy*, Proc. 52nd Annu. ACM SIGACT Sympos. Theory Comput., Chicago, IL, 2020, pp. 1139–1152.
9. I. Bárány, *On the power of linear dependencies*, Building bridges: between mathematics and computer science, vol. 19, Bolyai Society Mathematical Studies, Bolyai Society, Budapest, 2010, pp. 31–45.
10. I. Bárány and B. Doerr, *Balanced partitions of vector sequences*, Linear Algebra Appl. **414** (2006), no. 2–3, 464–469.
11. I. Bárány and V. Grinberg, *On some combinatorial questions in finite dimensional spaces*, Linear Algebra Appl. **41** (1981), 1–9.
12. J. Beck and T. Fiala, *“Integer-making” theorem*, Discrete Appl. Math. **3** (1981), no. 1, 1–8.

13. P. Billingsley, *Probability and measure*, 3rd ed., Wiley, Hoboken, NJ, 1995.
14. D. Dadush, A. Nikolov, K. Talwar, and N. Tomczak-Jaegermann, *Balancing vectors in any norm*, 2018 IEEE 59th Annu. Sympos. Found. Comput. Sci. (FOCS), IEEE, Piscataway, NJ, 2018, pp. 1–10.
15. A. Dvoretzky, *Problem*, Convexity, vol. 7, Proc. Sympos. Pure Math., American Mathematical Society, Providence, RI, 1963, pp. 496.
16. W. Feller, *An introduction to probability theory and its applications*, vol. 2, Wiley, Hoboken, NJ, 1971.
17. A. A. Giannopoulos, *On some vector balancing problems*, Studia Math. **122** (1997), 225–234.
18. E. D. Gluskin, *Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces*, Mathematics of the USSR-Sbornik **64** (1989), no. 1, 85.
19. V. Grinberg and S. V. Sevast'yanov, *Value of the Steinitz constant*, Funct. Anal. Appl. **14** (1980), no. 2, 125–126.
20. C. Harshaw, F. Sävje, D. A. Spielman, and P. Zhang, *Balancing covariates in randomized experiments with the Gram–Schmidt walk design*, J. Amer. Statist. Assoc. (2024), 1–13.
21. Z. Karnin and E. Liberty, *Discrepancy, coresets, and sketches in machine learning*, in A. Beygelzimer and D. Hsu (eds.), Proc. Thirty-Second Conf. Learn. Theory, vol. 99, Proc. Mach. Learn. Res. (PMLR), Long Beach, CA, 2019, pp. 1975–1993.
22. B. S. Kashin, *On one isometric operator in $L^2(0, 1)$* , C. R. Acad. Bulgare Sci. **38** (1985), no. 12, 1613–1615.
23. L. Lovász, J. Spencer, and K. Vesztegombi, *Discrepancy of set-systems and matrices*, European J. Combin. **7** (1986), no. 2, 151–160.
24. S. Lovett and R. Meka, *Constructive discrepancy minimization by walking on the edges*, SIAM J. Comput. **44** (2015), no. 5, 1573–1582.
25. B. Lund and A. Magazinov, *The sign-sequence constant of the plane*, Acta Math. Hungar. **151** (2017), 117–123.
26. J. Matousek, *Geometric discrepancy: an illustrated guide*, vol 18, Springer Science & Business Media, Berlin, 1999.
27. T. Oertel, J. Paat, and R. Weismantel, *A colorful Steinitz lemma with application to block-structured integer programs*, Math. Program. **204** (2024), no. 1–2, 677–702.
28. V. Reis and T. Rothvoss, *Vector balancing in Lebesgue spaces*, Random Structures Algorithms **62** (2022), 667–688.
29. R. Schneider, *Convex bodies: the Brunn–Minkowski theory*, 2nd ed., Encyclopedia of Mathematics and Its Applications, Cambridge University Press, Cambridge, 2013.
30. S. V. Sevast'yanov, *On the approximate solution of the problem of calendar planning*, Upravlyaemye Syst. **20** (1980), 49–63 (in Russian).
31. S. V. Sevast'yanov, *On some geometric methods in scheduling theory: a survey*, Discrete Appl. Math. **55** (1994), no. 1, 59–82.
32. S. V. Sevast'yanov, *Seven problems: so different yet close*, Algorithms — ESA'97: 5th Ann. Eur. Symp. Graz, Austria, September 15–17, 1997 Proceedings, Springer, Berlin, 2005, pp. 443–458.
33. J. Spencer, *Balancing games*, J. Combin. Theory Ser. B **23** (1977), no. 1, 68–74.
34. J. Spencer, *Balancing unit vectors*, J. Combin. Theory **30** (1981), 349–350.
35. J. Spencer, *Six standard deviations suffice*, Trans. Amer. Math. Soc. **289** (1985), no. 2, 679–706.
36. J. Spencer, *Balancing vectors in the max norm*, Combinatorica **6** (1986), 55–65.
37. J. Spencer, *Ten lectures on the probabilistic method*, SIAM, 1994.
38. R. M. Starr, *Quasi-equilibria in markets with non-convex preferences (Appendix 2: the Shapley–Folkman theorem, pp. 35–37)*, Econometrica **37** (1969), 25–38.
39. K. Swanepoel, *Balancing unit vectors*, J. Combin. Theory **89** (2000), 105–112.
40. R. Vershynin, *High-dimensional probability: an introduction with applications in data science*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2018.