The Online Steinitz Problem

Matteo Russo* Jiaye Wei*

Abstract

Inspired by Steinitz's celebrated result (1913) and recent online vector-balancing results of Kulkarni et al. (2024) and Aden-Ali (2026), we introduce the *online Steinitz problem*: T vectors in $\mathbb{R}^d$, each of ℓ_2-norm at most 1 and with total sum zero, arrive one at a time and must be assigned immediately and irrevocably to unused positions of an array. The goal is to keep every prefix sum of the resulting order small. We first obtain deterministic prefix-discrepancy bounds of $O(\log T)$ for $d = 1$ and $O(\sqrt{T})$ for $d \geq 2$. Our main contribution is a general reduction that converts deterministic and randomized online vector-balancing algorithms, which assign signs to arriving vectors, into online ordering algorithms. Combined with recent vector-balancing results, this yields a randomized algorithm with prefix discrepancy $O(\log T + \sqrt{d \log T})$ and running time $O(dT \log T)$. The guarantee holds with high probability against an oblivious adversary. The reduction also gives bounds for sparse vectors and extends to inputs in arbitrary compact convex bodies. Finally, we prove matching lower bounds in both the deterministic and randomized cases.

1 Introduction

The classical Steinitz lemma asks how to order a zero-sum family of bounded vectors so that every prefix sum remains small [Ste13]. For vectors $v_1, \ldots, v_T \in \mathbb{R}^d$ with $\|v_t\|_2 \leq 1$ and $\sum_{t=1}^{T} v_t = 0$, the goal is to prove there exists an ordering distributes cancellation throughout the sequence instead of postponing it until the end. Indeed, Steinitz [Ste13] proved that there is a permutation $\pi : [T] \to [T]$ such that

$$\max_{0 \leq k \leq T} \left\| \sum_{j=1}^{k} v_{\pi(j)} \right\|_2 \leq d.$$

Establishing an $O(\sqrt{d})$ upper bound for the Euclidean Steinitz constant has been a long-standing open problem. On 6 October 2026, OpenAI announced a proof of this bound [Ope26], which, if confirmed, would determine the constant depending on d up to absolute factors.[1] Beyond the importance of the mathematical question itself, the principle of the Steinitz lemma serves as a standard tool in discrepancy theory and integer programming, e.g., [Cho94, GS80, EW20].

We introduce the *online Steinitz problem* (OSP). In this problem, there are T initially empty output positions, and vectors $v_1, \ldots, v_T$ that arrive sequentially with $\|v_t\|_2 \leq 1$ and $\sum_{t=1}^{T} v_t = 0$, identically to the offline case. Upon seeing v_t, the algorithm must assign it immediately and irrevocably to an unused position $p_t \in [T]$, thereby building a permutation π on the fly such that $\pi(p_t) = t$. The completed output is therefore $v_{\pi(j)}$ and its *prefix discrepancy* by the end of the assignment is $\max_{0 \leq k \leq T} \| \sum_{j=1}^{k} v_{\pi(j)} \|_2$. Although the prefix discrepancy is an identical measure in both offline and the online case, the requirement of an immediate and irrevocable placement and the limited future knowledge leads to a substantially harder algorithmic task that we seek to investigate:

*EPFL, Switzerland, {matteo.russo, jiaye.wei}@epfl.ch

[1] The authors have not independently verified the announced proof, which none of this paper's results requires.

Question. *What is the worst-case online prefix discrepancy for immediate and irrevocable vector decisions, and how does this compare with the offline optimum?*

We study this question under the *oblivious adversarial* model, which fixes the entire input independently of the algorithm's randomness, and also under the *adaptive adversarial* model, which may choose each vector after observing the previous assignments. This means that, against the former kind of adversary, randomness may help, while against the latter it does not. We therefore write $\mathrm{OS}_{\mathrm{rand}}(d, T, \delta)$ for the smallest discrepancy bound that a randomized algorithm can guarantee with probability at least $1 - \delta$ on every fixed input. The corresponding optimal deterministic worst-case bound against adaptive inputs is denoted by $\mathrm{OS}_{\mathrm{det}}(d, T)$.

The problem is reminiscent of *online vector balancing* (OVB) [BJSS20, ALS21, KRR24, Ade26]: in that case, an OVB algorithm chooses a sign for each arriving vector and controls signed prefix sums in arrival order. On the other hand, an OSP algorithm chooses among all vacant positions, and one placement may affect many output prefixes, which are evaluated only after the stream ends, and not after each arrival. We defer the reader to the survey of Bansal and Nikolov [BN26] for an extensive treatment of algorithmic discrepancy in its various forms, including offline and online.

1.1 Our results

Despite their differences, our main upper-bound contribution is a general reduction from OSP to OVB. We first describe this general reduction, followed by its deterministic and randomized applications, further corollaries, and matching lower bounds.

A general reduction. In Section 2, we present a black-box reduction from online vector balancing to online Steinitz ordering. The reduction builds an OSP *router* from OVB subroutines, which we call *signers*. Its ordering guarantee is stated in terms of the signers' final vector and count bounds, or alternatively their subgaussian estimates. As a result, the same routing framework can use a range of deterministic or randomized signers, and any improvement in their balancing performance immediately yields stronger ordering bounds. The reduction also applies when the total sum of the input vectors is arbitrary and unknown in advance, and it ensures that every prefix remains close to its proportional share of this total. In addition, if the inputs lie in a known compact convex body K of affine dimension s, then any Euclidean proportional-prefix bound ρ implies that the residuals lie in $(s\rho/2)(K - K)$, where, as usual, $K - K = \{x - y : x, y \in K\}$. This extension does not require K to be symmetric, and it preserves deterministic and probabilistic guarantees under the same adversarial model.

Deterministic bounds. Using the OSP router given by the reduction instantiated with deterministic OVB signers, in Section 3, we show that the optimal deterministic discrepancy is

$$\mathrm{OS}_{\mathrm{det}}(d, T) \in \begin{cases} O(\log T), & d = 1, \\ O(\sqrt{T}), & d \geq 2. \end{cases}$$

As mentioned previously, both upper bounds hold against adaptive adversaries, and the higher-dimensional lower bound already holds in a fixed plane.

Tight bounds against an oblivious adversary. In Section 4, we instantiate the general reduction with Aden-Ali's randomized OVB signer [Ade26] and obtain a prefix discrepancy of

$$\mathrm{OS}_{\mathrm{rand}}(d, T, \delta) \in O\left(\log T + \sqrt{(d + \log(1/\delta)) \log T}\right), \tag{1.1}$$

with probability at least $1-\delta$. This guarantee is achieved efficiently, with the overall procedure having a running time of $O(dT\log T)$, and furthermore yields a guarantee of $O\left(\log T+\sqrt{\log T\cdot\log(1/\delta)}\right)$ in the ℓ_∞-norm. Furthermore, we can instantiate our reduction with a different OVB signer for zero-sum inputs in $[-1,1]^d$ with at most κ nonzero coordinates, and ℓ_2-norm as large as $\sqrt{\kappa}$. Specifically, the signer of Altschuler–Tikhomirov [AT26] gives $O(\sqrt{\kappa}\log T)$ coordinatewise discrepancy whenever the failure bound in Corollary 4.4 is at most δ.

Matching lower bounds. In Section 5, all hard inputs consist of an initial stream followed by identical vectors that cancel its total. An all-ones initial stream and Tijdeman–Wagner [TW80] give an $\Omega(\log T)$ lower bound; coordinate cycles and Roth's method [Rot54] give $\Omega(\sqrt{d_*\log(2+T/d_*)})$, where $d_*=\min\{d,T\}$. Both hold for every $0<\delta\le 1/2$. Their sum matches the upper bound of (4.3) whenever $T^{-\beta}\le\delta\le 1/2$ for fixed $\beta>0$. An adaptive initial stream in a fixed plane gives the matching deterministic $\Omega(\sqrt{T})$ lower bound for $d\ge 2$.

1.2 Related work

The Steinitz lemma and its extensions. In the offline setting, the classical Steinitz lemma [Ste13] guarantees that a zero-sum family of bounded vectors admits an ordering whose prefix sums are bounded independently of the number of vectors. Grinberg and Sevastyanov [GS80] established the bound d in every d-dimensional normed space. For the ℓ_2-norm, improving the dimension dependence to $O(\sqrt{d})$ remains a longstanding conjecture, with a matching $\Omega(\sqrt{d})$ lower bound. Recent work of Ambrus and Heck [AH26] relates this problem to ordering nearly unit vectors with arbitrary total sum, measuring each prefix against its proportional share of that total. Another direction concerns simultaneous rearrangements: Oertel et al. [OPW24] proved a colorful Steinitz lemma that permutes several sequences while controlling their combined prefixes independently of the number of sequences. Bárány [Bár24] improved their $40d^5$ bound to $4d-2$. Steinitz lemma also has algorithmic applications beyond discrepancy, as for instance, the work of Eisenbrand and Weismantel [EW20] which used it to establish proximity bounds and faster algorithms for integer programming.

Online vector balancing and scheduling. Online vector balancing replaces rearrangement by sequential sign choices. Bansal et al. [BJSS20] studied stochastic inputs drawn independently from an arbitrary distribution and obtained applications to geometric discrepancy. For oblivious adversarial inputs, Alweiss et al. [ALS21] introduced the self-balancing walk, giving efficient algorithms with logarithmic ℓ_∞ prefix discrepancy for Euclidean unit-ball vectors. Kulkarni et al. [KRR24] subsequently obtained a universal subgaussian guarantee for each signed prefix. This implies $O(\sqrt{\log T})$ maximum ℓ_∞ prefix discrepancy with high probability, matching their lower bound against oblivious adversaries. Aden-Ali [Ade26] obtained the same asymptotic guarantee with an $O(dT)$-time algorithm. For coordinate-bounded vectors with at most κ nonzero coordinates, Altschuler and Tikhomirov [AT26] obtained $O(\sqrt{\kappa})$ online ℓ_∞ prefix discrepancy above a near-logarithmic sparsity threshold. Online vector scheduling [IKKP19] likewise requires irrevocable allocation of arriving vectors, but distributes nonnegative demand vectors among machines to optimize their loads, whereas OSP assigns vectors to ordered positions and exploits cancellation to control all prefix sums.

Signing, rearrangement, and geometric discrepancy. The connection between signing and rearrangement predates their online versions. Bárány and Grinberg [BG81] proved that every prescribed sequence of unit vectors in a d-dimensional normed space admits offline signs for which all signed prefixes have norm at most $2d-1$. Chobanyan's transference theorem [Cho94] shows that a uniform bound for this

offline prefix-signing problem also bounds the Steinitz constant, without increasing the bound. However, this does not transfer to the online setting as argued in Section 2.5. Controlling only the final signed sum is a different problem. For example, Ambrus and Bozzai [AB24] proved sharp Euclidean bounds for colorful vector balancing, where one selects a vector from each prescribed family and controls the sum of the selected vectors. Our lower bounds use another connection to discrepancy: points record arrival times and assigned positions. Schmidt's planar rectangle-discrepancy theorem [Sch72] gives the deterministic scalar bound; Tijdeman–Wagner's strengthening [TW80] and a restricted form of Roth's orthogonal-function argument [Rot54] yield the oblivious bounds. For a broader account of these topics and their algorithmic connections, see the monograph manuscript of Bansal and Nikolov [BN26].

Online sorting and online TSP. Our placement model is also closely related to online sorting [AABK23] and its extension to online TSP [ABB+24, Ber25]: arriving numbers or points must be assigned immediately and irrevocably to unused array positions. It is worth noting that the work of Azar et al. [APV26] further studies the tradeoff between array size and competitiveness for online metric TSP, obtaining a deterministic $O(\log^3 T)$-competitive algorithm with $2T$ cells. Their placement constraint coincides with ours when the array has exactly T cells, but their objective controls distances between consecutive entries rather than cumulative vector sums. Their tree-based placement scheme organizes points by metric proximity, whereas our reduction uses online signing to balance the sums routed to subintervals. More generally, these problems minimize the total distance between consecutive entries, whereas OSP controls the norm of every prefix sum. Geometrically, their inputs are points to connect by a short path, while ours are steps to arrange into a walk that stays close to the origin through cancellation.

2 The signing-to-ordering reduction

The reduction repeatedly divides the output positions into two nearly equal intervals. At each division, an online signer proposes a child for the arriving vector, and the router can only override this choice when it has to avoid an interval that is already full. We first describe this algorithm, then state the signing guarantees that suffice and prove how its local splits control every output prefix.

2.1 The ordering algorithm

We first introduce some notation, beginning with the definition of residuals.

Residuals. For the completed ordering $v_\pi = (v_{\pi(1)}, \dots, v_{\pi(T)})$, write $\mathrm{tot} := \sum_{t=1}^{T} v_t$ and define the prefix residual for each $k \in [T]$ and the maximum residual by

$$\mathrm{res}_k(v_\pi) := \sum_{j=1}^{k} v_{\pi(j)} - \frac{k}{T}\mathrm{tot}, \qquad \mathrm{res}_{\max}(v_\pi) := \max_{0 \le k \le T} \|\mathrm{res}_k(v_\pi)\|_2. \tag{2.1}$$

These residuals measure deviations from proportional shares of the total and become ordinary prefix sums when $\mathrm{tot} = 0$. Hence, for some upper bound $U \ge 0$ achieved deterministically or in high probability, we have that, if $\mathrm{res}_{\max} \le U$ then $\mathrm{OS} \le U$, since $\mathrm{tot} = 0$. We next describe the position tree and the meta-algorithm needed for our reduction.

The position tree. Our algorithm assigns each arriving vector to an output position by routing it through a balanced binary tree. This tree has the output positions as its leaves, and so has height $h = \lceil \log_2 T \rceil$. Each node u represents a consecutive interval $I_u \subseteq [T]$ which consists of the output positions labeling the

Algorithm 1 Tree routing with local proposal rules

Require: Horizon T, local routines Sign_u, and vectors $v_1, \dots, v_T$ arriving one at a time.
Ensure: An online permutation π, with $\pi(p_t) = t$.

Build the balanced position tree with capacities $m_u = |I_u|$.
for each internal node u **do**
  Initialize Sign_u once.
  for $c \in \{u_L, u_R\}$ **do**
    $N_c \leftarrow 0, \quad X_c \leftarrow 0.$
for $t = 1, \dots, T$ **do**
  Receive v_t and set $u \leftarrow$ root.
  while u is not a leaf **do**
    $\varepsilon \leftarrow \mathsf{Sign}_u(v_t; X_{u_L}, X_{u_R}, m_{u_L} - N_{u_L}, m_{u_R} - N_{u_R}).$
    $\widehat{\varepsilon} \leftarrow \varepsilon.$ ▷ Proposed versus actual routing sign
    if $N_{u_L} = m_{u_L}$ **then**
      $\widehat{\varepsilon} \leftarrow -1.$
    else if $N_{u_R} = m_{u_R}$ **then**
      $\widehat{\varepsilon} \leftarrow +1.$
    if $\widehat{\varepsilon} = +1$ **then**
      $c \leftarrow u_L.$
    else
      $c \leftarrow u_R.$
    $N_c \leftarrow N_c + 1, \quad X_c \leftarrow X_c + v_t.$
    $u \leftarrow c.$
  $p_t \leftarrow \min I_u, \quad \pi(p_t) \leftarrow t.$

leaves in the subtree rooted at u. The node u has capacity $m_u := |I_u|$ with children u_L, u_R of capacities $\lceil m_u/2 \rceil$ and $\lfloor m_u/2 \rfloor$, respectively.

As illustrated in Algorithm 1 and Fig. 1, each internal node uses a persistent local signer Sign_u to propose $+1$ (left) or -1 (right). The router follows this proposal unless the requested child is full, in which case it uses the other child. Repeating the decision until a leaf is reached assigns the vector's output position immediately. During the descent, the algorithm records in N_c how many positions are already filled in child c, and in X_c the sum of the corresponding vectors. The arguments of Sign_u are the arriving vector v_t, the child sums X_{u_L}, X_{u_R}, and the remaining capacities $m_{u_L} - N_{u_L}, m_{u_R} - N_{u_R}$. Here the semicolon separates v_t from the routing statistics. In particular, the scalar rule in Section 3 uses these statistics directly, while, for a Euclidean signing subroutine, the node submits the augmented vector $(v_t, 1)/\sqrt{2}$, so the extra coordinate balances left and right counts along with the original vector coordinates.

Each local signer receives every vector reaching its node, even after one child fills, and updates its private state with its proposed sign. We stress that only the router overrides the destination. Moreover, we also note that randomized copies have independent tapes, independent of the input, and different rules may be used at different node sizes.

2.2 The reduction guarantee

To state the guarantee of Algorithm 1, we first identify which local splits contribute to the discrepancy of a given output prefix. Since the node intervals are consecutive, a prefix cuts through at most one interval at each depth. For each prefix $[k]$, where $0 \le k \le T$, the corresponding *boundary path* $P(k)$ consists of the internal nodes whose intervals meet both $[k]$ and $[T] \setminus [k]$, with $|P(k)| \le h$ and $P(0) = P(T) = \varnothing$.

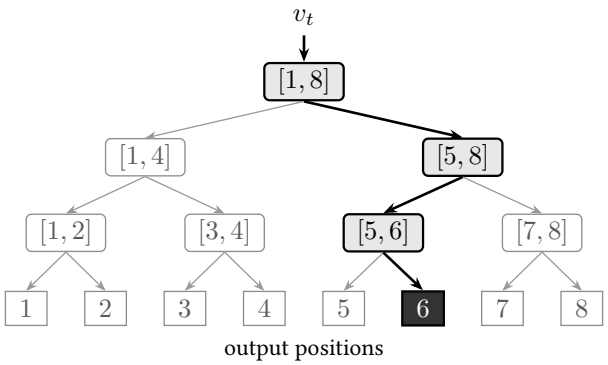


Figure 1: Routing an arriving vector v_t when $T = 8$. At each visited internal node, the local rule selects one of the two child intervals. The highlighted path assigns v_t to output position 6.

Every ancestor of a node in $P(k)$ also belongs to $P(k)$, hence these nodes form a path which is ordered by increasing depth. We will show that the prefix residual $\mathrm{res}_k(v_\pi)$ is a weighted sum of the split errors along this path. The reduction therefore depends on how well the local signers control these errors, which we quantify next.

Signers. An online signer returns $\varepsilon_i \in \{-1, +1\}$ upon receiving z_i. We say it has subgaussian scale $\Sigma \in [0, \infty)$ if, for every fixed stream $z_1, \ldots, z_m$ in the Euclidean unit ball B_2^D and every $\theta \in \mathbb{R}^D$,

$$\mathbb{E} \exp\left(\left\langle \theta, \sum_{i=1}^{m} \varepsilon_i z_i \right\rangle\right) \le \exp\left(\frac{\Sigma^2}{2}\|\theta\|_2^2\right). \tag{2.2}$$

The expectation is over its randomness, independent of the input, which gives Gaussian-type tails in every direction. Our randomized reduction uses a common scale Σ in dimension $d + 1$ for all node capacities, and needs only this final signed-sum guarantee, rather than a guarantee for every signed prefix.

The following theorem separates the property needed from the signer from the resulting ordering guarantee. Terminal vector and count bounds give a sum of local errors, while a subgaussian bound permits a sharper estimate for their combined effect.

Theorem 2.1 (Signing-to-ordering reduction)**.** *Algorithm 1 assigns every vector immediately to an unused position and does not need to know the input total. Randomized local signers use independent tapes, independent of the input.*

(i) ***Terminal vector and count bounds.*** *Suppose the input vectors have norm at most one in a fixed norm $\|\cdot\|$. At a node receiving $x_1, \ldots, x_m$, suppose the signer's proposed signs satisfy*

$$\left\|\sum_{i=1}^{m} \varepsilon_i x_i\right\| \le V(m), \qquad \left|\sum_{i=1}^{m} \varepsilon_i\right| \le W(m).$$

Whenever these inequalities hold at all internal nodes, every output prefix satisfies

$$\|\mathrm{res}_k(v_\pi)\| \le \sum_{u \in P(k)} \left(\frac{V(m_u) + W(m_u)}{2} + \mathbb{1}\{m_u \text{ odd}\}\right). \tag{2.3}$$

If they hold for every local stream, this is a deterministic guarantee against adaptive inputs. If, for each fixed length-m stream, they hold jointly with probability at least $1 - p(m)$, then for every oblivious input (2.3) holds simultaneously for all prefixes with probability at least $1 - \sum_u p(m_u)$.

(ii) ***Subgaussian terminal sums.*** *Suppose the inputs are in the Euclidean unit ball and the signer has scale $\Sigma \geq 1$ in the sense of (2.2), in dimension $d+1$ and for every node capacity. Then, for every oblivious input and $0 < \delta < 1$, with probability at least $1-\delta$,*

$$\mathrm{res}_{\max}(v_\pi) \leq C\Sigma\left(\log T + \sqrt{\log T\big(d + \log(1/\delta)\big)}\right), \tag{2.4}$$

where C is an absolute constant.

Both parts require only the terminal *guarantee for each local stream. For zero-sum inputs, the residuals are the ordinary prefix sums.*

Part (i) will give the deterministic Euclidean and sparse-vector bounds. Part (ii) will give the randomized Euclidean bound without paying a separate norm bound at every node. The sharper dimension-dependent bound follows by using deterministic routing near the leaves.

2.3 From local splits to output prefixes

We prove Theorem 2.1 by first relating the output prefixes to the errors of the local splits. The identity in Lemma 2.2 shows that each prefix depends only on its boundary path; Lemma 2.3 then bounds each split error in terms of the signer's proposed signs and the corrections needed to enforce capacities. Combining these two facts with the triangle inequality will prove part (i). For part (ii), we will use the conditional subgaussian estimates to bound the sum along the path.

Before continuing, we define quantities only needed in the analysis. Namely, we let $\mathrm{anc}(u)$ be the set of strict ancestors of u in the tree and, throughout, conditioning on $\mathrm{anc}(u)$ means conditioning on the complete random tapes of all nodes in $\mathrm{anc}(u)$. Furthermore, for the completed ordering,

$$\mathrm{tot}_u := \sum_{j \in I_u} v_{\pi(j)}, \qquad \mathrm{err}_u := \mathrm{tot}_{u_L} - \frac{m_{u_L}}{m_u}\mathrm{tot}_u. \tag{2.5}$$

Thus err_u is the left child's error relative to its proportional share of the parent total, and the right child's error is $-\mathrm{err}_u$. To track the same discrepancy within an interval, for $0 \leq s \leq m_u$ we set

$$F_u(s) := \sum_{j=\min I_u}^{\min I_u + s - 1} v_{\pi(j)} - \frac{s}{m_u}\mathrm{tot}_u.$$

This is the deviation of the first s positions in I_u from their proportional share of tot_u, so $F_{\mathrm{root}}(k) = \mathrm{res}_k(v_\pi)$ and $F_u(0) = F_u(m_u) = 0$. For $u \in P(k)$, we also set

$$\alpha_{u,k} := \min\left\{\frac{|I_u \cap [k]|}{m_{u_L}}, \frac{|I_u \setminus [k]|}{m_{u_R}}\right\} \in [0,1].$$

When the prefix ends in the left child, this coefficient is the fraction of that child included in the prefix; when it ends in the right child, it is the fraction of that child excluded from the prefix. At the boundary between the children, it equals one. The following identity shows that these fractions are exactly the weights with which the local split errors contribute to the prefix residual.

Lemma 2.2 (one-boundary-path identity)**.** *For every output prefix k, the boundary path and coefficients above satisfy*

$$\mathrm{res}_k(v_\pi) = \sum_{u \in P(k)} \alpha_{u,k}\mathrm{err}_u. \tag{2.6}$$

Proof. The cases $k = 0, T$ are immediate. At a node $u \in P(k)$, write $s = |I_u \cap [k]|$. If $s \le m_{u_L}$, the local prefix lies entirely in the left child, and adding and subtracting $(s/m_{u_L})\mathrm{tot}_{u_L}$ gives

$$F_u(s) = F_{u_L}(s) + \frac{s}{m_{u_L}}\mathrm{tot}_{u_L} - \frac{s}{m_u}\mathrm{tot}_u = F_{u_L}(s) + \frac{s}{m_{u_L}}\mathrm{err}_u.$$

Here $\alpha_{u,k} = s/m_{u_L}$, since $(m_u - s)/m_{u_R} \ge 1$. If instead $s = m_{u_L} + t$ with $0 \le t \le m_{u_R}$, the prefix contains the whole left child and the first t positions of the right child, so

$$\begin{aligned} F_u(m_{u_L} + t) &= F_{u_R}(t) + \mathrm{tot}_{u_L} + \frac{t}{m_{u_R}}\mathrm{tot}_{u_R} - \frac{m_{u_L} + t}{m_u}\mathrm{tot}_u \\ &= F_{u_R}(t) + \mathrm{err}_u + \frac{t}{m_{u_R}}\left(\mathrm{tot}_{u_R} - \frac{m_{u_R}}{m_u}\mathrm{tot}_u\right) = F_{u_R}(t) + \left(1 - \frac{t}{m_{u_R}}\right)\mathrm{err}_u. \end{aligned}$$

In this case $\alpha_{u,k} = 1 - t/m_{u_R}$, since $(m_{u_L} + t)/m_{u_L} \ge 1$. Starting with $F_{\mathrm{root}}(k) = \mathrm{res}_k(v_\pi)$, each substitution therefore adds the term $\alpha_{u,k}\mathrm{err}_u$ and leaves only the residual in the child crossed by the prefix boundary. More explicitly, when the recursion reaches a node u, it has established

$$\mathrm{res}_k(v_\pi) = \sum_{\substack{v \in P(k) \\ v \text{ a strict ancestor of } u}} \alpha_{v,k}\mathrm{err}_v + F_u(|I_u \cap [k]|).$$

The formulas above preserve this identity as we descend to the unique child that meets both the prefix and its complement. If the prefix ends between the two children, then $s = m_{u_L}$ and the remaining residual is $F_{u_L}(m_{u_L}) = 0$; a residual at a leaf is also zero. Thus the recursion visits exactly the nodes of $P(k)$ and leaves no residual term, proving (2.6). □

The path identity reduces the ordering problem to controlling err_u at each node. The signer supplies a proposed split, whereas err_u comes from the split after capacities have been enforced. The next lemma connects the two: it expresses err_u as a centered signed sum plus a correction controlled by the signer's count imbalance. It also records the conditional estimates needed later to combine these errors along a path.

Lemma 2.3 (one-node quota repair). *Let u be an internal node of capacity m, receiving Euclidean unit-ball vectors $x_1, \dots, x_m$, and write $\mu = m^{-1}\sum_i x_i$. Let ε_i be the proposed signs and $\hat{\varepsilon}_i$ the quota-corrected routing signs, where $+1$ means left, and put $q = \sum_i \varepsilon_i, Y = \sum_i \varepsilon_i(x_i - \mu)$ for succinctness. Then the local proportional error of (2.5) has a decomposition*

$$\mathrm{err}_u = \frac{1}{2}Y + R, \qquad \|R\|_2 \le |q| + 1.$$

Under the subgaussian hypothesis of Theorem 2.1, for a fixed input, conditionally on the routing tapes at the strict ancestors of u, for every unit vector $z \in \mathbb{R}^d$ and every $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda\langle z, Y\rangle/2} \mid \textit{strict ancestors}\right] \le e^{\Sigma^2\lambda^2/2}, \tag{2.7}$$

$$\mathbb{E}\left[e^{\lambda q} \mid \textit{strict ancestors}\right] \le e^{\Sigma^2\lambda^2}. \tag{2.8}$$

Proof. We first count the changes made by quota repair. Since the children receive exactly their prescribed numbers of vectors, the corrected signs satisfy $\sum_i \hat{\varepsilon}_i = \mathbb{1}\{m \text{ odd}\}$. Once a proposal is overridden, one child is full, so every later override changes a proposal in the same direction. Each change alters the signed count by two, and comparing the proposed count q with the corrected count $\mathbb{1}\{m \text{ odd}\}$ therefore gives

$$\text{number of overrides} = \frac{1}{2}|q - \mathbb{1}\{m \text{ odd}\}|. \tag{2.9}$$

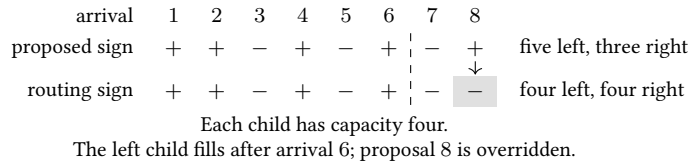


Figure 2: Enforcing capacities without changing the signer's internal decisions. The proposed signed count is 2, the required count is 0, and exactly one sign is overridden.

Fig. 2 illustrates this identity for even m and $q = 2$. To translate this into a bound on the split error, we simply use $m_{u_L}/m = (1 + \mathbb{1}\{m \text{ odd}\}/m)/2$ and $\mathrm{tot}_u = m\mu$ to write

$$\mathrm{err}_u = \frac{1}{2}\left(\sum_i \hat{\varepsilon}_i x_i - \mathbb{1}\{m \text{ odd}\}\mu\right) = \frac{1}{2}\sum_i \hat{\varepsilon}_i(x_i - \mu) = \frac{1}{2}Y + \underbrace{\frac{1}{2}\sum_i (\hat{\varepsilon}_i - \varepsilon_i)(x_i - \mu)}_{:=R},$$

Each overridden sign contributes one vector $\pm(x_i - \mu)$, whose norm is at most two because both x_i and μ lie in the unit ball. Together with (2.9), this yields $\|R\|_2 \le |q - \mathbb{1}\{m \text{ odd}\}| \le |q| + 1$.

We now turn to the conditional estimates, and fix the complete tapes of the strict ancestors of u. The original input is fixed, and ancestor routing never consults descendant signers, so these tapes determine the entire stream reaching u, while its private tape remains independent with its original distribution. We can therefore apply (2.2) to the augmented stream $(x_i, 1)/\sqrt{2}$, whose terminal signed sum is $(\sum_i \varepsilon_i x_i, q)/\sqrt{2}$. Taking $\theta = (\lambda/\sqrt{2})(z, -\langle z, \mu\rangle)$ gives

$$\left\langle \frac{\lambda}{\sqrt{2}}(z, -\langle z, \mu\rangle),\ \frac{1}{\sqrt{2}}\left(\sum_i \varepsilon_i x_i, q\right)\right\rangle = \frac{\lambda}{2}\langle z, Y\rangle, \qquad \|\theta\|_2^2 = \frac{\lambda^2}{2}\left(1 + \langle z, \mu\rangle^2\right) \le \lambda^2,$$

which substituted in (2.2) proves (2.7). Similarly, we may choose $\theta = (0, \dots, 0, \sqrt{2}\lambda)$, which has squared norm $2\lambda^2$ and inner product λq with the augmented signed sum, giving (2.8). □

We thus turn to the proof of the terminal vector and count bounds in Theorem 2.1.

Proof of Theorem 2.1 (i). First, note that the same quota calculation applies in any norm. Moreover, with the notation of Lemma 2.3, each override changes the signed vector sum by a vector of norm at most two. Hence (2.9) and the two terminal bounds assumed in part (i) of the statement give

$$\left\|\sum_i \hat{\varepsilon}_i x_i\right\| \le \left\|\sum_i \varepsilon_i x_i\right\| + |q - \mathbb{1}\{m \text{ odd}\}| \le V(m) + W(m) + \mathbb{1}\{m \text{ odd}\}.$$

We now have

$$\|\mathrm{err}_u\| = \left\|\frac{1}{2}\left(\sum_i \hat{\varepsilon}_i x_i - \frac{\mathbb{1}\{m \text{ odd}\}}{m}\mathrm{tot}_u\right)\right\| \le \frac{V(m) + W(m)}{2} + \mathbb{1}\{m \text{ odd}\}, \tag{2.10}$$

where the equality holds by definition of the exact proportional error err_u and the inequality by the fact that $\|\mathrm{tot}_u/m\| \le 1$. Applying the triangle inequality to Lemma 2.2 now gives

$$\|\mathrm{res}_k(v_\pi)\| \le \sum_{u \in P(k)} \alpha_{u,k}\|\mathrm{err}_u\| \le \sum_{u \in P(k)} \left(\frac{V(m_u) + W(m_u)}{2} + \mathbb{1}\{m_u \text{ odd}\}\right),$$

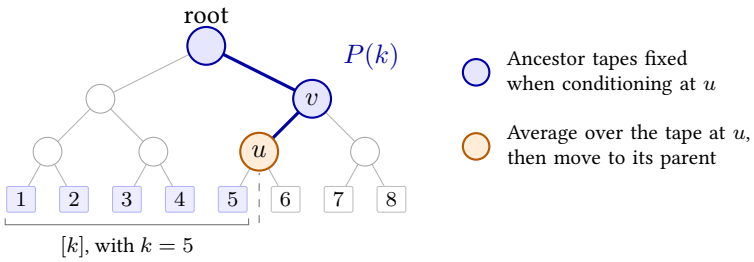


Figure 3: Conditioning from the deepest node upward, illustrated for $T = 8$ and $k = 5$. The highlighted path $P(k)$ contains the root, v, and u. When conditioning on the strict-ancestor tapes of u, the summands at the root and v are fixed, so only the exponential factor at u needs to be averaged. After removing that factor, the same argument applies at v, and then at the root.

which is (2.3). This argument holds for every realized input and collection of proposed signs, so a deterministic signer satisfying the terminal bounds on every stream also gives the claimed guarantee against adaptive inputs.

For the probabilistic statement, fix an oblivious input and say that a node fails if either of its two terminal bounds is violated. Conditioning on its strict-ancestor tapes fixes its entire input stream while leaving its own tape independent, just as in Lemma 2.3. The assumed fixed-stream guarantee thus bounds its conditional failure probability by $p(m_u)$. Taking expectations and then a union bound gives

$$\mathbb{P}(\exists \text{ node } u \text{ fails}) \le \sum_{u \text{ internal}} \mathbb{P}(\text{node } u \text{ fails}) = \sum_{u \text{ internal}} \mathbb{E}[\mathbb{P}(\text{node } u \text{ fails} \,|\, \mathrm{anc}(u))] \le \sum_{u \text{ internal}} p(m_u).$$

Outside this failure event, all local bounds hold and the pathwise argument above applies simultaneously to every prefix, which gives the first part of the theorem. □

2.4 Combining randomized errors

For part (ii) of Theorem 2.1, we use the decomposition in Lemma 2.3 to combine the centered signed sums before taking their norm, and then bound the quota corrections through the count imbalances. The errors along a path need not be independent, because ancestor decisions determine the streams reaching their descendants. The conditional estimates in (2.7) and (2.8) let us handle this dependence by conditioning successively from the deepest node to the root.

Lemma 2.4 (conditional composition). *Assume the subgaussian condition of Theorem 2.1, and write Y_u, q_u, R_u for the quantities in Lemma 2.3 at node u. There is an absolute constant C such that, for every prefix k and every $s > 0$,*

$$\mathbb{P}\left(\left\| \frac{1}{2} \sum_{u \in P(k)} \alpha_{u,k} Y_u \right\|_2 > C\Sigma \sqrt{h(d+s)} \right) \le e^{-s}, \tag{2.11}$$

$$\mathbb{P}\left(\sum_{u \in P(k)} \alpha_{u,k} |q_u| > C\Sigma \sqrt{h(h+s)} \right) \le e^{-s}. \tag{2.12}$$

Proof. The statement is immediate when $P(k)$ is empty, so assume otherwise. We first reduce the vector bound to a scalar tail estimate by choosing a $1/2$-net of the Euclidean unit sphere with at most e^{2d} points. For every nonzero vector x, a net point within distance $1/2$ of $x/\|x\|_2$ has inner product at least $\|x\|_2/2$

with x. Consequently, a union bound over the net proves (2.11), with $C = 4$, as soon as we show that every fixed unit vector z satisfies

$$\mathbb{P}\left(\left\langle z, \frac{1}{2}\sum_{v\in P(k)} \alpha_{v,k} Y_v \right\rangle > 2\Sigma\sqrt{h(d+s)}\right) \le e^{-2d-s}.$$

Indeed, a norm exceeding $4\Sigma\sqrt{h(d+s)}$ forces at least one net direction to exceed the displayed threshold, and $e^{2d}e^{-2d-s} = e^{-s}$. It thus remains to prove this directional estimate.

To this end, let us fix z and $\lambda > 0$, and note that the path and its coefficients depend only on the position tree and k, so they are deterministic. Let u be the deepest node of $P(k)$ so that every other node $v \in P(k) \setminus \{u\}$ is a strict ancestor of u (see Fig. 3). Since Y_v depends only on the tapes at v and its ancestors, conditioning on all strict-ancestor tapes of u fixes every term except the one at u. Its own tape remains independent, and (2.7), applied with $\lambda\alpha_{u,k}$, bounds its conditional exponential moment. Pulling out the fixed terms and then taking expectations gives

$$\begin{aligned}\mathbb{E}\exp\left(\frac{\lambda}{2}\sum_{v\in P(k)} \alpha_{v,k}\langle z, Y_v\rangle\right) &= \mathbb{E}\left[\exp\left(\frac{\lambda}{2}\sum_{v\in P(k)\setminus\{u\}} \alpha_{v,k}\langle z, Y_v\rangle\right)\mathbb{E}\left[e^{\lambda\alpha_{u,k}\langle z,Y_u\rangle/2}\,\middle|\,\mathrm{anc}(u)\right]\right] \\ &\le e^{\Sigma^2\lambda^2\alpha_{u,k}^2/2}\mathbb{E}\exp\left(\frac{\lambda}{2}\sum_{v\in P(k)\setminus\{u\}} \alpha_{v,k}\langle z, Y_v\rangle\right).\end{aligned}$$

Thus averaging over the tape at u removes its term from the sum, at the cost of the deterministic factor $e^{\Sigma^2\lambda^2\alpha_{u,k}^2/2}$. If any terms remain, the parent of u is now the deepest node, and the same calculation applies with that parent in place of u. Continuing upward to the root removes one term at a time and multiplies these factors, giving

$$\mathbb{E}\exp\left(\frac{\lambda}{2}\sum_{v\in P(k)} \alpha_{v,k}\langle z, Y_v\rangle\right) \le \exp\left(\frac{\Sigma^2\lambda^2}{2}\sum_{v\in P(k)}\alpha_{v,k}^2\right) \le \exp\left(\frac{h\Sigma^2\lambda^2}{2}\right).$$

The last inequality uses $0 \le \alpha_{v,k} \le 1$ and $|P(k)| \le h$. Chernoff's bound with $\lambda = 2\sqrt{d+s}/(\Sigma\sqrt{h})$ now yields

$$\mathbb{P}\left(\left\langle z, \frac{1}{2}\sum_{v\in P(k)} \alpha_{v,k} Y_v \right\rangle > 2\Sigma\sqrt{h(d+s)}\right) \le e^{-2(d+s)} \le e^{-2d-s},$$

which is precisely the directional estimate needed above.

For the count imbalances, we use the same conditioning argument, now with (2.8). For $\lambda > 0$, the inequality $e^{\lambda|x|} \le e^{\lambda x} + e^{-\lambda x}$ gives

$$\mathbb{E}\left[e^{\lambda\alpha_{u,k}|q_u|}\,\middle|\,\mathrm{anc}(u)\right] \le \mathbb{E}\left[e^{\lambda\alpha_{u,k}q_u} + e^{-\lambda\alpha_{u,k}q_u}\,\middle|\,\mathrm{anc}(u)\right] \le 2e^{\Sigma^2\lambda^2\alpha_{u,k}^2}.$$

Removing the nodes from deepest to root as before, and using $2^{|P(k)|} \le e^h$ and $\sum_{u\in P(k)}\alpha_{u,k}^2 \le h$, we obtain $\mathbb{E}\exp\left(\lambda\sum_{u\in P(k)}\alpha_{u,k}|q_u|\right) \le e^{h+h\Sigma^2\lambda^2}$. Markov's inequality with $\lambda = \sqrt{h+s}/(\Sigma\sqrt{h})$ therefore gives

$$\mathbb{P}\left(\sum_{u\in P(k)}\alpha_{u,k}|q_u| > 2\Sigma\sqrt{h(h+s)}\right) \le \exp\left(h + h\Sigma^2\lambda^2 - 2\lambda\Sigma\sqrt{h(h+s)}\right) = e^{-s},$$

which proves (2.12). □

Proof of Theorem 2.1 (ii). For $T = 1$ all residuals vanish, so assume $T \geq 2$. Fix k and use Lemmas 2.2 and 2.3 to write

$$\mathrm{res}_k(v_\pi) = \frac{1}{2} \sum_{u \in P(k)} \alpha_{u,k} Y_u + \sum_{u \in P(k)} \alpha_{u,k} R_u.$$

The first term is controlled by (2.11), while $0 \leq \alpha_{u,k} \leq 1$ and $|P(k)| \leq h$ give

$$\left\| \sum_{u \in P(k)} \alpha_{u,k} R_u \right\|_2 \leq \sum_{u \in P(k)} \alpha_{u,k} (|q_u| + 1) \leq h + \sum_{u \in P(k)} \alpha_{u,k} |q_u|.$$

Thus (2.12) controls the second term. Apply both estimates with $s = \log(2(T+1)/\delta)$ and take a union bound over the $T+1$ prefixes. The total failure probability is at most $2(T+1)e^{-s} = \delta$, so with probability at least $1 - \delta$,

$$\mathrm{res}_{\max}(v_\pi) \leq C\Sigma \left(\sqrt{h(d+s)} + \sqrt{h(h+s)} \right) + h. \tag{2.13}$$

Finally, $s \in O(h + \log(1/\delta))$ and $\Sigma \geq 1$, so (2.13) is at most $C\Sigma(h + \sqrt{h(d + \log(1/\delta))})$ after increasing the absolute constant, yielding (2.4). □

2.5 Some remarks on the reduction

Compact convex bodies. Both parts of Theorem 2.1 extend to inputs in a known compact convex body K of affine dimension $s \geq 1$. Let us consider any $0 \leq k \leq T$. Then, any Euclidean online router in dimension s with proportional-prefix radius ρ gives, after a fixed affine normalization,

$$\mathrm{res}_k(v_\pi) \in \frac{s\rho}{2}(K - K),$$

with the same deterministic or probabilistic guarantee and under the same adversarial model. Neither symmetry of K nor advance knowledge of the total is required.

Indeed, John's theorem in the relative affine space of K gives a centered ellipsoid $E \subseteq \mathrm{span}(K - K)$ and $c \in K$ with $c + E \subseteq K \subseteq c + sE$ and $2E \subseteq K - K$, where the second inclusion follows by subtracting $c - E$ from $c + E$. Fix a linear isomorphism Φ from $\mathrm{span}(K - K)$ onto $\mathbb{R}^s$ with $\Phi(E) = B_2^s$, and feed $\Phi(v_t - c)/s$ to the Euclidean router. These vectors lie in the unit ball, and translation cancels from proportional residuals:

$$\sum_{j=1}^{k} \frac{\Phi(v_{\pi(j)} - c)}{s} - \frac{k}{T} \sum_{t=1}^{T} \frac{\Phi(v_t - c)}{s} = \frac{1}{s} \Phi(\mathrm{res}_k(v_\pi)) \,.$$

Thus, whenever all normalized residuals have norm at most ρ,

$$\mathrm{res}_k(v_\pi) \in s\rho E \subseteq \frac{s\rho}{2}(K - K).$$

This implication holds for every realized run. Since the normalization is fixed before the stream, it preserves online decisions and the original deterministic or probabilistic guarantee, including its adversarial scope. If $s = 0$, every proportional residual is zero.

Comparison with offline transference. Chobanyan's transference theorem [Cho94] converts a uniform offline prefix-signing bound ζ into the same ordering bound. Indeed, choose an ordering of a zero-sum unit-ball sequence minimizing its maximum prefix norm ρ, and sign it with prefix bound ζ. Place the positively signed vectors first in their old order, followed by the negatively signed vectors in reverse order. Each new prefix is a positive-sign partial sum of the old order or, by the zero total, the negative of a negative-sign partial sum. Each is half a sum or difference of an old unsigned prefix and its signed prefix, so minimality gives $\rho \le (\rho + \zeta)/2$ and hence $\rho \le \zeta$. This argument chooses its initial ordering using the complete input. Our reduction uses terminal vector and count guarantees and commits immediately to positions, incurring losses from quota repair and accumulation along tree paths. As shown in Section 5, a loss is unavoidable already on the line: greedy signing has bound one, whereas deterministic online ordering requires $\Omega(\log T)$ discrepancy.

3 Deterministic applications

We apply Algorithm 1 with deterministic signing rules, first a scalar rule and then greedy vector signing. The scalar rule gives a stronger local invariant on the line; for greedy vector signing, we use the norm-and-count interface of Theorem 2.1 directly.

Theorem 3.1 (Deterministic OSP upper bound)**.** *For every $T \ge 2$, it holds that*

$$\mathrm{OS}_{\mathrm{det}}(d,T) \in \begin{cases} O(\log T), & d = 1, \\ O(\sqrt{T}), & d \ge 2. \end{cases}$$

These upper bounds hold against adaptive adversaries.

3.1 Scalar routing

We begin by giving a proof of the scalar upper bound in Theorem 3.1.

Proof of Theorem 3.1 for $d = 1$. We use the following local rule at every node of Algorithm 1, interpreting a left proposal as $+1$ and a right proposal as -1. Fix a node u. For the local argument, write L, R for the sums already sent to its children and r_L, r_R for their remaining capacities; once the node is full, $L = \mathrm{tot}_{u_L}$ and $R = \mathrm{tot}_{u_R}$. We maintain

$$|L + r_L y| + |R + r_R y| \le |L + R + (r_L + r_R)y| + 2 \qquad (y \in [-1, 1]). \tag{3.1}$$

Here $L + r_L y$ and $R + r_R y$ are the child totals if all remaining positions receive the same value y. The inequality fails exactly when one total exceeds 1 and the other is below -1. Initially $L = R = 0$, so the two totals have the same sign and the invariant holds.

If one child is full, the local rule proposes the other, so the meta-algorithm follows that proposal. Otherwise, propose a destination according to

$$\begin{cases} \text{left if } \dfrac{L+1}{r_L} \le \dfrac{R+1}{r_R}, \text{ right otherwise}, & L + R + (r_L + r_R)v_t \ge 0, \\ \text{left if } \dfrac{L-1}{r_L} \ge \dfrac{R-1}{r_R}, \text{ right otherwise}, & L + R + (r_L + r_R)v_t < 0. \end{cases}$$

Thus the hypothetical total determines which comparison the local rule uses; the meta-algorithm updates the chosen sum and capacity and continues from that child. For preservation, suppose both children have

space, the hypothetical total using v_t is nonnegative, and the rule chooses left. A subsequent completion with y would give child totals $L + v_t + (r_L - 1)y$ and $R + r_R y$.

There are only two possible violations of the invariant: first, suppose the new left total exceeds 1 and the right total is below -1. The left total changed by $v_t - y \le 2$, so its previous value $L + r_L y$ must exceed -1. Together with $R + r_R y < -1$, this gives

$$\frac{L+1}{r_L} > -y > \frac{R+1}{r_R},$$

contradicting the choice of the smaller ratio. Second, suppose the new left total is below -1 and the right total exceeds 1. The old invariant at y requires $L + r_L y \ge -1$, so the assignment must have decreased this total, meaning $v_t < y$. Therefore

$$L + r_L v_t = \underbrace{L + v_t + (r_L - 1)y}_{<-1} + \underbrace{(r_L - 1)(v_t - y)}_{\le 0} < -1.$$

Applying the old invariant at v_t now forces $R + r_R v_t \le 1$, and hence $L + R + (r_L + r_R)v_t < 0$, contradicting the assumed nonnegative total. Choosing right is symmetric, and the negative-total case follows by reversing all signs.

Forced routing also preserves the invariant. Indeed, if the left child is full, assigning v_t right and then filling its remaining positions with y gives the same totals as the previous hypothetical completion with common value $(v_t + (r_R - 1)y)/r_R \in [-1, 1]$. The old invariant applies directly; the other case is symmetric.

At completion, $r_L = r_R = 0$, so the invariant becomes

$$|\mathrm{tot}_{u_L}| + |\mathrm{tot}_{u_R}| \le |\mathrm{tot}_u| + 2.$$

Since nothing in the argument requires equal child capacities, the rule applies directly to the balanced tree for every T.

To conclude, fix an output prefix $[1, k]$. Starting from the root, split only the intervals in $P(k)$, as illustrated in Fig. 4. Each split increases the sum of absolute interval totals by at most 2. After these splits, the prefix is a union of complete intervals, so the triangle inequality gives

$$\left| \sum_{t=1}^{k} v_{\pi(t)} \right| \le |\mathrm{tot}_{\mathrm{root}}| + 2|P(k)| \le 2h.$$

The last inequality uses the zero root total and the tree height. The invariant holds for every arriving value, so the guarantee also holds against adaptive adversaries. □

3.2 Greedy signing in arbitrary dimension

The same routing algorithm, using a greedy signer at each node, controls proportional-prefix residuals even when the total is nonzero. Hence, the upper bound in Theorem 3.1 for $d \ge 2$ immediately follows since, for zero-sum inputs, the residuals are ordinary prefix sums.

Proposition 3.2 (Deterministic proportional routing)**.** *For every stream of Euclidean unit-ball vectors, Algorithm 1 with greedy signing of $(v_t, 1)/\sqrt{2}$ satisfies*

$$\mathrm{res}_{\max}(v_\pi) \le \frac{\sqrt{2}+1}{1 - \sqrt{2/3}} \sqrt{T}. \tag{3.2}$$

This bound holds against adaptive adversaries and requires no advance knowledge of the total.

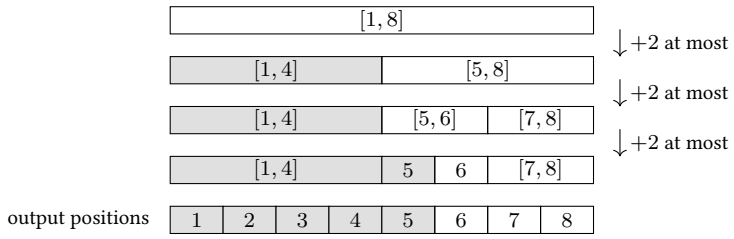


Figure 4: Analyzing $[1, 5]$ when $T = 8$. Only a partially included interval is split; shaded blocks are wholly inside the prefix. Each refinement costs at most 2 in the sum of absolute block totals, so three refinements bound this prefix by 6.

Proof. At an internal node u, let $v_{u,i}$ be its ith arriving vector and let $S_{u,i}$ be the signer's augmented signed sum after that arrival, initially $S_{u,0} = 0$. Choose $\varepsilon_{u,i} \in \{-1, +1\}$ so that

$$\varepsilon_{u,i} \left\langle S_{u,i-1}, \frac{(v_{u,i}, 1)}{\sqrt{2}} \right\rangle \le 0,$$

breaking ties in favor of $+1$, and update $S_{u,i} = S_{u,i-1} + \varepsilon_{u,i}(v_{u,i}, 1)/\sqrt{2}$. The update records the proposed sign even if the algorithm overrides its routing request. Expanding the squared norm gives

$$\|S_{u,i}\|_2^2 = \|S_{u,i-1}\|_2^2 + \frac{\|v_{u,i}\|_2^2 + 1}{2} + 2\varepsilon_{u,i} \left\langle S_{u,i-1}, \frac{(v_{u,i}, 1)}{\sqrt{2}} \right\rangle \le \|S_{u,i-1}\|_2^2 + 1 \le i.$$

Thus the vector and count imbalances obey the joint bound

$$\left\| \sum_{i=1}^{m_u} \varepsilon_{u,i} v_{u,i} \right\|_2^2 + \left| \sum_{i=1}^{m_u} \varepsilon_{u,i} \right|^2 = 2\|S_{u,m_u}\|_2^2 \le 2m_u.$$

The local estimate in the norm-and-count part of Theorem 2.1, followed by Cauchy–Schwarz, therefore gives

$$\|\mathrm{err}_u\|_2 \le \frac{1}{2} \left(\left\| \sum_{i=1}^{m_u} \varepsilon_{u,i} v_{u,i} \right\|_2 + \left| \sum_{i=1}^{m_u} \varepsilon_{u,i} \right| \right) + \mathbb{1}\{m_u \text{ odd}\} \le \sqrt{m_u} + \mathbb{1}\{m_u \text{ odd}\}.$$

The boundary-path conclusion of Theorem 2.1 now yields

$$\|\mathrm{res}_k(v_\pi)\|_2 \le \sum_{u \in P(k)} \|\mathrm{err}_u\|_2 \le (\sqrt{2} + 1)\sqrt{T} \sum_{\ell=0}^{\infty} (2/3)^{\ell/2} = \frac{\sqrt{2} + 1}{1 - \sqrt{2/3}} \sqrt{T}.$$

Indeed, every internal node has children of capacity at most $2m_u/3$, so a node at depth ℓ has capacity at most $T(2/3)^\ell$, and $P(k)$ contains at most one node at each depth. Every estimate holds for every local stream, which also proves the adaptive guarantee. □

4 Randomized applications

The reduction in Theorem 2.1 accepts any online signer satisfying its terminal guarantees, and, as in the deterministic case, this flexibility yields several randomized applications. As our main application, we use Aden-Ali's signer [Ade26] since it satisfies the subgaussian hypothesis in part (ii) and can be implemented efficiently. Moreover, we will derive a coordinatewise bound from it, and a sparse-vector application from the norm-and-count guarantees in part (i) and the Altschuler–Tikhomirov's signer [AT26]. Throughout this section, inputs are fixed independently of the algorithm's randomness, and all bounds allow arbitrary, initially unknown totals and become prefix-discrepancy bounds when the total is zero.

4.1 Efficient randomized ordering

Corollary 4.1 (Efficient signer). *For every $T \geq 2$, $d \geq 1$, and $0 < \delta < 1$, Algorithm 1 with the Aden-Ali signer uses $O(dT \log T)$ arithmetic operations and, for every fixed Euclidean unit-ball stream, satisfies, with probability at least $1 - \delta$,*

$$\mathrm{res}_{\max}(v_\pi) \in O\left(\log T + \sqrt{(d + \log(1/\delta)) \log T}\right).$$

Proof. We apply part (ii) of Theorem 2.1, whose assumption is the terminal moment-generating-function bound (2.2) in dimension $d+1$. Specifically, at a node receiving $x_1, \ldots, x_m$, we run an independent copy of Aden-Ali's signer on $z_i = (x_i, 1)/\sqrt{2}$ so that $\|z_i\|_2^2 \leq 1$. We keep this copy running throughout the local stream, recording its proposed signs even when the router overrides a destination to enforce the child capacities.

To verify the required assumption in (2.2), we observe that, for every fixed augmented stream, Aden-Ali's signer's guarantee [Ade26, Theorem 1 and Equation (5)] represents the terminal signed sum as $\sum_{i=1}^m \varepsilon_i z_i = G_1 + G_2 + G_3$, where $G_1, G_2, G_3 \sim N(0, I_{d+1})$ are Gaussian vectors that may be dependent. Then, for every $\theta \in \mathbb{R}^{d+1}$, Hölder's inequality and the Gaussian moment-generating function give

$$\mathbb{E}\exp\left(\left\langle \theta, \sum_{i=1}^m \varepsilon_i z_i \right\rangle\right) \leq \prod_{j=1}^3 \left(\mathbb{E}e^{3\langle \theta, G_j\rangle}\right)^{1/3} = \prod_{j=1}^3 \left(e^{9\|\theta\|_2^2/2}\right)^{1/3} = e^{9\|\theta\|_2^2/2}. \tag{4.1}$$

Hence, (2.2) is verified with $\Sigma = 3$, uniformly over all node capacities. Furthermore, conditioning on a node's strict-ancestor tapes fixes its input stream while leaving its own tape independent, and thus, (4.1) also supplies the conditional hypothesis used in Lemmas 2.3 and 2.4.

For completeness, we spell out the resulting substitutions in the reduction: set $h = \lceil \log_2 T \rceil$ and $s = \log(2(T+1)/\delta)$, and, with $\Sigma = 3$, the two estimates of Lemma 2.4 read, for each k,

$$\mathbb{P}\left(\left\|\frac{1}{2}\sum_{u \in P(k)} \alpha_{u,k} Y_u\right\|_2 > 3C\sqrt{h(d+s)}\right) \leq e^{-s},$$
$$\mathbb{P}\left(\sum_{u \in P(k)} \alpha_{u,k}|q_u| > 3C\sqrt{h(h+s)}\right) \leq e^{-s}.$$

A union bound over these two events and the $T+1$ prefixes has total failure probability at most $2(T+1)e^{-s} = \delta$. Outside this failure event, Lemmas 2.2 and 2.3 give, for every k,

$$\begin{aligned}\|\mathrm{res}_k(v_\pi)\|_2 &\leq \left\|\frac{1}{2}\sum_{u \in P(k)} \alpha_{u,k} Y_u\right\|_2 + \sum_{u \in P(k)} \alpha_{u,k}\|R_u\|_2 \leq \left\|\frac{1}{2}\sum_{u \in P(k)} \alpha_{u,k} Y_u\right\|_2 + h + \sum_{u \in P(k)} \alpha_{u,k}|q_u| \\ &\leq 3C\sqrt{h(d+s)} + h + 3C\sqrt{h(h+s)} \in O\left(\log T + \sqrt{(d + \log(1/\delta)) \log T}\right),\end{aligned}$$

where the last step uses $s = O(h + \log(1/\delta))$, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ and $h \in O(\log T)$.

Finally, a length-m signer in dimension $d+1$ uses $O(dm)$ arithmetic operations, and, since at each depth, the node capacities sum to at most T, and there are at most h internal levels, the overall running time $O(dT \log T)$. □

We next discuss an alternative signer and establish a sharper dimension dependence.

Using an inefficient signer. If computational efficiency is not required, the reduction also allows the Kulkarni–Reis–Rothvoss's signer in place of Aden-Ali's. For each fixed augmented unit-ball stream, its terminal guarantee is [KRR24, Theorem 1]

$$\mathbb{E}\exp\left(\frac{\langle\theta,\sum_i \varepsilon_i z_i\rangle^2}{100\|\theta\|_2^2}\right)\le 2 \qquad (\theta\neq 0). \tag{4.2}$$

At each node, sample an independent uniform sign $\tau\in\{-1,+1\}$ before processing its stream, and send $\tau\varepsilon_i$ as the proposals without changing the Kulkarni–Reis–Rothvoss signer's internal updates. The terminal sum is then symmetric and centered, while (4.2) is unchanged. Accordingly, we call this the *symmetrized Kulkarni–Reis–Rothvoss's signer*. By [KRR24, Lemma 11],

$$\mathbb{E}\exp\left(\left\langle\theta,\tau\sum_i \varepsilon_i z_i\right\rangle\right)\le e^{40\|\theta\|_2^2}.$$

Thus (2.2) holds with $\Sigma=\sqrt{80}$, and Theorem 2.1(ii) gives the same probability and discrepancy bound as Corollary 4.1, up to the absolute constant. Only the efficiency guarantee is lost, since the finite-horizon Kulkarni–Reis–Rothvoss's construction takes exponential time [KRR24, Sections 6–7].

Lemma 4.2 (Sharper dimension dependence)**.** *For every $T\ge 2$, $d\ge 1$, and $0<\delta<1$, there is a randomized router such that, for every fixed Euclidean unit-ball stream, with probability at least $1-\delta$,*

$$\mathrm{res}_{\max}(v_\pi)\in O\left(\min\left\{\sqrt{T},\ \log T+\sqrt{d_*\log(2+T/d_*)}+\sqrt{\log(2+T/d_*)\log(1/\delta)}\right\}\right), \tag{4.3}$$

where $d_*:=\min\{d,T\}$. With the Aden-Ali signer, the running time is $O(dT\log T)$. The same bound holds with the symmetrized Kulkarni–Reis–Rothvoss's signer, with a different absolute constant and without the running-time guarantee.

Proof. Use the randomized signer only at nodes strictly above a depth $0\le s\le h=\lceil\log_2 T\rceil$, and the deterministic router of Proposition 3.2 within each remaining interval. For $s\ge 1$, apply Lemma 2.4 with path length at most s and tail parameter $\log(2^{s+1}/\delta)$. A union bound over the two estimates at each of the at most 2^s-1 nontrivial interval endpoints bounds all their residuals by $C(s+\sqrt{s(d+\log(1/\delta))})$ with probability at least $1-\delta$. Inside an interval, a global residual equals the local proportional residual plus a convex combination of the two endpoint residuals. Since each interval has length at most $\lceil T/2^s\rceil$, Proposition 3.2 therefore gives

$$\mathrm{res}_{\max}(v_\pi)\in O\left(s+\sqrt{s(d+\log(1/\delta))}+\sqrt{T/2^s}\right)$$

on the same event, where the ceiling is absorbed into the constant because $s\le h$.

For $s=0$, this is the deterministic $C\sqrt{T}$ bound. When $d<T$, choose $s=\lceil\log_2(T/d)\rceil$. Then $T/2^s\le d$ and $s=O(\log(2+T/d))$. Comparing this choice with $s=0$ before the stream begins gives (4.3). For $d\ge T$, use $s=0$. The running time remains $O(dT\log T)$ with the Aden-Ali signer, since the total node capacity at each depth is at most T. The same probability argument applies to the symmetrized Kulkarni–Reis–Rothvoss's signer. Finally, if $\delta\ge T^{-\beta}$ for fixed $\beta>0$, then $\sqrt{\log(2+T/d_*)\log(1/\delta)}=O_\beta(\log T)$. □

4.2 Coordinatewise discrepancy

The next consequence uses the directional estimates in the proof of part (ii) of Theorem 2.1. A union bound over the relevant coordinate directions replaces the sphere net used for the ℓ_2-norm.

Corollary 4.3 (Coordinatewise ordering). *For every fixed Euclidean unit-ball stream and $0 < \delta < 1$, Algorithm 1, using either the symmetrized Kulkarni–Reis–Rothvoss's signer or the Aden-Ali signer throughout the full tree, satisfies, with probability at least $1 - \delta$,*

$$\max_{0\le k\le T} \|\mathrm{res}_k(v_\pi)\|_\infty \in O\left(\log T + \sqrt{\log T \cdot \log(1/\delta)}\right).$$

Proof. The proof of Corollary 4.1 verifies (2.2) with an absolute scale Σ for either signer. We use the same path decomposition and quota-repair estimate as in Theorem 2.1, but first identify which coordinates need a probabilistic estimate. By Cauchy–Schwarz and the unit-ball assumption,

$$\sum_{j=1}^{d}\left(\sum_{t=1}^{T}|v_{t,j}|\right)^2 \le T\sum_{t=1}^{T}\|v_t\|_2^2 \le T^2,$$

which implies that at most T^2 coordinates satisfy $\sum_t |v_{t,j}| > 1$. Every other coordinate has residual of magnitude at most one for every permutation: in a proportional residual, each input coordinate has coefficient either $1-k/T$ or $-k/T$, both of absolute value at most one. It is worth noting that the exceptional coordinates depend only on the fixed input and are used only in the analysis, and the algorithm does not need to identify them.

For any fixed unit direction z, the conditional moment calculation in the proof of Lemma 2.4, followed by Chernoff's bound, gives, for all $s > 0$,

$$\mathbb{P}\left(\frac{1}{2}\sum_{u\in P(k)}\alpha_{u,k}\langle z, Y_u\rangle > \Sigma\sqrt{2hs}\right) \le e^{-s}.$$

Apply this estimate to both signs of each exceptional coordinate direction. Also apply the count estimate (2.12) at each prefix. There are at most $(2T^2+1)(T+1)$ events in total, so choosing $s = \log((2T^2+1)(T+1)/\delta)$ and taking a union bound makes them all hold with probability at least $1-\delta$.

On this event, the centered path sum in every exceptional coordinate has magnitude at most $\Sigma\sqrt{2hs}$. The correction is controlled without a union bound over coordinates, because

$$\left\|\sum_{u\in P(k)}\alpha_{u,k}R_u\right\|_\infty \le \sum_{u\in P(k)}\alpha_{u,k}\|R_u\|_2 \le h + \sum_{u\in P(k)}\alpha_{u,k}|q_u|.$$

Combining the two estimates gives, at every prefix and in every exceptional coordinate, a residual of magnitude at most

$$\Sigma\sqrt{2hs} + h + C\Sigma\sqrt{h(h+s)} \le C\Sigma(h+\sqrt{hs}).$$

Now $s \in O(\log T + \log(1/\delta))$ and $h \in O(\log T)$. Absorbing the bound one for the remaining coordinates proves the claim. The running-time calculation is the same as in Corollary 4.1. □

4.3 Sparse vectors

Here we apply part (i) of Theorem 2.1 in the ℓ_∞ norm, with the input vectors that need only be coordinate-bounded, and their ℓ_2-norms may be as large as $\sqrt{\kappa}$. In particular, we use copies of the Altschuler–Tikhomirov signer [AT26].

Corollary 4.4 (OSP for sparse vectors). *For every $\eta > 0$, there is a constant $C_\eta > 0$ with the following property. Let $T \ge 2$, let $\kappa \in \{1,\dots,d\}$ be a sparsity bound known in advance, and let $0 < \delta < 1$. Write*

$h = \lceil \log_2 T \rceil$ *and suppose that* $\delta \geq C_\eta T h \exp(-(\kappa+1)/(C_\eta \log^{2+\eta}(e(\kappa+1))))$. *Then, for every fixed stream* $v_t \in [-1,1]^d$ *with* $\|v_t\|_0 \leq \kappa$, *Algorithm 1, using the Altschuler–Tikhomirov signer, satisfies, with probability at least* $1-\delta$,

$$\max_{0 \leq k \leq T} \|\mathrm{res}_k(v_\pi)\|_\infty \leq (C_\eta \sqrt{\kappa+1}+1)h \in O_\eta(\sqrt{\kappa}\log T).$$

Proof. Fix $\kappa \geq 1$ and $T \geq 2$. At a node receiving $x_1, \ldots, x_m$, append the quota coordinate and give $(x_i, 1)$ to an independent copy of the signer of Altschuler–Tikhomirov [AT26, Theorem 1.1], with announced sparsity $\kappa+1$. Thus, each augmented vector belongs to $[-1,1]^{d+1}$ and has at most $\kappa+1$ nonzero coordinates. Moreover, note that no Euclidean normalization is needed here, because we are using the norm-and-count part of the reduction with the ℓ_∞ norm. Next, for each fixed local stream, the signer's simultaneous-prefix guarantee implies, at the terminal prefix,

$$\left\| \sum_{i=1}^m \varepsilon_i x_i \right\|_\infty \leq C_\eta \sqrt{\kappa+1}, \qquad \left| \sum_{i=1}^m \varepsilon_i \right| \leq C_\eta \sqrt{\kappa+1},$$

except with probability at most

$$p(m) = C_\eta m \exp\left(-\frac{\kappa+1}{C_\eta \log^{2+\eta}(e(\kappa+1))} \right).$$

We have enlarged C_η, if necessary, so that the same constant works in both estimates. The first terminal bound comes from the original coordinates, and the second comes from the appended quota coordinate. Thus the two hypotheses of part (i) of Theorem 2.1 hold jointly with $V(m) = W(m) = C_\eta \sqrt{\kappa+1}$.

Although the stream reaching a node is random, it is determined by its strict-ancestor tapes. Conditional on those tapes, the node's own signer is an independent copy on a fixed stream: this is the conditioning argument in Theorem 2.1, so its failure bound applies at every node. At each depth, the node capacities sum to at most T, and there are at most h internal levels, so

$$\sum_{u \text{ internal}} p(m_u) \leq C_\eta T h \exp\left(-\frac{\kappa+1}{C_\eta \log^{2+\eta}(e(\kappa+1))} \right) \leq \delta.$$

With probability at least $1-\delta$, all local terminal bounds hold. On this single event, (2.3) gives, for every output prefix,

$$\|\mathrm{res}_k(v_\pi)\|_\infty \leq \sum_{u \in P(k)} \left(C_\eta \sqrt{\kappa+1} + \mathbb{1}\{m_u \text{ odd}\} \right) \leq (C_\eta \sqrt{\kappa+1}+1)h.$$

This gives the stated asymptotic bound as $h \in O(\log T)$. □

5 Lower bounds

In this section, we prove lower bounds that match our deterministic upper bounds and, for constant and polynomially small failure probabilities, our randomized upper bounds. All our hard inputs use the same zero-sum completion: stop an initial stream and fill the remaining arrivals with identical vectors that cancel its total. For the randomized bounds, the initial stream consists of coordinate cycles, stopped at a time chosen independently of the algorithm's random tape. For the deterministic bound in a plane, the initial stream instead depends on the algorithm's previous assignments.

Theorem 5.1 (Deterministic and randomized OSP lower bounds). *For $T \geq 2$ and $d \geq 1$, it holds that*

$$\mathrm{OS}_{\mathrm{det}}(d,T) \in \begin{cases} \Omega(\log T), & d = 1, \\ \Omega(\sqrt{T}), & d \geq 2, \end{cases}$$

and, for $0 < \delta \leq 1/2$ and $d_* := \min\{d, T\}$,

$$\mathrm{OS}_{\mathrm{rand}}(d,T,\delta) \in \Omega\left(\log T + \sqrt{d_* \log(T/d_*)}\right).$$

The deterministic bound for $d \geq 2$ already holds in a fixed plane. The oblivious hard inputs are independent of the algorithm's random tape.

5.1 The input construction

We first explain the construction of the hard inputs which will be used throughout this section. Fix $T \geq 2$. Consider the following input: after presenting $m \leq T/2$ vectors $v_1, \ldots, v_m$ with $\|v_i\|_2 \leq 1$, present $(T-m)$ copies of $-1/(T-m)\sum_{i=1}^m v_i$. These repeated vectors remain in the unit ball since $\|\sum_{i=1}^m v_i\|_2 \leq m \leq T-m$. Let $p_i \in [T]$ be the position assigned to v_i by the algorithm. Then the first k output positions contain exactly $k - |\{i \leq m : p_i \leq k\}|$ repeated vectors, hence the prefix residual is

$$\mathrm{res}_k(v_\pi) = \sum_{\substack{i \leq m \\ p_i \leq k}} v_i - \frac{k - |\{i \leq m : p_i \leq k\}|}{T-m} \sum_{i=1}^m v_i. \tag{5.1}$$

For the randomized lower bounds, suppose that $d \leq T/2$ and set $N = \lfloor T/(2d) \rfloor$. For each $s \in [N]$, the initial stream consists of s cycles through the coordinate vectors, giving the fixed input

$$I_s = \left(\underbrace{e_1, \ldots, e_d, \ldots, e_1, \ldots, e_d}_{s \text{ cycles}}, \underbrace{-\frac{s\mathbf{1}}{T-ds}, \ldots, -\frac{s\mathbf{1}}{T-ds}}_{T-ds \text{ copies}} \right), \tag{5.2}$$

where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^d$. The point is that the algorithm cannot know when the coordinate cycles will stop. Indeed, fix a random tape and let $p_{i,j}$ be the position assigned to e_j in cycle i when the algorithm runs on I_N. Clearly, these same positions describe its first s cycles on every I_s, since the inputs agree up to time ds.

5.2 Deterministic lower bounds

We prove a logarithmic lower bound in dimension one and a square-root lower bound in every dimension at least two.

Lemma 5.2 (Deterministic lower bounds). *There exists an absolute constant $c > 0$ such that, for every $T \geq 2$ and $d \geq 1$, every deterministic online ordering algorithm has a fixed zero-sum unit-ball input satisfying*

$$\mathrm{res}_{\max}(v_\pi) \geq \begin{cases} c \log T, & d = 1, \\ c\sqrt{T}, & d \geq 2. \end{cases}$$

For $d = 1$, the input can be chosen from the family (5.2); for $d \geq 2$, the input lies in a fixed plane.

Proof. On the line, let us take $N = \lfloor T/2 \rfloor$ and write $p_i = p_{i,1}$ for the common positions in the construction above. Let $c_0 > 0$ be the absolute constant in Schmidt's theorem [Sch72, Theorem 2]. Applied to the points $((i-1)/N, (p_i - 1)/T)$, it gives $x, y \in [0,1]$ such that

$$\left| \left| \left\{ i \le N : \frac{i-1}{N} < x,\ \frac{p_i - 1}{T} < y \right\} \right| - Nxy \right| \ge c_0 \log N.$$

Here the cardinality counts the points in $[0,x) \times [0,y)$, while Nxy is N times the area of this rectangle. We now set $s = \lceil Nx \rceil$ and $k = \lceil Ty \rceil$, and note that rounding the upper sides to s/N and k/T preserves the point count because $(i-1)/N < x$ if and only if $i \le s$ and $(p_i - 1)/T < y$ if and only if $p_i \le k$. In addition, this changes the area term by at most

$$0 \le \frac{sk}{T} - Nxy = N\left(\frac{s}{N} - x\right) y + s\left(\frac{k}{T} - y\right) \le 1 + \frac{N}{T} \le \frac{3}{2}.$$

The reverse triangle inequality therefore gives

$$\left| |\{ i \le s : p_i \le k \}| - \frac{sk}{T} \right| \ge c_0 \log N - \frac{3}{2} \ge c_0 \log N - 2.$$

On the input I_s, the completion identity (5.1) becomes

$$\mathrm{res}_k(v_\pi) = \frac{T}{T-s} \left(|\{ i \le s : p_i \le k \}| - \frac{sk}{T} \right). \tag{5.3}$$

Since $T/(T-s) \ge 1$, this proves the logarithmic lower bound for all sufficiently large T.

For the planar bound, we proceed in a similar way to Spencer [Spe77, Spe94], and present the proof for completeness. Indeed, let us assume $T \ge 8$, take $m = \lfloor T/4 \rfloor$, and focus on the output prefix $k = \lfloor T/2 \rfloor$. During the first m arrivals, let $\xi_t = +1$ if the algorithm assigns v_t to one of the first k positions, and $\xi_t = -1$ otherwise, and write $U_t = \sum_{i=1}^{t} v_i$, $Z_t = \sum_{i=1}^{t} \xi_i v_i$ with $U_0 = Z_0 = 0$. We choose the vectors so that the difference Z_t between the two groups grows, while their total U_t stays small enough that the completion cannot cancel this difference. At time t, choose a unit vector v_t in the plane spanned by e_1, e_2 perpendicular to Z_{t-1}, reversing its orientation if necessary so that $\langle U_{t-1}, v_t \rangle \le 0$. This choice uses only previous assignments, and whichever value of ξ_t the algorithm then produces,

$$\|Z_t\|_2^2 = \|Z_{t-1}\|_2^2 + 1, \qquad \|U_t\|_2^2 \le \|U_{t-1}\|_2^2 + 1.$$

Hence $\|Z_m\|_2 = \sqrt{m}$ and $\|U_m\|_2 \le \sqrt{m}$. We now complete the stream with $T-m$ copies of $-U_m/(T-m)$. Among the initial vectors, those assigned to the first k positions have sum $(U_m + Z_m)/2$ and number $(m + \sum_{t=1}^{m} \xi_t)/2$, so (5.1) gives

$$2\mathrm{res}_k(v_\pi) = Z_m + \frac{T - 2k + \sum_{t=1}^{m} \xi_t}{T-m} U_m.$$

Since $T - 2k \in \{0, 1\}$ and $|\sum_{t=1}^{m} \xi_t| \le m$, the triangle inequality yields

$$2\|\mathrm{res}_k(v_\pi)\|_2 \ge \left(1 - \frac{m+1}{T-m}\right) \sqrt{m} \ge \frac{\sqrt{m}}{2} \ge \frac{\sqrt{T}}{4\sqrt{2}},$$

where $(m+1)/(T-m) \le 1/2$ and $m \ge T/8$ for $T \ge 8$. Choosing $c > 0$ sufficiently small gives both bounds for large T; the remaining bounded values follow from the one-cycle bound below after further decreasing c. All vectors lie in the same fixed plane, and although the initial stream was described adaptively, for any deterministic algorithm its assignments can be simulated in advance, producing a fixed input with the claimed cost. □

Remark 5.3. *It is worth noting that the one-cycle input I_1, using $\min\{d, \lfloor T/2 \rfloor\}$ coordinates, has discrepancy at least $c\sqrt{\min\{d,T\}}$ under every ordering, for an absolute constant $c > 0$. Indeed, consider the one-cycle input I_1 in dimension $d \le T/2$. For any ordering, linearly interpolate between consecutive prefix sums and stop during a repeated vector, or at its endpoint, when exactly half of the repeated copies have been traversed. Their contribution is then $-1/2$ in each coordinate, while each basis vector has either appeared in full or has not appeared yet, so every coordinate of the interpolated sum is $-1/2$ or $1/2$. Its norm is therefore $\sqrt{d}/2$, and convexity of the norm implies that at least one adjacent genuine prefix has norm at least as large.*

5.3 Randomized lower bounds

We now show that randomized algorithms can also have a large prefix sum on some fixed input. We start by proving the first term $\Omega(\log T)$ of the randomized lower bound, which holds already in dimension one.

Lemma 5.4 (Randomized lower bound for $d = 1$)**.** *There exists an absolute constant $c > 0$ such that, for every $T \ge 2$ and every randomized online ordering algorithm, some fixed input from the scalar family (5.2) satisfies*

$$\mathbb{P}(\mathrm{res}_{\max}(v_\pi) > c \log T) > \frac{1}{2}.$$

Proof. Let $d = 1$ and $N = \lfloor T/2 \rfloor$. The family (5.2) becomes

$$I_s = \left(\underbrace{1, \dots, 1}_{s \text{ copies}}, \underbrace{-\frac{s}{T-s}, \dots, -\frac{s}{T-s}}_{T-s \text{ copies}} \right), \qquad s \in [N].$$

Let us write $\mathrm{res}_{\max}(I_s)$ for the maximum residual of the algorithm's output on I_s, and let us fix a random tape so that the algorithm's decisions are deterministic, with p_i being the output position assigned to the ith arriving copy of 1 on I_N. Using the same tape, its first s assignments on I_s are also $p_1, \dots, p_s$, since both inputs begin with s copies of 1. For $s \in [N]$, denote the discrepancy over output prefixes as

$$\Delta_s := \max_{0 \le k \le T} \left| |\{i \le s : p_i \le k\}| - \frac{sk}{T} \right|.$$

By (5.3), we have $\Delta_s = (T-s)/T \cdot \mathrm{res}_{\max}(I_s) \le \mathrm{res}_{\max}(I_s)$. To apply Tijdeman–Wagner's theorem on an infinite sequence in $[0,1)$, we first append infinitely many zeros after $((p_i - 1)/T)_{i=1}^N$. By Tijdeman–Wagner's theorem [TW80, Theorem 1], there exists an absolute constant $c_0 > 0$ such that for every fixed tape,

$$\left| \left\{ 1 \le s \le N : \Delta_s + \frac{1}{2} < c_0 \log N \right\} \right| \le 2N^{5/6}.$$

Here the term $1/2$ comes from restricting interval endpoints to multiples of $1/T$. Now, for any $y \in [0,1)$, we round y upwards to $\lceil Ty \rceil / T$, and, since all points lie on this grid, the intervals $[0, y)$ and $[0, \lceil Ty \rceil / T)$ contain the same points. The term sy changes by at most $s/T \le 1/2$, so the discrepancy over arbitrary intervals $[0, y)$ is at most $\Delta_s + 1/2$.

Let c_0 be the constant in Tijdeman–Wagner's theorem and set $c_1 = \min\{c_0/4, 1/26\}$. If $c_1 \log T < 1/2$, the input I_1 gives the desired bound with probability one. Otherwise, $\log T \ge 13$, so $N > 2^{12}$ and $\log N > 1/2 \cdot \log T$, giving $c_1 \log T + 1/2 \le 2c_1 \log T < c_0 \log N$. This gives that, for any s with $\mathrm{res}_{\max}(I_s) \le c_1 \log N$,

$$\Delta_s + \frac{1}{2} \le \mathrm{res}_{\max}(I_s) + \frac{1}{2} \le c_1 \log N + \frac{1}{2} < c_0 \log N.$$

Therefore for each random tape, there are at most $2N^{5/6}$ indices with $\mathrm{res}_{\max}(I_s) \le c_1 \log N$. Averaging over the algorithm's random tape gives

$$\frac{1}{N}\sum_{s=1}^{N} \mathbb{P}(\mathrm{res}_{\max}(I_s) > c_1 \log N) \ge 1 - 2N^{-\frac{1}{6}} > \frac{1}{2},$$

so there exists a fixed I_s which has maximum residual greater than $c \log N$ with probability greater than $1/2$. Hence, some fixed input I_s satisfies

$$\mathbb{P}(\mathrm{res}_{\max}(I_s) > c \log T) > \frac{1}{2},$$

for $c = c_1/2$, which concludes the proof. □

The following lemma derives the second term of the randomized lower bound for any dimension.

Lemma 5.5 (Randomized lower bound in arbitrary dimension). *There exists an absolute constant $c > 0$ such that the following holds. Let $T \ge 2$, $d \ge 1$, and $0 < \delta < 1$. If a randomized online ordering algorithm satisfies*

$$\mathbb{P}(\mathrm{res}_{\max}(v_\pi) \le \rho) \ge 1 - \delta$$

on every fixed zero-sum unit-ball input, then $\rho \ge c(1-\delta)\sqrt{d_* \log(T/d_*)}$, where $d_* := \min\{d, T\}$.

Proof. When $d > T/2$, the one-cycle input in Lemma 5.2 has cost $\Omega(\sqrt{d_*})$ under every ordering and, since $\log(2 + T/d_*) \in O(1)$, this proves the claim for sufficiently small $c > 0$.

Therefore we assume that $d \le T/2$ and let $N = \lfloor T/(2d) \rfloor$. The expected fraction of inputs $I_1, \dots, I_N$ on which the algorithm succeeds is at least $1 - \delta$, and so we fix a tape that succeeds on a set $S \subseteq [N]$ with probability $p := |S|/N \ge 1 - \delta$. For its positions $p_{i,j}$, define for each $j = 1, \dots, d$ and $(x, y) \in [0,1]^2$,

$$D_j(x,y) := \left| \left\{ i \in [N] : \frac{i-1}{N} < x,\ \frac{p_{i,j}-1}{T} < y \right\} \right| - Nxy,$$

and for fixed $s \in [N]$ and $0 \le k \le T$, write $a_j := D_j\left(s/N, k/T\right) = |\{i \le s : p_{i,j} \le k\}| - sk/T$. The identity in (5.1) can thus be rewritten as

$$\mathrm{res}_k(v_\pi) = (a_j)_{j=1}^d + \frac{s}{T - ds}\left(\sum_{j=1}^{d} a_j\right)\mathbf{1}.$$

Note that these two terms have nonnegative inner product since

$$\left\langle (a_j)_{j=1}^d,\ \frac{s}{T-ds}\left(\sum_{\ell=1}^{d} a_\ell\right)\mathbf{1} \right\rangle = \frac{s}{T-ds}\left(\sum_{\ell=1}^{d} a_\ell\right)\left(\sum_{j=1}^{d} a_j\right) = \frac{s}{T-ds}\left(\sum_{j=1}^{d} a_j\right)^2 \ge 0,$$

and therefore, for every $s \in S$ and $0 \le k \le T$, success on I_s implies

$$\sum_{j=1}^{d} D_j\left(\frac{s}{N}, \frac{k}{T}\right)^2 \le \|\mathrm{res}_k(v_\pi)\|_2^2 \le \rho^2.$$

We now define the set $E = \bigcup_{s \in S}((s-1)/N, s/N)$, so that each successful stopping time contributes an interval of length $1/N$ and $|E| = |S|/N = p$. We extend the discrepancy bound at successful grid points to these intervals to show that

$$\int_E \int_0^1 \sum_{j=1}^{d} D_j(x,y)^2 \, dy \, dx \le p\left(\rho + \frac{3\sqrt{d}}{2}\right)^2. \tag{5.4}$$

For any $1 \le s \le N$, $1 \le k \le T$, and $(x, y) \in ((s-1)/N, s/N) \times ((k-1)/T, k/T)$, we have

$$\frac{i-1}{N} < x \iff i \le s, \qquad \frac{p_{i,j}-1}{T} < y \iff p_{i,j} \le k.$$

Thus (x, y) and the upper corner $(s/N, k/T)$ count exactly the same points, so only the area term changes:

$$0 \le D_j(x,y) - D_j\left(\frac{s}{N}, \frac{k}{T}\right) = \frac{sk}{T} - Nxy = N\left(\frac{s}{N} - x\right)y + s\left(\frac{k}{T} - y\right) \le 1 + \frac{s}{T} \le \frac{3}{2}.$$

When $s \in S$, success bounds the discrepancy norm at the grid corner by ρ, and the triangle inequality gives

$$\left(\sum_{j=1}^{d} D_j(x,y)^2\right)^{1/2} \le \left(\sum_{j=1}^{d} D_j\left(\frac{s}{N}, \frac{k}{T}\right)^2\right)^{1/2} + \frac{3\sqrt{d}}{2} \le \rho + \frac{3\sqrt{d}}{2}.$$

This holds almost everywhere on $E \times [0,1]$; squaring and integrating over this set of area p proves (5.4).

In the following claim, we prove a lower bound of the contribution of each D_j to the integral (5.4), i.e., a lower bound of $\int_E \int_0^1 D_j(x,y)^2 \, dy \, dx$ for any each coordinate j.

Claim 5.6. *Let $N \ge 1$ and let $(u_i, v_i) \in [0,1]^2$ for $i = 1, \dots, N$. For $(x, y) \in [0,1]^2$, define*

$$D(x,y) := |\{i \in [N] : u_i < x,\ v_i < y\}| - Nxy.$$

Then for every $E \subseteq [0,1]$ which is a union of intervals, with $p = |E| > 0$,

$$\int_E \int_0^1 D(x,y)^2 \, dy \, dx \ge Cp^3 \log(N+2),$$

where $C > 0$ is an absolute constant.

Proof of Claim 5.6. We apply Roth's empty-rectangle argument [Rot54] on $E \times [0,1]$, letting $n = \lceil \log_2(2N) \rceil$ for succinctness. First we construct two recursive partitions of $[0,1]$, one for each coordinate.

- For the x-coordinate, start with $[0,1]$. Split each interval I at a point z into a left child I_- and a right child I_+ such that
$$|E \cap I_-| = |E \cap I_+| = \frac{|E \cap I|}{2}.$$
Thus the split divides the length of the part of E inside I equally between the children; the children themselves need not have equal lengths. Such a point z exists because the length of E to its left varies continuously with z. Let $\mathcal{I}_a$ denote the collection of intervals after a rounds of splitting. It contains 2^a intervals, each satisfying $|E \cap I| = p2^{-a}$.
- For the y-coordinate, repeatedly split each interval at its midpoint. Let $\mathcal{J}_b$ denote the collection after b rounds. It contains 2^b intervals, each of length 2^{-b}. We construct both partitions through level $n+1$, so that every interval used below also has two children.

For each integer $a = 0, \dots, n$, consider all rectangles $I \times J$ with $I \in \mathcal{I}_a$ and $J \in \mathcal{J}_{n-a}$, which satisfy $|E \cap I| = p2^{-a}$, $|J| = 2^{-(n-a)}$. There are $2^a 2^{n-a} = 2^n$ such rectangles, with disjoint interiors, and, since the point set defining D has N points, at most $N \le 2^{n-1}$ rectangles contain a point in their interior. We select exactly 2^{n-1} rectangles whose interiors contain no points, which over all a, selects $(n+1)2^{n-1}$ rectangles. Next, for each selected rectangle $I \times J$, let us write I_-, I_+ for the left and right children of I, and J_-, J_+ for the lower and upper halves of J. We then define

$$h_{I,J}(x,y) = \begin{cases} +1, & (x,y) \in (I_- \times J_-) \cup (I_+ \times J_+), \\ -1, & (x,y) \in (I_- \times J_+) \cup (I_+ \times J_-), \\ 0, & (x,y) \notin I \times J, \end{cases}$$

with values on the boundaries that may be set to zero. Applying the Cauchy–Schwarz inequality to D and $\sum_{I,J} h_{I,J}$ gives

$$\int_E \int_0^1 D(x,y)^2 \, dy \, dx \geq \frac{\left(\int_E \int_0^1 D(x,y) \sum_{I,J} h_{I,J}(x,y) \, dy \, dx \right)^2}{\int_E \int_0^1 \left(\sum_{I,J} h_{I,J}(x,y) \right)^2 \, dy \, dx}, \tag{5.5}$$

where the sum is over all selected rectangles (I, J). We will deal with the two terms denominator and numerator separately. First we compute the denominator: since the lengths of $E \cap I_-$ and $E \cap I_+$ are equal and so are J_- and J_+, for every fixed y and every fixed x, we have

$$\int_E h_{I,J}(x,y) \, dx = \int_0^1 h_{I,J}(x,y) \, dy = 0, \qquad \int_E \int_0^1 h_{I,J}(x,y)^2 \, dy \, dx = |E \cap I| \, |J| = p 2^{-n}.$$

The functions $h_{I,J}$ associated with the selected rectangles are pairwise orthogonal, since either the rectangles have disjoint interiors, or the finer x-sign has integral zero over E while the coarser sign is constant on that interval. Hence

$$\int_E \int_0^1 \left(\sum_{I,J} h_{I,J}(x,y) \right)^2 \, dy \, dx = (n+1) 2^{n-1} p 2^{-n} = \frac{p(n+1)}{2}. \tag{5.6}$$

We next bound the numerator: for an empty rectangle, every point has at least one coordinate outside the corresponding interval's interior. Its counting indicator is constant almost everywhere in that coordinate, so its integral against $h_{I,J}$ vanishes. Consequently, only the term $-Nxy$ in D contributes:

$$-\int_E \int_0^1 D(x,y) h_{I,J}(x,y) \, dy \, dx = N \int_{E \cap I} \int_J xy h_{I,J}(x,y) \, dy \, dx.$$

To bound this expression, let z be the splitting point of I and put $q = p 2^{-a-1}$, the common length of $E \cap I_-$ and $E \cap I_+$. Then

$$\int_{E \cap I_+} x \, dx - \int_{E \cap I_-} x \, dx = \int_{E \cap I_+} (x - z) \, dx + \int_{E \cap I_-} (z - x) \, dx \geq 2 \int_0^q t \, dt = q^2 = \frac{p^2 2^{-2a}}{4}.$$

The inequality holds because, among sets of length q on either side of z, the interval adjacent to z minimizes the integral of the distance from z. For the two equal halves of J,

$$\int_{J_+} y \, dy - \int_{J_-} y \, dy = \frac{|J|^2}{4}.$$

The definition of $h_{I,J}$ therefore gives

$$\begin{aligned} -\int_E \int_0^1 D(x,y) h_{I,J}(x,y) \, dy \, dx &= N \left(\int_{E \cap I_+} x \, dx - \int_{E \cap I_-} x \, dx \right) \left(\int_{J_+} y \, dy - \int_{J_-} y \, dy \right) \\ &\geq \frac{N p^2 2^{-2a} |J|^2}{16} = \frac{N p^2 2^{-2n}}{16}. \end{aligned}$$

Summing the preceding bounds and using $2^n < 4N$, we obtain

$$-\int_E \int_0^1 D(x,y) \sum_{I,J} h_{I,J}(x,y) \, dy \, dx \geq (n+1) 2^{n-1} \frac{N p^2 2^{-2n}}{16} = \frac{N p^2 (n+1)}{32 \cdot 2^n} \geq \frac{p^2 (n+1)}{128}.$$

The integral is negative, so squaring gives

$$\left(\int_E \int_0^1 D(x,y) \sum_{I,J} h_{I,J}(x,y)\, dy\, dx\right)^2 \geq \frac{p^4(n+1)^2}{128^2}. \tag{5.7}$$

Substituting (5.6) and (5.7) into (5.5) gives, for an absolute constant $C > 0$,

$$\int_E \int_0^1 D(x,y)^2\, dy\, dx \geq Cp^3(n+1) \geq Cp^3 \log(N+2),$$

as required. □

We are now ready to conclude the proof of Lemma 5.5. First apply Claim 5.6 to the points defining D_j and E, we have

$$\int_E \int_0^1 D_j(x,y)^2\, dy\, dx \geq Cp^3 \log(N+2). \tag{5.8}$$

We then sum up (5.8) over j and apply (5.4) to get

$$p\left(\rho + \frac{3\sqrt{d}}{2}\right)^2 \geq Cdp^3 \log(N+2).$$

The one-cycle input forces $\rho \geq \sqrt{d}/2$, since its cost is at least $\sqrt{d}/2$ under every ordering and the success probability is positive. Hence $\rho + 3\sqrt{d}/2 \leq 4\rho$. Using $p \geq 1 - \delta$ and $N = \lfloor T/(2d) \rfloor$, we obtain

$$\rho \geq c_0 p \sqrt{d \log\left(2 + \frac{T}{d}\right)} \geq c_0(1-\delta)\sqrt{d \log\left(2 + \frac{T}{d}\right)}$$

for an absolute constant $c_0 > 0$. Eventually, we may choose $c \leq c_0$ sufficiently small to include the case $d > T/2$ discussed at the very beginning, which will complete the proof. □

Proof of Theorem 5.1. The deterministic assertions follow from Lemma 5.2. For $0 < \delta \leq 1/2$, a uniform randomized guarantee must satisfy both the scalar lower bound of Lemma 5.4, embedded in the first coordinate, and the dimension lower bound of Lemma 5.5. Since $1 - \delta \geq 1/2$ and the maximum of two nonnegative quantities is at least half their sum, these bounds give

$$\rho \geq c\left(\log T + \sqrt{d_* \log(2 + T/d_*)}\right),$$

for an absolute constant $c > 0$, as claimed. □

Acknowledgments

The authors thank Rom Pinchasi for the fruitful discussions and initial ideas for the deterministic algorithm. The authors used ChatGPT to assist with mathematical arguments. Specifically, the initial quota correction argument in the reduction was suggested by ChatGPT-5.5 and subsequently refined by the authors. The lower-bound argument was suggested by ChatGPT-6 Astra. The authors take full responsibility for all the content of this work.

References

[AABK23] Anders Aamand, Mikkel Abrahamsen, Lorenzo Beretta, and Linda Kleist. Online sorting and translational packing of convex polygons. In *Proceedings of SODA*, pages 1806–1833, 2023.

[AB24] Gergely Ambrus and Rainie Bozzai. Colorful vector balancing. *Mathematika*, 70(4):e12274, 2024.

[ABB+24] Mikkel Abrahamsen, Ioana O. Bercea, Lorenzo Beretta, Jonas Klausen, and László Kozma. Online sorting and online TSP: Randomized, stochastic, and high-dimensional. In *Proceedings of ESA*, volume 308 of *LIPIcs*, pages 5:1–5:15, 2024.

[Ade26] Ishaq Aden-Ali. Optimal online discrepancy minimization in linear time, 2026.

[AH26] Gergely Ambrus and Rainie Heck. A note on the Steinitz lemma. *Mathematika*, 72(2):e70085, 2026.

[ALS21] Ryan Alweiss, Yang P. Liu, and Mehtaab Sawhney. Discrepancy minimization via a self-balancing walk. In *STOC*, pages 14–20. ACM, 2021.

[APV26] Yossi Azar, Debmalya Panigrahi, and Or Vardi. Online metric TSP: beyond the $\sqrt{n}$ barrier. In *ICALP*, volume 374 of *LIPIcs*, pages 18:1–18:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2026.

[AT26] Dylan J. Altschuler and Konstantin Tikhomirov. Online Beck–Fiala down to logarithmic sparsity, 2026.

[Bár24] Imre Bárány. A matrix version of the Steinitz lemma. *Journal für die reine und angewandte Mathematik*, 809:261–267, 2024.

[Ber25] Christian Bertram. Online metric TSP. In *ESA*, volume 351 of *LIPIcs*, pages 80:1–80:9. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2025.

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

[BJSS20] Nikhil Bansal, Haotian Jiang, Sahil Singla, and Makrand Sinha. Online vector balancing and geometric discrepancy. In *STOC*, pages 1139–1152. ACM, 2020.

[BN26] Nikhil Bansal and Aleksandar Nikolov. Discrepancy theory: An algorithmic and geometric perspective, 2026. Manuscript for an unpublished monograph.

[Cho94] Sergej Chobanyan. Convergence a.s. of rearranged random series in Banach space and associated inequalities. In Jørgen Hoffmann-Jørgensen, James Kuelbs, and Michael B. Marcus, editors, *Probability in Banach Spaces, 9*, volume 35 of *Progress in Probability*, pages 3–29. Birkhäuser, Boston, MA, 1994.

[EW20] Friedrich Eisenbrand and Robert Weismantel. Proximity results and faster algorithms for integer programming using the Steinitz lemma. *ACM Transactions on Algorithms*, 16(1):1–14, 2020.

[GS80] V. S. Grinberg and S. V. Sevastyanov. Value of the Steinitz constant. *Functional Analysis and Its Applications*, 14(2):125–126, 1980.

[IKKP19] Sungjin Im, Nathaniel Kell, Janardhan Kulkarni, and Debmalya Panigrahi. Tight bounds for online vector scheduling. *SIAM Journal on Computing*, 48(1):93–121, 2019.

[KRR24] Janardhan Kulkarni, Victor Reis, and Thomas Rothvoss. Optimal online discrepancy minimization. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1832–1840. ACM, 2024.

[Ope26] OpenAI. Sharing AI progress in mathematics. Online announcement, 2026. 6 October 2026; see result 097 in the accompanying catalogue.

[OPW24] Timm Oertel, Joseph Paat, and Robert Weismantel. A colorful Steinitz lemma with application to block-structured integer programs. *Mathematical Programming*, 204:677–702, 2024.

[Rot54] Klaus F. Roth. On irregularities of distribution. *Mathematika*, 1(2):73–79, 1954.

[Sch72] Wolfgang M. Schmidt. Irregularities of distribution, VII. *Acta Arithmetica*, 21(1):45–50, 1972.

[Spe77] Joel Spencer. Balancing games. *Journal of Combinatorial Theory, Series B*, 23(1):68–74, 1977.

[Spe94] Joel Spencer. *Ten Lectures on the Probabilistic Method*. Society for Industrial and Applied Mathematics, 2nd edition edition, 1994.

[Ste13] Ernst Steinitz. Bedingt konvergente reihen und konvexe systeme. *Journal für die reine und angewandte Mathematik*, 143:128–176, 1913.

[TW80] Robert Tijdeman and Gerold Wagner. A sequence has almost nowhere small discrepancy. *Monatshefte für Mathematik*, 90(4):315–329, 1980.