The Price of Geometry for Private Distance-Sensitive Filters: ℓ_∞ versus ℓ_2

Paul Cesaretti @
The Graduate Center, City University of New York, USA

Abstract

A distance-sensitive set representation stores a set S in a metric space and must accept points within distance r of S and reject points farther than cr, each with error probability at most α. Under differential privacy, even exact membership costs a privacy loss of $\ln(1/\alpha)$; we ask how much *more* the geometry of the space forces. Earlier lower bounds cover only restricted mechanisms, such as partition-based ones or those whose outputs contain no scattered false positives ("dust"). We prove lower bounds for every mechanism.

In ℓ_∞^d with $c < 3$ the extra cost is $\Theta_c(d)$, which proves a recent conjecture in this range; when $\lfloor cr \rfloor - r + 1$ divides $2r$, a randomly shifted grid is optimal up to $\tilde{O}(\alpha)$ per dimension. In ℓ_2^d the extra cost is instead $\Theta_c(\min(d, \ln(1/\alpha)))$: a new mechanism that outputs a single very large ball shows that dimensions beyond about $\ln(1/\alpha)$ are free, so partition-based constructions are optimal in ℓ_∞ but not in ℓ_2. The same arguments bound the space of every correct representation of k-sets, private or not: $\Omega(k(\log(1/\alpha) + d))$ bits in ℓ_∞^d with $c < 3$, and $\Omega_c(k(\log(1/\alpha) + \min(d, \log(1/\alpha))))$ bits in every ℓ_p^d.

2012 ACM Subject Classification Security and privacy → Data anonymization and sanitization; Theory of computation → Data structures design and analysis

Keywords and phrases differential privacy, distance-sensitive filters, approximate membership, lower bounds, isoperimetry, concentration of measure

Acknowledgements The author used Claude to perform grammar revision, spelling checks, paraphrasing and rewording. Additionally, Claude was consulted to run numerical checks on arguments for missing cases and sign errors, and the author took suggestions from Claude for revising proofs, proof strategies, and drafting LaTeX figures. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the content of the publication.

1 Introduction

A *filter* stores a set S compactly and answers membership queries, allowing one-sided or two-sided errors [7, 10]. *Distance-sensitive* filters [21, 17] generalize this to a metric space (U, d): for an approximation factor $c > 1$, a query q should be accepted if $d(q, S) \le r$ and rejected if $d(q, S) > cr$, and may be answered arbitrarily in between.

Motivation. A distance-sensitive filter answers the question "is some element of S close to q?", which is the natural query when records are noisy or approximate, and which arises precisely where the records are sensitive. In privacy-preserving record linkage, for instance, hospitals encode patient identifiers as Bloom filters of character bigrams and link records whose encodings have high Dice similarity, so that spelling errors are tolerated [29]. One of the motivating queries for distance-sensitive filters is "is there a DNA sequence close to this one in your database?" [21], yet even aggregate genomic statistics can reveal that an individual took part in a study [20]. Near-duplicate detection, similarly, checks a submission against an archive that may itself be confidential, such as a collection of manuscripts under review. In each of these settings the filter is useful because it can be handed to others and queried without contacting the data holder, and for the same reason releasing it leaks S.

Unmodified Bloom filters offer only a "better than nothing" form of privacy [6], and the encodings used for record linkage have been inverted by constraint-satisfaction and frequency attacks [23, 25]. More fundamentally, whoever holds a filter can query every point of U, and the accept region of an accurate filter traces S to within distance cr. Cryptographic protocols such as private set intersection protect S, but only by requiring the data holder to take part in every query. Differential privacy instead gives a guarantee that survives publication: the released representation, and anything computed from it, changes little when one element of S is added or removed.

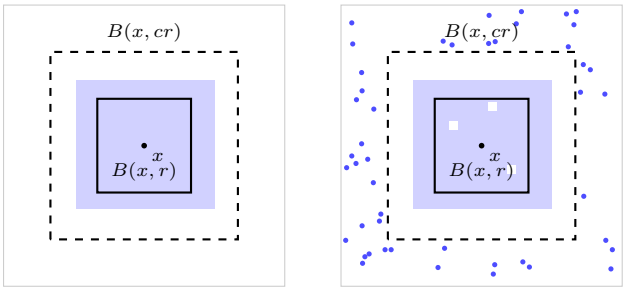
Background. Patel, Persiano, Seo and Yeo [28] initiated the study of *differentially private* (DP) [15] set representations, in which neighbouring inputs differ by one element and the released representation must hide which. For exact membership with error α, they show that a privacy loss of $\varepsilon \approx \ln(1/\alpha)$ is both necessary and achievable. Differentially private distance-sensitive set representations (*DPDS-representations*) were introduced in [12], together with a construction for Hamming space, and the partition framework of [11] extends them to general metrics. That framework releases only the occupancy of a public *admissible* partition, one whose cells have diameter at most $(c-1)r$, and achieves $\varepsilon = \ln(2\kappa/\alpha)$ [11, Thm. 3.4], where the *straddle number* κ is the largest number of cells that a ball of radius r can meet. For partitions of ℓ_∞^d at constant c, the secluded-partition bound of Vander Woude et al. [30, 31] forces $\kappa \geq (\frac{c+1}{c-1})^d$, so every partition-based representation has $\varepsilon = \Omega(d)$ [11, Thm. 5.4]. Whether *every* representation must pay this geometric cost was left open as a conjecture [11, Conj. 8.3].

The question. Fix the universe $U = \mathbb{Z}_N^d$ with a toroidal ℓ_p metric, a radius r, a gap $c > 1$ and an error α, and let ε^* be the least privacy loss of any (c, r, α)-correct mechanism (Section 2). We call

$$\Gamma \;:=\; \varepsilon^* - \ln\frac{1}{\alpha}.$$

the *geometric cost*, and ask how it depends on d, c, α and the norm. Since the baseline $\ln\frac{1}{\alpha}$ is paid even without geometry, Γ isolates the price of answering distance queries rather than membership queries. For partition-based representations of ℓ_∞^d the answer is known to be $\Gamma \geq d\ln\frac{c+1}{c-1}$ (Section 6); what stands in the way of extending such a bound to *every* representation is the form of the correctness guarantee itself.

Why this is hard: dust. Correctness is a *per-query* guarantee: each fixed far point is accepted with probability at most α. A single output may therefore accept about an α fraction of the far points around the data point (Figure 1); in ℓ_∞^{20} with $r = 1$, $c = 2$ and $\alpha = 0.01$, that amounts to about 10^{12} scattered false positives. Earlier lower-bound arguments avoid this *dust* by forbidding false negatives [17], by assuming that the output is clean around each data point, or by restricting the mechanism, for instance to partitions. With clean outputs, the constraints imposed by two nearby data points are rigid, and Brunn–Minkowski-type isoperimetry then forces $\Gamma \geq d\ln\frac{c+1}{c-1}$. Dust turns these hard constraints into soft ones that need hold only for most test points, and a mechanism could conceivably exploit the resulting slack to escape the geometry. We show that in ℓ_∞ it essentially cannot, whereas in ℓ_2 it can, up to a barrier of $\ln(1/\alpha)$. It has been suggested that a proof of the conjecture needs a fingerprinting-style argument that does not assume the global consistency of a fixed partition [11], and our test families are arguments of this kind: each one examines a single draw of the accept set on many queries at once and assumes nothing about its structure.



(a) clean output (b) per-query correct output (dust)

Figure 1 The main difficulty, in ℓ_∞^2 with $c = 2$: the near region $B(x, r)$ (solid) and the boundary of $B(x, cr)$ (dashed). (a) A *clean* output accepts all of $B(x, r)$ and nothing beyond cr; earlier lower bounds assume this. (b) A *per-query correct* output may also have holes in $B(x, r)$ and accepted points (dust) beyond cr, since each fixed point may be wrong with probability α. Lower bounds must survive this dust.

1.1 Our results

Throughout, $s = s(c, r) := \lfloor cr \rfloor - r + 1$ denotes the largest grid-cell side whose r-dilation stays within distance cr of every point of the cell, and $L(c, r)$ denotes the box parameter of Lemma 10; it satisfies $L(c, r) \le \lfloor 2r/s \rfloor$, with equality when $s \mid 2r$, and $L(c, r) \ge 1$ essentially iff $c < 3$. Subscripts on asymptotic notation, as in O_c, Ω_c and Θ_c, indicate that the hidden constants depend only on c. We state the main results informally here and give their formal versions in Theorems 14 and 26 and Corollary 28.

▶ **Theorem 1** (ℓ_∞: every dimension costs; informal)**.** *Let $1 < c < 3$ and $L = L(c, r) \ge 1$. Every (ε, δ)-DP mechanism that is (c, r, α)-correct on the d-dimensional ℓ_∞ grid, with $\delta \le \frac{1}{4}$ and α small enough, has geometric cost*

$$\Gamma \;\ge\; d \ln(L + 1) - \tilde{O}(\alpha d) - O(1).$$

A randomly shifted grid achieves $\Gamma \le d \ln(1 + \frac{2r}{s})$, which equals $d \ln(L+1)$ when $s \mid 2r$; hence $\Gamma = \Theta_c(d)$.

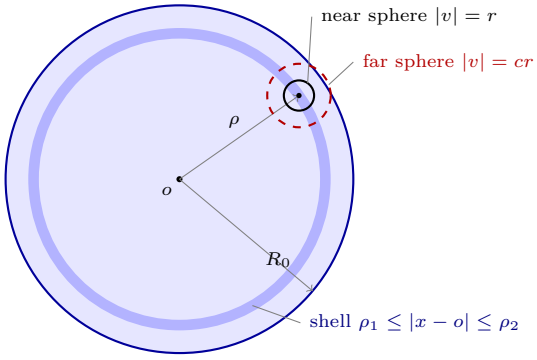
The condition $s \mid 2r$ holds, for example, whenever $r = 1$ and $c \in (1, 3)$. In particular, Theorem 1 proves the conjecture of [11] for every $c < 3$ and every α below a constant $\alpha_0(c)$, on the conjecture's bounded universe $[0, m]^d$ (Corollary 15). In this range, the lower bound that [11] proves for partition-based representations therefore holds for every representation.

▶ **Theorem 2** (ℓ_2: the cost saturates at $\ln(1/\alpha)$; informal)**.** *Let $c > 1$ be fixed, and let α be small and r large enough. In d-dimensional ℓ_2, every (ε, δ)-DP, (c, r, α)-correct mechanism with $\delta \le \frac{1}{4}$ has*

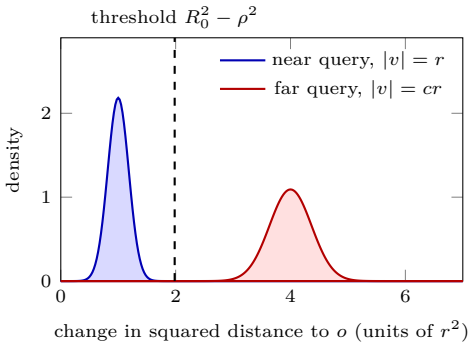
$$\Gamma \;\ge\; g_2(c) \min\left(d, \log_2 \frac{1}{4\alpha}\right) - O_c(1 + \log d),$$

and once $d \gg \ln \frac{1}{\alpha}$ also $\Gamma \ge \big(\min(1, \frac{2}{c^2-1}) - o(1)\big) \ln \frac{1}{\alpha} - O(1)$. Conversely, a ball blanket achieves $\Gamma \le d \ln \frac{c+1}{c-1}$, and a new big-ball mechanism achieves $\Gamma \le \frac{4c}{(c-1)^2} \ln \frac{2}{\alpha} + O_c(1)$ in every dimension $d \ge \frac{5c}{c-1} \ln \frac{2}{\alpha}$. Hence $\Gamma = \Theta_c\big(\min(d, \ln \frac{1}{\alpha})\big)$.

The big-ball bound in Theorem 2 does not depend on d. It comes from the *big-ball mechanism* (Section 4.3), which outputs a Euclidean ball of radius about $r\sqrt{d}\,(c-1)/\sqrt{8 \ln(2/\alpha)}$,



(a) placement (schematic)



(b) exact distributions, $d = 1000$

Figure 2 The big-ball mechanism. (a) The output is a ball $B(o, R_0)$ of radius about $r\sqrt{d}\,(c-1)/\sqrt{8\ln(2/\alpha)}$, with x in a thin shell just inside its boundary (not to scale). (b) A query $x + v$ changes the squared distance to o by $|v|^2 + 2\rho\langle U, v\rangle$ for a uniformly random direction U. In high dimension the cross term concentrates, so the threshold $R_0^2 - \rho^2$ separates near ($|v| = r$) from far ($|v| = cr$) queries up to probability α. Exact densities for $d = 1000$, $c = 2$, $\alpha = 10^{-6}$.

Table 1 Bounds on the geometric cost $\Gamma = \varepsilon^* - \ln\frac{1}{\alpha}$. Lower bounds hold for every mechanism.

	lower bound on Γ (every mechanism)	upper bound on Γ
ℓ_∞^d, $c < 3$	$d\ln(L+1) - \tilde{O}(\alpha d)$	$d\ln(1+\frac{2r}{s})$ (equal if $s \mid 2r$)
ℓ_2^d, $d \lesssim \ln\frac{1}{\alpha}$	$g_2(c)\,d - O(\log d)$	$d\ln\frac{c+1}{c-1}$
ℓ_2^d, $d \gg \ln\frac{1}{\alpha}$	$\max\{g_2(c)\log_2\frac{1}{4\alpha},\ \min(1, \frac{2}{c^2-1})\ln\frac{1}{\alpha}\}$	$\frac{4c}{(c-1)^2}\ln\frac{2}{\alpha} + O(1)$
ℓ_p^d, $p < \infty$	$g_p(c)\min(d, \log_2\frac{1}{4\alpha})$ (also spectral, Rem. 23)	open for $p \neq 2$

placed so that the data point lies in a thin shell just inside its boundary. Such a ball is far too large to be a cell of any admissible partition (Section 6), and every realization misclassifies some near and some far points; concentration of measure on the sphere nevertheless keeps the error of every *fixed* query below α (Figure 2). This is exactly the kind of mechanism that dust permits, and its existence shows that per-query errors are worth a great deal in high-dimensional ℓ_2, whereas Theorem 1 shows that they are worth almost nothing in ℓ_∞. Table 1 summarizes the bounds, and Figure 3 plots them for $c = 2$.

Space. Every lower bound in this paper comes from a *test family* (Definition 8), and through an information-complexity direct sum [13, 5] the same test families also bound the space of filters that need not be private (Section 5).

▶ **Theorem 3** (Space; informal). *For α small enough, every (c, r, α)-correct representation of k-sets, private or not, uses $\Omega(k(\log\frac{1}{\alpha} + d))$ bits in ℓ_∞^d with $c < 3$, and, for r large enough, $\Omega_c(k(\log\frac{1}{\alpha} + \min(d, \log\frac{1}{\alpha})))$ bits in every ℓ_p^d with $p < \infty$.*

In ℓ_∞ the bound is $k\big(\log_2\frac{1}{\alpha} + d\log_2(L+1)\big)(1 - o(1))$ bits (Corollary 28). This matches the $O(k\log(\kappa/\alpha))$ space of partition-based representations [11] when $s \mid 2r$, and shows that, in ℓ_∞ with $c < 3$, representations of arbitrary internal structure also need the $\log\kappa$ term.

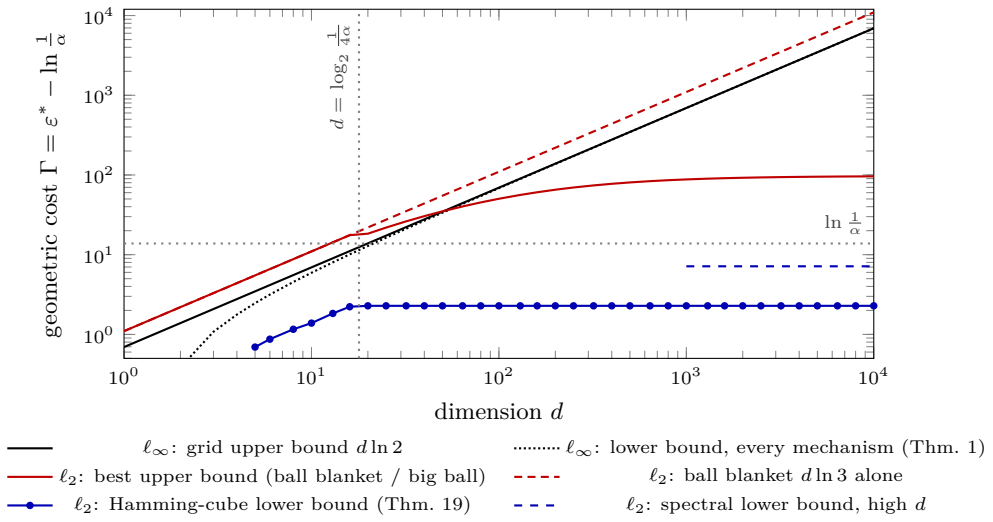


Figure 3 The dichotomy for $c = 2$, $r = 1$, $\alpha = 10^{-6}$. In ℓ_∞ the lower bound of Theorem 1 matches the grid and grows linearly in d. In ℓ_2 the best upper bound (ball blanket, then big ball, computed exactly) flattens out, and the lower bounds (Hamming cubes, then the spectral bound of Section 4.2 with $M = 8$) are linear in d only up to $d \approx \log_2 \frac{1}{4\alpha}$. The remaining gap is discussed in Section 6.

1.2 Techniques

All of our lower bounds follow one template (Lemma 9): for every possible accept set A we choose a set $G(A)$ of centres that "look correct", and we need two properties:

(T1) under input $\{x\}$, $x \in G(A)$ with probability $1 - \beta$;
(T2) deterministically, $|G(A)| \le \omega |A|$.

Averaging the DP guarantee over x then gives $\varepsilon \ge \ln \frac{1}{\alpha} + \ln \frac{1}{\omega} - O(1)$, so the whole difficulty lies in establishing (T2) in the presence of dust. Figure 5 in the appendix summarizes the template and its three instantiations.

ℓ_∞: robust box inequality. We test a centre on a sparse box of $(L+1)^d$ near points with spacing h, together with the points pushed one step beyond each face, which are far. If two good boxes overlap, then on every axis line through a shared point some point is pushed for one box and lies inside the other, so the conflicts between overlapping boxes sit at *single points on axis lines*. By counting these conflicts, charging each error a bounded number of times, and applying Fano's inequality together with subadditivity of entropy across coordinates, we show that dust changes the hard-case bound $(L+1)^{-d}$ only by a factor $e^{O(\beta \log(1/\beta))d}$ (Theorem 11).

ℓ_2: why boxes fail, and what replaces them. In ℓ_p with $p < \infty$ the corners of a ball are not near points, and a sign-averaging (cotype) argument shows that box tests fit into only $O_c(1)$ coordinates (Remark 17). We use two substitutes instead. The first is *Hamming subcubes*: on a cube $\{0, h\}^D$ the ℓ_p distance is a function of Hamming distance, so the hard case is solved exactly by Kleitman's diameter theorem [22], Harper's theorem [19] and normalized matching (Lemma 18), while dust is handled by a union bound over the $2^D \approx 1/\alpha$ cube points. The second is a *two-scale spectral inequality*, which replaces worst-case test points by a random walk run for times $t_1 < t_2$. Positivity, self-adjointness and log-convexity of the heat semigroup bound the number of centres whose time-t_1 walk is mostly accepted and

whose time-t_2 walk is mostly rejected (Theorem 21). This bound is dimension-free and needs no structure of A; Figure 6 in the appendix illustrates the test.

1.3 Related work

Filters and their space lower bounds go back to [7, 10], and distance-sensitive filters were introduced in [21]. Goswami et al. [17] prove a geometric space lower bound for Hamming-space filters *without false negatives*, via Harper's theorem; their argument requires the data to lie in the erosion of the accept set with certainty, which a DP filter cannot guarantee. Patel et al. [28] give tight privacy–utility and space bounds for DP exact membership, which have no geometric term. Earlier private variants of Bloom filters either randomize the bits without a utility guarantee for membership queries [1] or use a Bloom encoding as a local-DP report rather than as the protected object [16]. Our title echoes the line of work on the geometry of differential privacy [18, 26], which determines, up to polylogarithmic factors, the least error with which any mechanism can answer a given workload of linear queries. There, the relevant geometry is that of the convex body spanned by the queries, error is mean squared at constant ε, and the lower bounds come from volume (packing) arguments under pure DP and from hereditary discrepancy under approximate DP. Here, by contrast, the geometry is that of the metric space, accuracy is a per-query error probability α, and we bound the privacy loss as $\alpha \to 0$. A packing argument relies on the whole output being accurate for its input with constant probability; under a per-query guarantee that is a for-all event, which need not have constant probability, and what remains is dust, for which test families are our substitute. DPDS constructions appear in [12, 11]. Differentially private approximate near-neighbour *counting* [2, 3] uses locality-sensitive filters but addresses counting rather than the decision problem, and secluded partitions [30, 31] lower-bound the straddle number of partition-based constructions but not that of general mechanisms. Finally, our spectral inequality is in the spirit of hypercontractive lower bounds for locality-sensitive hashing [24, 27], although the regime differs: we need near-acceptance close to 1 and far-acceptance close to 0 simultaneously.

1.4 Limitations

Our results leave several gaps. In ℓ_∞ the box tests exist only for $c < 3$ (Lemma 10), and when $s \nmid 2r$ our per-dimension constant $\ln(L+1)$ falls short of the grid's $\ln(1+2r/s)$, although in the continuum limit the shortfall is at most $\ln(1+\frac{c-1}{2})$ per dimension. In ℓ_2 the order $\Theta_c(\min(d, \ln\frac{1}{\alpha}))$ is settled but the constants are not: at $c=2$ the ratio $\varepsilon^*/\ln\frac{1}{\alpha}$ lies between 1.67 and 9, and both of our lower-bound techniques stop at $2\ln\frac{1}{\alpha}$ (Section 6). For ℓ_p with $p \notin \{2, \infty\}$ we prove the Hamming-cube bound, which grows linearly in d up to $d \approx \log_2\frac{1}{\alpha}$, but we do not know whether the cost saturates. Finally, our upper bounds are single-point mechanisms, whereas our lower bounds hold for every k.

2 Model and the transfer lemma

Let (U, d) be a metric space; in this paper $U = \mathbb{Z}_N^d$ with the toroidal ℓ_p metric ($1 \le p \le \infty$). Write $B(q, \rho) = \{u \in U : d(u, q) \le \rho\}$. Sets $S, S' \subseteq U$ of size at most k are *neighbouring*, written $S \sim S'$, if one is obtained from the other by adding a single element.

▶ **Definition 4** (DPDS-representation)**.** *Let $k \ge 1$, $r > 0$, $c > 1$ and $\alpha \in (0, \frac{1}{2})$. Consider a pair* (Rep, Query) *of randomized algorithms with access to a public random string* R *drawn independently of the data. Here* $\mathsf{Rep}(S; \mathsf{R})$ *maps a set $S \subseteq U$ with $|S| \le k$ to a string* Π*, and*

$\mathsf{Query}(\Pi, q; \mathsf{R}) \in \{0,1\}$. *The pair is an* (ε, δ)-differentially private (c, r, α)-distance-sensitive set representation *((c, r, α)-DPDS-representation) if:*

(i) Completeness. *If* $d(u, q) \le r$ *for some* $u \in S$, *then* $\Pr[\mathsf{Query}(\Pi, q; \mathsf{R}) = 1] \ge 1 - \alpha$.
(ii) Soundness. *If* $d(u, q) > cr$ *for every* $u \in S$, *then* $\Pr[\mathsf{Query}(\Pi, q; \mathsf{R}) = 1] \le \alpha$.
(iii) Privacy. *For all* $S \sim S'$ *and measurable* T,

$$\Pr[(\mathsf{Rep}(S; \mathsf{R}), \mathsf{R}) \in T] \le e^{\varepsilon} \Pr[(\mathsf{Rep}(S'; \mathsf{R}), \mathsf{R}) \in T] + \delta.$$

The probabilities in (i) and (ii) are taken over R *and the coins of both algorithms, separately for each query* q. *A pair satisfying (i) and (ii) is* (c, r, α)-correct. *Its* space *is* $|\Pi|$ *in bits;* R *is not counted.*

▶ Remark 5 (On the model). With Query deterministic, Definition 4 is the *non-oblivious* guarantee of [11, Defs. 2.2 and 2.3]: each query is fixed independently of the public data R, such as a randomly drawn partition, and of Π. Allowing randomized queries is only for generality. This is the weakest per-query guarantee, so all our lower bounds apply to oblivious and non-oblivious representations alike. Our upper bounds are pure DP, use no public data and have deterministic queries, so they are oblivious.

Two features of the definition are essential. First, correctness is per query. The *for-all* guarantee, under which a single draw is correct on every query, forces e^{ε} to grow with the packing number of U [11, Thm. 6.3], by a packing argument as in [18]; per-query correctness is precisely what allows dust. Second, the error is two-sided, and this is unavoidable: under pure DP, randomized response on the membership bit forces both error rates to be $\Omega(e^{-\varepsilon})$ (cf. [28]).

▶ Remark 6 (Why $\mathbb{Z}_N^d$ with the toroidal ℓ_p metric). The torus has two properties that our arguments use. It is symmetric: translations act transitively by isometries, so all points look alike. This lets Lemma 9 average over a uniformly random centre with no boundary terms, and it makes ε^* depend only on (d, p, r, c, α), not on where the data lie. It is also finite. Our arguments count points (a test family compares $|G(A)|$ with $|A|$, and centres are drawn uniformly), and a finite universe makes both meaningful; letting $r \to \infty$ recovers the continuum.

The restriction costs little for lower bounds. The ℓ_∞ box tests are local, within distance $3r$ of their centre, so Theorem 14 transfers from the torus to a bounded box $[0, m]^d$ (Corollary 15), and from there to $\mathbb{R}^d$, since a mechanism on $\mathbb{R}^d$ restricts to a fine lattice in a box. As for the norms, the ℓ_p metrics are those of prior work: Hamming space [17, 12] is ℓ_1 on $\{0,1\}^d$, and partition-based constructions exist in every ℓ_p [11]. Moreover ℓ_∞ and ℓ_2 are the two ends of our dichotomy.

▶ Remark 7 (Accept sets). Sample the coins of Query once and use them for all queries. This defines a random *accept set*

$$A = \{q : \mathsf{Query}(\Pi, q; \mathsf{R}) = 1\},$$

the accept region of the output with the query coins fixed. Each query's acceptance probability is unchanged, so A satisfies (i) and (ii) query by query. Moreover A is a post-processing of (Π, R), so it inherits (iii). All our arguments concern the random set A; no structure of Rep is assumed.

We also call Rep, together with Query, a *mechanism*. A mechanism for k-sets restricts to one for sets of size at most one, with the same parameters. So all privacy lower bounds

below may, and do, assume $k = 1$, and they then hold for every k. We write A_x and $A_\emptyset$ for the accept sets on inputs $\{x\}$ and $\emptyset$. Since every point is far from $\emptyset$, $\mathbb{E}|A_\emptyset| \le \alpha|U|$.

Every privacy lower bound in this paper comes from one argument. DP controls a single event at a time: an event of probability p under the input $\{x\}$ has probability at least $e^{-\varepsilon}(p-\delta)$ under the neighbouring input $\emptyset$. Correctness supplies such an event. A correct output must behave recognizably around its data point, accepting what is near and rejecting what is far, while under $\emptyset$ the accept set is small: $\mathbb{E}|A_\emptyset| \le \alpha|U|$. A *test family* makes this precise. It is a rule that reads only the accept set and returns the set $G(A)$ of centres around which A looks correct. The rule must tolerate dust, so that the true data point passes with high probability, yet be strict enough that few points pass by accident. The second requirement is a deterministic counting bound, $|G(A)| \le \omega|A|$ for every set A, and it is where all the geometry enters. Lemma 9 turns it into $\varepsilon \ge \ln\frac{1}{\alpha} + \ln\frac{1}{\omega} - O(1)$.

▶ **Definition 8** (Test family). *A map $A \mapsto G(A) \subseteq U$ is a (β,ω)-*test family at error α' *if:*

(T1) *for every $x \in U$ and every random set A that is (c,r,α')-correct for the input $\{x\}$, $\Pr[x \in G(A)] \ge 1-\beta$;*
(T2) *$|G(A)| \le \omega|A|$ for every $A \subseteq U$.*

▶ **Lemma 9** (Transfer). *If a (β,ω)-test family at error α exists and $\beta+\delta<1$, then every (ε,δ)-DP, (c,r,α)-correct mechanism satisfies*

$$\varepsilon \;\ge\; \ln\frac{1}{\alpha} + \ln\frac{1}{\omega} + \ln(1-\beta-\delta).$$

Proof. For every x, DP applied to the event $\{x \in G(A)\}$ gives $1-\beta \le \Pr[x \in G(A_x)] \le e^{\varepsilon}\Pr[x \in G(A_\emptyset)] + \delta$. Averaging over uniform $x \in U$,

$$\frac{1}{|U|}\sum_x \Pr[x \in G(A_\emptyset)] = \frac{\mathbb{E}|G(A_\emptyset)|}{|U|} \le \omega\frac{\mathbb{E}|A_\emptyset|}{|U|} \le \omega\alpha.$$

Hence $1-\beta-\delta \le e^{\varepsilon}\omega\alpha$. ◀

The lemma reduces privacy lower bounds to a deterministic counting problem: (T2) for *arbitrary* sets A, dust included. Taking $G(A) = A$ (so $\omega = 1$) recovers the baseline $\varepsilon \ge \ln\frac{1}{\alpha} - O(1)$. The geometric cost is the extra factor $1/\omega$.

3 ℓ_∞: a robust box inequality

The lower bound in ℓ_∞ comes from a single test that a correct output must pass around its data point. Place a sparse grid of $(L+1)^d$ points with spacing h inside the near region of a centre x, and add the points one grid step beyond each face. Because the ℓ_∞ ball is itself a cube, the grid can reach its faces while every pushed point lies in the far region. A correct output must accept the grid and reject the pushed points. If it did so exactly, the boxes of two centres that pass the test could not overlap, since an overlap puts a pushed point of one inside the box of the other. Each passing centre would then own $(L+1)^d$ accepted points, so $|G(A)| \le (L+1)^{-d}|A|$, and Lemma 9 would give a geometric cost of $d\ln(L+1)$. With $h = s$ the test mirrors the grid mechanism of Proposition 13: the box spacing is the cell side, and when $s \mid 2r$, $(L+1)^d = (1+2r/s)^d$ is the grid's straddle number. The difficulty is dust. A per-query correct output may miss a few box points and accept a few pushed points, and then the boxes of passing centres can overlap. Section 3.2 shows that every overlap forces an error at a single point on an axis line (Figure 4), and that dust changes the bound by a factor of only $e^{O(\beta\log(1/\beta))d}$.

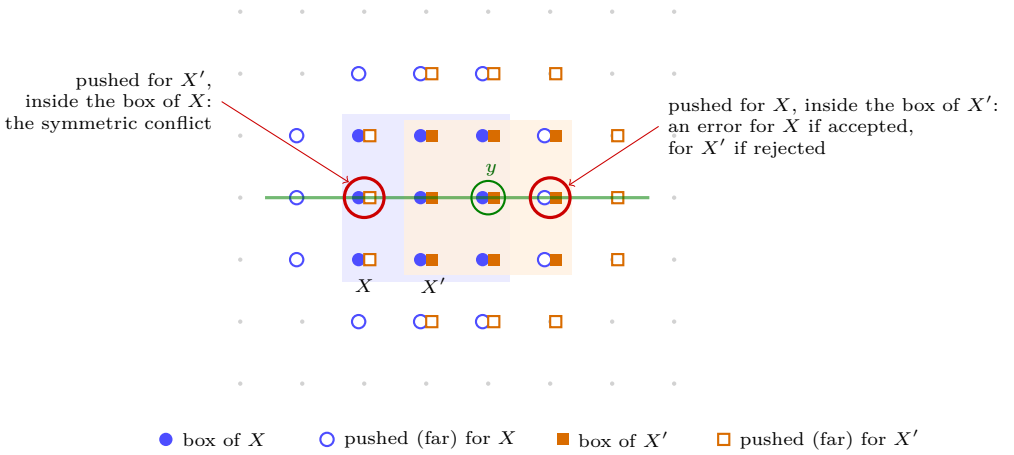


Figure 4 Conflicts sit at single points on axis lines ($L = 2$, lattice units, $X' = X + (1, 0)$). The boxes of X and X' share y. On the axis line through y, the circled point on the right is pushed for X but inside the box of X', so whatever A does there, one of the two centres errs; the circled point on the left is the symmetric conflict.

3.1 Box tests

Fix integers $L, h \geq 1$ and a. For a centre x let

$$\mathrm{box}(x) = x + a\mathbf{1} + h\{0, \dots, L\}^d,$$

$$\mathrm{push}(x) = \big\{x + a\mathbf{1} + hu : \ u_j \in \{-1, L+1\} \text{ for exactly one } j, \ \ u_i \in \{0, \dots, L\} \text{ for } i \neq j\big\}.$$

The triple (L, h, a) is *admissible* if every box point is near x and every pushed point is far from x.

▶ **Lemma 10.** *Let $L(c, r)$ be the largest L for which an admissible triple exists. Then*

$$L(c, r) = \max\left\{L \geq 1 : \ \left\lceil \frac{2(s+r)}{L+2} \right\rceil \leq \left\lfloor \frac{2r}{L} \right\rfloor \right\} \ \leq \ \left\lfloor \frac{2r}{s} \right\rfloor.$$

Equality holds when $s \mid 2r$, with $h = s$ and $a = -r$. In particular $L(c, r) \geq 1$ requires $s \leq 2r$, hence $c < 3$; and for fixed $c < 3$, $L(c, r) \to \lceil \frac{2}{c-1} \rceil - 1$ as $r \to \infty$.

The proof is in Appendix B. Within a residue class modulo h, the map $y \mapsto (y - \text{offset})/h$ identifies the box and pushed points of all centres of that class with points of $\tilde{U} = \mathbb{Z}^d_{N/h}$. In these coordinates

$$\mathrm{box}_X = X + \{0, \dots, L\}^d,$$

and push_X consists of the points with exactly one coordinate equal to $X_j - 1$ or $X_j + L + 1$ and all others in the box range. In the hard case, with every box point accepted and no pushed point accepted, the boxes of good centres are disjoint and $|G| \leq (L+1)^{-d}|A|$.

3.2 The robust box inequality

Let $\mathrm{h}(t) = -t \ln t - (1-t)\ln(1-t)$, and define

$$\psi_L(t) = \mathrm{h}(t) + t \ln L \ \text{ for } 0 \leq t \leq \tfrac{L}{L+1}, \qquad \psi_L(t) = \ln(L+1) \ \text{ for } t > \tfrac{L}{L+1}.$$

The two pieces agree at $t = \frac{L}{L+1}$. On $[0, \frac{L}{L+1}]$, $\psi_L'(t) = \ln \frac{(1-t)L}{t} \geq 0$ and $\psi_L'' < 0$. So ψ_L is concave and nondecreasing on $[0, \infty)$, and $\lambda \mapsto e^{-d\psi_L(\lambda)}$ is convex and nonincreasing.

▶ **Theorem 11** (Robust box inequality). *Let $L \geq 1$, $n \geq 2L+3$, $\tilde{A} \subseteq \mathbb{Z}_n^d$ and $\eta, \theta \in [0,1)$. Let G be the set of $X \in \mathbb{Z}_n^d$ with $|\mathrm{box}_X \setminus \tilde{A}| \leq \eta(L+1)^d$ and $|\mathrm{push}_X \cap \tilde{A}| \leq \theta|\mathrm{push}_X|$. Let $\beta := \frac{L(\eta+2\theta/(L+1))}{1-\eta}$. Then*

$$|G| \;\leq\; \frac{e^{d\psi_L(\beta)}}{1-\eta}\,(L+1)^{-d}\,|\tilde{A}|.$$

Proof sketch; details in Appendix C. For $y \in \tilde{A}$ let M_y be the good centres whose box contains y; their offsets $y - X \in \{0,\dots,L\}^d$ are distinct. *Conflicts:* on the j-th axis line through y, two members whose offsets differ in coordinate j disagree at a single point, which is pushed for one and inside the box of the other, so one of them makes an error there (Figure 4). The number of such member–error pairs is at least $|M_y|(1-p_{\max}^{(j)})$, where $p^{(j)}$ is the law of the j-th offset of a uniform member. *Entropy:* since the offsets are distinct, $\ln|M_y| \leq \sum_j H(p^{(j)})$, and Fano's inequality bounds each term by $\psi_L(1-p_{\max}^{(j)})$, so many members force many conflicts. *Charging:* each error is counted at most L times per coordinate, so the average conflict rate per member and coordinate is at most β. *Jensen:* convexity of $\lambda \mapsto e^{-d\psi_L(\lambda)}$ then gives $|\tilde{A}| \geq \#\{y : M_y \neq \emptyset\} \geq (1-\eta)(L+1)^d|G|\,e^{-d\psi_L(\beta)}$. ◀

3.3 The ℓ_∞ theorem

▶ **Theorem 12.** *Let (L,h,a) be admissible with $L \geq 1$, $h \mid N$ and $N/h \geq 2L+3$. Let $M > 0$ satisfy $\frac{2}{M} + \delta < 1$ and $M\alpha < 1$, and put $\beta_M := \frac{LM\alpha(1+2/(L+1))}{1-M\alpha}$. Every (ε,δ)-DP, (c,r,α)-correct mechanism on $(\mathbb{Z}_N^d, \ell_\infty)$ satisfies*

$$\varepsilon \;\geq\; \ln\frac{1}{\alpha} + d\big(\ln(L+1) - \psi_L(\beta_M)\big) + \ln\big((1-\tfrac{2}{M}-\delta)(1-M\alpha)\big).$$

Proof. Let $G(A)$ be the set of x that miss at most an $M\alpha$ fraction of $\mathrm{box}(x)$ and accept at most an $M\alpha$ fraction of $\mathrm{push}(x)$.

(T1): if A is correct for $\{x\}$, the expected missing fraction and the expected accepted pushed fraction are each at most α. By Markov's inequality, $\Pr[x \notin G(A)] \leq \frac{2}{M}$.

(T2): apply Theorem 11 in each of the h^d residue classes with $\eta = \theta = M\alpha$, and sum. This gives $\omega = e^{d\psi_L(\beta_M)}(L+1)^{-d}/(1-M\alpha)$.

Lemma 9 finishes the proof. ◀

The matching upper bound is a randomly shifted grid.

▶ **Proposition 13** (Randomly shifted grid). *If $s \geq 1$, $s \mid N$ and $N > s+2r$, there is a pure-DP, (c,r,α)-correct mechanism for sets of size at most one with $\varepsilon \leq \ln\frac{1}{\alpha} + d\ln(1+\frac{2r}{s})$.*

Proof sketch; details in Appendix D. On $\{x\}$ the mechanism outputs, with probability $1-\frac{\alpha}{2}$, the cell of side s containing x in a uniformly shifted grid, dilated by r. This set contains $B(x,r)$ and lies within distance $\lfloor cr \rfloor$ of x, so this branch makes no errors. A rare branch shared with the input $\emptyset$ handles the remaining probability, and comparing the probability of each output under $\{x\}$ and $\emptyset$ gives $e^\varepsilon \leq (1+\frac{2r}{s})^d/\alpha$. ◀

Up to an additive $\ln 2$, Proposition 13 is what the partition framework [11] gives with a grid of side s: when $s \mid 2r$, that grid has straddle number $(1+\frac{2r}{s})^d$.

We can now state Theorem 1 formally.

▶ **Theorem 14** (Formal version of Theorem 1). *Let $1 < c < 3$, let $r \geq 1$ be an integer with $L = L(c, r) \geq 1$, and let N be large and divisible as required by Theorem 12 and Proposition 13. Every (ε, δ)-DP, (c, r, α)-correct mechanism on $(\mathbb{Z}_N^d, \ell_\infty)$ with $\delta \leq \frac{1}{4}$ and $\alpha \leq \frac{1}{40L}$ satisfies*

$$\varepsilon \;\geq\; \ln \frac{1}{\alpha} + d\Big(\ln(L+1) - 20L\alpha \ln \frac{1}{\alpha}\Big) - 1.$$

Conversely, a randomly shifted grid is pure DP with $\varepsilon \leq \ln \frac{1}{\alpha} + d \ln(1 + \frac{2r}{s})$. Hence, whenever $s \mid 2r$,

$$d \ln\Big(1 + \frac{2r}{s}\Big) - \frac{40r}{s}\,\alpha d \ln \frac{1}{\alpha} - 1 \;\leq\; \Gamma \;\leq\; d \ln\Big(1 + \frac{2r}{s}\Big).$$

Proof. For the lower bound, apply Theorem 12 with an admissible triple attaining $L = L(c, r)$ and $M = 8$. Then $\beta_8 \leq 20L\alpha$, and $\psi_L(\beta_8) \leq \beta_8 \ln \frac{eL}{\beta_8} \leq 20L\alpha \ln \frac{1}{\alpha}$ for $\alpha \leq \frac{1}{40L}$; moreover $\ln\big((1 - \frac{1}{4} - \delta)(1 - 8\alpha)\big) \geq \ln \frac{2}{5} \geq -1$. The upper bound is Proposition 13. When $s \mid 2r$, Lemma 10 gives $L = 2r/s$, so $\ln(L+1) = \ln(1 + \frac{2r}{s})$ and $20L = \frac{40r}{s}$. ◀

▶ **Corollary 15** (Bounded box, $c < 3$). *Let $1 < c < 3$ and $L_c = \lceil \frac{2}{c-1} \rceil - 1$, so that $L_c \geq 1$, and let $U = [0, m]^d$ under ℓ_∞ with $m \geq 24r$. There is a constant $\alpha_0(c) > 0$, given in Appendix A, such that every (ε, δ)-DP, (c, r, α)-correct mechanism on U with $\delta \leq \frac{1}{4}$ and $\alpha \leq \alpha_0(c)$ satisfies*

$$\varepsilon \;\geq\; \ln \frac{1}{\alpha} + \frac{d}{2} \ln \frac{3(L_c + 1)}{4} - 1.$$

In particular, the geometric cost is $\Omega(d)$.

The proof is in Appendix A, which also gives an explicit bound for every $m > 6r$. It discretizes U to a fine lattice and runs the box tests only at centres at distance at least $3r$ from the boundary, where they stay inside U. Theorem 11 still applies, and averaging over this interior costs $d \ln \frac{m}{m-6r}$, which is at most $d \ln \frac{4}{3}$ when $m \geq 24r$.

▶ Remark 16 (On the hypotheses of Corollary 15). Numerically, $\alpha_0(c) \approx 1.6 \cdot 10^{-3}$ for $2 \leq c < 3$, and $\alpha_0(1.5) \approx 1.4 \cdot 10^{-3}$. Some condition on m/r is necessary: if $m \leq cr$, no point of U is far from any other, and a mechanism that outputs all of U or nothing, by randomized response on whether S is empty, is correct with $\varepsilon = \ln \frac{1-\alpha}{\alpha}$ and no geometric cost. The conjecture asks for constant α. We cover every constant $\alpha \leq \alpha_0(c)$; constants in $(\alpha_0(c), \frac{1}{2})$, and $c \geq 3$, remain open.

4 ℓ_2: lower bounds for every mechanism, and saturation

This section proves Theorem 26. We first explain why the box tests of Section 3 do not carry over to ℓ_2. We then give two replacements, Hamming subcubes (Section 4.1) and a two-scale spectral test (Section 4.2), and finally the big-ball mechanism (Section 4.3).

▶ Remark 17 (Why box tests fail in ℓ_p). Suppose box points $x + \sum_j \sigma_j a_j$ ($\sigma \in \{\pm 1\}^{d'}$) are within ℓ_2-distance r of x, and points pushed in one coordinate, $x + \sum_j \sigma_j a_j + 2\sigma_k a_k$, are beyond cr. By the triangle inequality $\|a_k\| > \frac{(c-1)r}{2}$ for every k, and averaging over signs gives $r^2 \geq \mathbb{E}_\sigma \|\sum_j \sigma_j a_j\|^2 = \sum_j \|a_j\|^2$. So $d' < \frac{4}{(c-1)^2}$, and the cotype inequality gives $d' = O_c(1)$ in any ℓ_p with $p < \infty$. In ℓ_∞, which has no finite cotype, there is no such limit; that is why Section 3 reaches every coordinate. Equidistant boxes still give an additive term for c near 1 (Appendix I), but a bound growing with d needs a different structure. Figure 7 in the appendix quantifies the limit.

4.1 Hamming subcubes

For $b, b' \in \{0,1\}^D$ write $|b \oplus b'|$ for their Hamming distance and $B(j) = B_D(j)$ for the Hamming ball of radius j.

▶ **Lemma 18** (Exact hard-case constant on the cube). *Let $1 \le R < F \le D$, $\Delta = F - 1 - R$ and $m = \lceil \Delta/2 \rceil$. For $A \subseteq \{0,1\}^D$ let $G(A)$ be the set of x with $B(x,R) \subseteq A$ and $A \cap \{y : |x \oplus y| \ge F\} = \emptyset$. Then*

$$|G(A)| \;\le\; \omega_D(R,F)\,|A|, \qquad \omega_D(R,F) := \frac{|B(m)|}{|B(m+R)|},$$

with equality for $A = B(m+R)$ when Δ is even.

Proof sketch; details in Appendix E. If $x, x' \in G$ were at distance at least $F-R$, flipping up to R coordinates of x on which they agree would produce a point accepted because of x yet far from x'. So $\operatorname{diam} G \le \Delta$, and Kleitman's diameter theorem [22] gives $|G| \le |B(m)|$. Moreover A contains the R-neighbourhood of G, and Harper's theorem [19], normalized matching [8] and log-concavity of $j \mapsto |B(j)|$ show that $|G|/|N_R(G)| \le |B(m)|/|B(m+R)|$. ◀

▶ **Theorem 19** (Hamming-cube restriction). *Let $1 \le p < \infty$, and let h, D be integers with $1 \le D \le d$ and $2h \mid N$. Put $R = \lfloor (r/h)^p \rfloor$ and $F = \lfloor (cr/h)^p \rfloor + 1$, and suppose $R \ge 1$, $F \le D$ and $2^D\alpha + \delta < 1$. Every (ε,δ)-DP, (c,r,α)-correct mechanism on $(\mathbb{Z}_N^d, \ell_p)$ satisfies*

$$\varepsilon \;\ge\; \ln\frac{1}{\alpha} + \ln\frac{|B_D(m+R)|}{|B_D(m)|} + \ln\bigl(1 - 2^D\alpha - \delta\bigr), \qquad m = \Bigl\lceil \frac{F-1-R}{2} \Bigr\rceil.$$

Proof. Partition U into cubes $Q = \{v + hb : b \in \{0,1\}^D\}$, varying the first D coordinates with the others fixed. Since $h \le N/2$, the ℓ_p distance between $v+hb$ and $v+hb'$ is $h|b \oplus b'|^{1/p}$. So Hamming distance at most R means near, and at least F means far.

Let $G(A)$ be the set of x that answer every near and every far point of their own cube correctly. Then (T1) holds with $\beta = 2^D\alpha$, by a union bound over at most 2^D tests, and (T2) holds with $\omega = \omega_D(R,F)$, by Lemma 18 applied in each cube. Apply Lemma 9. ◀

▶ **Corollary 20** (Asymptotics). *Let*

$$g_p(c) = \sup_{0 < x \le 1/(c^p+1)} \Bigl[\mathrm{h}\bigl(\tfrac{(c^p+1)x}{2}\bigr) - \mathrm{h}\bigl(\tfrac{(c^p-1)x}{2}\bigr) \Bigr].$$

For $D = D_\alpha$ and $r \ge r_0(c,D)$, Theorem 19 gives $\varepsilon \ge \ln\frac{1}{\alpha} + g_p(c)D - O_c(\log D) - \ln 2$ when $\delta \le \frac14$. Moreover $g_p(c) \uparrow \ln 2$ as $c \downarrow 1$, and $g_p(c) = (0.5569\ldots + o(1))c^{-p}$ as $c \to \infty$, where the constant is $\max_u 2u \ln\frac{1-u}{u}$.

Proof. Choose h so that $R \approx xD$; this is possible once r is large enough that consecutive values of $(r/h)^p$ are closer than $1/c^p$. Then $F = c^pR + O(1)$, and $\binom{D}{k} \ge e^{D\mathrm{h}(k/D)}/(D+1)$ while $|B_D(k)| \le e^{D\mathrm{h}(k/D)}$ for $k \le D/2$. The asymptotics follow from $\mathrm{h}(u+\frac{x}{2}) - \mathrm{h}(u-\frac{x}{2}) \approx x \ln\frac{1-u}{u}$ with $x = 2u/c^p$. ◀

For example, $g_2(1.5) = 0.254$, $g_2(2) = 0.140$ and $g_1(2) = 0.288$. For $d \le \log_2 \frac{1}{4\alpha}$ this bound is linear in d. For larger d it is $(1 + g_p(c)/\ln 2 - o(1))\ln\frac{1}{\alpha}$, which grows with $\ln\frac{1}{\alpha}$ but no longer with d.

4.2 A two-scale spectral inequality

Let $(T_t)_{t\ge 0}$ be a symmetric Markov semigroup on a finite set U: $T_t g(x) = \mathbb{E}\, g(X_t^x)$ for a Markov process with symmetric transition kernel. Each T_t is then self-adjoint and positive on $\ell^2(U)$, and $T_{s+t} = T_s T_t$. On $\mathbb{Z}_N^d$ we use the continuous-time simple random walk, in which each coordinate jumps ± 1 at total rate 1.

▶ **Theorem 21** (Two-scale inequality). *Let $A \subseteq U$, $f = \mathbf{1}_A$, $0 < t_1 < t_2$, $s = t_2 - t_1$, $\eta, \theta \in [0,1)$, and*

$$G = \{x : \ T_{t_1} f(x) \ge 1 - \eta \text{ and } T_{t_2} f(x) \le \theta\}.$$

Then

$$|G| \ \le \ \frac{\theta^{\gamma}}{(1-\eta)^{2+\gamma}}\, |A|, \qquad \gamma = \min\Big(1, \frac{2t_1}{s}\Big).$$

Proof. *Far side.* Pointwise $T_{t_1} f \ge (1-\eta)\mathbf{1}_G$, and T_s is positive, so

$$\theta |G| \ \ge \ \sum_{x \in G} T_{t_2} f(x) = \langle \mathbf{1}_G, T_s T_{t_1} f\rangle \ \ge \ (1-\eta)\langle \mathbf{1}_G, T_s \mathbf{1}_G\rangle = (1-\eta)\|T_{s/2}\mathbf{1}_G\|^2.$$

Interpolation. By the spectral theorem $t \mapsto \ln \|T_t g\|^2$ is convex and nonincreasing. Interpolating between $t = 0$ and $t = s/2$ (or using monotonicity if $t_1 > s/2$) gives

$$\|T_{t_1}\mathbf{1}_G\|^2 \le |G|\Big(\frac{\theta}{1-\eta}\Big)^{\gamma}.$$

Near side. By Cauchy–Schwarz and $\|f\|^2 = |A|$,

$$(1-\eta)|G| \le \langle T_{t_1}\mathbf{1}_G, f\rangle \le \|T_{t_1}\mathbf{1}_G\|\, |A|^{1/2}.$$

Combining the three displays proves the claim. ◀

▶ **Theorem 22** (Privacy from a semigroup). *Let (U,d) be a finite metric space with a symmetric Markov semigroup as above, fix $0 < t_1 < t_2$ and let γ be as in Theorem 21. Let*

$$\nu_1 = \max_x \Pr[d(x, X_{t_1}^x) > r], \qquad \nu_2 = \max_x \Pr[d(x, X_{t_2}^x) \le cr].$$

For M with $\eta = M(\alpha + \nu_1) < 1$, $\theta = M(\alpha + \nu_2) < 1$ and $\frac{2}{M} + \delta < 1$, every (ε, δ)-DP, (c, r, α)-correct mechanism satisfies

$$\varepsilon \ \ge \ \ln\frac{1}{\alpha} + \gamma \ln\frac{1}{\theta} + \ln\Big(1 - \frac{2}{M} - \delta\Big) + (2+\gamma)\ln(1-\eta).$$

Proof. Take $G(A)$ as in Theorem 21.

(T1): if A is correct for $\{x\}$, then $\mathbb{E}[1 - T_{t_1}\mathbf{1}_A(x)] \le \alpha + \nu_1$ and $\mathbb{E}[T_{t_2}\mathbf{1}_A(x)] \le \alpha + \nu_2$. By Markov's inequality, $\Pr[x \notin G(A)] \le \frac{2}{M}$.

(T2) is Theorem 21. Apply Lemma 9. ◀

On $(\mathbb{Z}_N^d, \ell_2)$, take $t_1 = (1-\varepsilon')r^2/d$ and $t_2 = (1+\varepsilon')c^2 r^2/d$. Then $\|X_t\|^2$ concentrates around dt. If $r^2 \ge 2d \ln\frac{4d}{\alpha}$, $d \ge 41\varepsilon'^{-2}\ln\frac{2}{\alpha}$ and $N \ge 5cr$, then $\nu_1, \nu_2 \le \alpha$ (Lemma 29, Appendix F), and Theorem 22 with $M = 8$ gives

$$\varepsilon \ \ge \ (1 + \gamma_{\varepsilon'})\ln\frac{1}{\alpha} - 4, \qquad \gamma_{\varepsilon'} = \min\Big(1, \frac{2(1-\varepsilon')}{c^2(1+\varepsilon') - (1-\varepsilon')}\Big) \ \xrightarrow[\varepsilon' \to 0]{} \ \min\Big(1, \frac{2}{c^2 - 1}\Big).$$

This is the high-dimensional lower bound of Theorem 26. For $c < \sqrt{3}$, fix ε' with $c^2(1+\varepsilon') \le 3(1-\varepsilon')$, so that $\gamma_{\varepsilon'} = 1$; then $\varepsilon \ge 2\ln\frac{1}{\alpha} - 4$ once $d \ge 41\varepsilon'^{-2}\ln\frac{2}{\alpha}$.

▶ Remark 23 (Other metrics). Theorems 21 and 22 need only a symmetric Markov semigroup whose kernels at times t_1 and t_2 concentrate inside $B(r)$ and outside $B(cr)$, respectively. The ratio t_2/t_1 then sets γ. In ℓ_p with Gaussian-like coordinates ($r \gg d^{1/p}$), $t_2/t_1 \to c^2$; for a sparse walk ($\ln \frac{1}{\alpha} \ll r^p \ll d$), $t_2/t_1 \to c^p$; and in Hamming space at gap C the bit-flip semigroup gives $\gamma = \min(1, \frac{2}{C-1})$. The concentration step must be redone in each case.

4.3 The big-ball mechanism

▶ **Theorem 24** (Big ball). *Let $\alpha \le \frac{1}{100}$, $\Lambda = \ln \frac{2}{\alpha}$ and $d \ge \frac{5c\Lambda}{c-1}$. Set*

$$t_0 = \sqrt{2\Lambda/d}, \qquad \rho_2 = \frac{(c-1)r}{2t_0(1+1/(4\Lambda))}, \qquad \rho_1 = \rho_2(1-\tfrac{1}{d}), \qquad R_0^2 = \rho_2^2 + r^2 + 2\rho_2 r t_0,$$

and let $N > 2(R_0+\rho_2)$. Consider the following mechanism, which outputs the lattice points of the Euclidean ball $B(o,R_0)$. On input $\{x\}$, with probability $1-\frac{\alpha}{2}$ draw o uniformly from the shell $\{\rho_1 \le |o-x| \le \rho_2\}$; otherwise run the mechanism for $\emptyset$. On input $\emptyset$, with probability $\min(1, \alpha N^d/\operatorname{vol} B(R_0))$ draw o uniformly from $[0,N)^d$; otherwise output $\emptyset$. It is (c,r,α)-correct and pure ε-DP with

$$\varepsilon \le \ln\frac{1}{\alpha} + \frac{4c}{(c-1)^2}\Big(\ln\frac{2}{\alpha}+1\Big) + 0.46.$$

Proof sketch; details in Appendix G. Write $o = x - \rho U$ with U uniform on the sphere. For $y = x + v$,

$$|y-o|^2 = \rho^2 + |v|^2 + 2\rho\langle U, v\rangle.$$

The cross term is about $\pm 2\rho|v|t_0$ except with probability $e^{-dt_0^2/2} = \frac{\alpha}{2}$, by the spherical cap bound [4]. Therefore near points ($|v| \le r$) stay inside $B(o,R_0)$, while far points ($|v| > cr$) leave it; this is what the choice $\rho_2 \approx \frac{(c-1)r}{2t_0}$ ensures. The privacy loss is $\ln\frac{1}{\alpha} + \ln\frac{\operatorname{vol} B(R_0)}{\operatorname{vol}(\text{shell})}$, and

$$\ln\frac{\operatorname{vol} B(R_0)}{\operatorname{vol}(\text{shell})} = \frac{d}{2}\ln\frac{R_0^2}{\rho_2^2} + O(1) \le \frac{d}{2}\Big(\frac{r^2}{\rho_2^2} + \frac{2rt_0}{\rho_2}\Big) + O(1),$$

which evaluates to $\frac{4c}{(c-1)^2}(\Lambda+1)$ because $\frac{d}{2}t_0^2 = \Lambda$. ◀

▶ **Proposition 25** (Ball blanket). *Let $R_1 = \frac{(c+1)r}{2}$ and $q = \min(1, \alpha N^d/\operatorname{vol} B(R_1))$. On input $\emptyset$, with probability q output $B(o,R_1)$ for o uniform in $[0,N)^d$, and otherwise output $\emptyset$. On input $\{x\}$, with probability $1-\frac{\alpha}{2}$ output $B(o,R_1)$ for o uniform in $B(x, \frac{(c-1)r}{2})$, and otherwise run the mechanism for $\emptyset$. In any norm, this mechanism is (c,r,α)-correct and pure ε-DP with $\varepsilon \le \ln\frac{1}{\alpha} + d\ln\frac{c+1}{c-1}$.*

Proof. For $|y-x| \le r$, $|y-o| \le r + \frac{(c-1)r}{2} = \frac{(c+1)r}{2}$. For $|y-x| > cr$, $|y-o| > cr - \frac{(c-1)r}{2} = \frac{(c+1)r}{2}$. So the main branch has no errors. The privacy ratio is $\operatorname{vol} B(\frac{(c+1)r}{2})/\operatorname{vol} B(\frac{(c-1)r}{2}) = (\frac{c+1}{c-1})^d$, up to the $\frac{\alpha}{2}$ mixing. ◀

4.4 The ℓ_2 theorem

We can now state Theorem 2 formally.

▶ **Theorem 26** (Formal version of Theorem 2). *Fix $c > 1$, $\alpha \le \frac{1}{100}$ and $\delta \le \frac{1}{4}$, let $D_\alpha = \min(d, \lfloor \log_2 \frac{1}{4\alpha} \rfloor)$, and let $r \ge r_0(c, D_\alpha)$ and N be large enough. Every (ε, δ)-DP, (c, r, α)-correct mechanism on $(\mathbb{Z}_N^d, \ell_2)$ satisfies*

$$\varepsilon \;\ge\; \ln \frac{1}{\alpha} + g_2(c)\, D_\alpha - O_c(\log D_\alpha) - 1,$$

where $g_2(c) > 0$ is the constant of Corollary 20. If moreover $r^2 \ge 2d \ln \frac{4d}{\alpha}$, then also $\varepsilon \ge \big(1 + \min(1, \frac{2}{c^2-1}) - o(1)\big) \ln \frac{1}{\alpha} - 4$ as $d/\ln \frac{1}{\alpha} \to \infty$. Conversely, there are pure-DP mechanisms with

$$\varepsilon \le \ln \frac{1}{\alpha} + d \ln \frac{c+1}{c-1} \;\text{ for all } d, \qquad \varepsilon \le \ln \frac{1}{\alpha} + \frac{4c}{(c-1)^2}\Big(\ln \frac{2}{\alpha} + 1\Big) + 0.46 \;\text{ for } d \ge \frac{5c \ln(2/\alpha)}{c-1}.$$

Consequently, $\Gamma = \Theta_c\big(\min(d, \ln \frac{1}{\alpha})\big)$ in ℓ_2.

Proof. The first lower bound is Corollary 20 with $p = 2$, the second is the discussion following Theorem 22, and the upper bounds are Proposition 25 and Theorem 24. For the Θ statement, when $d < \frac{5c \ln(2/\alpha)}{c-1}$ the blanket bound is $O_c(\min(d, \ln \frac{1}{\alpha}))$, and otherwise the big ball's bound is. The lower bound is $\Omega_c(D_\alpha) = \Omega_c(\min(d, \ln \frac{1}{\alpha}))$. ◀

5 Space lower bounds for all representations

Privacy is not needed here. A *representation* of k-sets is a pair (Rep, Query) as in Definition 4 that is (c, r, α)-correct, i.e. satisfies (i) and (ii) but not necessarily (iii).

▶ **Theorem 27** (Space transfer). *Suppose a (β, ω)-test family at error 2α exists. Let Π be a (c, r, α)-correct representation of k-sets, and let S consist of k i.i.d. uniform points with $k|B(cr)| \le \alpha|U|$. Then, with mutual information measured in bits,*

$$\mathbb{E}|\Pi| \;\ge\; I(S; \Pi \mid \mathsf{R}) \;\ge\; k\Big[(1-\beta)\Big(\log_2 \frac{1}{3\alpha} + \log_2 \frac{1}{\omega}\Big) - 1 - \frac{1}{e \ln 2}\Big].$$

The proof (Appendix H) combines a direct sum with a list-decoding form of Fano's inequality. By the chain rule, $I(S; \Pi \mid \mathsf{R}) = \sum_i I(x_i; \Pi \mid \mathsf{R}, x_{<i})$. Each term is the information cost of a single-point protocol, in which $x_{<i}$ becomes public randomness and $x_{>i}$ private randomness; that protocol has error at most $\alpha + (k-1)|B(cr)|/|U| \le 2\alpha$. Fano's inequality is then applied to the candidate set $G(A)$, which contains x_i with probability $1 - \beta$ and has expected size at most $3\omega\alpha|U|$.

▶ **Corollary 28** (Formal version of Theorem 3).

1. *In ℓ_∞ with $c < 3$ and $\alpha \le \frac{1}{100L^2}$,*

$$\frac{\mathbb{E}|\Pi|}{k} \;\ge\; (1 - 2\sqrt{\alpha})\Big(\log_2 \frac{1}{\alpha} + d \log_2(L+1)\Big) - 5Ld\sqrt{\alpha} \log_2 \frac{1}{\alpha} - 5.$$

2. *In ℓ_2, under the conditions of Lemma 29,*

$$\frac{\mathbb{E}|\Pi|}{k} \;\ge\; \big(1 + \gamma_{\varepsilon'} - o(1)\big) \log_2 \frac{1}{\alpha}.$$

3. *In every ℓ_p with $p < \infty$ and for $r \ge r_0(c, D')$,*

$$\frac{\mathbb{E}|\Pi|}{k} \;\ge\; \Big(\log_2 \frac{1}{\alpha} + \frac{g_p(c)}{\ln 2} D' - O(\log D')\Big)(1 - o(1)), \qquad D' = \min\Big(d, \log_2 \frac{1}{2\alpha \ln(1/\alpha)}\Big).$$

Proof. For part 1, apply Theorem 27 with the test family of Theorem 12 at error 2α and $M = \alpha^{-1/2}$. Parts 2 and 3 use the families of Theorem 22, with $M = \ln \frac{1}{\alpha}$, and of Theorem 19, with $2^D \cdot 2\alpha \le 1/\ln \frac{1}{\alpha}$. ◀

6 Discussion

Relation to the partition framework.

The partition framework of [11] pays $\varepsilon = \ln(2\kappa/\alpha)$, and its admissible cells have diameter at most $(c-1)r$. In *any* norm such a partition has

$$\kappa \;\geq\; \Big(\frac{c+1}{c-1}\Big)^d.$$

By the isodiametric and Brunn–Minkowski inequalities, each cell C satisfies $\mathrm{vol}(C + B(r)) \geq (\frac{c+1}{c-1})^d \,\mathrm{vol}(C)$, and averaging over a uniformly random query point gives the bound. Outside ℓ_∞, this improves the simple volume bound $(\frac{2}{c-1})^d$. In ℓ_∞ with $c < 3$ the framework is optimal. When $s \mid 2r$, Theorem 14 shows that this cost is unavoidable for *every* representation, up to $\tilde{O}(\alpha d)$, so the restriction to partitions in the lower bound of [11, Thm. 5.4] can be dropped for $c < 3$. Whether obliviousness itself costs privacy is left open in [11]; in this range it costs essentially nothing, since our lower bounds cover non-oblivious representations and a fixed grid partition, which is oblivious, matches them up to $\ln 2 + \tilde{O}(\alpha d)$.

In ℓ_2, by contrast, the framework is far from optimal. The big ball uses no public data, so it is oblivious; like every per-query construction, each realization misclassifies some queries, which a user reading the released centre can locate. It beats every partition-based representation once $d \gg \ln \frac{1}{\alpha}$, by a factor of order $d/\ln \frac{1}{\alpha}$ in the geometric cost. The comparison also shows that doubling dimension does not determine the cost. At $c = 2$, partitions give $\varepsilon \leq 5 \ln \lambda + \ln \frac{2}{\alpha}$ in any λ-doubling metric [11, Thm. 4.7], yet ℓ_∞^d and ℓ_2^d, which both have doubling dimension $\Theta(d)$, have geometric costs $\Theta(d)$ and $\Theta(\min(d, \ln \frac{1}{\alpha}))$ respectively.

The $2\ln(1/\alpha)$ barrier.

Both ℓ_2 lower-bound techniques stop at $\varepsilon \approx 2 \ln \frac{1}{\alpha}$. A union bound over $T \lesssim 1/\alpha$ test points can never certify $\omega < 1/T$, and the spectral argument loses through Cauchy–Schwarz with $\|\mathbf{1}_A\|^2 = |A|$. The big ball pays $(1 + \frac{4c}{(c-1)^2}) \ln \frac{2}{\alpha}$. At $c = 2$ the two bounds give $\frac{5}{3} \leq \varepsilon^* / \ln \frac{1}{\alpha} \leq 9$ in the limit $\alpha \to 0$, $d \to \infty$; numerically optimizing the big ball's parameters lowers the upper end to about 8. Narrowing this gap requires an argument that uses the 0/1 structure of accept sets more fully. Heuristically the big ball is near-optimal among radially symmetric designs, which suggests a symmetrization approach.

Open problems.

Four questions remain. In ℓ_∞, the case $c \geq 3$ and the exact continuum constant $\ln \frac{c+1}{c-1}$ for $c < 3$ both require test boxes whose pushed points lie $m \geq 2$ steps beyond the faces. There, good centres can form legitimate clusters, and the conflict counting of Theorem 11 no longer applies. In ℓ_2 the order of the geometric cost is settled, but its constant is not, and neither is its dependence on c: our lower bounds decay like c^{-2}, while the big ball pays order c^{-1}. For ℓ_p with $p \notin \{2, \infty\}$ we do not know whether the cost saturates at all, since the big ball relies on the rotation invariance of ℓ_2. Finally, our upper bounds are single-point mechanisms. A k-set version of the big ball with matching space would also settle the space complexity in high-dimensional ℓ_2.

References

1 Mohammad Alaggan, Sébastien Gambs, and Anne-Marie Kermarrec. BLIP: Non-interactive differentially-private similarity computation on Bloom filters. In *Stabilization, Safety, and Security of Distributed Systems (SSS)*, volume 7596 of *LNCS*. Springer, 2012. doi:10.1007/978-3-642-33536-5_20.

2 Alexandr Andoni, Piotr Indyk, Sepideh Mahabadi, and Shyam Narayanan. Differentially private approximate near neighbor counting in high dimensions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

3 Martin Aumüller, Fabrizio Boninsegna, and Francesco Silvestri. Differentially private high-dimensional approximate range counting, revisited. In *6th Symposium on Foundations of Responsible Computing (FORC)*, volume 329 of *LIPIcs*, pages 15:1–15:24, 2025.

4 Keith Ball. An elementary introduction to modern convex geometry. In *Flavors of Geometry*, volume 31 of *MSRI Publications*, pages 1–58. Cambridge University Press, 1997.

5 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

6 Giuseppe Bianchi, Lorenzo Bracciale, and Pierpaolo Loreti. "better than nothing" privacy with Bloom filters: To what extent? In *Privacy in Statistical Databases (PSD)*, volume 7556 of *LNCS*, pages 348–363. Springer, 2012.

7 Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426, 1970.

8 Béla Bollobás. *Combinatorics: Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability*. Cambridge University Press, 1986.

9 Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

10 Larry Carter, Robert W. Floyd, John Gill, George Markowsky, and Mark N. Wegman. Exact and approximate membership testers. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing (STOC)*, pages 59–65, 1978.

11 Paul Cesaretti. Differentially private distance-sensitive representations for general metrics. Cryptology ePrint Archive, Paper 2026/2788, 2026. URL: https://eprint.iacr.org/2026/2788.

12 Paul Cesaretti. Differentially private distance-sensitive set representations. In Luca Geatti and Carla Piazza, editors, *Proceedings of the 27th Italian Conference on Theoretical Computer Science, ICTCS 2026, Udine, Italy, September 7-9, 2026*, volume 4269 of *CEUR Workshop Proceedings*, pages 222–235. CEUR-WS.org, 2026. URL: https://ceur-ws.org/Vol-4269/paper26.pdf.

13 Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 270–278, 2001.

14 Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

15 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference (TCC)*, volume 3876 of *LNCS*, pages 265–284, 2006.

16 Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*, pages 1054–1067, 2014.

17 Mayank Goswami, Rasmus Pagh, Francesco Silvestri, and Johan Sivertsen. Distance sensitive Bloom filters without false negatives. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 257–269, 2017. arXiv:1607.05451.

18 Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 705–714, 2010.

19 Lawrence H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1(3):385–393, 1966.

20 Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V. Pearson, Dietrich A. Stephan, Stanley F. Nelson, and David W. Craig. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, 4(8):e1000167, 2008.

21 Adam Kirsch and Michael Mitzenmacher. Distance-sensitive Bloom filters. In *Proceedings of the 8th Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 41–50, 2006.

22 Daniel J. Kleitman. On a combinatorial conjecture of Erdős. *Journal of Combinatorial Theory*, 1(2):209–214, 1966.

23 Mehmet Kuzu, Murat Kantarcioglu, Elizabeth Durham, and Bradley Malin. A constraint satisfaction cryptanalysis of Bloom filters in private record linkage. In *Privacy Enhancing Technologies (PETS)*, volume 6794 of *LNCS*. Springer, 2011. doi:10.1007/978-3-642-22263-4_13.

24 Rajeev Motwani, Assaf Naor, and Rina Panigrahy. Lower bounds on locality sensitive hashing. *SIAM Journal on Discrete Mathematics*, 21(4):930–935, 2007.

25 Frank Niedermeyer, Simone Steinmetzer, Martin Kroll, and Rainer Schnell. Cryptanalysis of basic Bloom filters used for privacy preserving record linkage. *Journal of Privacy and Confidentiality*, 6(2), 2014. doi:10.29012/jpc.v6i2.640.

26 Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: The sparse and approximate cases. In *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC)*, pages 351–360, 2013.

27 Ryan O'Donnell, Yi Wu, and Yuan Zhou. Optimal lower bounds for locality-sensitive hashing (except when q is tiny). *ACM Transactions on Computation Theory*, 6(1):5:1–5:13, 2014.

28 Sarvar Patel, Giuseppe Persiano, Joon Young Seo, and Kevin Yeo. Differentially private set representations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2501.16680.

29 Rainer Schnell, Tobias Bachteler, and Jörg Reiher. Privacy-preserving record linkage using Bloom filters. *BMC Medical Informatics and Decision Making*, 9:41, 2009. doi:10.1186/1472-6947-9-41.

30 Jason Vander Woude, Peter Dixon, A. Pavan, Jamie Radcliffe, and N. V. Vinodchandran. Geometry of rounding, 2022. arXiv:2211.02694.

31 Jason Vander Woude, Peter Dixon, A. Pavan, Jamie Radcliffe, and N. V. Vinodchandran. Geometry of rounding: Near optimal bounds and a new neighborhood Sperner's lemma, 2023. arXiv:2304.04837.

A Proof of Corollary 15

Write $L = L_c$. We prove the following explicit bound: for every $m > 6r$, every (ε, δ)-DP, (c, r, α)-correct mechanism on $U = [0, m]^d$ with $\delta \le \frac{1}{4}$ and $\alpha \le \frac{1}{40L}$ satisfies

$$\varepsilon \;\ge\; \ln\frac{1}{\alpha} + d\Big(\ln(L+1) - \ln\frac{m}{m-6r} - 20L\alpha\ln\frac{1}{\alpha}\Big) - 1. \tag{1}$$

Corollary 15 follows. Let $\alpha_0(c)$ be the largest $\alpha \le \frac{1}{40L}$ with $20L\alpha\ln\frac{1}{\alpha} \le \frac{1}{2}\ln\frac{3(L+1)}{4}$; it is well defined because $\alpha\ln\frac{1}{\alpha}$ is increasing on $(0, \frac{1}{e})$. For $m \ge 24r$ we have $\ln\frac{m}{m-6r} \le \ln\frac{4}{3}$, and for $\alpha \le \alpha_0(c)$ the last term is at most $\frac{1}{2}\ln\frac{3(L+1)}{4}$, so (1) gives $\varepsilon \ge \ln\frac{1}{\alpha} + \frac{d}{2}\ln\frac{3(L+1)}{4} - 1$.

Set $M = 8$. The proof of (1) has five steps.

Step 1: an admissible box at a large radius. Since $L < \frac{2}{c-1}$, we have $L + 2 - cL > 0$. Fix an integer

$$r' \;\geq\; r_1(c) := \frac{L(L+4)}{2(L+2-cL)},$$

and let $s' = \lfloor cr' \rfloor - r' + 1$. Then $s' + r' \leq cr' + 1$, so the real interval

$$\Big[\frac{2(cr'+1)}{L+2},\; \frac{2r'}{L}\Big]$$

contains every h with $\frac{2(s'+r')}{L+2} \leq h \leq \frac{2r'}{L}$. Its length is $\frac{2r'(L+2-cL)}{L(L+2)} - \frac{2}{L+2} \geq 1$ by the choice of r', so it contains an integer h. By the proof of Lemma 10, there is then an integer a for which (L, h, a) is admissible at radius r'; the condition $h \geq s'$ is implied by the other two. Every test point of a centre x lies within ℓ_∞-distance $r' + h \leq 3r'$ of x, because $a \geq -r'$, $a + Lh \leq r'$ and $h \leq 2r'/L \leq 2r'$.

Step 2: discretize. Put $\eta = r/r'$, $n_0 = \lfloor m/\eta \rfloor$ and $\Lambda = \eta\{0, \ldots, n_0\}^d \subseteq U$. Restrict data sets and queries to Λ and rescale distances by $1/\eta$. This gives an (ε, δ)-DP mechanism on $B = \{0, \ldots, n_0\}^d$ with the ℓ_∞ metric, and it is (c, r', α)-correct: for $u, q \in \Lambda$, $d(u, q) \leq r$ iff $\|(u - q)/\eta\|_\infty \leq r'$, and likewise for cr. On input $\emptyset$ every point of B is accepted with probability at most α, so $\mathbb{E}|A_\emptyset| \leq \alpha|B|$.

Step 3: a test family on the interior. Let $I = \{x \in B : 3r' \leq x_i \leq n_0 - 3r' \text{ for all } i\}$. It is nonempty because $n_0 + 1 > m/\eta = mr'/r > 6r'$. By Step 1, $\mathrm{box}(x) \cup \mathrm{push}(x) \subseteq B$ for every $x \in I$. Let $G(A)$ be the set of $x \in I$ that miss at most an $M\alpha$ fraction of $\mathrm{box}(x)$ and accept at most an $M\alpha$ fraction of $\mathrm{push}(x)$. (T1): as in the proof of Theorem 12, $\Pr[x \in G(A_x)] \geq 1 - \frac{2}{M}$ for every $x \in I$. Its tests are near and far points of x inside B, so correctness applies to them.

(T2): choose $n \geq n_0 + 1$ with $h \mid n$ and $n/h \geq 2L + 3$, and view $B \subseteq \mathbb{Z}_n^d$. For $x \in I$ the tests of x in the torus coincide with its tests in B, since they stay inside B and do not wrap around. So $G(A)$ is contained in the set of torus centres that pass their tests, and Theorem 11, summed over the h^d residue classes, gives

$$|G(A)| \leq \omega|A|, \qquad \omega = \frac{e^{d\psi_L(\beta_M)}}{1 - M\alpha}(L+1)^{-d}.$$

Step 4: average over the interior. For $x \in I$, DP applied to the event $\{x \in G(A)\}$ gives $1 - \frac{2}{M} \leq e^\varepsilon \Pr[x \in G(A_\emptyset)] + \delta$. Summing over $x \in I$,

$$|I|\Big(1 - \frac{2}{M} - \delta\Big) \;\leq\; e^\varepsilon\, \mathbb{E}|G(A_\emptyset)| \;\leq\; e^\varepsilon \omega\, \mathbb{E}|A_\emptyset| \;\leq\; e^\varepsilon \omega\alpha|B|.$$

Hence

$$\varepsilon \;\geq\; \ln\frac{1}{\alpha} + \ln\frac{1}{\omega} + \ln\Big(1 - \frac{2}{M} - \delta\Big) - \ln\frac{|B|}{|I|}.$$

Step 5: constants. We have $|B|/|I| = \big(\frac{n_0+1}{n_0+1-6r'}\big)^d$. The map $t \mapsto \frac{t}{t-6r'}$ is decreasing and $n_0 + 1 > mr'/r$, so

$$\frac{n_0+1}{n_0+1-6r'} < \frac{m}{m-6r}.$$

With $M = 8$, $\delta \leq \frac{1}{4}$ and $\alpha \leq \frac{1}{40L}$, exactly as for Theorem 14, we have $\beta_8 \leq 20L\alpha \leq \frac{1}{2}$, $\psi_L(\beta_8) \leq 20L\alpha \ln\frac{1}{\alpha}$, and $\ln(1 - \frac{2}{8} - \delta) + \ln(1 - 8\alpha) \geq \ln\frac{1}{2} + \ln\frac{4}{5} \geq -1$. Together with Steps 4 and 5, this gives (1).

B Proof of Lemma 10

On the lattice, "far" means ℓ_∞-distance at least $\lfloor cr \rfloor + 1 = s + r$. A box point's coordinate offsets lie in $[a, a + Lh]$, and a pushed point has one offset equal to $a - h$ or $a + (L + 1)h$, with the rest in $[a, a + Lh]$. So admissibility means

$$-r \le a, \qquad a + Lh \le r, \qquad a - h \le -(s + r), \qquad a + (L + 1)h \ge s + r.$$

An integer a satisfying all four exists iff

$$\max\big(-r,\ s + r - (L + 1)h\big) \le \min\big(r - Lh,\ h - s - r\big).$$

Checking the four pairs, this holds iff $Lh \le 2r$, $h \ge s$ and $(L + 2)h \ge 2(s + r)$; the second condition is implied by the other two. So an admissible h exists iff $\lceil \frac{2(s+r)}{L+2} \rceil \le \lfloor \frac{2r}{L} \rfloor$.

Since $h \ge s$ and $Lh \le 2r$, we get $L \le 2r/s$. When $s \mid 2r$, the choice $h = s$, $a = -r$ meets every condition. $L \ge 1$ forces $s \le 2r$, i.e. $\lfloor cr \rfloor \le 3r - 1$, so $c < 3$.

As $r \to \infty$ with c fixed, $s/r \to c - 1$. The admissible h form an interval of length $r(\frac{2}{L} - \frac{2c}{L+2}) + O(1)$. This is positive for large r iff $L < \frac{2}{c-1}$. If $\frac{2}{c-1}$ is an integer, then $L = \frac{2}{c-1}$ fails because $s > (c - 1)r$. Hence $L(c, r) \to \lceil \frac{2}{c-1} \rceil - 1$.

C Proof of Theorem 11

We use the notation of Theorem 11. For $y \in \tilde{A}$ let $M_y = \{X \in G : y \in \mathrm{box}_X\}$, and for $X \in M_y$ let $e(X) = y - X \in \{0, \dots, L\}^d$. This is well defined because $n > L$, and $X \mapsto e(X)$ is injective on M_y. An *error* of X is a missing box point or an accepted pushed point.

Step 1: conflicts on axis lines. Fix y and a coordinate j, and split M_y into classes $M_{y,0}, \dots, M_{y,L}$ by the value of e_j. Consider the points $y + te_j$ on the j-line through y; their other coordinates are in the box range of every member. For a member of class a, the box covers $t \in [-a, L - a]$, and its pushed points on this line are at $t = L + 1 - a$ and $t = -a - 1$. So for $a \ge 1$ the point $t = L + 1 - a$ is pushed for $M_{y,a}$ and inside the box of every member of $M_{y,<a}$; whether or not it is in $\tilde{A}$, it is an error for all of $M_{y,a}$ or for all of $M_{y,<a}$. Symmetrically, for $a \le L - 1$ the point $t = -a - 1$ is an error for all of $M_{y,a}$ or for all of $M_{y,>a}$. These positions are $t \in \{1, \dots, L\}$ and $t \in \{-L, \dots, -1\}$, which are $2L$ distinct points of $\mathbb{Z}_n$ because $n \ge 2L + 3$. Let $\mathrm{conf}_j(y)$ be the number of member–error pairs at these positions, and let $p^{(j)}$ be the law of e_j for a uniform member. Then

$$\mathrm{conf}_j(y) \ \ge \ \sum_{a \ge 1} \min(|M_{y,a}|, |M_{y,<a}|) + \sum_{a \le L-1} \min(|M_{y,a}|, |M_{y,>a}|) \ \ge \ |M_y|\,\big(1 - p^{(j)}_{\max}\big).$$

For the last inequality, let a^* index a largest class. Every class $a > a^*$ has $|M_{y,<a}| \ge |M_{y,a}|$, and every class $a < a^*$ has $|M_{y,>a}| \ge |M_{y,a}|$.

Step 2: entropy. Let $E = e(X)$ for X uniform on M_y. Since e is injective, $H(E) = \ln |M_y|$. By Fano's inequality, with the mode as estimator on an alphabet of size $L + 1$, $H(p^{(j)}) \le \psi_L(1 - p^{(j)}_{\max})$. Step 1 gives $\sum_j (1 - p^{(j)}_{\max}) \le d\lambda_y$. So subadditivity, concavity and monotonicity of ψ_L give

$$\ln |M_y| \le \sum_j H(p^{(j)}) \le d\,\psi_L(\lambda_y), \qquad \lambda_y := \frac{\sum_j \mathrm{conf}_j(y)}{d\,|M_y|}.$$

Step 3: charging. Each pair (X, z), with z an error of X, is counted in $\mathrm{conf}_j(y)$ only if $y \in \mathrm{box}_X$ lies on the j-line through z. A missing point z of X is counted at most L times

per coordinate, once for each other box point of X on its line, so at most dL times in total. An accepted pushed point z of X lies outside the box only in its pushed coordinate j_0, so $j = j_0$, and the candidate y are the $L+1$ box points of X on that line. A lower pushed point is a position only for classes $a \le L-1$, and an upper one only for $a \ge 1$, so z is counted at most L times. Hence

$$\sum_y \sum_j \mathrm{conf}_j(y) \le |G|\, dL\big(\eta(L+1)^d + 2\theta(L+1)^{d-1}\big).$$

Moreover $\sum_y |M_y| = \sum_{X\in G} |\mathrm{box}_X \cap \tilde{A}| \ge (1-\eta)(L+1)^d|G|$. So the $|M_y|$-weighted average of λ_y is at most β.

Step 4: Jensen. Apply Jensen's inequality to the convex nonincreasing function $\lambda \mapsto e^{-d\psi_L(\lambda)}$, with weights proportional to $|M_y|$, and then use the bound β on the weighted average:

$$|\tilde{A}| \;\ge\; \sum_{y:M_y\neq\emptyset} |M_y| \cdot \frac{1}{|M_y|} \;\ge\; \sum_y |M_y|\, e^{-d\psi_L(\lambda_y)} \;\ge\; \Big(\sum_y |M_y|\Big) e^{-d\psi_L(\beta)}.$$

Combining with Step 3 proves the theorem.

D Proof of Proposition 13

On $\{x\}$. With probability $1-\frac{\alpha}{2}$, draw a uniform shift, let $C \ni x$ be the cell of side s of the shifted grid, and output $A = C + \{-r,\dots,r\}^d$. Otherwise run the mechanism for $\emptyset$.

On $\emptyset$. With probability $q = \alpha N^d/(s+2r)^d$, output $C + \{-r,\dots,r\}^d$ for a uniformly random cell C; otherwise output $\emptyset$.

Correctness. A contains $B(x,r)$. The farthest point of A from x is at distance at most $s-1+r = \lfloor cr \rfloor$, so every far point is rejected. The only errors come from the $\frac{\alpha}{2}$ mixture. Under $\emptyset$ every point is accepted with probability exactly α.

Privacy. The output is determined by the cell's corner. Its probability is

$$\Big(1-\frac{\alpha}{2}\Big)s^{-d} + \frac{\alpha}{2}\,qN^{-d} \;\text{ under } \{x\} \qquad \text{and} \qquad qN^{-d} \;\text{ under } \emptyset,$$

giving $e^{\varepsilon} \le (1+\frac{2r}{s})^d/\alpha$. The reverse ratio is at most $2/\alpha$.

E Proof of Lemma 18

Diameter. Let $x, x' \in G$ with $t = |x \oplus x'| \ge F - R$. Flip $\min(R, D-t)$ coordinates of x on which x and x' agree, obtaining y. Then $|x \oplus y| \le R$, so $y \in A$, while $|x' \oplus y| = t + \min(R, D-t) \ge \min(F, D) = F$. This contradicts $x' \in G$. Hence $\operatorname{diam} G \le \Delta$.

Size. By Kleitman's theorem [22], $|G| \le |B(m)|$ if $\Delta = 2m$, and $|G| \le 2|B_{D-1}(m-1)| \le |B(m)|$ if $\Delta = 2m-1$.

Expansion. $A \supseteq N_R(G) = \bigcup_{x\in G} B(x,R)$. Let $|B(j)| \le |G| \le |B(j+1)|$; then $j+1 \le m$ unless $|G| = |B(m)|$. By Harper's theorem [19], $|N_R(G)| \ge |N_R(S)|$, where $S = B(j) \cup \mathcal{F}$ is the initial segment of the simplicial order of size $|G|$ and $\mathcal{F}$ lies on level $j+1$. Now $N_R(S)$ contains $B(j+R)$ and, disjointly, the upper shadow of $\mathcal{F}$ on level $j+1+R$. By the normalized matching property of the Boolean lattice [8], that shadow has size at least $|\mathcal{F}|\binom{D}{j+1+R}/\binom{D}{j+1}$. Hence $|G|/|N_R(G)|$ is at most a linear-fractional function of $|\mathcal{F}| \in [0, \binom{D}{j+1}]$, maximized at an endpoint. The endpoints give $\frac{|B(j)|}{|B(j+R)|}$ and $\frac{|B(j+1)|}{|B(j+1+R)|}$.

Monotonicity. The partial sums $j \mapsto |B(j)|$ of the log-concave binomial sequence are log-concave, so $j \mapsto |B(j)|/|B(j+R)|$ is nondecreasing. Both endpoints are therefore at most $\omega_D(R, F)$.

Equality for even $\Delta = 2m$: if $|x| \le m$ and $|y| \le m + R$, then $|x \oplus y| \le 2m + R = F - 1$, so $B(m) \subseteq G(B(m + R))$.

F Concentration for the walk

▶ **Lemma 29.** *Let* $\alpha \le \frac{1}{2}$, $0 < \varepsilon' \le \frac{1}{2}$, $t_1 = (1-\varepsilon')\frac{r^2}{d}$, $t_2 = (1+\varepsilon')\frac{c^2 r^2}{d}$ *and* $\Lambda_d = \ln \frac{4d}{\alpha}$. *Suppose*

$$\text{(a) } t_1 \ge \Lambda_d, \qquad \text{(b) } d \ge 41\,\varepsilon'^{-2} \ln \frac{2}{\alpha}, \qquad \text{(c) } N \ge 5cr.$$

Then the continuous-time simple random walk on $(\mathbb{Z}_N^d, \ell_2)$ *has* $\nu_1 \le \frac{5}{8}\alpha$ *and* $\nu_2 \le \frac{5}{8}\alpha$. *Condition (a) holds whenever* $r^2 \ge 2d \ln \frac{4d}{\alpha}$.

Proof. Let $X = (X_1, \dots, X_d) \in \mathbb{Z}^d$ be the unwrapped walk at time t; the walk on the torus is X mod N. Put $Y_i = X_i^2$ and $S_t = \sum_i Y_i = \|X\|^2$. The toroidal distance from the start is at most $\|X\|$, and it equals $\|X\|$ when every $|X_i| < N/2$. Hence

$$\nu_1 \le \Pr[S_{t_1} > r^2], \qquad \nu_2 \le \Pr[S_{t_2} \le c^2 r^2] + \Pr\big[\max_i |X_i| \ge N/2 \text{ at time } t_2\big].$$

Both t_1 and t_2 are at least $\Lambda_d \ge \ln 8 > 1$.

One coordinate. Each X_i is a difference of two independent Poisson$(t/2)$ variables. So $\mathbb{E}Y_i = t$ and $\mathbb{E}Y_i^2 = 3t^2 + t \le 4t^2$. Moreover $\mathbb{E}e^{\theta X_i} = e^{t(\cosh\theta - 1)} \le e^{t(e^\theta - 1 - \theta)}$ for $\theta \ge 0$, which is the moment generating function of a centred Poisson(t) variable. Bennett's inequality [9, Sec. 2.7] and symmetry therefore give

$$\Pr[|X_i| \ge v] \;\le\; 2\exp\Big(-\frac{v^2}{2(t + v/3)}\Big) \qquad (v \ge 0). \tag{2}$$

For $v \le 3t$ this is at most $2e^{-v^2/(4t)}$, and the right-hand side of (2) is decreasing in v.

Upper tail at time t_1. Write $t = t_1$ and $r^2 = (1+u)dt$ with $u = \frac{\varepsilon'}{1-\varepsilon'} \in [\varepsilon', 1]$.

Truncation. By (2) at $v = 3t$ and (a), $\Pr[\max_i Y_i > 9t^2] \le 2d\,e^{-9t/4} \le 2d\big(\frac{\alpha}{4d}\big)^{9/4} \le \frac{\alpha}{8}$.

Moments. Let $\tilde{Y}_i = \min(Y_i, 9t^2)$. For $y \le 9t^2$, $\Pr[\tilde{Y}_i \ge y] \le 2e^{-y/(4t)}$, so $\mathbb{E}\tilde{Y}_i^k \le 2k!(4t)^k$, and $\mathbb{E}\tilde{Y}_i^2 \le \mathbb{E}Y_i^2 \le 4t^2$.

Moment generating function. For $\theta = \frac{u}{32t}$ we have $4t\theta = \frac{u}{8} \le \frac{1}{8}$, and

$$\mathbb{E}e^{\theta\tilde{Y}_i} \le 1 + \theta\mathbb{E}\tilde{Y}_i + 2t^2\theta^2 + \sum_{k\ge 3} 2(4t\theta)^k \le 1 + \theta\mathbb{E}\tilde{Y}_i + \frac{u^2}{512}\Big(1 + \frac{16u}{7}\Big).$$

Hence $\mathbb{E}e^{\theta(\tilde{Y}_i - \mathbb{E}\tilde{Y}_i)} \le \exp\big(\frac{23u^2}{3584}\big)$.

Chernoff. Since $\mathbb{E}\tilde{Y}_i \le t$,

$$\Pr\Big[\sum_i \tilde{Y}_i > r^2\Big] \le \Pr\Big[\sum_i (\tilde{Y}_i - \mathbb{E}\tilde{Y}_i) \ge udt\Big] \le \exp\Big(-\frac{u^2 d}{32} + \frac{23u^2 d}{3584}\Big) \le e^{-u^2 d/41}.$$

On the event $\max_i Y_i \le 9t^2$ we have $S_t = \sum_i \tilde{Y}_i$. So $\nu_1 \le \frac{\alpha}{8} + e^{-\varepsilon'^2 d/41} \le \frac{\alpha}{8} + \frac{\alpha}{2}$ by (b).

Lower tail at time t_2. For $Y \ge 0$ and $\theta \ge 0$, $e^{-\theta Y} \le 1 - \theta Y + \frac{\theta^2 Y^2}{2}$, so $\mathbb{E}e^{-\theta(Y-\mathbb{E}Y)} \le e^{\theta^2 \mathbb{E}Y^2/2}$. The Chernoff bound with the best θ gives, for independent nonnegative Y_i,

$$\Pr\Big[\sum_i Y_i \le \sum_i \mathbb{E}Y_i - \lambda\Big] \le \exp\Big(-\frac{\lambda^2}{2\sum_i \mathbb{E}Y_i^2}\Big).$$

With $t = t_2$, $c^2 r^2 = dt/(1+\varepsilon')$, so $\lambda = u' dt$ with $u' = \frac{\varepsilon'}{1+\varepsilon'} \ge \frac{2}{3}\varepsilon'$. Using $\sum_i \mathbb{E}Y_i^2 \le 4dt^2$, this gives $\Pr[S_{t_2} \le c^2 r^2] \le e^{-u'^2 d/8} \le e^{-\varepsilon'^2 d/18} \le \frac{\alpha}{2}$ by (b).

No wrap-around at time t_2. By (b), $d \ge 164 \ln 4$, and then $\Lambda_d = \ln\frac{2}{\alpha} + \ln(2d) \le \frac{d}{164} + \ln(2d) \le \frac{d}{2}$. Put $v_0 = \sqrt{8 t_2 \Lambda_d}$. Then $v_0 \le \sqrt{4(1+\varepsilon')}\, cr < 2.5\, cr \le N/2$ by (c). Also $v_0 \le \sqrt{8}\, t_2$, because $\Lambda_d \le t_2$, so $t_2 + v_0/3 \le 2t_2$. By (2) and monotonicity, $\Pr[|X_i| \ge N/2] \le 2e^{-2\Lambda_d} = \frac{\alpha^2}{8d^2}$. A union bound gives at most $\frac{\alpha}{8}$. So $\nu_2 \le \frac{\alpha}{2} + \frac{\alpha}{8}$. ◀

G Proof of Theorem 24

Write $o = x - \rho U$ with U uniform on S^{d-1} and $\rho \in [\rho_1, \rho_2]$. We use $\Pr[U_1 \ge t] \le e^{-dt^2/2}$ [4].

Near queries. For $y = x + v$ with $|v| = s \le r$, $y \notin B(o, R_0)$ requires $U_1 > a(\rho, s) = \frac{R_0^2 - \rho^2 - s^2}{2\rho s}$, which is decreasing in both ρ and s. At (ρ_2, r) it equals t_0, so such y is rejected with probability at most $\frac{\alpha}{2}$.

Far queries. For $|v| = s > cr$, $y \in B(o, R_0)$ requires $U_1 \ge b(\rho, s) = \frac{\rho^2 + s^2 - R_0^2}{2\rho s}$, which is increasing in s. Moreover $b(\rho, cr) \ge t_0$ iff $f(\rho) := \rho^2 - 2\rho c r t_0 + c^2 r^2 - R_0^2 \ge 0$. Now f is increasing for $\rho \ge crt_0$, and $\rho_1 \ge crt_0$ follows from $d \ge 5c\Lambda/(c-1)$ and $\Lambda \ge \ln 200$. Using $\rho_2^2 - \rho_1^2 \le 2\rho_2^2/d$, the inequality $f(\rho_1) \ge 0$ follows from $(c^2-1)r^2 \ge 2rt_0(c+1)\rho_2 + 2\rho_2^2/d$, which is exactly what the factor $1 + \frac{1}{4\Lambda}$ in ρ_2 ensures. So far queries are accepted with probability at most $\frac{\alpha}{2}$.

Mixing and wrap-around. Mixing with weight $\frac{\alpha}{2}$ keeps both errors at most α, and $N > 2(R_0 + \rho_2)$ rules out wrap-around.

Privacy. The output is a function of o. The density ratio is

$$e^{\varepsilon} = \frac{(1-\alpha/2)\,\mathrm{vol}\, B(R_0)}{\alpha\ \mathrm{vol(shell)}} + \frac{\alpha}{2} \le \frac{\mathrm{vol}\, B(R_0)}{\alpha\ \mathrm{vol(shell)}},$$

and the reverse ratio is at most $\frac{2}{\alpha}$.

Cost. Using $(1 - \frac{1}{d})^d \le e^{-1}$ and $\ln(1+z) \le z$,

$$\ln \frac{\mathrm{vol}\, B(R_0)}{\mathrm{vol(shell)}} = \frac{d}{2} \ln \frac{R_0^2}{\rho_2^2} - \ln\big(1 - (1 - \tfrac{1}{d})^d\big) \le \frac{d}{2}\Big(\frac{r^2}{\rho_2^2} + \frac{2rt_0}{\rho_2}\Big) + 0.46.$$

With $\varphi = 1 + \frac{1}{4\Lambda}$ and $\frac{d}{2} \cdot 4t_0^2 = 4\Lambda$, this is

$$4\Lambda\varphi\, \frac{\varphi + c - 1}{(c-1)^2} + 0.46 \ \le\ \frac{4c\Lambda\varphi^2}{(c-1)^2} + 0.46 \ \le\ \frac{4c(\Lambda+1)}{(c-1)^2} + 0.46.$$

H Proof of Theorem 27

Bob's query coins R_B are independent of everything else, so including them does not change the information. Since Π depends on the points only through the set S, the chain rule gives

$$I(S; \Pi \mid \mathsf{R}) = \sum_i I(x_i; \Pi \mid \mathsf{R}, x_{<i}).$$

Fix i and view $x_{<i}$ as public randomness and $x_{>i}$ as private randomness. The accept set A of the resulting single-point protocol is $(c, r, 2\alpha)$-correct for $\{x_i\}$. Near points of x_i are near S. A far point of x_i is within cr of another point with probability at most $(k-1)|B(cr)|/|U| \le \alpha$.

Apply the candidate-set lemma (a list-decoding form of Fano's inequality [14]): if X is uniform on U, $L(Y) \subseteq U$ and $p = \Pr[X \in L(Y)]$, then

$$H(X \mid Y) \le \ln 2 + p \ln \frac{\mathbb{E}|L(Y)|}{p} + (1-p) \ln |U|.$$

Use it with $L = G(A)$, so that $p \ge 1-\beta$ and $\mathbb{E}|G(A)| \le \omega \mathbb{E}|A| \le \omega(|B(cr)| + 2\alpha|U|) \le 3\omega\alpha|U|$. This gives

$$I(x_i; \Pi \mid \mathsf{R}, x_{<i}) \;\ge\; p \ln \frac{1}{3\omega\alpha} + p \ln p - \ln 2.$$

Finally, $p \ln p \ge -\frac{1}{e}$, and for a prefix-free encoding $\mathbb{E}|\Pi| \ln 2 \ge H(\Pi \mid \mathsf{R}) \ge I(S; \Pi \mid \mathsf{R})$. In this appendix entropies use natural logarithms; dividing by $\ln 2$ gives the statement in bits.

I Equidistant boxes in ℓ_p

Let $d' \le d$ with $d'(c^p - 1) < 3^p - 1$, and let h be even with $\frac{h}{2} d'^{1/p} \le r < \frac{h}{2}(d' - 1 + 3^p)^{1/p}/c$; such h exists for all large r. Use the box $X + h\{0,1\}^{d'}$ with centre $x = X + \frac{h}{2}\mathbf{1}$ on the first d' coordinates. Every box point is at ℓ_p-distance exactly $\frac{h}{2} d'^{1/p} \le r$, and a point pushed in one coordinate is at distance $\frac{h}{2}(d' - 1 + 3^p)^{1/p} > cr$. This is the $L = 1$ test of Section 3, so Theorems 11 and 12 apply on each d'-slice. The result is

$$\Gamma \;\ge\; \min\left(d, \left\lceil \frac{3^p - 1}{c^p - 1} \right\rceil - 1\right)\left(\ln 2 - \tilde{O}(\alpha)\right) - O(1).$$

For example, in ℓ_2 at $c = 1.1$ this gives 38 dimensions.

J Additional figures

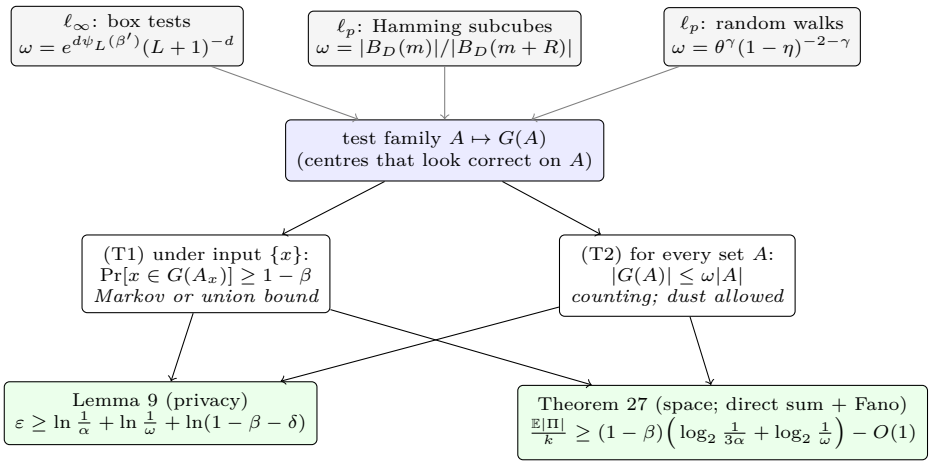


Figure 5 The template behind every lower bound in the paper. A test family selects, for every accept set A, the centres that look correct. Property (T1) holds under the true input, by Markov's inequality or a union bound. Property (T2) is a deterministic counting inequality that must hold for every set A, dust included. Together they give the privacy bound (Lemma 9) and the space bound (Theorem 27). The top row lists the three instantiations used in the paper.

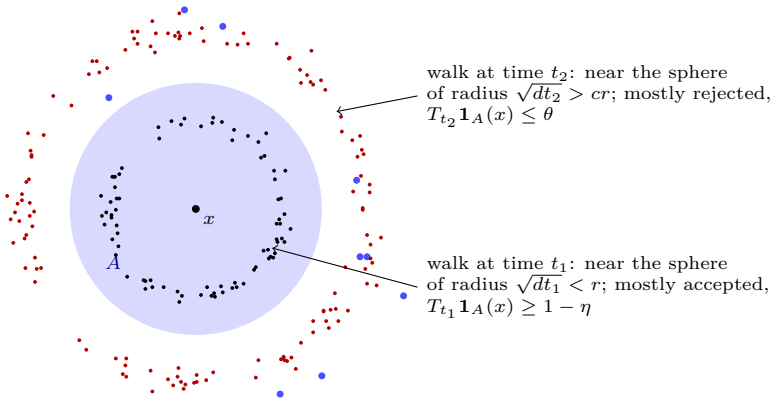
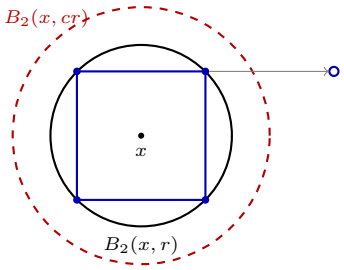
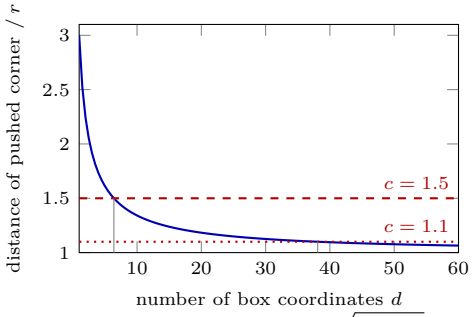


Figure 6 The two-scale spectral test (schematic). In high dimension a random walk from x at time t concentrates near the sphere of radius $\sqrt{dt}$. At time t_1 the walk lies within distance r of x, and at time t_2 beyond distance cr, so a correct output accepts the first cloud and rejects the second. Because $T_{t_2} = T_{t_2-t_1}T_{t_1}$, centres passing both tests have small self-correlation at time $t_2 - t_1$. Log-convexity of $t \mapsto \|T_t\mathbf{1}_G\|^2$ and Cauchy–Schwarz then bound their number (Theorem 21).



(a) box corners at distance r; pushing one coordinate moves a corner to $r\sqrt{1+8/d}$



(b) the test works only while $\sqrt{1+8/d} > c$, i.e. for $d < 8/(c^2-1)$ coordinates

Figure 7 Why the box tests of Section 3 reach only $O_c(1)$ coordinates in ℓ_2 (Remark 17 and Appendix I). An equidistant box of 2^d corners fits in $B_2(x, r)$ only with half-width $r/\sqrt{d}$. Pushing one coordinate by one box width then moves a corner only to distance $r\sqrt{1+8/d}$, which is beyond cr only while $d < 8/(c^2-1)$: 6 coordinates at $c = 1.5$, 38 at $c = 1.1$. In ℓ_∞ the corners of the box are themselves at distance r, and no such limit arises.