Electronic Colloquium on Computational Complexity, Report No. 101 (2023)

Directed Poincaré Inequalities and L^1 Monotonicity Testing of Lipschitz Functions

Renato Ferreira Pinto Jr.*
University of Waterloo
r4ferrei@uwaterloo.ca

Abstract

We study the connection between directed isoperimetric inequalities and monotonicity testing. In recent years, this connection has unlocked breakthroughs for testing monotonicity of functions defined on discrete domains. Inspired the rich history of isoperimetric inequalities in continuous settings, we propose that studying the relationship between directed isoperimetry and monotonicity in such settings is essential for understanding the full scope of this connection.

Hence, we ask whether directed isoperimetric inequalities hold for functions $f : [0,1]^n \to \mathbb{R}$, and whether this question has implications for monotonicity testing. We answer both questions affirmatively. For Lipschitz functions $f : [0,1]^n \to \mathbb{R}$, we show the inequality $d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_1\right]$, which upper bounds the L^1 distance to monotonicity of f by a measure of its "directed gradient". A key ingredient in our proof is the *monotone rearrangement* of f, which generalizes the classical "sorting operator" to continuous settings. We use this inequality to give an L^1 monotonicity tester for Lipschitz functions $f : [0,1]^n \to \mathbb{R}$, and this framework also implies similar results for testing real-valued functions on the hypergrid.

*Partly funded by an NSERC Canada Graduate Scholarship Doctoral Award.

ISSN 1433-8092

1 Introduction

In property testing, algorithms must make a decision about whether a function $f : \Omega \to R$ has some property $\mathcal{P}$, or is *far* (under some distance metric) from having that property, using a small number of queries to f. One of the most well-studied problems in property testing is *monotonicity testing*, the hallmark case being that of testing monotonicity of Boolean functions on the Boolean cube, $f : \{0,1\}^n \to \{0,1\}$. We call f monotone if $f(x) \le f(y)$ whenever $x \preceq y$, i.e. $x_i \le y_i$ for every $i \in [n]$.

A striking trend emerging from this topic of research has been the connection between monotonicity testing and *isoperimetric inequalities*, in particular directed analogues of classical results such as Poincaré and Talagrand inequalities. We preview that the focus of this work is to further explore this connection by establishing directed isoperimetric inequalities for functions $f : [0,1]^n \to \mathbb{R}$ with continuous domain and range, and as an application obtain monotonicity testers in such settings. Before explaining our results, let us briefly summarize the connection between monotonicity testing and directed isoperimetry.

For a function $f : \{0,1\}^n \to \mathbb{R}$, let $d_1^{\mathsf{const}}(f)$ denote its L^1 distance to any constant function $g : \{0,1\}^n \to \mathbb{R}$, and for any point x, define its discrete gradient $\nabla f(x) \in \mathbb{R}^n$ by $(\nabla f(x))_i := f(x^{i\to 1}) - f(x^{i\to 0})$ for each $i \in [n]$, where $x^{i\to b}$ denotes the point x with its i-th coordinate set to b. Then the following inequality[1] is usually called the Poincaré inequality on the Boolean cube (see e.g. [O'D14]): for every $f : \{0,1\}^n \to \{0,1\}$,

$$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_1\right] . \tag{1}$$

(Here and going forward, we write $f \lesssim g$ to denote that $f \le cg$ for some universal constant c, and similarly for $f \gtrsim g$. We write $f \approx g$ to denote that $f \lesssim g$ and $g \lesssim f$.)

Now, let $d_1^{\mathsf{mono}}(f)$ denote the L^1 distance from f to any monotone function $g : \{0,1\}^n \to \mathbb{R}$, and for each point x let $\nabla^- f(x)$, which we call the *directed gradient* of f, be given by $\nabla^- f(x) := \min\{\nabla f(x), 0\}$. Then [CS16] were the first to notice that the main ingredient of the work of [GGL+00], who gave a monotonicity tester for Boolean functions on the Boolean cube with query complexity $O(n/\epsilon)$, was the following "directed analogue" of (1)[2]: for every $f : \{0,1\}^n \to \{0,1\}$,

$$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_1\right] . \tag{2}$$

The tester of [GGL+00] is the "edge tester", which samples edges of the Boolean cube uniformly at random and rejects if any sampled edge violates monotonicity. Inequality (2) shows that, if f is far from monotone, then many edges are violating, so the tester stands good chance of finding one.

In their breakthrough work, [CS16] gave the first monotonicity tester with $o(n)$ query complexity by showing a directed analogue of Margulis's inequality. This was improved by [CST14], and eventually the seminal paper of [KMS18] resolved the problem of (nonadaptive) monotonicity testing of Boolean functions on the Boolean cube, up to polylogarithmic factors, by giving a tester with query complexity $\widetilde{O}(\sqrt{n}/\epsilon^2)$. The key ingredient was to show a directed analogue of *Talagrand's inequality*. Talagrand's inequality gives that, for every $f : \{0,1\}^n \to \{0,1\}$,

$$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_2\right] .$$

[1] The left-hand side is usually written $\mathrm{Var}[f]$ instead; for Boolean functions, the two quantities are equivalent up to a constant factor, and writing $d_1^{\mathsf{const}}(f)$ is more consistent with the rest of our presentation.

[2] Typically the left-hand side would be the distance to a *Boolean* monotone function, rather than any real-valued monotone function, but the two quantities are equal; this may be seen via a maximum matching of violating pairs of f, see [FLN+02].

Compared to (1), this replaces the ℓ^1-norm of the gradient with its ℓ^2-norm. [KMS18] showed the natural directed analogue[3] up to polylogarithmic factors, which were later removed by [PRW22]: for every $f : \{0,1\}^n \to \{0,1\}$,

$$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_2\right] .$$

Since then, directed isoperimetric inequalities have also unlocked results in monotonicity testing of Boolean functions on the hypergrid [BCS18, BCS22, BKKM22, BCS23] (see also [BCS20, HY22]) and real-valued functions on the Boolean cube [BKR20].

Our discussion so far has focused on isoperimetric (*Poincaré-type*) inequalities on *discrete* domains. On the other hand, a rich history in geometry and functional analysis, originated in continuous settings, has established an array of isoperimetric inequalities for functions defined on continuous domains, as well as an impressive range of connections to topics such as partial differential equations [Poi90], Markov diffusion processes [BGL14], probability theory and concentration of measure [BL97], optimal transport [BS16], polynomial approximation [Ver99], among others. (See Appendix A for a brief background on Poincaré-type inequalities.)

As a motivating starting point, we note that for suitably smooth (Lipschitz) functions $f : [0,1]^n \to \mathbb{R}$, an L^1 Poincaré-type inequality holds [BH97]:

$$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_2\right] . \tag{3}$$

Thus, understanding the full scope of the connection between classical isoperimetric inequalities, their directed counterparts, and monotonicity seems to suggest the study of the continuous setting. In this work, we ask: do *directed* Poincaré-type inequalities hold for functions f with continuous domain and range? And if so, do such inequalities have any implications for monotonicity testing? We answer both questions affirmatively: Lipschitz functions $f : [0,1]^n \to \mathbb{R}$ admit a directed L^1 Poincaré-type inequality (Theorem 1.2), and this inequality implies an upper bound on the query complexity of testing monotonicity of such functions with respect to the L^1 distance (Theorem 1.4). (We view L^1 as the natural distance metric for the continuous setting; see Section 1.3 for a discussion.) This framework also yields results for L^1 testing monotonicity of real-valued functions on the hypergrid $f : [m]^n \to \mathbb{R}$. Our testers are *partial derivative testers*, which naturally generalize the classical *edge testers* [GGL+00, CS13] to continuous domains.

We now introduce our model, and then summarize our results.

1.1 L^p-testing

Let (Ω, Σ, μ) be a probability space (typically for us, the unit cube or hypergrid with associated uniform probability distribution). Let $R \subseteq \mathbb{R}$ be a range, and $\mathcal{P}$ a property of functions $g : \Omega \to R$. Given a function $f : \Omega \to \mathbb{R}$, we denote the L^p distance of f to property $\mathcal{P}$ by $d_p(f, \mathcal{P}) := \inf_{g \in \mathcal{P}} d_p(f, g)$, where $d_p(f,g) := \mathop{\mathbb{E}}_{x \sim \mu}\left[|f(x) - g(x)|^p\right]^{1/p}$. For fixed domain Ω, we write $d_p^{\mathsf{const}}(f)$ for the L^p distance of f to the property of constant functions, and $d_p^{\mathsf{mono}}(f)$ for the L^p distance of f to the property of monotone functions. (See Definition 2.2 for a formal definition contemplating e.g. the required measurability and integrability assumptions.)

Definition 1.1 (L^p-testers)**.** Let $p \geq 1$. For probability space (Ω, Σ, μ), range $R \subseteq \mathbb{R}$, property $\mathcal{P} \subseteq L^p(\Omega, \mu)$ of functions $g : \Omega \to R$, and proximity parameter $\epsilon > 0$, we say that randomized algorithm A is an *L^p-tester for $\mathcal{P}$* with query complexity q if, given *oracle access* to an unknown input function $f : \Omega \to R \in L^p(\Omega, \mu)$, A makes at most q oracle queries and 1) accepts with probability at least $2/3$ if $f \in \mathcal{P}$; 2) rejects with probability at least $2/3$ if $d_p(f, \mathcal{P}) > \epsilon$.

[3] In fact, they require a *robust* version of this inequality, but we omit that discussion for simplicity.

We say that A has *one-sided error* if it accepts functions $f \in \mathcal{P}$ with probability 1, otherwise we say it has *two-sided error*. It is *nonadaptive* if it decides all of its queries in advance (i.e. before seeing output from the oracle), and otherwise it is *adaptive*. We consider two types of oracle:

Value oracle: Given point $x \in \Omega$, this oracle outputs the value $f(x)$.

Directional derivative oracle: Given point $x \in \Omega$ and vector $v \in \mathbb{R}^n$, this oracle outputs the derivative of f along v at point x, given by $\frac{\partial f}{\partial v}(x) = v \cdot \nabla f(x)$, as long as f is differentiable at x. Otherwise, it outputs a special symbol $\perp$.

A directional derivative oracle is weaker than a full first-order oracle, which would return the entire gradient [BV04], and it seems to us like a reasonable model for the high-dimensional setting; for example, obtaining the full gradient costs n queries, rather than a single query. This type of oracle has also been studied in optimization research, e.g. see [CWZ21]. For our applications, only the *sign* of the result will matter, in which case we remark that, for sufficiently smooth functions (say, functions with bounded second derivatives) each directional derivative query may be simulated using two value queries on sufficiently close together points.

Our definition (with value oracle) coincides with that of [BRY14a] when the range is $R = [0, 1]$. On the other hand, for general R, we keep the distance metric unmodified, whereas [BRY14a] normalize it by the magnitude of R. Intuitively, we seek testers that are efficient even when f may take large values as the dimension n grows; see Section 1.3.3 for more details.

1.2 Results and main ideas

1.2.1 Directed Poincaré-type inequalities

Our first result is a directed Poincaré inequality for Lipschitz functions $f : [0, 1]^n \to \mathbb{R}$, which may be seen as the continuous analogue of inequality (2) of [GGL+00].

Theorem 1.2. *Let $f : [0, 1]^n \to \mathbb{R}$ be a Lipschitz function with monotone rearrangement f^*. Then*

$$d_1^{\mathsf{mono}}(f) \approx \mathbb{E}\left[|f - f^*|\right] \lesssim \mathbb{E}\left[\|\nabla^- f\|_1\right] . \tag{4}$$

As hinted in the statement, a crucial tool for this result is the *monotone rearrangement* f^* of f. We construct f^* by a sequence of axis-aligned rearrangements $R_1, \dots, R_n$; each R_i is the *non-symmetric monotone rearrangement* operator along dimension i, which naturally generalizes the *sorting* operator of [GGL+00] to the continuous case. For each coordinate $i \in [n]$, the operator R_i takes f into an equimeasurable function $R_i f$ that is monotone in the i-th coordinate, at a "cost" $\mathbb{E}\left[|f - R_i f|\right]$ that is upper bounded by $\mathbb{E}\left[|\partial_i^- f|\right]$, where $\partial_i^- f := (\nabla^- f)_i$ is the directed partial derivative along the i-th coordinate. We show that each application R_i can only decrease the "cost" associated with further applications R_j, so that the total cost of obtaining f^* (i.e. the LHS of (4)) may be upper bounded, via the triangle inequality, by the sum of all directed partial derivatives, i.e. the RHS of (4).

A technically simpler version of this argument also yields a directed Poincaré inequality for real-valued functions on the hypergrid. We also note that Theorems 1.2 and 1.3 are both tight up to constant factors.

Theorem 1.3. *Let $f : [m]^n \to \mathbb{R}$ and let f^* be its monotone rearrangement. Then*

$$d_1^{\mathsf{mono}}(f) \approx \mathbb{E}\left[|f - f^*|\right] \lesssim m\mathbb{E}\left[\|\nabla^- f\|_1\right] .$$

Inequality \ Setting		Discrete $\{0,1\}^n \to \{0,1\}$	Discrete $\{0,1\}^n \to \mathbb{R}$	Continuous $[0,1]^n \to \mathbb{R}$
(L^1, ℓ^1)-Poincaré	$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_1\right]$	* [Tal93]	* [Tal93]	* [BH97]
	$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_1\right]$	[GGL+00]	Theorem 1.3	Theorem 1.2
(L^1, ℓ^2)-Poincaré	$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_2\right]$	* [Tal93]	[Tal93]	[BH97]
	$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_2\right]$	[KMS18]	?	Conjecture 1.8

Table 1: Classical and directed Poincaré-type inequalities on discrete and continuous domains. Cells marked with * indicate inequalities that follow from another entry in the table.

Table 1 places our results in the context of existing classical and directed inequalities. In that table and going forward, for any $p, q \geq 1$ we call the inequalities

$$d_p^{\mathsf{const}}(f)^p \lesssim \mathbb{E}\left[\|\nabla f\|_q^p\right] \qquad \text{and} \qquad d_p^{\mathsf{mono}}(f)^p \lesssim \mathbb{E}\left[\|\nabla^- f\|_q^p\right]$$

a *classical* and *directed* (L^p, ℓ^q)-*Poincaré inequality*, respectively. Note that the L^p notation refers to the space in which we take norms, while ℓ^q refers to the geometry in which we measure gradients. In this paper, we focus on the L^1 inequalities. See also Appendix A for an extended version of Table 1 including other related hypergrid inequalities shown in recent work.

We also note that we have ignored in our discussion the issues of *robust* inequalities, which seem essential for some of the testing applications (see [KMS18]), and the distinction between *inner* and *outer boundary*, whereby some inequalities on Boolean f may be made stronger by setting $\nabla f(x) = 0$ when $f(x) = 0$ (see e.g. [Tal93]). We refer the reader to the original works for the strongest version of each inequality and a detailed treatment of these issues.

1.2.2 Testing monotonicity on the unit cube and hypergrid

Equipped with the results above, we give a monotonicity tester for Lipschitz functions $f : [0,1]^n \to \mathbb{R}$, and the same technique yields a tester for functions on the hypergrid as well. The testers are parameterized by an upper bound L on the best Lipschitz constant of f in ℓ^1 geometry, which we denote $\mathsf{Lip}_1(f)$ (see Definition 2.1 for a formal definition).

Both of our testers are *partial derivative testers*. These are algorithms which only have access to a directional derivative oracle and, moreover, their queries are promised to be axis-aligned vectors. In the discrete case, these are usually called *edge testers* [GGL+00, CS13].

Theorem 1.4. *There is a nonadaptive partial derivative L^1 monotonicity tester for Lipschitz functions $f : [0,1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \leq L$ with query complexity $O\left(\frac{nL}{\epsilon}\right)$ and one-sided error.*

Similarly, there is a nonadaptive partial derivative L^1 monotonicity tester for functions $f : [m]^n$ satisfying $\mathsf{Lip}_1(f) \leq L$ with query complexity $O\left(\frac{nmL}{\epsilon}\right)$ and one-sided error.

The testers work by sampling points x and coordinates $i \in [n]$ uniformly at random, and using directional derivative queries to reject if $\partial_i^- f(x) < 0$. Their correctness is shown using Theorems 1.2 and 1.3, which imply that, when f is ϵ-far from monotone in L^1-distance, the total magnitude of its negative partial derivatives must be large—and since each partial derivative is at most L by assumption, the values $\partial_i^- f(x)$ must be strictly negative in a set of large measure, which the tester stands good chance of hitting with the given query complexity.

1.2.3 Testing monotonicity on the line

The results above, linking a Poincaré-type inequality with a monotonicity tester that uses partial derivative queries and has linear dependence on n, seem to suggest a close parallel with the case of the edge tester on the Boolean cube [GGL+00, CS13]. On the other hand, we also show a strong separation between Hamming and L^1 testing. Focusing on the simpler problem of monotonicity testing *on the line*, we show that the tight query complexity of L^1 monotonicity testing Lipschitz functions grows with the square root of the size of the (continuous or discrete) domain:

Theorem 1.5. *There exist nonadaptive L^1 monotonicity testers for Lipschitz functions $f : [0, m] \to \mathbb{R}$ and $f : [m] \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ with query complexity $\widetilde{O}\left(\sqrt{mL/\epsilon}\right)$. The testers use value queries and have one-sided error.*

This result (along with the near-tight lower bounds in Section 1.2.4) is in contrast with the case of Hamming testing functions $f : [m] \to \mathbb{R}$, which has sample complexity $\Theta(\log m)$ [EKK+98, Fis04, BRY14b, Bel18]. Intuitively, this difference arises because a Lipschitz function may violate monotonicity with rate of change L, so the area under the curve may grow quadratically on violating regions. The proof is in fact a reduction to the Hamming case, using the Lipschitz assumption to establish a connection between the L^1 and Hamming distances to monotonicity.

1.2.4 Lower bounds

We give two types of lower bounds: under no assumptions about the tester and for constant n, we show that the dependence of Theorem 1.4 on L/ϵ is close to optimal[4]. We give stronger bounds for the special case of partial derivative testers (such as the ones from Theorem 1.4), essentially showing that our analysis of the partial derivative tester is tight.

Theorem 1.6. *Let n be a constant. Any L^1 monotonicity tester (with two-sided error, and adaptive value and directional derivative queries) for Lipschitz functions $f : [0, 1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ requires at least $\Omega\left((L/\epsilon)^{\frac{n}{n+1}}\right)$ queries.*

Similarly, any L^1 monotonicity tester (with two-sided error and adaptive queries) for functions $f : [m]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ requires at least $\Omega\left(\min\left\{(mL/\epsilon)^{\frac{n}{n+1}}, m^n\right\}\right)$ queries.

Notice that the bounds above cannot be improved beyond logarithmic factors, due to the upper bounds for the line in Theorem 1.5. It also follows that adaptivity (essentially) does not help with L^1 monotonicity testing on the line, matching the situation for Hamming testing [Fis04, CS14, Bel18].

Theorem 1.6 is obtained via a "hole" construction, which hides a non-monotone region of f inside an ℓ^1-ball B of radius r. We choose r such the violations of monotonicity inside B are large enough to make f ϵ-far from monotone, but at the same time, the ball B is hard to find using few queries. However, this construction has poor dependence on n.

To lower bound the query complexity of partial derivative testers with better dependence on n, we employ a simpler "step" construction, which essentially chooses a coordinate i and hides a small negative-slope region on every line along coordinate i. These functions are far from monotone, but a partial derivative tester must correctly guess both i and the negative-slope region to detect them. We conclude that Theorem 1.4 is optimal for partial derivative testers on the unit cube, and optimal for edge testers on the hypergrid for constant ϵ and L:

[4] Note that one may always multiply the input values by $1/L$ to reduce the problem to the case with Lipschitz constant 1 and proximity parameter ϵ/L, so this is the right ratio to look at.

Domain	**Hamming testing** $f : \Omega \to \mathbb{R}$	L^1**-testing** (prior works) $f : \Omega \to \mathbb{R}$, $\mathsf{Lip}_1(f) \le L$	L^1**-testing** (this work) $f : \Omega \to \mathbb{R}$, $\mathsf{Lip}_1(f) \le L$
$\Omega = [0,1]^n$	Infeasible	$\widetilde{O}\left(\frac{n^2 L}{\epsilon}\right)$ (*) [BRY14a]	$O\left(\frac{nL}{\epsilon}\right)$ p.d.t.
		—	$\Omega\left(\left(\frac{L}{\epsilon}\right)^{\frac{n}{n+1}}\right)$ const. n $\Omega\left(\frac{nL}{\epsilon}\right)$ p.d.t.
$\Omega = [m]^n$	$O\left(\frac{n \log m}{\epsilon}\right)$ [CS13]	$\widetilde{O}\left(\frac{n^2 mL}{\epsilon}\right)$ (*) [BRY14a]	$O\left(\frac{nmL}{\epsilon}\right)$ p.d.t.
	$\Omega\left(\frac{n \log(m) - \log(1/\epsilon)}{\epsilon}\right)$ [CS14]	$\widetilde{\Omega}\left(\frac{L}{\epsilon}\right)$ n.a. 1-s. [BRY14a] $\Omega(n \log m)$ n.a. [BRY14b]	$\Omega\left(\left(\frac{mL}{\epsilon}\right)^{\frac{n}{n+1}}\right)$ const. n $\Omega(nm)$ p.d.t.

Table 2: Query complexity bounds for testing monotonicity on the unit cube and hypergrid. Upper bounds are for nonadaptive (n.a.) algorithms with one-sided error (1-s.), and lower bounds are for adaptive algorithms with two-sided error, unless stated otherwise. For L^1-testing, the upper bounds derived from prior works (*) are specialized to the Lipschitz case by us; see the text for details. Our lower bounds hold either for constant (const.) n, or for partial derivative testers (p.d.t.).

Theorem 1.7. *Any partial derivative L^1 monotonicity tester for Lipschitz functions $f : [0,1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ (with two-sided error and adaptive queries) requires at least $\Omega(nL/\epsilon)$ queries.*

For sufficiently small constant ϵ and constant L, any partial derivative L^1 monotonicity tester for functions $f : [m]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ (with two-sided error and adaptive queries) requires at least $\Omega(nm)$ queries.

Table 2 summarizes our upper and lower bounds for testing monotonicity on the unit cube and hypergrid, along with the analogous Hamming testing results for intuition and bounds for L^1 testing from prior works. See Section 1.3.3 and Appendices B and C for a discussion and details of how prior works imply the results in that table, since to our knowledge the problem of L^1 monotonicity testing parameterized by the Lipschitz constant has not been explicitly studied before. See also Section 7 for a broader overview of prior works on a spectrum of monotonicity testing models.

1.3 Discussion and open questions

1.3.1 Stronger directed Poincaré inequalities?

Classical Poincaré inequalities are usually of the ℓ^2 form, which seems natural e.g. due to basis independence. On the other hand, in the directed setting, the weaker ℓ^1 inequalities (as in [GGL+00] and Theorems 1.2 and 1.3) have more straightforward proofs than ℓ^2 counterparts such as [KMS18]. A perhaps related observation is that monotonicity is *not* a basis-independent concept, since it is defined in terms of the standard basis. It is not obvious whether directed ℓ^2 inequalities ought to hold in every (real-valued, continuous) setting. Nevertheless, in light of the parallels and context established thus far, we are hopeful that such an equality does hold. Otherwise, we believe that the reason should be illuminating. For now, we conjecture:

Conjecture 1.8. *For every Lipschitz function $f : [0,1]^n \to \mathbb{R}$, it holds that*

$$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_2\right] .$$

Accordingly, we also ask whether an L^1 tester with $O(\sqrt{n})$ complexity exists, presumably with a dependence on the $\mathsf{Lip}_2(f)$ constant rather than $\mathsf{Lip}_1(f)$ since ℓ^2 is the relevant geometry above.

1.3.2 Query complexity bounds

Our lower bounds either have weak dependence on n, or only apply to a specific family of algorithms (partial derivative testers). Previous works have established tester-independent lower bounds with strong dependence on n by using reductions from communication complexity [BBM12, BRY14b], whose translation to the continuous setting is not obvious[5], by reduction to comparison-based testers [CS14], whose connection to L^1 testing setting seems less immediate, or directly via a careful construction [Bel18]. We believe that finding strong tester-independent lower bounds for L^1 testing Lipschitz functions on the unit cube is an interesting direction for further study.

We also remark that even a tight lower bound matching Theorem 1.4 may not rule out testers with better dependence on n if, for example, such a tester were parameterized by $\mathsf{Lip}_2(f)$, which can be a factor of $\sqrt{n}$ larger than $\mathsf{Lip}_1(f)$. We view the possibility of better testers on the unit cube, or otherwise a conceptual separation with [KMS18], as an exciting direction for future work.

1.3.3 Relation to prior work on L^p-testing

[BRY14a] initiated the systematic study of L^p-testing and, most relevant to the present work, established the first (and, to our knowledge, only) results on L^p testing of the monotonicity property, on the hypergrid and on the discrete line. While our models are broadly compatible, a subtle but crucial distinction must be explained.

[BRY14a] focused their exposition on the case of functions $f : \Omega \to [0, 1]$, and in this regime, L^1 testing can only be easier than Hamming testing, which they show via a reduction based on Boolean threshold functions. On the other hand, for functions with other ranges, say $f : \Omega \to [a, b]$, their definition normalizes the notion of distance by a factor of $\frac{1}{b-a}$. In our terminology, letting $r := b-a$ and $g := f/r$, it follows that $d_1(g) = d_1(f)/r$, so testing f with proximity parameter ϵ reduces to testing g with proximity parameter ϵ/r. For Hamming testers with query complexity that depends linearly on $1/\epsilon$, this amounts to paying a factor of r in the reduction to the Boolean case[6]. This loss is indeed necessary, because by the same reasoning, testing g with proximity parameter ϵ reduces to testing f with proximity parameter $r\epsilon$. Therefore the problems of testing f with proximity parameter ϵ and testing f/r with proximity parameter ϵ/r have the same query complexity.

In this work, we do not normalize the distance metric by r; we would like to handle functions f that may take large values as the dimension n grows, as long as f satisfies a Lipschitz assumption, and our goal is to beat the query complexity afforded by the reduction to the Boolean case. We derive these benchmarks by assuming that the input f is Lipschitz, and inferring an upper bound on r based on the Lipschitz constant and the size of the domain. Combined with the hypergrid tester of [BRY14a] and a discretization argument for the unit cube inspired by [BCS20, HY22], we establish benchmarks for our testing problem. See Appendix B for details.

With the discussion above in mind, it is instructive to return to Table 2. We note that our upper bounds have polynomially smaller dependence on n than the benchmarks, suggesting that our use

[5]Note that there is no obvious reduction from testing on the hypergrid to testing on the unit cube—one idea is to simulate the unit cube tester on a multilinear interpolation of the function defined on the hypergrid, but the challenge is that simulating each query to the unit cube naively requires an exponential number of queries to the hypergrid.

[6]This factor can also be tracked explicitly in the characterization of the L^1 distance to monotonicity of [BRY14a]: it arises in Lemmas 2.1 and 2.2, where an integral from 0 to 1 must be changed to an integral from a to b, so the best threshold function is only guaranteed to be ϵ/r-far from monotone.

of the Lipschitz assumption—via the directed Poincaré inequalities in Theorems 1.2 and 1.3—exploits useful structure underlying the monotonicity testing problem (whereas the benchmark testers must work for every function with bounded range, not only the Lipschitz ones). Our lower bounds introduce an almost-linear dependence on the hypergrid length m; intuitively, this dependence is not implied by the previous bounds in [BRY14a, BRY14b] because those construct the violations of monotonicity via Boolean functions, whereas our constructions exploit the fact that a Lipschitz function can "keep growing" along a given direction, which exacerbates the L^1 distance to monotonicity in the region where that happens. Our lower bounds for partial derivative testers show that the analysis of our algorithms is essentially tight, so new (upper or lower bound) ideas are required to establish the optimal query complexity for arbitrary testers.

On the choice of L^1 distance and Lipschitz assumption. We briefly motivate our choice of distance metric and Lipschitz assumption. For continuous range and domain, well-known counterexamples rule out testing with respect to Hamming distance: given any tester with finite query complexity, a monotone function may be made far from monotone by arbitrarily small, hard to detect perturbations. Testing against L^1 distance is then a natural choice, since this metric takes into account the magnitude of the change required to make a function monotone ([BRY14a] also discuss connections with learning and approximation theory). However, an arbitrarily small region of the input may still have disproportionate effect on the L^1 distance if the function is arbitrary, so again testing is infeasible. Lipschitz continuity seems like a natural enough assumption which, combined with the choice of L^1 distance, makes the problem tractable. Another benefit is that Lipschitz functions are differentiable almost everywhere by Rademacher's theorem, so the gradient is well-defined almost everywhere, which enables the connection with Poincaré-type inequalities.

Organization. Section 2 introduces definitions and conventions that will be used throughout the paper. In Section 3 we prove our directed Poincaré inequalities on the unit cube and hypergrid, and in Section 4 we give our L^1 monotonicity testers for these domains. Section 5 gives the upper bound for testing functions on the line, and in Section 6 we prove our lower bounds. Finally, in Section 7 we give a broader overview of prior works on monotonicity testing for the reader's convenience.

2 Preliminaries

In this paper, $\mathbb{N}$ denotes the set of strictly positive integers $\{1, 2, \dots\}$. For $m \in \mathbb{N}$, we write $[m]$ to denote the set $\{i \in \mathbb{N} : i \le m\}$. For any $c \in \mathbb{R}$, we write c^+ for $\max\{0, c\}$ and c^- for $-\min\{0, c\}$. We denote the closure of an open set $B \subset \mathbb{R}^n$ by $\overline{B}$.

For a (continuous or discrete) measure space (Ω, Σ, ν) and measurable function $f : \Omega \to \mathbb{R}$, we write $\int_\Omega f \, \mathrm{d}\nu$ for the Lebesgue integral of f over this space. Then for $p \ge 1$, the space $L_p(\Omega, \nu)$ is the set of measurable functions f such that $|f|^p$ is Lebesgue integrable, i.e. $\int_\Omega |f|^p \, \mathrm{d}\nu < \infty$, and we write the L^p norm of such functions as $\|f\|_{L^p} = \|f\|_{L^p(\nu)} = \left(\int_\Omega |f|^p \, \mathrm{d}\nu\right)^{1/p}$. We will write ν to denote the Lebesgue measure when $\Omega \subset \mathbb{R}^n$ is a continuous domain (in which case we will simply write $L^p(\Omega)$ for $L^p(\Omega, \nu)$) and the counting measure when $\Omega \subset \mathbb{Z}^n$ is a discrete domain, and reserve μ for the special case of probability measures.

2.1 Lipschitz functions and L^p distance

We first define Lipschitz functions with respect to a choice of ℓ^p geometry.

Definition 2.1. Let $p \geq 1$. We say that $f : \Omega \to \mathbb{R}$ is (ℓ^p, L)-*Lipschitz* if, for every $x, y \in \Omega$, $|f(x) - f(y)| \leq L\|x - y\|_p$. We say that f is Lipschitz if it is (ℓ^p, L)-Lipschitz for any L (in which case this also holds for any other choice of ℓ^q), and in this case we denote by $\mathsf{Lip}_p(f)$ the best possible Lipschitz constant:

$$\mathsf{Lip}_p(f) := \inf_L \{f \text{ is } (\ell^p, L)\text{-Lipschitz}\} .$$

It follows that $\mathsf{Lip}_p(f) \leq \mathsf{Lip}_q(f)$ for $p \leq q$.

We now formally define L^p distances, completing the definition of L^p-testers from Section 1.1.

Definition 2.2 (L^p-distance)**.** Let $p \geq 1$, let $R \subseteq \mathbb{R}$, and let (Ω, Σ, μ) be a probability space. For a property $\mathcal{P} \subseteq L^p(\Omega, \mu)$ of functions $g : \Omega \to R$ and function $f : \Omega \to R \in L^p(\Omega, \mu)$, we define the distance from f to $\mathcal{P}$ as $d_p(f, \mathcal{P}) := \inf_{g \in \mathcal{P}} d_p(f, g)$, where

$$d_p(f, g) := \|f - g\|_{L^p(\mu)} = \mathop{\mathbb{E}}_{x \sim \mu} [|f(x) - g(x)|^p]^{1/p} .$$

For $p = 0$, we slightly abuse notation and, taking $0^0 = 0$, write $d_0(f, g)$ for the Hamming distance between f and g weighted by μ (and $\mathcal{P}$ may be any set of measurable functions on (Ω, Σ, μ)).

In our applications, we will always take μ to be the uniform distribution over Ω[7]. As a shorthand, when (Ω, Σ, μ) is understood from the context and $R = \mathbb{R}$, we will write

1. $d_p^{\mathsf{const}}(f) := d_p(f, \mathcal{P}^{\mathsf{const}})$ where $\mathcal{P}^{\mathsf{const}} := \{f : \Omega \to \mathbb{R} \in L^p(\Omega, \mu) : f = c, c \in \mathbb{R}\}$; and
2. $d_p^{\mathsf{mono}}(f) := d_p(f, \mathcal{P}^{\mathsf{mono}})$ where $\mathcal{P}^{\mathsf{mono}} := \{f : \Omega \to \mathbb{R} \in L^p(\Omega, \mu) : f \text{ is monotone}\}$.

Going forward, we will also use the shorthand $d_p(f) := d_p^{\mathsf{mono}}(f)$.

2.2 Directed partial derivatives and gradients

We first consider functions on continuous domains. Let B be an open subset of $\mathbb{R}^n$, and let $f : B \to \mathbb{R}$ be Lipschitz. Then by Rademacher's theorem f is differentiable almost everywhere in B. For each $x \in B$ where f is differentiable, let $\nabla f(x) = (\partial_1 f(x), \ldots, \partial_n f(x))$ denote its gradient, where $\partial_i f(x)$ is the partial derivative of f along the i-th coordinate at x. Then, let $\partial_i^- := \min\{0, \partial_i\}$, i.e. for every x where f is differentiable we have $\partial_i^- f(x) = -(\partial_i f(x))^-$. We call ∂_i^- the *directed partial derivative* operator in direction i. Then we define the *directed gradient* operator by $\nabla^- := (\partial_1^-, \ldots, \partial_n^-)$, again defined on every point x where f is differentiable.

Now considering the hypergrid domains, let $f : [m]^n \to \mathbb{R}$. Fix $x \in [m]^n$ and $i \in [n]$, and write e_i for the i-th basis vector, i.e. e_i takes value 1 in its i-th component and 0 elsewhere. We then define the (discrete) partial derivative of f along the i-th coordinate at x by $\partial_i f(x) := f(x + e_i) - f(x)$ if $x_i < m$, and $\partial_i f(x) := 0$ if $x_i = m$. We then define its discrete gradient by $\nabla := (\partial_1, \ldots, \partial_n)$. Their directed counterparts are defined as above: $\partial_i^- := \min\{0, \partial_i\}$ and $\nabla^- := (\partial_1^-, \ldots, \partial_n^-)$.

Note that this definition for the discrete discrete gradient on the hypergrid is slightly different from how we introduced the discrete gradient on the Boolean cube in the opening (c.f. inequality (1)) and its use in Table 1, where we allowed each edge (x, y) to "contribute" to both $\partial_i f(x)$ and $\partial_i f(y)$. In contrast, the definition above (which we will use going forward) only allows the "contribution" to $\partial_i f(x)$, since on domain $[m]^n$ with $m = 2$, the point y falls under the case $y_i = m$, so $\partial_i f(y) := 0$.

[7]More precisely: when $\Omega = [0, 1]^n$, μ will be the Lebesgue measure on Ω (with associated σ-algebra Σ), and when $\Omega = [m]^n$, μ will be the uniform distribution over Ω (with the power set of Ω as the σ-algebra Σ).

The definition we choose seems more natural for the hypergrid settings, but we also remark that for ℓ^1 inequalities, the choice does not matter up to constant factors (i.e. each edge is counted once or twice). For ℓ^2 inequalities, this choice is related to the issues of inner/outer boundaries and robust inequalities [Tal93, KMS18].

3 Directed Poincaré inequalities for Lipschitz functions

In this section, we establish Theorems 1.2 and 1.3. We start with the one-dimensional case, i.e. functions on the line, and then generalize to higher dimensions. In each subsection, we will focus our presentation on the setting where the domain is continuous (corresponding to our results for the unit cube), and then show how the same proof strategy (more easily) yields analogous results for discrete domains (corresponding to our results for the hypergrid).

3.1 One-dimensional case

Let $m > 0$, let $I := (0, m)$, and let $f : \overline{I} \to \mathbb{R}$ be a measurable function. We wish to show that $\|f - f^*\|_{L^1} \lesssim m \|\partial^- f\|_{L^1}$, where f^* is the monotone rearrangement of f. We first introduce the monotone rearrangement, and then show this inequality using an elementary calculus argument.

3.1.1 Monotone rearrangement

Here, we introduce the (non-symmetric, non-decreasing) monotone rearrangement of a one-dimensional function. We follow the definition of [Kaw85], with the slight modification that we are interested in the *non-decreasing* rearrangement, whereas most of the literature usually favours the non-increasing rearrangement. The difference is purely syntactic, and our choice more conveniently matches the convention in the monotonicity testing literature. Up to this choice, our definition also agrees with that of [BS88, Chapter 2], and we refer the reader to these two texts for a comprehensive treatment.

We define the (lower) *level sets* of $f : \overline{I} \to \mathbb{R}$ as the sets

$$\overline{I}_c := \left\{ x \in \overline{I} : f(x) \le c \right\}$$

for all $c \in \mathbb{R}$. For nonempty measurable $S \subset \mathbb{R}$ of finite measure, the *rearrangement* of S is the set

$$S^* := [0, \nu(S)]$$

(recall that ν stands for the Lebesgue measure here), and we define $\emptyset^* := \emptyset$. For a level set $\overline{I}_c$, we write $\overline{I}_c^*$ to mean $\left(\overline{I}_c\right)^*$.

Definition 3.1. The *monotone rearrangement* of f is the function $f^* : \overline{I} \to \mathbb{R}$ given by

$$f^*(x) := \inf \left\{ c \in \mathbb{R} : x \in \overline{I}_c^* \right\} . \tag{5}$$

Note that f^* is always a non-decreasing function.

We note two well-known properties of the monotone rearrangement: equimeasurability and order preservation. Two functions f, g are called *equimeasurable* if $\nu\{f \ge c\} = \nu\{g \ge c\}$ for every $c \in \mathbb{R}$. A mapping $u \mapsto u^*$ is called *order preserving* if $f(x) \le g(x)$ for all $x \in \overline{I}$ implies $f^*(x) \le g^*(x)$ for all $x \in \overline{I}$. See [BS88, Chapter 2, Proposition 1.7] for a proof of the following:

Fact 3.2. *Let $f : \overline{I} \to \mathbb{R}$ be a measurable function. Then f and f^* are equimeasurable.*

Fact 3.3. *The mapping $f \mapsto f^*$ is order preserving.*

3.1.2 Absolutely continuous functions and the one-dimensional Poincaré inequality

Let $f : \overline{I} \to \mathbb{R}$ be absolutely continuous. It follows that f has a derivative ∂f almost everywhere (i.e. outside a set of measure zero), $\partial f \in L^1(I)$ (i.e. its derivative is Lebesgue integrable), and

$$f(x) = f(0) + \int_0^x \partial f(t)\,\mathrm{d}t$$

for all $x \in \overline{I}$. It also follows that $\partial^- f \in L^1(I)$.

We may now show our one-dimensional inequality:

Lemma 3.4. *Let $f : \overline{I} \to \mathbb{R}$ be absolutely continuous. Then $\|f - f^*\|_{L^1} \le 2m \|\partial^- f\|_{L^1}$.*

Proof. Let $S := \{x \in \overline{I} : f^*(x) > f(x)\}$, and note that S is a measurable set because f, f^* are measurable functions (the latter by Fact 3.2). Moreover, since f and f^* are equimeasurable (by the same result), we have $\int f\,\mathrm{d}\nu = \int f^*\,\mathrm{d}\nu$ and therefore

$$\begin{aligned}\|f - f^*\|_{L^1} &= \int_I |f - f^*|\,\mathrm{d}\nu = \int_S (f^* - f)\,\mathrm{d}\nu + \int_{I\setminus S} (f - f^*)\,\mathrm{d}\nu \\ &= \int_S (f^* - f)\,\mathrm{d}\nu + \left(\int_I (f - f^*)\,\mathrm{d}\nu - \int_S (f - f^*)\,\mathrm{d}\nu\right) = 2\int_S (f^* - f)\,\mathrm{d}\nu\,.\end{aligned}$$

Hence our goal is to show that

$$\int_S (f^* - f)\,\mathrm{d}\nu \le m \left\|\partial^- f\right\|_{L^1}\,.$$

Let $x \in \overline{I}$. We claim that there exists $x' \in [0, x]$ such that $f(x') \ge f^*(x)$. Suppose this is not the case. Then since f is continuous on $[0, x]$, by the extreme value theorem it attains its maximum and therefore there exists $c < f^*(x)$ such that $f(y) \le c$ for all $y \in [0, x]$. Thus $[0, x] \subseteq \overline{I}_c$, so $\nu\left(\overline{I}_c\right) \ge x$ and hence $x \in \overline{I}_c^*$. Then, by Definition 3.1, $f^*(x) \le c < f^*(x)$, a contradiction. Thus the claim is proved.

Now, let $x \in S$ and fix some $x' \in [0, x]$ such that $f(x') \ge f^*(x)$. Since f is absolutely continuous, we have

$$f^*(x) - f(x) \le f(x') - f(x) = -\int_{x'}^x \partial f(t)\,\mathrm{d}t \le -\int_0^m \partial^- f(t)\,\mathrm{d}t = \left\|\partial^- f\right\|_{L^1}\,.$$

The result follows by applying this estimate to all x:

$$\int_S (f^* - f)\,\mathrm{d}\nu \le \int_S \left\|\partial^- f\right\|_{L^1}\mathrm{d}\nu = \nu(S)\left\|\partial^- f\right\|_{L^1} \le m\left\|\partial^- f\right\|_{L^1}\,. \qquad \square$$

3.1.3 Discrete case

Let $m \in \mathbb{N}$ and let $I := [m]$. We may define the monotone rearrangement $f^* : I \to \mathbb{R}$ of $f : I \to \mathbb{R}$ as in Definition 3.1 by identifying $\overline{I}$ with I and writing $S^* := [|S|]$ for each finite $S \subset \mathbb{N}$. More directly, f^* is the function such that $f^*(1) \le f^*(2) \le \cdots \le f^*(m)$ is the *sorted sequence* of the values $f(1), f(2), \ldots, f(m)$. It is easy to show that the directed version of Lemma 3.4 holds, and in fact one may simply repeat the proof of that lemma.

Lemma 3.5. *Let $f : [m] \to \mathbb{R}$. Then $\|f - f^*\|_{L^1} \le 2m \|\partial^- f\|_{L^1}$.*

3.2 Multidimensional case

In the continuous case, we ultimately only require an inequality on the unit cube $[0,1]^n$. However, we will first work in slightly more generality and consider functions defined on a *box* in $\mathbb{R}^n$, defined below. This approach makes some of the steps more transparent, and also gives intuition for the discrete case of the hypergrid.

Definition 3.6. Let $a \in \mathbb{R}^n_{>0}$. The *box of size* a is the closure $\overline{B} \subset \mathbb{R}^n$ of $B = (0, a_1) \times \cdots \times (0, a_n)$.

Going forward, $\overline{B} \subset \mathbb{R}^n$ will always denote such a box.

Notation. For $x \in \mathbb{R}^n$, $y \in \mathbb{R}$ and $i \in [n]$, we will use the notation x^{-i} to denote the vector in $\mathbb{R}^{[n]\setminus\{i\}}$ obtained by removing the i-th coordinate from x (note that the indexing is not changed), and we will write (x^{-i}, y) as a shorthand for the vector $(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n) \in \mathbb{R}^n$. We will also write x^{-i} directly to denote any vector in $\mathbb{R}^{[n]\setminus\{i\}}$. For function $f : \overline{B} \to \mathbb{R}$ and $x^{-i} \in \mathbb{R}^{[n]\setminus\{i\}}$, we will write $f_{x^{-i}}$ for the function given by $f_{x^{-i}}(y) = f(x^{-i}, y)$ for all $(x^{-i}, y) \in \overline{B}$. For any set $D \in \mathbb{R}^n$, we will denote by D^{-i} the projection $\{x^{-i} : x \in D\}$, and extend this notation in the natural way to more indices, e.g. D^{-i-j}.

Definition 3.7 (Rearrangement in direction i)**.** Let $f : \overline{B} \to \mathbb{R}$ be a measurable function and let $i \in [n]$. The *rearrangement of f in direction i* is the function $R_i f : \overline{B} \to \mathbb{R}$ given by

$$(R_i f)_{x^{-i}} := (f_{x^{-i}})^* \tag{6}$$

for all $x^{-i} \in \left(\overline{B}\right)^{-i}$. We call each R_i the *rearrangement operator in direction i*.

We may put (6) in words as follows: on each line in direction i determined by point x^{-i}, the restriction of $R_i f$ to that line is the monotone rearrangement of the restriction of f to that line.

Proposition 3.8. *Let $\overline{B}$ be the box of size $a \in \mathbb{R}^n$, and let $f : \overline{B} \to \mathbb{R}$ be Lipschitz continuous. Then for each $i \in [n]$,*

$$\|f - R_i f\|_{L^1} \le 2a_i \left\|\partial_i^- f\right\|_{L^1} .$$

Proof. Since f is Lipschitz continuous, each $f_{x^{-i}} : [0, a_i] \to \mathbb{R}$ is Lipschitz continuous and *a fortiori* absolutely continuous. The result follows from Lemma 3.4, using Tonelli's theorem to choose the order of integration. □

A key ingredient in our multi-dimensional argument is that the rearrangement operator preserves Lipschitz continuity:

Lemma 3.9 ([Kaw85, Lemma 2.12])**.** *If $f : \overline{B} \to \mathbb{R}$ is Lipschitz continuous (with Lipschitz constant L), then $R_i f$ is Lipschitz continuous (with Lipschitz constant $2L$).*

We are now ready to define the (multidimensional) monotone rearrangement f^*:

Definition 3.10. Let $f : \overline{B} \to \mathbb{R}$ be a measurable function. The *monotone rearrangement* of f is the function

$$f^* := R_n R_{n-1} \cdots R_1 f .$$

We first show that f^* is indeed a monotone function:

Proposition 3.11. *Let $f : \overline{B} \to \mathbb{R}$ be Lipschitz continuous. Then f^* is monotone.*

Proof. Say that $g : \overline{B} \to \mathbb{R}$ is *monotone in direction i* if $g_{x^{-i}}$ is non-decreasing for all $x^{-i} \in \left(\overline{B}\right)^{-i}$. Then g is monotone if and only if it is monotone in direction i for every $i \in [n]$. Note that $R_i f$ is monotone in direction i by definition of monotone rearrangement. Therefore, it suffices to prove that if f is monotone in direction j, then $R_i f$ is also monotone in direction j.

Suppose f is monotone in direction j, and suppose $i < j$ without loss of generality. Let $a \in \mathbb{R}^n$ be the size of B. Let $x^{-j} \in \left(\overline{B}\right)^{-j}$ and $0 \le y_1 < y_2 \le a_j$, so that $(x^{-j}, y_1), (x^{-j}, y_2) \in \overline{B}$. We need to show that $(R_i f)(x^{-j}, y_1) \le (R_i f)(x^{-j}, y_2)$. Let $\overline{I}_i := [0, a_i]$. For each $k \in \{1, 2\}$, let $g_k : \overline{I}_i \to \mathbb{R}$ be given by

$$g_k(z) := f(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_{j-1}, y_k, x_{j+1}, \dots, x_n) .$$

Note that

$$g_k^*(z) = (R_i f)(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_{j-1}, y_k, x_{j+1}, \dots, x_n)$$

for every $z \in \overline{I}_i$, and therefore our goal is to show that $g_1^*(x_i) \le g_2^*(x_i)$. But f being monotone in direction j means that $g_1(z) \le g_2(z)$ for all $z \in \overline{I}_i$, so by the order preserving property (Fact 3.3) of the monotone rearrangement we get that $g_1^*(x_i) \le g_2^*(x_i)$, concluding the proof. □

It is well-known that the monotone rearrangement is a non-expansive operator. Actually a stronger fact holds, as we note below.

Proposition 3.12 ([CT80])**.** *Let $m > 0$ and let $f, g \in L^1[0, m]$. Then f^*, g^* satisfy*

$$\int_{[0,m]} (f^* - g^*)^- \, \mathrm{d}\nu \le \int_{[0,m]} (f - g)^- \, \mathrm{d}\nu$$

and

$$\int_{[0,m]} |f^* - g^*| \, \mathrm{d}\nu \le \int_{[0,m]} |f - g| \, \mathrm{d}\nu .$$

The result above is stated for functions on the interval. Taking the integral over the box B and repeating for each operator R_i yields the non-expansiveness of our monotone rearrangement operator, as also noted by [Kaw85]:

Corollary 3.13. *Let $f, g \in L^1(\overline{B})$. Then $\|f^* - g^*\|_{L^1} \le \|f - g\|_{L^1}$.*

We show that the rearrangement operator can only make the norm of the directed partial derivatives smaller, i.e. decrease the violations of monotonicity, which is the key step in this proof.

Proposition 3.14. *Let $f : \overline{B} \to \mathbb{R}$ be Lipschitz continuous and let $i, j \in [n]$. Then $\left\|\partial_j^- (R_i f)\right\|_{L^1} \le \left\|\partial_j^- f\right\|_{L^1}$.*

Proof. We may assume that $i \ne j$, since otherwise the LHS is zero. We will use the following convention for variables names: $w \in \mathbb{R}^n$ will denote points in B; $z \in \mathbb{R}^{[n] \setminus \{i,j\}}$ will denote points in B^{-i-j}; $x \in \mathbb{R}$ will denote points in $(0, a_i)$ (indexing the i-th dimension); and $y \in \mathbb{R}$ will denote points in $(0, a_j)$ (indexing the j-th dimension). For each $i \in [n]$, let e_i denote the i-th basis vector.

Since f is Lipschitz, so is $R_i f$ by Lemma 3.9. By Rademacher's theorem, these functions are differentiable almost everywhere. Therefore, let $D \subseteq B$ be a measurable set such that f and $R_i f$

are differentiable in D and $\nu(D) = \nu(B)$. We have

$$\begin{aligned}
\left\|\partial_j^-(R_i f)\right\|_{L^1} &= \int_D \left|\partial_j^-(R_i f)\right| \mathrm{d}\nu \\
&= \int_D \left[\lim_{h\to 0} \left(\frac{(R_i f)(w + he_j) - (R_i f)(w)}{h}\right)^-\right] \mathrm{d}\nu(w) \\
&\overset{(BC1)}{=} \lim_{h\to 0} \int_D \left(\frac{(R_i f)(w + he_j) - (R_i f)(w)}{h}\right)^- \mathrm{d}\nu(w) \\
&\overset{(D1)}{=} \lim_{h\to 0} \int_B \left(\frac{(R_i f)(w + he_j) - (R_i f)(w)}{h}\right)^- \mathrm{d}\nu(w) \\
&\overset{(T1)}{=} \lim_{h\to 0} \int_{B^{-i-j}} \int_{(0,a_j)} \int_{(0,a_i)} \left(\frac{(R_i f)(z, y + h, x) - (R_i f)(z, y, x)}{h}\right)^- \mathrm{d}\nu(x)\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(z) \\
&\leq \lim_{h\to 0} \int_{B^{-i-j}} \int_{(0,a_j)} \int_{(0,a_i)} \left(\frac{f(z, y + h, x) - f(z, y, x)}{h}\right)^- \mathrm{d}\nu(x)\,\mathrm{d}\nu(y)\,\mathrm{d}\nu(z) \\
&\overset{(T2)}{=} \lim_{h\to 0} \int_B \left(\frac{f(w + he_j) - f(w)}{h}\right)^- \mathrm{d}\nu(w) \\
&\overset{(D2)}{=} \lim_{h\to 0} \int_D \left(\frac{f(w + he_j) - f(w)}{h}\right)^- \mathrm{d}\nu(w) \\
&\overset{(BC2)}{=} \int_D \left[\lim_{h\to 0} \left(\frac{f(w + he_j) - f(w)}{h}\right)^-\right] \mathrm{d}\nu(w) \\
&= \int_D \left|\partial_j^- f\right| \mathrm{d}\nu \\
&= \left\|\partial_j^- f\right\|_{L^1} .
\end{aligned}$$

Equalities (BC1) and (BC2) hold by the bounded convergence theorem, which applies because the difference quotients are uniformly bounded by the Lipschitz constants of $R_i f$ and f (respectively), and because $R_i f$ and f are differentiable in D (which gives pointwise convergence of the limits). Equalities (D1) and (D2) hold again by the uniform boundedness of the difference quotients, along with the fact that $\nu(B \setminus D) = 0$. Equalities (T1) and (T2) hold by Tonelli's theorem. Finally, the inequality holds by Proposition 3.12, since $(R_i f)(z, y + h, \cdot)$ is the monotone rearrangement of $f(z, y + h, \cdot)$ and $(R_i f)(z, y, \cdot)$ is the monotone rearrangement of $f(z, y, \cdot)$. □

We are now ready to prove our directed (L^1, ℓ^1)-Poincaré inequality.

Theorem 3.15. *Let B be the box of size $a \in \mathbb{R}^n$ and let $f : \overline{B} \to \mathbb{R}$ be Lipschitz continuous. Then*

$$\|f - f^*\|_{L^1} \leq 2 \sum_{i=1}^{n} a_i \left\|\partial_i^- f\right\|_{L^1} .$$

Proof. We have

$$\begin{aligned}
\|f - f^*\|_{L^1} &\le \sum_{i=1}^{n} \|R_{i-1}\cdots R_1 f - R_i \cdots R_1 f\|_{L^1} && \text{(Triangle inequality)} \\
&\le 2\sum_{i=1}^{n} a_i \left\|\partial_i^-(R_{i-1}\cdots R_1 f)\right\|_{L^1} && \text{(Lemma 3.9 and Proposition 3.8)} \\
&\le 2\sum_{i=1}^{n} a_i \left\|\partial_i^- f\right\|_{L^1} && \text{(Lemma 3.9 and Proposition 3.14)}\,.
\end{aligned}$$

□

Setting $B = (0,1)^n$ yields the inequality portion of Theorem 1.2:

Corollary 3.16. *Let $B = (0,1)^n$ and let $f : \overline{B} \to \mathbb{R}$ be Lipschitz continuous. Then*

$$\mathbb{E}\left[|f - f^*|\right] = \|f - f^*\|_{L^1} \le 2\int_B \|\nabla^- f\|_1 \,\mathrm{d}\nu = 2\mathbb{E}\left[\|\nabla^- f\|_1\right]\,.$$

To complete the proof of Theorem 1.2, we need to show that $d_1(f) \approx \mathbb{E}\left[|f - f^*|\right]$, i.e. that the monotone rearrangement is "essentially optimal" as a target monotone function for f. The inequality $d_1(f) \le \mathbb{E}\left[|f - f^*|\right]$ is clear from the fact that f^* is monotone. The inequality in the other direction follows from the non-expansiveness of the rearrangement operator, with essentially the same proof as that of [KMS18] for the Boolean cube:

Proposition 3.17. *Let $f : [0,1]^n \to \mathbb{R}$ be Lipschitz continuous. Then $\mathbb{E}\left[|f - f^*|\right] \le 2d_1(f)$.*

Proof. Let $g \in L^1([0,1]^n)$ be any monotone function. It follows that $g^* = g$. By Corollary 3.13, we have that $\|f^* - g^*\|_{L^1} \le \|f - g\|_{L^1}$. Using the triangle inequality, we obtain

$$\|f - f^*\|_{L^1} \le \|f - g\|_{L^1} + \|g - f^*\|_{L^1} = \|f - g\|_{L^1} + \|f^* - g^*\|_{L^1} \le 2\|f - g\|_{L^1}\,.$$

The claim follows by taking the infimum over the choice of g. □

Tightness of the inequality. To check that Corollary 3.16 is tight up to constant factors, it suffices to take the linear function $f : [0,1]^n \to \mathbb{R}$ given by $f(x) = 1 - x_1$ for all $x \in [0,1]^n$. Then f^* is given by $f^*(x) = x_1$, so $\mathbb{E}\left[f - f^*\right] = 1/2$ while $\mathbb{E}\left[\|\nabla^- f\|_1\right] = 1$, as needed.

3.2.1 Discrete case

The proof above carries over to the case of the hypergrid almost unmodified, as we now outline. We now consider functions $f : [m]^n \to \mathbb{R}$, so the box B is replaced with $[m]^n$ and its dimensions a_i are all replaced with the length m of the hypergrid. We define the rearrangement in direction i, $R_i f$, as in Definition 3.7 by sorting the restrictions of f to each line along direction i. We also define f^* as in Definition 3.10 by subsequent applications of each operator R_i. Then Proposition 3.8 carries over by applying the one-dimensional Lemma 3.5, and the proof of Proposition 3.11 carries over unmodified.

The non-expansiveness properties Proposition 3.12 and Corollary 3.13 also carry over unmodified, and the key Proposition 3.14 carries over with a more immediate proof: the use of Proposition 3.12

remains the same, but rather than expanding the definition of derivative and reasoning about the limit, the discrete argument boils down to showing the inequality

$$\int_{[m]^n} ((R_i f)(w+e_j) - (R_i f)(w))^- \, \mathrm{d}\nu(w) \le \int_{[m]^n} (f(w+e_j) - f(w))^- \, \mathrm{d}\nu(w)\,,$$

which follows immediately from the discrete version of Proposition 3.12 by summing over all lines in direction i. Then, the hypergrid version of Theorem 3.15 follows by the same application of the triangle inequality, and we conclude the inequality portion of Theorem 1.3:

Theorem 3.18. *Let $f : [m]^n \to \mathbb{R}$. Then $\mathbb{E}\left[|f - f^*|\right] \le 2m\mathbb{E}\left[\|\nabla^- f\|_1\right]$.*

The discrete version of Proposition 3.17 follows identically, and we state it here for convenience:

Proposition 3.19. *Let $f : [m]^n \to \mathbb{R}$. Then $\mathbb{E}\left[|f - f^*|\right] \le 2d_1(f)$.*

Finally, the tightness of Theorem 3.18 is mostly easily verified for the following step function: letting m be even for simplicity, define $f : [m]^n \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & \text{if } x_1 \le m/2\,, \\ 0 & \text{if } x_1 > m/2\,. \end{cases}$$

Then f^* is obtained by flipping this function along the first coordinate, or equivalently swapping the values 1 and 0 in the definition above. Thus $\mathbb{E}\left[|f - f^*|\right] = 1$. On the other hand, $\|\nabla^- f\|_1$ takes value 1 on exactly one point in each line along the first coordinate, and 0 elsewhere. Hence $\mathbb{E}\left[\|\nabla^- f\|_1\right] = 1/m$, as needed.

4 Applications to monotonicity testing

In this section, we use the directed Poincaré inequalities on the unit cube and hypergrid to show that the natural partial derivative tester (or edge tester) attains the upper bounds from Theorem 1.4.

Let Ω denote either $[0,1]^n$ or $[m]^n$, and let $q(\Omega, L, \epsilon)$ denote the query complexity of testers for (ℓ^1, L)-Lipschitz functions on these domains, as follows:

$$q([0,1]^n, L, \epsilon) := \Theta\left(\frac{nL}{\epsilon}\right) \qquad \text{and} \qquad q([m]^n, L, \epsilon) := \Theta\left(\frac{nmL}{\epsilon}\right)\,.$$

The tester is given in Algorithm 1. It is clear that this algorithm is a nonadaptive partial derivative tester, and that it always accepts monotone functions. It suffices to show that it rejects with good probability when $d_1(f) > \epsilon$.

Algorithm 1 L^1 monotonicity tester for Lipschitz functions using partial derivative queries

Input: Partial derivative oracle access to Lipschitz function $f : \Omega \to \mathbb{R}$.
Output: Accept if f is monotone, reject if $d_1(f) > \epsilon$.
Requirement: $\mathsf{Lip}_1(f) \le L$.

procedure PARTIALDERIVATIVETESTER(f, Ω, L, ϵ)
  repeat $q(\Omega, L, \epsilon)$ **times**
    Sample $x \in \Omega$ uniformly at random.
    Sample $i \in [n]$ uniformly at random.
    Reject if $\partial_i f(x) < 0$.
  end repeat
  Accept.

Lemma 4.1. *Let Ω be one of $[0,1]^n$ or $[m]^n$, and let $f : \Omega \to \mathbb{R}$ be a Lipschitz function satisfying* $\mathsf{Lip}_1(f) \le L$. *Suppose $d_1(f) > \epsilon$. Then Algorithm 1 rejects with probability at least $2/3$.*

Proof. **Continuous case.** Suppose $\Omega = [0,1]^n$. Let $D \subseteq [0,1]^n$ be a measurable set such that f is differentiable on D and $\mu(D) = 1$, which exists by Rademacher's theorem. For each $i \in [n]$, let $S_i := \{x \in D : \partial_i f(x) < 0\}$. A standard argument gives that each $S_i \subset \mathbb{R}^n$ is a measurable set. We claim that

$$\sum_{i=1}^{n} \mu(S_i) > \frac{\epsilon}{2L}\,.$$

Suppose this is not the case. By the Lipschitz continuity of f, we have that $|\partial_i f(x)| \le L$ for every $x \in D$ and $i \in [n]$, and therefore

$$2\sum_{i=1}^{n} \mathbb{E}\left[\left|\partial_i^- f\right|\right] \le 2L \sum_{i=1}^{n} \mu(S_i) \le \epsilon\,.$$

On the other hand, the assumption that $d_1(f) > \epsilon$ and Corollary 3.16 yield

$$\epsilon < \mathbb{E}\left[|f - f^*|\right] \le 2\mathbb{E}\left[\|\nabla^- f\|_1\right] = 2\sum_{i=1}^{n} \mathbb{E}\left[\left|\partial_i^- f\right|\right]\,,$$

a contradiction. Therefore the claim holds.

Now, the probability that one iteration of the tester rejects is the probability that $x \in S_i$ when x and i are sampled uniformly at random. This probability is

$$\mathbb{P}\left[\text{Iteration rejects}\right] = \sum_{j=1}^{n} \mathbb{P}_i\left[i = j\right] \mathbb{P}_x\left[x \in S_j\right] = \sum_{j=1}^{n} \frac{1}{n} \cdot \mu(S_j) > \frac{\epsilon}{2nL}\,.$$

Thus $\Theta\left(\frac{nL}{\epsilon}\right)$ iterations suffice to reject with high constant probability.

Discrete case. Suppose $\Omega = [m]^n$. The proof proceeds the same way, but we give it explicitly for convenience. For each $i \in [n]$, let $S_i := \{x \in [m]^n : \partial_i f(x) < 0\}$. We then claim that

$$\sum_{i=1}^{n} \mu(S_i) > \frac{\epsilon}{2mL}\,.$$

Indeed, if this is not the case, then since $|\partial_i f(x)| \le L$ for every i and x, we get that

$$2\sum_{i=1}^{n} \mathbb{E}\left[\left|\partial_i^- f\right|\right] \le 2L \sum_{i=1}^{n} \mu(S_i) \le \frac{\epsilon}{m}\,.$$

On the other hand, the assumption that $d_1(f) > \epsilon$ and Theorem 3.18 yield

$$\frac{\epsilon}{m} < \frac{1}{m} \cdot \mathbb{E}\left[|f - f^*|\right] \le \frac{1}{m} \cdot 2m\mathbb{E}\left[\|\nabla^- f\|_1\right] = 2\sum_{i=1}^{n} \mathbb{E}\left[\left|\partial_i^- f\right|\right] ,$$

a contradiction. Thus the claim holds, and the probability that one iteration of the tester rejects is

$$\mathbb{P}\left[\text{Iteration rejects}\right] = \sum_{j=1}^{n} \mathbb{P}_i\left[i = j\right] \mathbb{P}_x\left[x \in S_j\right] = \sum_{j=1}^{n} \frac{1}{n} \cdot \mu(S_j) > \frac{\epsilon}{2nmL} .$$

Thus $\Theta\left(\frac{nmL}{\epsilon}\right)$ iterations suffice to reject with high constant probability. ☐

5 L^1-testing monotonicity on the line

In this section, we show the upper bounds for L^1 monotonicity testing on the line from Theorem 1.5. The main idea is to reduce from L^1 testing to Hamming testing by using the Lipschitz constant to show that, if the L^1 distance to monotonicity is large, then the Hamming distance to monotonicity must be somewhat large as well; combined with the Hamming testers of [EKK+98, Bel18], this yields an L^1 tester for the discrete line $[m]$.

To obtain a tester for the continuous line $[0, m]$, we furthermore apply a discretization strategy inspired by the domain reduction and downsampling ideas from [BCS20, HY22]. The idea is that, given ϵ and L, we may impose a fine enough grid on $[0, m]$ such that the function defined on that grid preserves the L^1 distance to monotonicity compared to the continuous function; again, the Lipschitz assumption is essential for this step.

In this section, we will follow the convention of denoting functions on continuous domains by f, g, and those on discrete domains by $\overline{f}, \overline{g}$. Depending on the context, it will be clear whether $\overline{f}$ is an arbitrary function or one obtained by discretizing a particular function f. We will also write "f is L-Lipschitz" without specifying the ℓ^p geometry, since all choices are equivalent in one dimension.

Lemma 5.1 (Discretization preserves distance to monotonicity). *Let $m, L, \epsilon > 0$ and let $f : [0, m] \to \mathbb{R}$ be an L-Lipschitz function. Let the* discretized function *$\overline{f} : [m'] \to \mathbb{R}$, for suitable choice of $m' = \Theta(mL/\epsilon)$, be given by $\overline{f}(i) = f(\delta i)$ for each $i \in [m']$, where $\delta := m/m'$. Then if $d_1(f) > \epsilon$, we have $d_1(\overline{f}) > \epsilon/4$.*

Proof. Let $m' \in [cmL/\epsilon, 2cmL/\epsilon]$ be an integer, where c is a sufficiently large universal constant.[8] Let $\overline{f} : [m'] \to \mathbb{R}$ be the function given in the statement, and suppose $d_1(f) > \epsilon$.

Let $\overline{g} : [m'] \to \mathbb{R}$ be the monotone rearrangement of $\overline{f}$. It is easy to check that $\overline{g}$ is Lipschitz with at most the Lipschitz constant of $\overline{f}$. Let $g : [0, m] \to \mathbb{R}$ be the following piecewise linear function whose discretization is $\overline{g}$: for each $i \in [m']$ we set $g(\delta i) = \overline{g}(i)$, and g is the linear spline induced by these points elsewhere (and constant in the segment $[0, \delta]$). Then clearly g is monotone, and thus $d_1(f, g) > \epsilon$. Moreover, g is L-Lipschitz, since its steepest slope is the same as that of $\overline{g}$

[8] We may assume that $mL/\epsilon > 1$, otherwise the problem is trivial: the maximum L^1 distance from monotonicity attainable by an L-Lipschitz function is $\frac{1}{m} \cdot \frac{m \cdot mL}{2} = mL/2$. Therefore the given interval does contain an integer.

up to the coordinate changes.[9] Hence, we have

$$\begin{aligned}
\epsilon < d_1(f,g) &= \frac{1}{m}\int_0^m |f(x)-g(x)|\,\mathrm{d}x = \frac{1}{m}\sum_{i=1}^{m'}\int_{(i-1)\delta}^{i\delta}|f(x)-g(x)|\,\mathrm{d}x \\
&= \frac{1}{m}\sum_{i=1}^{m'}\int_{(i-1)\delta}^{i\delta}|(f(i\delta)\pm L\delta)-(g(i\delta)\pm L\delta)|\,\mathrm{d}x \qquad \text{(Lipschitz property)} \\
&\le \frac{1}{m}\sum_{i=1}^{m'}\int_{(i-1)\delta}^{i\delta}\left[\left|\overline{f}(i)-\overline{g}(i)\right|+2L\delta\right]\mathrm{d}x \\
&= \frac{1}{m}\left[2m'L\delta^2+\delta\sum_{i=1}^{m'}\left|\overline{f}(i)-\overline{g}(i)\right|\right] = \frac{2mL}{m'}+\frac{1}{m'}\sum_{i=1}^{m'}\left|\overline{f}(i)-\overline{g}(i)\right| \le \frac{2\epsilon}{c}+d_1(\overline{f},\overline{g}),
\end{aligned}$$

where we used the notation $a \pm b$ to denote any number in the interval $[a-b, a+b]$.

We may set $c \ge 4$ so that $2\epsilon/c \le \epsilon/2$. Therefore, we obtain $d_1(\overline{f},\overline{g}) > \epsilon/2$. Since $\overline{g}$ is the monotone rearrangement of $\overline{f}$, Proposition 3.19 implies that $d_1(\overline{f},\overline{g}) \le 2d_1(\overline{f})$. We conclude that $d_1(\overline{f}) > \epsilon/4$, as desired. □

Observation 5.2. *The function $\overline{f}$ defined in Lemma 5.1 is ϵ-Lipschitz: since $m' \ge mL/\epsilon$, we have*

$$\left|\overline{f}(i)-\overline{f}(i+1)\right| = |f(\delta i) - f(\delta(i+1))| \le L\delta = Lm/m' \le \epsilon\,.$$

Lemma 5.3 (Far in L^1 distance imples far in Hamming distance)**.** *Let $\overline{f} : [m'] \to \mathbb{R}$ be an L'-Lipschitz function. Then $d_0(\overline{f}) \ge \sqrt{\frac{d_1(\overline{f})}{m'L'}}$.*

Proof. Let $S \subseteq [m']$ be a set such that 1) $|S| = d_0(\overline{f})m'$ and 2) it suffices to change $\overline{f}$ on inputs in S to obtain a monotone function; note that S exists by definition of Hamming distance. Write S as the union of maximal, pairwise disjoint contiguous intervals, $S = I_1 \cup \cdots \cup I_k$.

We define a monotone function $\overline{g} : [m'] \to \mathbb{R}$ as follows. For each $i \in S$, set $i^* \in [m'] \setminus S$ as follows: if there exists $j \in [m'] \setminus S$ such that $j > i$, pick the smallest such j; otherwise, pick the largest $j \in [m'] \setminus S$. In other words, i^* is obtained by picking a direction (right if possible, otherwise left) and choosing the first point outside the interval I_k that contains i. Now, define $\overline{g}$ by

$$\overline{g}(i) = \begin{cases} \overline{f}(i) & \text{if } i \notin S \\ \overline{f}(i^*) & \text{if } i \in S\,. \end{cases}$$

We first claim that $\overline{g}$ is monotone. Indeed the sequence of values $(f(i))_{i\in[m']\setminus S}$ (taken in order of increasing i) is monotone by our first assumption on S, and since $\overline{g}$ is obtained by extending some of these values into flat regions, the resulting function is also monotone. Therefore we can

[9] Formally, if f is L-Lipschitz, then $\overline{f}$ is L'-Lipschitz for $L' = Lm/m'$, hence so is its monotone rearrangement $\overline{g}$. Then since the steepest slope of g must come from two vertices of the spline, g is Lipschitz with Lipschitz constant $L'm'/m = L$.

upper bound the L^1 distance of $\overline{f}$ to monotonicity by

$$\begin{aligned}
d_1(\overline{f}) \le d_1(\overline{f}, \overline{g}) &= \frac{1}{m'} \sum_{i=1}^{m'} \left|\overline{f}(i) - \overline{g}(i)\right| = \frac{1}{m'} \sum_{j=1}^{k} \sum_{i \in I_j} \left|\overline{f}(i) - \overline{f}(i^*)\right| \\
&= \frac{1}{m'} \sum_{j=1}^{k} \sum_{i \in I_j} \left|\left(\overline{f}(i^*) \pm L'|i - i^*|\right) - \overline{f}(i^*)\right| && \text{(Lipschitz property)} \\
&\le \frac{L'}{m'} \sum_{j=1}^{k} \sum_{i \in I_j} |i - i^*| \le \frac{L'}{m'} \sum_{j=1}^{k} \sum_{i \in I_j} |I_j| = \frac{L'}{m'} \sum_{j=1}^{k} |I_j|^2 \\
&\le \frac{L'}{m'} \cdot |S|^2 && \text{(Since } |I_1| + \cdots + |I_k| = |S|\text{)} \\
&= L' d_0(\overline{f})^2 m' .
\end{aligned}$$

The claim follows. ☐

Combining the two lemmas with the classical Hamming monotonicity tester of [EKK+98], the following theorem establishes Theorem 1.5 for the continuous domain $[0, m]$:

Theorem 5.4. *There exists a nonadaptive one-sided L^1 monotonicity tester for L-Lipschitz functions $f : [0, m] \to \mathbb{R}$ with query complexity $O\left(\sqrt{\frac{mL}{\epsilon}} \log\left(\frac{mL}{\epsilon}\right)\right)$.*

Proof. The tester works as follows. It first fixes $m' = \Theta(mL/\epsilon)$ as given by Lemma 5.1. Let $\overline{f} : [m'] \to \mathbb{R}$ be the discretization defined therein ($\overline{f}$ is not explicitly computed upfront, but will rather be queried as needed). The algorithm then simulates the (nonadaptive, one-sided) monotonicity tester of [EKK+98] on the function $\overline{f}$ with proximity parameter $\epsilon' = \Theta\left(\sqrt{\frac{\epsilon}{mL}}\right)$ (the constant may easily be made explicit), producing $f(\delta i) = f(im/m')$ whenever the simulation queries $\overline{f}(i)$. The algorithm returns the result produced by the simulated tester. The query complexity claim follows from the fact that the tester of [EKK+98] has query complexity $O\left(\frac{1}{\epsilon'} \log m'\right)$.

We now show correctness. When f is monotone, so is $\overline{f}$, so the algorithm will accept since the tester of [EKK+98] has one-sided error. Now, suppose $d_1(f) > \epsilon$. Then $d_1(\overline{f}) > \epsilon/4$ by Lemma 5.1. Moreover, since $\overline{f}$ is ϵ-Lipschitz by Observation 5.2, Lemma 5.3 implies that

$$d_0(\overline{f}) \ge \sqrt{\frac{d_1(\overline{f})}{m'\epsilon}} > \sqrt{\frac{1}{4m'}} = \Omega\left(\sqrt{\frac{\epsilon}{mL}}\right) .$$

Since this is the proximity parameter ϵ' used to instantiate the [EKK+98] tester, the algorithm will reject with high constant probability, as needed. ☐

Lemma 5.3 itself also implies Theorem 1.5 for the discrete domain $[m]$. This time, we use the Hamming tester of [Bel18] to obtain a slightly more precise query complexity bound[10].

Theorem 5.5. *There exists a nonadaptive one-sided L^1 monotonicity tester for L-Lipschitz functions $\overline{f} : [m] \to \mathbb{R}$ with query complexity $O\left(\sqrt{\frac{mL}{\epsilon}} \log\left(\frac{m\epsilon}{L}\right)\right)$ when $\epsilon/L \ge 4/m$, and $O(m)$ otherwise.*

[10]One may check that this refinement would have no effect in Theorem 5.4

Proof. The tester sets $\epsilon' := \sqrt{\frac{\epsilon}{mL}}$, and then runs the (nonadaptive, one-sided) Hamming monotonicity tester of [Bel18] on the line $[m]$ with proximity parameter ϵ'. That tester has query complexity $O\left(\frac{1}{\epsilon'}\log(\epsilon' m)\right)$ when $\epsilon' \geq 2/m$ and (trivially) $O(m)$ otherwise, which gives the claimed upper bounds. It remains to show correctness.

When $\overline{f}$ is monotone, the algorithm will accept since the tester of [Bel18] has one-sided error. Now, suppose $d_1(\overline{f}) > \epsilon$. Then Lemma 5.3 yields

$$d_0(\overline{f}) > \sqrt{\frac{\epsilon}{mL}} = \epsilon' ,$$

so the tester of [Bel18] will reject with high constant probability. □

6 Lower bounds

In this section, we prove our lower bounds for testing monotonicity on the unit cube and on the hypergrid. We first show our general lower bounds based on a "hole" construction, which hides a monotonicity violating region inside a randomly placed ℓ^1-ball; these bounds imply near tightness of our upper bounds for testing on the line from Section 5. Then we give our lower bounds for partial derivative testers, which show that the analysis of our tester in Section 4 is tight.

Definition 6.1 (ℓ^1-ball)**.** Let Ω be one of $\mathbb{R}^n$ or $\mathbb{Z}^n$, let $x \in \Omega$ and let $r > 0$ be a real number. The *ℓ^1-ball of radius r centered at x* is the set $B_1^n(r, x) := \{y \in \Omega : \|x - y\|_1 \leq r\}$. We will also write $B_1^n(r) := B_1^n(r, 0)$.

It will be clear from the context whether the domain should be taken to be continuous or discrete, i.e. whether $B_1^n(r, c)$ should be understood under $\Omega = \mathbb{R}^n$ or $\Omega = \mathbb{Z}^n$.

We give the following simple bounds on the volume of continuous and discrete ℓ^1-balls. Since we do not require particularly tight bounds, we opt for a simple formulation and elementary proof.

Proposition 6.2. *There exist functions $c_1, c_2 : \mathbb{N} \to \mathbb{R}_{>0}$ satisfying the following. Let $n \in \mathbb{N}$. Let Ω be one of $\mathbb{R}^n$ or $\mathbb{Z}^n$. Let $r \in \mathbb{R}$ satisfy $r > 0$ if $\Omega = \mathbb{R}^n$, and $r \geq 1$ if $\Omega = \mathbb{Z}^n$. Then*

$$c_1(n) r^n \leq \nu\left(B_1^n(r)\right) \leq c_2(n) r^n .$$

Proof. First suppose $\Omega = \mathbb{R}^n$. Then we have the following formula for the area of the ℓ^1-ball of radius r (see e.g. [Wan05]):

$$\nu\left(B_1^n(r)\right) = \frac{(2r)^n}{n!} .$$

The result follows by letting $c_1(n) \leq 2^n/n!$ and $c_2(n) \geq 2^n/n!$.

Now, suppose $\Omega = \mathbb{Z}^n$, and suppose r is an integer without loss of generality (because since $r \geq 1$, there exist integers within a factor of 2 above and below r). We proceed by an inductive argument. For $n = 1$, the volume is

$$\nu\left(B_1^1\right) = 1 + 2\sum_{d=1}^{r} 1 = 1 + 2r ,$$

so the claim holds by letting $c_1(1) \le 2$ and $c_2(n) \ge 3$. Assuming the claim for some $n \in \mathbb{N}$, we have

$$\nu\left(B_1^{n+1}(r)\right) = |\{x = (x_1, \dots, x_n, x_{n+1}) : \|x\|_1 \le r\}| = \sum_{y=-r}^{r} \left|\left\{x' = (x'_1, \dots, x'_n) : \|x'\|_1 \le r - |y|\right\}\right|$$
$$= \nu\left(B_1^n(r)\right) + 2\sum_{d=1}^{r} \nu\left(B_1^n(r-d)\right) .$$

Since the last expression is at most $3r \cdot \nu\left(B_1^n(r)\right)$, using the inductive hypothesis we conclude

$$\nu\left(B_1^{n+1}(r)\right) \le 3r \cdot c_2(n) r^n = 3c_2(n) r^{n+1} \le c_2(n+1) r^{n+1} ,$$

the last inequality as long as $c_2(n+1) \ge 3c_2(n)$.

For the lower bound, we consider two cases. Note that $r - d \ge r/2$ for at least $\lfloor r/2 \rfloor$ values of d. When $r \ge 4$, we have $\lfloor r/2 \rfloor \ge r/3$, and then

$$\nu\left(B_1^{n+1}(r)\right) \ge c_1(n) r^n + 2\sum_{d=1}^{r} c_1(n)(r-d)^n > \frac{2r}{3} \cdot c_1(n) \left(\frac{r}{2}\right)^n \ge c_1(n+1) r^{n+1} ,$$

the last inequality as long as $c_1(n+1) \le \frac{2}{3} \cdot 2^{-n} \cdot c_1(n)$. On the other hand, if $r < 4$, the bound follows easily for small enough $c_1(n+1)$, since

$$\nu\left(B_1^{n+1}(r)\right) \ge c_1(n) r^n + 2\sum_{d=1}^{r} c_1(n)(r-d)^n > \frac{c_1(n) r^{n+1}}{r} > \frac{c_1(n)}{4} r^{n+1} . \qquad \square$$

Remark 6.3. Note that the constants $c_1(n)$ and $c_2(n)$ in Proposition 6.2 have poor dependence on n, and in particular this is tight in the continuous case. This fact is essentially the reason why this construction is only efficient for constant dimension n.

We now prove our tester-independent lower bounds. Note that there exists a tester for (ℓ^1, L)-Lipschitz functions with proximity parameter ϵ if and only if there exists a tester for $(\ell^1, 1)$-Lipschitz functions with proximity parameter ϵ/L (the reduction consists of simply rescaling the input values). Therefore it suffices to prove the theorems for the case $L = 1$. The following two theorems establish the continuous and discrete cases of Theorem 1.6.

Theorem 6.4 (Lower bound for constant n on the unit cube)**.** *Let $n \in \mathbb{N}$ be a constant. Any L^1 monotonicity tester (with two-sided error, and adaptive value and directional derivative queries) for Lipschitz functions $f : [0,1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le 1$ requires at least $\Omega\left((1/\epsilon)^{\frac{n}{n+1}}\right)$ queries.*

Proof. We construct a family of functions that are ϵ-far from monotone in L^1 distance such that any deterministic algorithm cannot reliably distinguish between a function chosen uniformly at random from this family and the constant-0 function with fewer than the announced number of queries; then, the claim will follow from Yao's principle.

Each such function f is constructed as follows. Let $c \in [0,1]^n$ be a point such that the ball $B_1^n(r, c)$ is completely inside $[0,1]^n$, for radius r to be chosen below. Then f takes value 0 everywhere outside $B_1^n(r, c)$, and inside this ball, it takes value

$$f(x) = -r + \|x - c\|_1$$

for each $x \in B_1^n(r,c)$. Then $\mathsf{Lip}_1(f) = 1$. We now lower bound $d_1(f)$, its distance to monotonicity. Fix any $x' \in [0,1]^{n-1}$ and consider the line of points (y, x') for $y \in [0,1]$, i.e. the line along the first coordinate with remaining coordinates set to x'. Suppose this line intersects $B_1^n(r,c)$. Then this intersection occurs on some interval $[a,b]$ of y-values, and on this interval, f first decreases from $f(a,x') = 0$ to $f\left(\frac{a+b}{2}, x'\right) = -\frac{b-a}{2}$ at rate 1, and then increases at rate 1 back to $f(b,x') = 0$. Any monotone function g is in particular monotone over this line, and it is easy to see that this requires total change to f proportional to the area under this curve:

$$\int_0^1 \left|f(y,x') - g(y,x')\right| \mathrm{d}y \gtrsim \int_0^1 \left|f(y,x')\right| \mathrm{d}y \,.$$

Now, since this holds for any line intersecting $B_1^n(r,c)$, and the collection of such lines gives a partition of $B_1^n(r,c)$, the total distance between f and any monotone function g is lower bounded (up to a constant) by the L^1-norm of f:

$$\int_{[0,1]^n} |f - g| \,\mathrm{d}\nu \gtrsim \int_{[0,1]^n} |f| \,\mathrm{d}\nu \,,$$

and since this holds for any choice of g, we conclude that

$$d_1(f) \gtrsim \int_{[0,1]^n} |f| \,\mathrm{d}\nu \,.$$

We now note that this last expression is half the volume of an ℓ^1-ball in dimension $n+1$: for each point $x \in B_1^n(r,c)$, the contribution to the integrand is $|f(x)| = r - \|x - c\|_1$, corresponding to the measure of points (x, z') for all $0 \le z' \le z$ where $z = r - \|x - c\|_1$, so that the point $(x,z) \in \mathbb{R}^{n+1}$ satisfies $\|(x,z) - (c,0)\|_1 = r$. In other words, the points (x,z') are the points of $B_1^{n+1}(r,(c,0))$ with nonnegative last coordinate. Conversely, all such points contribute to the integral above. Therefore, since n is a constant, using Proposition 6.2 and writing ν^{n+1} for the Lebesgue measure on $\mathbb{R}^{n+1}$, we have

$$d_1(f) \gtrsim \int_{[0,1]^n} |f| \,\mathrm{d}\nu \gtrsim \nu^{n+1}\left(B_1^{n+1}(r)\right) \gtrsim r^{n+1} \,.$$

We wish this last quantity to be at least $\Omega(\epsilon)$, so (recalling n is a constant) it suffices to set

$$r \approx \epsilon^{\frac{1}{n+1}} \,.$$

We have established that each function f, for this choice of r and any choice of c, is ϵ-far from monotone as desired. Our family of functions from which f will be drawn will be given by choices of c such that the balls $B_1^n(r,c)$ are disjoint, so that each query may only rule out one such choice (because queries outside $B_1^n(r,c)$ take value 0). How many disjoint balls $B_1^n(r,c)$ can we fit inside $[0,1]^n$? It suffices to divide $[0,1]^n$ into a grid of n-dimensional cells of side $2r$, each of which can contain one ball. The number of such cells is at least (up to a constant factor)

$$(1/r)^n \gtrsim (1/\epsilon)^{\frac{n}{n+1}} \,.$$

Therefore to distinguish some f uniformly drawn from this family from the constant-0 function with constant probability, any deterministic algorithm must have query complexity at least $\Omega\left((1/\epsilon)^{\frac{n}{n+1}}\right)$. □

Theorem 6.5 (Lower bound for constant n on the hypergrid)**.** *Let $n \in \mathbb{N}$ be a constant. Any L^1 monotonicity tester (with two-sided error and adaptive queries) for functions $f : [m]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le 1$ requires at least $\Omega\left(\min\left\{(m/\epsilon)^{\frac{n}{n+1}}, m^n\right\}\right)$ queries.*

Proof. We proceed similarly to Theorem 6.4, with small changes for the discrete setting (essentially corresponding to the requirement that $r \ge 1$ in the discrete case of Proposition 6.2).

We will again construct functions f based on balls $B_1^n(r, c)$ for suitable choices of r and c. For fixed r and c, f takes value 0 outside the ball and, for each $x \in B_1^n(r, c)$,

$$f(x) = -r + \|x - c\|_1 ,$$

so that $\mathsf{Lip}_1(f) = 1$. Again by a line restriction argument, for any monotone function g we have

$$\int_{[m]^n} |f - g| \, \mathrm{d}\nu \gtrsim \int_{[m]^n} |f| \, \mathrm{d}\nu ,$$

and thus

$$d_1(f) \gtrsim \frac{1}{m^n} \int_{[m]^n} |f| \, \mathrm{d}\nu , \tag{7}$$

the normalizing factor due to Definition 2.2.

When $\epsilon \le 1/m^n$, this construction boils down to setting $f(x) = -1$ at a single point x, which requires $\Omega(m^n)$ queries to identify. Now, assume $\epsilon > 1/m^n$.

Again we may identify the integrand of (7) with points on half of $B_1^{n+1}(r, (c, 0))$. As long as $r \ge 1$ and since n is a constant, Proposition 6.2 implies that

$$d_1(f) \gtrsim \frac{r^{n+1}}{m^n} .$$

Thus to have $d_1(f) \ge \epsilon$, it suffices (since n is a constant) to set

$$r \approx m^{\frac{n}{n+1}} \epsilon^{\frac{1}{n+1}} ,$$

and indeed this gives $r \ge 1$ since $\epsilon > 1/m^n$. Then, our functions f are given by choices of c placed on the hypergrid $[m]^n$ inside disjoint cells of side $2r$, of which there are at least (up to a constant factor)

$$\left(\frac{m}{r}\right)^n \gtrsim \left(\frac{m}{\epsilon}\right)^{\frac{n}{n+1}} ,$$

and thus any deterministic algorithm requires $\Omega\left((m/\epsilon)^{\frac{n}{n+1}}\right)$ queries to distinguish a uniformly chosen f from this family from the constant-0 function. □

The construction for the partial derivative tester lower bounds is simpler: we start with a "step" one-dimensional construction which is flat everywhere except for a small region of negative slope, and then copy this function onto every line along a randomly chosen coordinate i. Then a partial derivative tester must correctly guess both i and the negative-slope region to detect such functions. The following two theorems establish the continuous and discrete cases of Theorem 1.7.

Theorem 6.6 (Lower bound for partial derivative testers on the unit cube)**.** *Any partial derivative L^1 monotonicity tester for Lipschitz functions $f : [0, 1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le 1$ (with two-sided error and adaptive queries) requires at least $\Omega(n/\epsilon)$ queries.*

Proof. Let $\epsilon \le 1/6$. For any $z \in \left[\frac{1}{3}, \frac{2}{3} - \epsilon\right]$, let $g_z : [0,1] \to \mathbb{R}$ be the function given by

$$g(x) = \begin{cases} \epsilon & \text{if } x < z\,, \\ \epsilon - (x - z) & \text{if } z \le x \le z + \epsilon\,, \\ 0 & \text{if } x > z + \epsilon\,. \end{cases}$$

Note that g_z is Lipschitz with $\mathsf{Lip}_1(g) = 1$. Moreover, we claim that $d_1(g_z) \gtrsim \epsilon$. Indeed, for any $x \in [0, 1/3]$, we have that $g_z(x) = \epsilon$ and $g_z(2/3 + x) = 0$. On the other hand, for any monotone function $h : [0,1] \to \mathbb{R}$ we must have $h(x) \le h(2/3 + x)$. Thus, for any such h we have $|g_z(x) - h(x)| + |g_z(2/3 + x) - h(2/3 + x)| \ge \epsilon$. Since this holds for all $x \in [0, 1/3]$, we conclude that for any such h we must have $\mathbb{E}\left[|g_z - h|\right] \ge \epsilon/3$, proving the claim.

Now, for any $i \in [n]$ and $z \in \left[\frac{1}{3}, \frac{2}{3} - \epsilon\right]$, let $f_{i,z} : [0,1]^n \to \mathbb{R}$ be given by copying g_z onto f along every line in direction i, i.e. setting $f_{i,z}(x) = g_z(x_i)$ for every $x \in [0,1]^n$. Note that $\mathsf{Lip}_1(f) = 1$ (since its partial derivatives are 0 along non-i coordinates), and $d_1(f) \gtrsim \epsilon$ (since the lines in direction i partition the domain).

We construct a set of $\Omega(n/\epsilon)$ functions $f_{i,z}$ as follows. First, i can be any of the coordinates in $[n]$. Then let $z_1, \dots, z_k$ be given by $z_j = \frac{1}{3} + k\epsilon$ for $k = \Omega(1/\epsilon)$, such that for each j we have $z_j \in \left[\frac{1}{3}, \frac{2}{3} - \epsilon\right]$ and, moreover, for distinct $j, \ell \in [k]$, the regions where f_{i,z_j} and f_{i,z_ℓ} take non-zero slope are disjoint. It follows that each partial derivative query may only rule out one such $f_{i,z}$, so any partial derivative tester that distinguishes an $f_{i,z}$ chosen uniformly at random from the constant-0 function must make at least $\Omega(n/\epsilon)$ queries. □

The argument for the hypergrid is similar, except that the construction cannot be made to occupy an arbitrarily small region of the domain when the domain is discrete. We opt to keep the argument simple and give a proof for constant parameter ϵ.

Theorem 6.7 (Lower bound for edge testers on the hypergrid)**.** *For sufficiently small constant ϵ, any partial derivative L^1 monotonicity tester for functions $f : [m]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le 1$ (with two-sided error and adaptive queries) requires at least $\Omega(nm)$ queries.*

Proof. Let m be a multiple of 3 for simplicity. For each $z \in \left\{\frac{m}{3} + 1, \dots, \frac{2m}{3}\right\}$, define $g_z : [m] \to \mathbb{R}$ by

$$g_z(x) = \begin{cases} 1 & \text{if } x < z\,, \\ 0 & \text{if } z \ge z\,. \end{cases}$$

Then $\mathsf{Lip}_1(g_z) = 1$ and, as before, we have $d_1(g_z) = \Omega(1)$. Then for each $i \in [n]$ and $z \in \left[\frac{m}{3} + 1, \frac{2m}{3}\right]$, we let $f_{i,z} : [m]^n \to \mathbb{R}$ be given by $f_{i,z}(x) = g_z(x_i)$ for each $x \in [m]^n$; it follows that $\mathsf{Lip}_1(f) = 1$ and $d_1(f) = \Omega(1)$. Note that there are $\Omega(nm)$ such functions. Moreover, each partial derivative query may only rule out one such $f_{i,z}$, and therefore any edge tester that distinguishes an $f_{i,z}$ chosen uniformly at random from the constant-0 function must make at least $\Omega(nm)$ queries. □

7 Overview of prior works on monotonicity testing

We first summarize results on testing monotonicity with respect to the Hamming distance.

Boolean-valued functions. Among the early works on this problem, [GGL+00] gave testers for functions on the hypergrid $[m]^n$ with query complexities $O(n \log(m)/\epsilon)$ and $O((n/\epsilon)^2)$; note that the latter bound is independent of m, and the query complexity of testers with this property was subsequently improved to $O((n/\epsilon)\log^2(n/\epsilon))$ by [DGL+99] and to $O((n/\epsilon)\log(n/\epsilon))$ by [BRY14a]. For functions on the Boolean cube $\{0,1\}^n$, [CS16] gave the first $o(n)$ tester, subsequently improved by [CST14], culminating in the $\widetilde{O}(\sqrt{n}/\epsilon^2)$ tester of [KMS18], which essentially resolved the question for nonadaptive testers. Whether adaptivity helps in monotonicity testing is still an open question; see the lower bounds below, and also [CS19].

Returning to hypergrid domains $[m]^n$, [BCS18, BCS20] established first testers with $o(n)$ query complexity and, via a domain reduction technique, also obtained $o(n)$ testers for product distributions on $\mathbb{R}^n$ (and the alternative proof of [HY22] improves the number of *samples* drawn by the tester when the distribution is unknown). Subsequent works [BCS22, BKKM22] attained the optimal dependence on n at the cost of a dependence on m, with upper bounds of the form $\widetilde{O}(\sqrt{n}\,\mathrm{poly}(m))$. Most recently, [BCS23] gave a tester with query complexity $O(n^{1/2+o(1)}/\epsilon^2)$, which is almost optimal for nonadaptive algorithms, and again extends to product measures on $\mathbb{R}^n$.

Real-valued functions. [EKK+98] gave a tester with query complexity $O(\log(m)/\epsilon)$ for real-valued functions on the line $[m]$; the tight query complexity of this problem was more recently shown to be $\Theta(\log(\epsilon m)/\epsilon)$ [Bel18]. As for functions on the hypergrid $[m]^n$, [GGL+00, DGL+99] also gave testers for larger ranges, but the query complexity depends on the size of the range. Then, [CS13] gave a nonadaptive tester with one-sided error and (optimal) query complexity $O(n\log(m)/\epsilon)$. On the Boolean cube, [BKR20] gave a tester with query complexity $\widetilde{O}\left(\min\left\{r\sqrt{n}/\epsilon^2, n/\epsilon\right\}\right)$ for real-valued functions f with image size r, and showed that this is optimal (for constant ϵ) for nonadaptive testers with one-sided error.

Lower bounds. We briefly summarize the known lower bounds for these problems; all lower bounds listed are for testers with two-sided error unless noted otherwise. For Boolean functions on the Boolean cube $\{0,1\}^n$, there is a near-optimal lower bound of $\widetilde{\Omega}(\sqrt{n})$ for nonadaptive testers [CWX17], which improves on prior results of [FLN+02, CST14, CDST15]. For adaptive testers, [BB16] gave the first polynomial lower bound of $\widetilde{\Omega}(n^{1/4})$, since improved to $\widetilde{\Omega}(n^{1/3})$ by [CWX17].

Turning to real-valued functions, [Fis04] combined Ramsey theory arguments with a result of [EKK+98] to show a $\Omega(\log m)$ lower bound for adaptive testers on the line $[m]$. On the Boolean cube, [BCGM12] gave a $\Omega(n/\epsilon)$ nonadaptive one-sided lower bound, and [BBM12] gave an adaptive lower bound of $\Omega(n)$. On the hypergrid, [BRY14b] gave a nonadaptive lower bound of $\Omega(n\log m)$ by communication complexity arguments, [CS14] showed the optimal lower bound of $\Omega(n\log(m)/\epsilon - \log(1/\epsilon)/\epsilon)$ for adaptive testers using Ramsey theory (which involves functions with large range), and [Bel18] gave an alternative proof of this bound that does not use Ramsey theory.

L^p-testing. Finally, moving from Hamming testers to L^p testers, and assuming functions with range $[0,1]$, [BRY14a] (who formally introduced this model) gave nonadaptive L^p monotonicity testers with one-sided error on the hypergrid $[m]^n$ with query complexity $O((n/\epsilon^p)\log(n/\epsilon^p))$—note this is independent of m, bypassing the Hamming testing lower bound—and a lower bound of $\Omega((1/\epsilon^p)\log(1/\epsilon^p))$ for nonadaptive testers with one-sided error; on the line, they showed there is an $O(1/\epsilon^p)$ nonadaptive tester with one-sided error and a matching lower bound for adaptive testers with two-sided error. They also gave a reduction from L^p monotonicity testing to Hamming testing of Boolean functions for nonadaptive one-sided testers, so in particular L^1 testing functions with range $[0,1]$ is no harder than Hamming testing functions with Boolean range.

We also remark that our problem, which is parameterized by the upper bound L on the Lipschitz constant of input functions, lies under the umbrella of parameterized property testing, and refer to [PRV17] for an introduction to, and results on this type of tester.

Acknowledgments. We thank Eric Blais for helpful discussions throughout the course of this project, and for comments and suggestions on preliminary versions of this paper.

References

[AD04] Gabriel Acosta and Ricardo Durán. An optimal poincaré inequality in L^1 for convex domains. *Proceedings of the American Mathematical Society*, 132(1):195–202, 2004.

[BB16] Aleksandrs Belovs and Eric Blais. A polynomial lower bound for testing monotonicity. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, page 1021–1032. ACM, Jun 2016.

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *computational complexity*, 21:311–358, 2012.

[BCGM12] Jop Briët, Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Monotonicity testing and shortest-path routing on the cube. *Combinatorica*, 32(1):35–53, 2012.

[BCS18] Hadley Black, Deeparnab Chakrabarty, and Comandur Seshadhri. A $o(d) \cdot \mathrm{polylog}\, n$ monotonicity tester for boolean functions over the hypergrid $[n]^d$. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2133–2151. SIAM, 2018.

[BCS20] Hadley Black, Deeparnab Chakrabarty, and Comandur Seshadhri. Domain reduction for monotonicity testing: A $o(d)$ tester for boolean functions in d-dimensions. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1975–1994. SIAM, 2020.

[BCS22] Hadley Black, Deeparnab Chakrabarty, and C Seshadhri. Directed isoperimetric theorems for boolean functions on the hypergrid and an $\widetilde{O}(n\sqrt{d})$ monotonicity tester. *arXiv preprint arXiv:2211.05281*, 2022.

[BCS23] Hadley Black, Deeparnab Chakrabarty, and C Seshadhri. A $d^{1/2+o(1)}$ monotonicity tester for boolean functions on d-dimensional hypergrids. *arXiv preprint arXiv:2304.01416*, 2023.

[Beb03] Mario Bebendorf. A note on the poincaré inequality for convex domains. *Zeitschrift für Analysis und ihre Anwendungen*, 22(4):751–756, 2003.

[Bel18] Aleksandrs Belovs. Adaptive lower bound for testing monotonicity on the line. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 103. Springer, 2014.

[BH97] Sergey G Bobkov and Christian Houdré. Isoperimetric constants for product probability measures. *The Annals of Probability*, pages 184–205, 1997.

[BKKM22] Mark Braverman, Subhash Khot, Guy Kindler, and Dor Minzer. Improved monotonicity testers via hypercube embeddings. *arXiv preprint arXiv:2211.09229*, 2022.

[BKR20] Hadley Black, Iden Kalemaj, and Sofya Raskhodnikova. Isoperimetric inequalities for real-valued functions with applications to monotonicity testing. *arXiv preprint arXiv:2011.09441*, 2020.

[BL96] Dominique Bakry and Michel Ledoux. Lévy–gromov's isoperimetric inequality for an infinite dimensional diffusion generator. *Inventiones mathematicae*, 123(2):259–281, 1996.

[BL97] Sergey Bobkov and Michel Ledoux. Poincaré's inequalities and talagrand's concentration phenomenon for the exponential distribution. *Probability Theory and Related Fields*, 107:383–400, 1997.

[Bob97] Sergey G Bobkov. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in gauss space. *The Annals of Probability*, 25(1):206–214, 1997.

[Bor75] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.

[BRY14a] Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. L^p-testing. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 164–173, 2014.

[BRY14b] Eric Blais, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Lower bounds for testing properties of functions over hypergrid domains. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 309–320. IEEE, 2014.

[BS88] Colin Bennett and Robert C Sharpley. *Interpolation of operators*. Academic press, 1988.

[BS16] Lorenzo Brasco and Filippo Santambrogio. A note on some poincaré inequalities on convex sets by optimal transport methods. In *Geometric Properties for Parabolic and Elliptic PDE's: GPPEPDEs, Palinuro, Italy, May 2015 4*, pages 49–63. Springer, 2016.

[BV04] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CDST15] Xi Chen, Anindya De, Rocco A. Servedio, and Li-Yang Tan. Boolean function monotonicity testing requires (almost) $n^{1/2}$ non-adaptive queries. In *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, page 519–528. ACM, Jun 2015.

[CS13] Deeparnab Chakrabarty and C. Seshadhri. Optimal bounds for monotonicity and lipschitz testing over hypercubes and hypergrids. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 419–428, 2013.

[CS14] Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory Of Computing*, 10(17):453–464, 2014.

[CS16] D. Chakrabarty and C. Seshadhri. An $o(n)$ monotonicity tester for boolean functions over the hypercube. *SIAM Journal on Computing*, 45(2):461–472, Jan 2016.

[CS19] D Chakrabarty and C Seshadhri. Adaptive boolean monotonicity testing in total influence time. *Innovations in Theoretical Computer Science (ITCS)*, 2019.

[CST14] Xi Chen, Rocco A. Servedio, and Li-Yang Tan. New algorithms and lower bounds for monotonicity testing. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, page 286–295, Oct 2014.

[CT80] Michael G Crandall and Luc Tartar. Some relations between nonexpansive and order preserving mappings. *Proceedings of the American Mathematical Society*, 78(3):385–390, 1980.

[CWX17] Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond talagrand functions: new lower bounds for testing monotonicity and unateness. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 523–536, 2017.

[CWZ21] Shuyu Cheng, Guoqiang Wu, and Jun Zhu. On the convergence of prior-guided zeroth-order optimization algorithms. *Advances in Neural Information Processing Systems*, 34:14620–14631, 2021.

[DGL+99] Yevgeniy Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. In *RANDOM/APPROX*, pages 97–108. Springer, 1999.

[EKK+98] Funda Ergün, Sampath Kannan, S Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 259–268, 1998.

[EKLM22] Ronen Eldan, Guy Kindler, Noam Lifshitz, and Dor Minzer. Isoperimetric inequalities made simpler. *arXiv preprint arXiv:2204.06686*, 2022.

[Fis04] Eldar Fischer. On the strength of comparisons in property testing. *Information and Computation*, 189(1):107–116, 2004.

[FLN+02] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 474–483, 2002.

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 3(20):301–337, 2000.

[Gro75] Leonard Gross. Logarithmic sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[HY22] Nathaniel Harms and Yuichi Yoshida. Downsampling for testing and learning in product distributions. In *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[Kaw85] Bernhard Kawohl. *Rearrangements and convexity of level sets in PDE*, volume 1150 of *Lecture Notes in Mathematics*. Springer-Verlag Berlin, Heidelberg, 1985.

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric-type theorems. *SIAM Journal on Computing*, 47(6):2238–2276, 2018.

[KN15] Nikolay Kuznetsov and Alexander Nazarov. Sharp constants in the poincaré, steklov and related inequalities (a survey). *Mathematika*, 61(2):328–344, 2015.

[Lat03] Rafał Latała. On some inequalities for gaussian measures. *arXiv preprint math/0304343*, 2003.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[Poi90] Henri Poincaré. Sur les équations aux dérivées partielles de la physique mathématique. *American Journal of Mathematics*, pages 211–294, 1890.

[PRV17] Ramesh Krishnan S Pallavoor, Sofya Raskhodnikova, and Nithin Varma. Parameterized property testing of functions. *ACM Transactions on Computation Theory (TOCT)*, 9(4):1–19, 2017.

[PRW22] Ramesh Krishnan S Pallavoor, Sofya Raskhodnikova, and Erik Waingarten. Approximating the distance to monotonicity of boolean functions. *Random Structures & Algorithms*, 60(2):233–260, 2022.

[PW57] LE Payne and HF Weinberger. Lower bounds for vibration frequencies of elastically supported membranes and plates. *Journal of the Society for Industrial and Applied Mathematics*, 5(4):171–182, 1957.

[ST78] Vladimir N Sudakov and Boris S Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics*, 9(1):9–18, 1978.

[Tal93] Michel Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis' graph connectivity theorem. *Geometric & Functional Analysis GAFA*, 3(3):295–314, 1993.

[Ver99] Rüdiger Verfürth. A note on polynomial approximation in sobolev spaces. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(4):715–719, 1999.

[Wan05] Xianfu Wang. Volumes of generalized unit balls. *Mathematics Magazine*, 78(5):390–395, 2005.

A Background on isoperimetric inequalities

Let us trace an extremely brief history of developments that are most relevant to our study of isoperimetric inequalities. We start with the original work of Poincaré [Poi90], which yields inequalities of the type $\|f - \mathbb{E}[f]\|_p \le C(\Omega)\|\nabla f\|_p$ for sufficiently smooth domains Ω and sufficiently integrable functions f[11]. The optimal constant $C(\Omega)$, also called the *Poincaré constant* of Ω, depends on properties of this domain[12], and often the goal is to establish the sharp constant for families of domains Ω. [PW57] and [AD04] (see also [Beb03]) showed that, for convex domains, $C(\Omega)$ is essentially upper bounded by the diameter of Ω, and this bound is tight in general. However, for specific structured domains such as the product domain $[0,1]^n$, the diameter characterization

[11] More precisely, for f in the appropriate Sobolev space.

[12] For example, it is characterized by the first nontrivial eigenvalue of the Laplacian operator on smooth bounded Ω. There are additional considerations relating the assumptions made of f on the boundary $\partial\Omega$ and the (Dirichlet or Neumann) boundary conditions associated with the Laplacian; see [KN15] for a survey.

Inequality \ Setting		Discrete $\{0,1\}^n \to \{0,1\}$	Discrete $\{0,1\}^n \to \mathbb{R}$	Continuous $[0,1]^n \to \mathbb{R}$
(L^1, ℓ^1)-Poincaré	$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_1\right]$	* [Tal93]	* [Tal93]	* [BH97]
	$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_1\right]$	[GGL+00]	Theorem 1.3	Theorem 1.2
(L^1, ℓ^2)-Poincaré	$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_2\right]$	* [Tal93]	[Tal93]	[BH97]
	$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_2\right]$	[KMS18]	?	Conjecture 1.8
Related inequalities	$d_0^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\phi f\|_2\right]$	For $f : \{0,1\}^n \to \mathbb{R}$ [BKR20]		
	$d_0^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\phi^- f\|_2\right]$	For $f : \{0,1\}^n \to \mathbb{R}$ [BKR20]		
	$d_0^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\phi^- f\|_2\right]$	For $f : [m]^n \to \{0,1\}$ [BCS22, BKKM22]		

Table 3: Classical and directed functional inequalities on discrete and continuous domains. Cells marked with * indicate inequalities that follow from another entry in the table. For simplicity, logarithmic factors in the inequalities are ignored.

falls short of yielding a dimension-free inequality (see also the literature on logarithmic Sobolev inequalities [Gro75]).

Making progress on this front in the discrete setting, the landmark work of Talagrand [Tal93] established inequalities like the above for domain $\Omega = \{0,1\}^n$, with $C = C(\Omega)$ independent of n, and established connections with earlier works of Margulis on graph connectivity and Pisier on probability in Banach spaces. (More recently, Fourier-analytic proofs of Talagrand's inequality have also been given [EKLM22].) In continuous settings, similar results were first established for the Gaussian measure in connection with the Gaussian isoperimetric inequality [Bob97, BL96, ST78, Bor75, Lat03]. Tying back to our present settings of interest, Bobkov and Houdré [BH97] showed that a dimension-independent Poincaré-type inequality also holds for product measures in $\mathbb{R}^n$, including the uniform measure on $[0,1]^n$, as shown in (3).

As introduced in the opening, it is these dimension-free Poincaré inequalities for discrete and continuous product measures whose directed analogues have implications for the structure of monotone functions and therefore for property testing [GGL+00, CS16, KMS18]. To enrich the summary laid out in Table 1, we present additional related inequalities recently shown by [BKR20, BCS22, BKKM22] in Table 3, and briefly explain them here. These inequalities have unlocked algorithmic results for testing monotonicity of real-valued functions on the Boolean cube, and Boolean-valued functions on the hypergrid, as summarized in Section 7.

Define the vector-valued operators ϕ and ϕ^- on functions $f : [m]^n \to \mathbb{R}$ as follows: for each $x \in [m]^n$ and $i \in [n]$,

$$(\phi f(x))_i := \mathbb{1}\left[\exists y : (x \preceq_i y \text{ or } y \preceq_i x) \text{ and } f(x) \neq f(y)\right],$$
$$(\phi^- f(x))_i := \mathbb{1}\left[(\exists y : x \preceq_i y, f(x) > f(y)) \text{ or } (\exists y : y \preceq_i x, f(y) > f(x))\right],$$

where we write $x \preceq_i y$ if $x_j = y_j$ for every $j \neq i$, and $x_i \leq y_i$. Compared to the gradient, these operators 1) are only sensitive to the order relation between function values (which suits the setting of Hamming testing); and 2) capture "long range" violations of monotonicity (accordingly, the corresponding hypergrid testers are not edge testers). See Section 1.2.1 for a remark on the nuances of inner/outer boundaries and robust inequalities.

B Upper bounds from [BRY14a] applied to Lipschitz functions

In this section, we show how the L^1 monotonicity testing upper bounds from [BRY14a] imply testers with query complexity $\widetilde{O}\left(\frac{n^2 L}{\epsilon}\right)$ on the unit cube and $\widetilde{O}\left(\frac{n^2 mL}{\epsilon}\right)$ on the hypergrid for functions f satisfying $\mathsf{Lip}_1(f) \le L$. We start with the case of the hypergrid, and first state the upper bound of [BRY14a] for functions with arbitrary range of size r, which without loss of generality (by translation invariance) we denote $[0, r]$:

Theorem B.1 ([BRY14a])**.** *There exists an L^1 monotonicity tester for functions $f : [m]^n \to [0, r]$ that uses $O\left(\frac{rn}{\epsilon} \log \frac{rn}{\epsilon}\right)$ value queries. The tester is nonadaptive and has one-sided error.*

As explained in Section 1.3.3, the extra factor of r compared to the bounds stated in [BRY14a] accounts for the conversion between range $[0, r]$ and range $[0, 1]$, which affects the proximity parameter ϵ by a factor of r: testing functions with range $[0, r]$ for proximity parameter ϵ is equivalent to testing functions with range $[0, 1]$ for proximity parameter ϵ/r.

We thus obtain the following L^1 monotonicity tester for Lipschitz functions on the hypergrid:

Corollary B.2. *There is an L^1 monotonicity tester for functions $f : [m]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ that uses $O\left(\frac{n^2 mL}{\epsilon} \log\left(\frac{nmL}{\epsilon}\right)\right)$ value queries. The tester is nonadaptive and has one-sided error.*

Proof. The algorithm simulates the tester from Theorem B.1 with parameters $r = Lmn$ and ϵ, which gives the announced query complexity. As for correctness, note that since $\mathsf{Lip}_1(f) \le L$, for any $x, y \in [m]^n$ we have $|f(x) - f(y)| \le L\|x - y\|_1 \le Lmn$. Thus f has range of size at most Lmn, so the reduction to the tester from Theorem B.1 is correct. □

We outline how this result implies a tester for Lipschitz functions on the unit cube as well, via the domain reduction or downsampling principle of [BCS20, HY22]. Even though the tester above has query complexity that depends on m, the main observation is that, given an (ℓ^1, L)-Lipschitz function on $[0, 1]^n$, we may discretize it into an $(\ell^1, L/m)$-Lipschitz function on an arbitrarily fine hypergrid $[m]^n$. Then the term $m(L/m) = L$ remains fixed for any choice of m, so the complexity of the tester above does not depend on m in this reduction. Finally, by setting m large enough, we may upper bound the error introduced by the discretization.

Corollary B.3. *There is an L^1 monotonicity tester for functions $f : [0, 1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ that uses $O\left(\frac{n^2 L}{\epsilon} \log\left(\frac{nL}{\epsilon}\right)\right)$ value queries. The tester is nonadaptive and has one-sided error.*

Proof sketch. For sufficiently large m to be chosen below, the algorithm imposes a hypergrid $[m]^n$ uniformly on $[0, 1]^n$. More precisely, we define a function $f' : [m]^n \to \mathbb{R}$ via $f'(x) = f(x/m)$ for all $x \in [m]^n$. Note that $\mathsf{Lip}_1(f') \le L/m$. Then, the algorithm simulates the tester from Corollary B.2 on f' with proximity parameter $\Omega(\epsilon)$, and returns its result. Note that the query complexity is as desired, so it remains to show correctness. If f is monotone so is f', in which case the algorithm accepts. Now assume that $d_1(f) > \epsilon$, and we need to show that $d_1(f') \gtrsim \epsilon$.

Let $g' : [m]^n \to \mathbb{R}$ be any monotone function. We obtain a monotone function $g : [0, 1]^n \to \mathbb{R}$ as follows: for each $x \in [0, 1]^n$, let $\overline{x}$ be obtained by rounding up each coordinate x_i to a positive integral multiple of $1/m$. Then $x' := m\overline{x} \in [m]^n$, and we set $g(x) := g'(x')$. It follows that g is monotone, and by the triangle inequality,

$$\mathbb{E}\left[f-g\right] \le \mathbb{E}\left[f'-g'\right] + \max_{x\in[0,1]^n} |f(x)-f(\overline{x})| \le \mathbb{E}\left[f'-g'\right] + \frac{Ln}{m}\,.$$

In the first inequality, we separately account for the cost of turning each value of f into the value at its corresponding "rounded up" point (accounted for by the second summand), and then the cost of turning each equal-sized, constant-valued cell into the value of g on that cell, and these values agree with f' and g' (accounted for by the first summand). In the second inequality, we use the fact that any point x satisfies $\|x-\overline{x}\|_1 \le \frac{1}{m}\cdot n$, along with the Lipschitz assumption on f.

Therefore, by setting $m > \frac{10Ln}{\epsilon}$, we obtain $\frac{Ln}{m} < \frac{\epsilon}{10}$, and since the inequality above holds for every monotone function g', we conclude that $d_1(f') \ge d_1(f) - \epsilon/10 > 9\epsilon/10$, as desired. ☐

Remark B.4. One may wonder whether the reductions above could yield more efficient testers if combined with Hamming testers for Boolean functions on the hypergrid with better dependence on n (via the reduction from L^1 testing to Hamming testing of [BRY14a], which is behind Theorem B.1), since e.g. the tester of [BCS23] has query complexity $\widetilde{O}(n^{1/2+o(1)}/\epsilon^2)$. However, it seems like this is not the case, i.e. Corollary B.2 has the best query complexity of any reduction that upper bounds the size of the range of f by Lmn. The reason is as follows: [KMS18] showed that any nonadaptive, one-sided pair tester for the Boolean cube with query complexity $O(n^\alpha/\epsilon^\beta)$ must satisfy $\alpha + \frac{\beta}{2} \ge \frac{3}{2}$, so hypergrid testers must also satisfy this as well as $\beta \ge 1$, assuming query complexity independent of m. Then, given a hypergrid tester for Boolean functions with query complexity $\widetilde{O}(n^\alpha/\epsilon^\beta)$, our reduction via the inferred range size $r = Lnm$ gives an L^1 tester with asymptotic query complexity at least $\frac{n^\alpha}{(\epsilon/r)^\beta} = \frac{n^{\alpha+\beta/2} n^{\beta/2}(mL)^\beta}{\epsilon^\beta} \ge \frac{n^2 mL}{\epsilon}$.

C Lower bound from [BRY14b] applied to L^1 testing

We briefly explain how the $\Omega(n \log m)$ nonadaptive lower bound of [BRY14b] for Hamming testing monotonicity of functions $f : [m]^n \to \mathbb{R}$ (with sufficiently small constant ϵ) also applies to L^1-testing functions satisfying $\mathsf{Lip}_1(f) \le O(1)$. The construction of [BRY14b] relies on two main ingredients: *step functions* and *Walsh functions*.

Let $m = 2^\ell$ for simplicity. For each $i \in \{0,\dots,m\}$ the i-th step function $s_i : [2^\ell] \to [2^{\ell-i}]$ is given by

$$s_i(x) = \left\lfloor \frac{x-1}{2^i} \right\rfloor + 1\,.$$

In words, $s_i(x)$ increases by 1 after every 2^i consecutive elements (called a *block* of size 2^i).

The Walsh functions are defined as follows. For each $i \in [\ell]$, the function $w_i : [2^\ell] \to \{\pm 1\}$ is given by

$$w_i(x) = (-1)^{\mathsf{bit}_i(x-1)}\,,$$

where the operator bit_i extracts the i-th bit of its input (indexed from least to most significant). Then for each $S \subseteq [\ell]$, the function $w_S : [2^\ell] \to \{\pm 1\}$ is given by

$$w_S(x) = \prod_{i\in S} w_i(x)\,.$$

These two types of functions are defined on the line, and they are extended into a multidimensional construction on the hypergrid as follows. Given a vector $\boldsymbol{i} \in [\ell]^n$, the step function

$s_{\boldsymbol{i}} : [2^\ell]^n \to [n2^\ell]$ is given by

$$s_{\boldsymbol{i}}(x_1, \dots, x_n) = \sum_{j=1}^{n} s_{\boldsymbol{i}_j}(x_j)\,,$$

and given a vector $\boldsymbol{S} = (\boldsymbol{S}_1, \dots, \boldsymbol{S}_n)$ of subsets of $[\ell]$, the Walsh function $w_{\boldsymbol{S}} : [2^\ell]^n \to \{\pm 1\}$ is given by

$$w_{\boldsymbol{S}}(x_1, \dots, x_n) = \prod_{j=1}^{n} w_{\boldsymbol{S}_j}(x_j)\,.$$

Then, [BRY14b] use a communication complexity argument (namely a reduction from the AugmentIndex problem) to show that (nonadaptive) Hamming testing monotonicity of functions $h_{\boldsymbol{i},\boldsymbol{S}} : [2^\ell]^n \to \mathbb{N}$ of the form

$$h_{\boldsymbol{i},\boldsymbol{S}}(x) = 2s_{\boldsymbol{i}}(x) + w_{\boldsymbol{S}}(x)\,,$$

for appropriate choices of $\boldsymbol{i}$ and $\boldsymbol{S}$, requires at least $\Omega(n \log m)$ queries. Therefore, to show that (nonadaptive) L^1 testing monotonicity of $(\ell^1, O(1))$-Lipschitz functions also requires at least this number of queries, it suffices to show that every such function $h = h_{\boldsymbol{i},\boldsymbol{S}}$ satisfies

1. $\mathsf{Lip}_1(h) \le O(1)$; and
2. If $d_0(h) \ge \epsilon$, then $d_1(h) \gtrsim \epsilon$.

The first property follows from the definitions of the step and Walsh functions: let $x, y \in [2^\ell]^n$ be such that $\|x - y\|_1 = 1$. Then let $j \in [\ell]$ be the coordinate such that $|x_j - y_j| = 1$ and $x_k = y_k$ for $k \ne j$. Then

$$|h_{\boldsymbol{i},\boldsymbol{S}}(x) - h_{\boldsymbol{i},\boldsymbol{S}}(y)| \le 2\left|s_{\boldsymbol{i}_j}(x_j) - s_{\boldsymbol{i}_j}(y_j)\right| + |w_{\boldsymbol{S}}(x) - w_{\boldsymbol{S}}(y)| \le 2 + 2 = 4\,,$$

the first inequality because x and y agree on every coordinate except for j, and the second inequality because the step function $s_{\boldsymbol{i}_j}$ changes by at most 1 on adjacent inputs, and the Walsh functions only take values ± 1. Thus $\mathsf{Lip}_1(h_{\boldsymbol{i},\boldsymbol{S}}) \le 4$, as desired.

As for the second property, note that if $d_0(h) \ge \epsilon$, then there exists a matching of the form $(x^i, y^i)_i$ where for each i we have $x^i \preceq y^i$ and $h(x^i) > h(y^i)$ (i.e. x^i, y^i form a violating pair), such that at least an ϵ-fraction of the points $[m]^n$ belong to this matching (see [FLN+02]). Now, for each such violating pair x^i, y^i, it follows that $h(x^i) - h(y^i) \ge 1$, since h is integer-valued. Therefore for any monotone function $h' : [m]^n \to \mathbb{R}$, it must be the case that $\left|h(x^i) - h'(x^i)\right| + \left|h(y^i) - h'(y^i)\right| \ge 1$. Since this is true for disjoint pairs x^i, y^i covering an ϵ-fraction of the domain, it follows that $d_1(h, h') \ge \epsilon/2$ for any monotone function h'. Hence $d_1(h) \ge \epsilon/2$, and we are done.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il