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

Renato Ferreira Pinto Jr.
University of Waterloo, Canada

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.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms; Theory of computation → Lower bounds and information complexity; Mathematics of computing → Mathematical analysis

Keywords and phrases Monotonicity testing, property testing, isoperimetric inequalities, Poincaré inequalities

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.61

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/abs/2307.02193` [24]

Funding *Renato Ferreira Pinto Jr.*: Partly funded by an NSERC Canada Graduate Scholarship Doctoral Award.

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

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. [31]): 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 [17] were the first to notice that the main ingredient of the work of [27], 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 [27] 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, [17] gave the first monotonicity tester with $o(n)$ query complexity by showing a directed analogue of Margulis's inequality. This was improved by [20], and eventually the seminal paper of [30] 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 $f : \{0,1\}^n \to \{0,1\}$,

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

Compared to (1), this replaces the ℓ^1-norm of the gradient with its ℓ^2-norm. [30] showed the natural directed analogue[3] up to polylogarithmic factors, which were later removed by [32]: 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 [7, 5, 16, 6] (see also [8, 28]) and real-valued functions on the Boolean cube [9].

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

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 [26].

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

defined on continuous domains, as well as an impressive range of connections to topics such as partial differential equations [33], Markov diffusion processes [1], probability theory and concentration of measure [12], optimal transport [15], polynomial approximation [35], among others.

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 [13]:

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

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* [27, 18] 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 $d_p(f,g) := \mathbb{E}_{x\sim\mu}\left[|f(x) - g(x)|^p\right]^{1/p}$. 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 $d_p(f, \mathcal{P}) > \epsilon$.*

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 [14], 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 [21]. 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 [4] when the range is $R = [0, 1]$. On the other hand, for general R, we keep the distance metric unmodified, whereas [4] 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 [27].

▶ **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 [27] 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*

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

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.

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.

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	$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_1\right]$	* [34]	* [34]	* [13]
	$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_1\right]$	[27]	Theorem 1.3	Theorem 1.2
(L^1, ℓ^2)-Poincaré	$d_1^{\mathsf{const}}(f) \lesssim \mathbb{E}\left[\|\nabla f\|_2\right]$	* [34]	[34]	[13]
	$d_1^{\mathsf{mono}}(f) \lesssim \mathbb{E}\left[\|\nabla^- f\|_2\right]$	[30]	?	Conjecture 1.8

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 [30]), 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. [34]). 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* [27, 18].

▶ **Theorem 1.4.** *There is a nonadaptive partial derivative L^1 monotonicity tester for Lipschitz functions $f : [0,1]^n \to \mathbb{R}$ satisfying $\mathsf{Lip}_1(f) \le L$ 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 $\mathsf{Lip}_1(f) \le L$ 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 [27, 18]. 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 $\Theta(\log m)$ [23, 25, 11, 2]. 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 [25, 19, 2].

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:

► **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.*

[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.

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 full version of the paper for details. Our lower bounds hold either for constant (const.) n, or for partial derivative testers (p.d.t.).

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	$\widetilde{O}\left(\frac{n^2 L}{\epsilon}\right)$ (*) [4]	$O\left(\frac{nL}{\epsilon}\right)$ 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.
$\Omega = [m]^n$	$O\left(\frac{n \log m}{\epsilon}\right)$ [18]	$\widetilde{O}\left(\frac{n^2 mL}{\epsilon}\right)$ (*) [4]	$O\left(\frac{nmL}{\epsilon}\right)$ p.d.t.
	$\Omega\left(\frac{n\log(m)-\log(1/\epsilon)}{\epsilon}\right)$ [19]	$\widetilde{\Omega}\left(\frac{L}{\epsilon}\right)$ n.a. 1-s. [4] $\Omega(n \log m)$ n.a. [11]	$\Omega\left(\left(\frac{mL}{\epsilon}\right)^{\frac{n}{n+1}}\right)$ const. n $\Omega(nm)$ p.d.t.

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 the full version of the paper 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.

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 [27] and Theorems 1.2 and 1.3) have more straightforward proofs than ℓ^2 counterparts such as [30]. 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 [10, 11], whose translation to the continuous setting is not obvious[5], by reduction to comparison-based testers [19], whose connection to L^1 testing setting seems less immediate, or directly via a careful construction [2]. 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 [30], as an exciting direction for future work.

1.3.3 Relation to prior work on L^p-testing

[4] 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.

[4] 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 [4] and a discretization argument for the unit cube inspired by [8, 28], we establish benchmarks for our testing problem. See the full version of the paper for further 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 of the Lipschitz assumption – via the directed Poincaré inequalities

[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 [4]: 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.

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 [4, 11] 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 ([4] 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 notation. In Section 3 we prove our directed Poincaré inequality on the unit cube, and in Section 4 we give our L^1 monotonicity tester for this domain. The analogous versions for the discrete case of the hypergrid, as well as the proofs of our results for testing on the line (Section 1.2.3) and lower bounds (Section 1.2.4) may be found in the full version of the paper.

2 Preliminaries

For integer $m \geq 1$, we write $[m]$ to denote the set $\{i \in \mathbb{Z} : 1 \leq i \leq m\}$. 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 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 $p \geq 1$, 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 on $\mathbb{R}^n$ (the dimension n being clear from context), and simply write $L^p(\Omega)$ for $L^p(\Omega, \nu)$; we will 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:*

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

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

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.

3 Directed Poincaré inequalities for Lipschitz functions

In this section, we establish Theorem 1.2. We start with the one-dimensional case, i.e. functions on the line, and then generalize to higher dimensions. See the full version of the paper for the discrete case of 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.

---

[7] More precisely: when $\Omega = [0, 1]^n$, μ will be the Lebesgue measure on Ω (with associated σ-algebra Σ).

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 [29], 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 [3, 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 [3, 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 $\int f \, d\nu = \int f^* \, d\nu$ and therefore

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

Hence our goal is to show that

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

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

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

The result follows by applying this estimate to all x:

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

3.2 Multidimensional case

Although we ultimately only require an inequality on the unit cube $[0, 1]^n$, 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 3.5.** *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 3.6** (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 3.7.** *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 $i \in [n]$,*

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

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 multidimensional argument is that the rearrangement operator preserves Lipschitz continuity:

▶ **Lemma 3.8** ([29, 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 3.9.** *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 3.10.** *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 $0 \leq y_1 < y_2 \leq a_j$, so that $(x^{-j}, y_1), (x^{-j}, y_2) \in \overline{B}$. We need to show that $(R_i f)(x^{-j}, y_1) \leq (R_i f)(x^{-j}, y_2)$. 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

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

Note that

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

for every $z \in \overline{I}_i$, and therefore our goal is to show that $g_1^*(x_i) \leq g_2^*(x_i)$. But f being monotone in direction j means that $g_1(z) \leq g_2(z)$ for all $z \in \overline{I}_i$, so by the order preserving property (Fact 3.3) of the monotone rearrangement we get that $g_1^*(x_i) \leq g_2^*(x_i)$, 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 3.11** ([22]). *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 [29]:

▶ **Corollary 3.12.** *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 3.13.** *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 3.8. 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

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

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 3.11, 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 3.14.** *Let B be the box of size $a \in \mathbb{R}^n$ and let $f : \overline{B} \to \mathbb{R}$ be Lipschitz continuous. Then*

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

Proof. We have

$$\begin{aligned}
\|f - f^*\|_{L^1} &\le \sum_{i=1}^{n} \|R_{i-1} \cdots R_1 f - R_i \cdots R_1 f\|_{L^1} && \text{(Triangle inequality)} \\
&\le 2 \sum_{i=1}^{n} a_i \left\|\partial_i^- (R_{i-1} \cdots R_1 f)\right\|_{L^1} && \text{(Lemma 3.8 and Proposition 3.7)} \\
&\le 2 \sum_{i=1}^{n} a_i \left\|\partial_i^- f\right\|_{L^1} && \text{(Lemma 3.8 and Proposition 3.13)} .
\end{aligned}$$

◀

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

▶ **Corollary 3.15.** *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 [30] for the Boolean cube:

▶ **Proposition 3.16.** *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 3.12, 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. ◀

To check that Corollary 3.15 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.

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

Input: Partial derivative oracle access to Lipschitz function $f : [0,1]^n \to \mathbb{R}$.
Output: Accept if f is monotone, reject if $d_1(f) > \epsilon$.
Requirement: $\mathsf{Lip}_1(f) \le L$.
procedure PartialDerivativeTester(f, L, ϵ)
  repeat $\Theta(nL/\epsilon)$ **times**
    Sample $x \in [0,1]^n$ uniformly at random.
    Sample $i \in [n]$ uniformly at random.
    Reject if $\partial_i f(x) < 0$.
  end repeat
  Accept.

4 Application to monotonicity testing

In this section, we use the directed Poincaré inequality on the unit cube to show that the natural partial derivative tester attains the upper bound from Theorem 1.4. As noted in the introduction, we refer to the full version of the paper for the case of the hypergrid.

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 $d_1(f) > \epsilon$.

▶ **Lemma 4.1.** *Let $f : [0,1]^n \to \mathbb{R}$ 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. 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 3.15 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. ◀

References

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

2 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.

3 Colin Bennett and Robert C Sharpley. *Interpolation of operators*. Academic press, 1988.

4 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.

5 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*, 2022. arXiv:2211.05281.

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

7 Hadley Black, Deeparnab Chakrabarty, and Comandur Seshadhri. A $o(d) \cdot \operatorname{polylog} n$ 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.

8 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.

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

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

11 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.

12 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.

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

14 Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

15 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.

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

17 D. Chakrabarty and C. Seshadhri. An $o(n)$ monotonicity tester for boolean functions over the hypercube. *SIAM Journal on Computing*, 45(2):461–472, January 2016. doi:10.1137/13092770X.

18 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.

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

20 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*, pages 286–295, October 2014. doi:10.1109/FOCS.2014.38.

21 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.

22 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.

23 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.

24 Renato Ferreira Pinto Jr. Directed poincaré inequalities and L^1 monotonicity testing of lipschitz functions. *arXiv preprint*, 2023. `arXiv:2307.02193`.

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

26 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.

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

28 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.

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

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

31 Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

32 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.

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

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

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