

ALTERNATIVE PROOF OF KEITH–ZHONG SELF-IMPROVEMENT AND CONNECTIVITY

Sylvester Eriksson-Bique

UCLA, Department of Mathematics
520 Portola Plaza, Los Angeles CA 90095, U.S.A.; syerikss@math.ucla.edu

Abstract. We find a new proof for the celebrated theorem of Keith and Zhong that a $(1, p)$ -Poincaré inequality self-improves to a $(1, p - \epsilon)$ -Poincaré inequality. The paper consists of a novel characterization of Poincaré inequalities and then uses it to give an entirely new proof which is closely related to Muckenhoupt-weights. This new characterization, and the alternative proof, demonstrate a formal similarity between Muckenhoupt-weights and Poincaré inequalities. The proofs we give are short and somewhat more direct. With them we can give the first completely transparent bounds for the quantity of self-improvement and the constants involved. We observe that the quantity of self-improvement is, for large p , directly proportional to p , and inversely proportional to a power of the doubling constant and the constant in the Poincaré inequality. The proofs can be localized and thus we obtain more transparent proofs of the self-improvement of local Poincaré inequalities.

1. Introduction

1.1. Self-improvement of Poincaré inequalities. Our goal is two-fold. On the one hand, we wish to reprove a result by Keith and Zhong on the self-improvement of Poincaré inequalities [12], and to give explicit bounds for the quantity of self-improvement. Prior to Keith’s and Zhong’s result it was common to assume a $(1, q)$ -Poincaré inequality for some $q < p$ when proving statements involving functions in the Sobolev space with an exponent p . The result of Keith and Zhong replaces this assumption with a more natural assumption of a $(1, p)$ -Poincaré inequality, and thus is widely applied in the study of analysis on metric measure spaces. Despite its significance, its proof has remained somewhat mysterious to many outside of a small community of experts. In order to remedy this situation, we aim to give a more direct and transparent proof, that is based on new ideas of iteration and curve fragments. These ideas may become useful in studying other self-improvement phenomena as well.

On the other hand, our goal is to draw attention to an intimate connection between the theory of Muckenhoupt-weights (see [16]) and Poincaré inequalities. It is well-known, that the results of self-improvement for Muckenhoupt-weights and Poincaré inequalities bear striking similarity. However, that this similarity extends to the level of proofs and definitions is surprising. When the underlying metric space is $X = \mathbf{R}$, Muckenhoupt weights coincide with those doubling measures permitting Poincaré inequalities [3]. In a general metric space the question is much more subtle, but we describe a sense in which a Poincaré inequality can be characterized by a Muckenhoupt-type condition “along some curves”.

<https://doi.org/10.5186/aasfm.2019.4424>

2010 Mathematics Subject Classification: Primary 30L99, 42B25, 39B72.

Key words: Poincaré inequality, self-improvement, metric spaces, PI-spaces, analysis on metric spaces, connectivity, Muckenhoupt-weights.

To state the result, we will need the following terminology. For simplicity, we will consistently work with proper metric measure spaces (X, d, μ) equipped with locally finite measures μ such that $0 < \mu(B(x, r)) < \infty$ for all open balls $B(x, r) \subset X$.

Definition 1.1. A proper metric measure space (X, d, μ) equipped with a Radon measure μ is said to be *D-doubling* if for all $0 < r$ and any $x \in X$ we have

$$(1.2) \quad \frac{\mu(B(x, 2r))}{\mu(B(x, r))} \leq D.$$

We say that (X, d, μ) is *D-doubling up to scale r_0* if the same holds for all $r \in (0, r_0)$.

The average of a measurable function $f: X \rightarrow \mathbf{R}$ on a metric measure space (X, d, μ) over a measurable set A , with $0 < \mu(A) < \infty$, is denoted by

$$f_A := \int_A f d\mu := \frac{1}{\mu(A)} \int_A f d\mu,$$

when it makes sense, and its local (upper) Lipschitz constant is defined as

$$\text{Lip } f(x) := \limsup_{y \rightarrow x, y \neq x} \frac{|f(x) - f(y)|}{d(x, y)}.$$

If $B = B(x, r)$ is a ball, we denote $CB = B(x, Cr)$ (despite the ambiguity that a ball as a set may not be uniquely defined by a center and a radius).

Definition 1.3. Let $1 \leq p < \infty$ be given. A proper metric measure space (X, d, μ) with a Radon measure μ and $\text{supp}(\mu) = X$ is said to satisfy a $(1, p)$ -Poincaré inequality (with constants (C, C_{PI})) if for all Lipschitz functions f and all $x \in X$, $0 < r$ we have for $B = B(x, r)$

$$\int_B |f - f_B| d\mu \leq C_{PI} r \left(\int_{CB} (\text{Lip } f)^p d\mu \right)^{\frac{1}{p}}.$$

We say that (X, d, μ) satisfies a $(1, p)$ -Poincaré inequality (with constants (C, C_{PI})) up to scale $r_0 > 0$ if the same holds for all $r \in (0, r_0)$. If X is *D-doubling* and satisfies a $(1, p)$ -Poincaré inequality, then it is called a *PI-space*.

This inequality could be expressed in different generalities, but we choose this simple expression as it is sufficient. For a detailed discussion of these issues we refer to [11, 6, 7].

By an application of Hölder's inequality, we can see that for smaller p the $(1, p)$ -Poincaré inequality becomes stronger. Thus, the following theorem of Keith and Zhong is called a self-improvement result. For a more detailed discussion of the background we refer to [12, 2]. For the original proof, see [12], or its presentation in [9].

Theorem 1.4. (Keith–Zhong [12]) Assume $p > 1$. Let (X, d, μ) be a proper *D-doubling metric measure space* with a $(1, p)$ -Poincaré inequality with constants (C, C_{PI}) . There exists a positive constant $\epsilon(D, p, C_{PI}) > 0$ such that for any $\epsilon \in (0, \epsilon(D, p, C_{PI}))$ the space admits a $(1, p - \epsilon)$ -Poincaré inequality with constants $C' = C'(D, C_{PI}, \epsilon, C)$, $C'_{PI} = C'_{PI}(D, p, C_{PI}, \epsilon, C)$.

Our proof gives the following bound, which shows that the quantity of self-improvement is independent of the inflation factor C ,

$$\epsilon(D, p, C_{PI}) \geq \frac{p}{(2^{13p+3} C_{PI}^p D^{3p+4})^{\frac{1}{p-1}}}.$$

Letting $p \rightarrow \infty$, we obtain the asymptotic estimate for the improvement $\frac{p}{2^{12}C_{PI}D^3}$. This is, naturally, not a tight bound. This estimate means that for larger p the improvement in Keith–Zhong becomes larger, and in fact is linearly proportional to it for large p . We remark, that sharp bounds for the self-improvement of Muckenhoupt-weights have been studied in [10], as well as the references mentioned therein.

Another reproof has been concurrently developed by other authors in [13]. Their methods yield more general insights into self-improvement phenomena, while this write up is restricted to classical Poincaré inequalities. Also, a careful examination of their paper seems to lead to similar bounds for the self-improvement.

We would also like to mention the recent unpublished work of Lukáš Malý on types of Lorentz-Poincaré inequalities without self-improvement, and general conditions for self-improvement for various types of Poincaré inequalities.

1.2. Proof techniques and characterizations of Poincaré inequalities.

We were motivated to re-investigate the beautiful and insightful proof of the Keith–Zhong result [12] for a few reasons. Firstly, the original proof is somewhat non-intuitive. It proceeds by an abstract argument estimating distributions of certain maximal functions, where the relationships between different estimates is only revealed at the very end. This makes the argument somewhat indirect. As a consequence, extracting bounds from their proof seems very complicated. This was done in [9], but the bounds seem to deteriorate for large exponents p . The bounds we obtain below are much sharper.

On the other hand, we have worked on more general applications of “self-improvement”-type methods, where much of the machinery of the original proof of Keith and Zhong become unnecessary [5]. Our goal is to understand whether the framework of [5] could be used to provide an easier proof of the Keith–Zhong result. This framework is based on tools such as iteration and the idea of “refilling” curves. However, to achieve this goal we need new techniques, because the paper in [5] does not give sharp characterizations of Poincaré inequalities. More precisely, while those results are sharp in general, for several classes of spaces better results can be obtained, and thus we needed to develop an understanding of different characterizations.

These characterizations come in the flavor of Muckenhoupt-type conditions. Thus, an additional motivation of this paper is to study the formal similarity between Poincaré inequalities and Muckenhoupt-weights. This similarity was alluded to in our prior paper [5], but we wish to make this formal analogy more precise. In the process, we obtain a new characterization of Poincaré-inequalities that clarifies the dependence of the exponent. This relationship to Muckenhoupt-weights has been previously observed in [3] as a way of characterizing measures on $\mathbf{R}$ which admit Poincaré inequalities. Thus, our results can be thought of as weaker and higher dimensional analogues of such characterizations.

Theorem 1.5. *For a proper metric measure space (X, d, μ) which is D -doubling the following conditions are equivalent.*

- PI_p : X satisfies a $(1, p)$ -Poincaré inequality.
- $PtPI_p$: X satisfies a pointwise Poincaré inequality: There are constants (C, C_{PPI}) such that for every continuous f and any upper gradient g for f and all $x, y \in X$ with $d(x, y) = r$ the following estimate holds:

$$|f(x) - f(y)| \leq C_{PPI}d(x, y) (\mathcal{M}_{p, Cr}g(x) + \mathcal{M}_{p, Cr}g(y)).$$

$A_p C$: X is A_p -connected: There are constants (C, C_A) such that for every non-negative lower semi-continuous g , and any $x, y \in X$ with $d(x, y) = r$, there is a Lipschitz curve γ connecting x to y with $\text{Len}(\gamma) \leq C d(x, y)$ and

$$\int_{\gamma} g \, ds \leq C_A d(x, y) (\mathcal{M}_{p, Cr} g(x) + \mathcal{M}_{p, Cr} g(y)).$$

The constants denoted C with or without subscripts in the various statements can be different and depend quantitatively on each other.

Recall, that for a locally integrable and measurable function $f \in L^p_{\text{loc}}$ we define

$$\mathcal{M}_{p,s} f(x) := \sup_{r \in (0,s)} \left(\int_{B(x,r)} f^p \, d\mu \right)^{\frac{1}{p}}.$$

This result is closely related to a lemma by Heinonen and Koskela [8, Lemma 5.1]. The novelty is on the new notion of A_p -connectivity that arises. It reduces the problem of proving a Poincaré inequality to finding a single curve, with controlled length and integral. However, the difficulty is to do this for an arbitrary pair of points and every function. In a sense, this notion of connectivity is nothing other than a reformulated modulus estimate involving Riesz kernels from Keith [11]. Here, the modulus condition is reformulated as a problem of finding curves with small integrals. This point, while present in some work, seems to not have been fully utilized, and doesn't appear explicitly in prior literature. Formally, the task of constructing a single curve is much easier than constructing “thick” curve families, which traditionally is involved in proving modulus estimates.

The task of constructing a curve can be done iteratively, which is the core idea in [5], and is reformulated here. This idea involves both the notion of “level” and “scale”. The iteration is started by constructing an initial curve, where some bad behavior occurs only on some small set. By replacing the portions in this bad set, we obtain a better curve than initially expected. The replacing is done at a smaller scale. The badness corresponds to the size of $\mathcal{M}_{p,s} g$, where s is the scale we are working at. This size of $\mathcal{M}_{p,s} g$ is also referred to as a level, and will be a definite amount larger than the initial level. If a “good” level can be chosen, such that the size of the next scale is small enough compared to it, then we can obtain an absorbable lower order term.

This idea of absorbing a term from a higher level and smaller scale is included indirectly in [12], and forms the core of many good- λ -type inequalities. The estimate, which involves the level in a scale invariant way, leads to the definition of an α -function. This function describes the connectivity of the space and naturally encodes the iteration procedure. A similar function appear in [15], and our terminology is motivated by theirs. In [5] the iteration is done differently. There, the desired curve is directly constructed via an infinite recursion and limiting process, where at each step some “gaps” or undefined portions of the curve are refilled. Here, we can avoid both the use of “gaps” and the use of an infinite recursion. In a sense, the new function measures connectivity at various levels in a scale invariant way.

Finally, we remark, that our methods are local, and thus we obtain the following transparent local version of self-improvement.

Theorem 1.6. Assume $p > 1$. Let (X, d, μ) be a proper metric measure space, which is D -doubling up to scale r_D with a $(1, p)$ -Poincaré inequality with constants (C, C_{PI}) up to scale r_{PI} . There exists a $\epsilon(D, p, C_{PI}) > 0$ such that for any $\epsilon \in$

$(0, \epsilon(D, p, C_{PI}))$ the space admits a $(1, p - \epsilon)$ -Poincaré inequality with constants $C' = C''(D, C_{PI}, \epsilon, C)$, $C'_{PI} = C''_{PI}(D, p, C_{PI}, \epsilon, C)$ up to scale

$$r_0 \leq \min \left\{ \frac{r_{PI}}{4}, \frac{r_D}{20C} \right\}.$$

We end this introduction with an intuitive, and informal, reason for Theorem 1.4 to hold true. This intuition is abstractly present in the proofs of this paper. Given a function g , the A_p -connectivity from Theorem 1.5 implies the existence of curves γ connecting x and y with $d(x, y) = r$ a bound of the form

$$\int_{\gamma} g \, ds \leq C_A d(x, y) (\mathcal{M}_{p, Cr} g(x) + \mathcal{M}_{p, Cr} g(y)).$$

However, this bound is not optimal for all g . In the case where g is supported on a very small set, or is highly concentrated, then a much better curve can be obtained. Namely, for any $\delta \in (0, 1)$, there is an $\epsilon \in (0, 1)$ such that if $E = \text{supp}(g)$ and $\mathcal{M}_{p, Cr} 1_E(x) + \mathcal{M}_{p, Cr} 1_E(y) < \epsilon$, then in fact we could get a bound roughly of the form

$$\int_{\gamma} g \, ds \leq \delta d(x, y) (\mathcal{M}_{p, Cr} g(x) + \mathcal{M}_{p, Cr} g(y)).$$

Here, the constant C_A can be replaced with the much smaller δ . Thus, highly concentrated “obstacle” functions g are in fact easier to avoid. Quantifying this leads to the self-improvement phenomenon, since if g is not highly concentrated, then the L^q and L^p -norms become comparable. On the other hand, if g is highly concentrated, then the previous sketch of an argument shows that the curve integrals are much smaller than expected.

Acknowledgments. I thank my adviser Professor Bruce Kleiner for discussing similar topics, especially in relation to the previous paper [5]. I also thank Professor Juha Kinnunen and Antti Vähäkangas for discussing their related work in [13] and giving feedback on the presentation of the current paper, and for presenting many interesting problems related to this work. I also thank Nageswari Shanmugalingam for encouraging us to rethink the proofs from an earlier version, which improved the presentation. Finally, the paper has benefited from a careful reading by the referee and his many corrections and comments. This research has been supported by NSF graduate fellowship DGE-1342536 and NSF grant DMS-1704215.

2. Preliminary lemmas

Throughout this paper we will assume that (X, d, μ) is a proper metric measure space equipped with a Radon measure μ . By a curve $\gamma: I \rightarrow X$ we mean a continuous function whose domain $I \subset \mathbf{R}$ is compact. The length of an interval I is denoted $|I|$. The length of a curve is defined as

$$(2.1) \quad \text{Len}(\gamma) := \sup_{x_1 \leq \dots \leq x_n \in I} \sum_{i=1}^{n-1} d(\gamma(x_{i+1}), \gamma(x_i)).$$

A curve γ is called *rectifiable* if $\text{Len}(\gamma) < \infty$. Most of the time we will focus on *Lipschitz curves*, i.e. those for which there exists a $L \in (0, \infty)$ such that $d(\gamma(a), \gamma(b)) \leq L|b - a|$ for any $a, b \in I$. The smallest L for which this inequality is satisfied is also called the Lipschitz constant of γ and is denoted $\text{LIP}(\gamma)$. If γ is assumed to be Lipschitz, we have $\text{Len}(\gamma) \leq \text{LIP}(\gamma)|I|$. In fact, any curve can be reparametrized

by length as $\gamma^*: [0, \text{Len}(\gamma)] \rightarrow X$. This makes the curve 1-Lipschitz, and such that $\text{Len}(\gamma|_{[a,b]}) = |b - a|$ whenever $0 < a < b < \text{Len}(\gamma)$ [1].

For rectifiable curves one can define a curve integral according to [1], and which is defined for any bounded/signed Borel function. In fact, if γ is a rectifiable curve, and $\gamma^*: [0, \text{Len}(\gamma)] \rightarrow X$ is its length-reparametrization, the integral can be defined as

$$\int_{\gamma} g \, ds := \int_0^{\text{Len}(\gamma)} g(\gamma^*(t)) \, dt,$$

when the right-hand side makes sense.

A metric space (X, d) is called $(L-)$ quasiconvex if for every $x, y \in X$, there exists a rectifiable curve γ connecting x to y with $\text{Len}(\gamma) \leq Ld(x, y)$. A space that is 1-quasiconvex is called geodesic. We recall, that a curve $\gamma: I \rightarrow X$ is said to connect a pair of points x, y if $\gamma(\min(I)) = x, \gamma(\max(I)) = y$.

If f is a continuous function on X , we call a non-negative Borel function g an *upper gradient* for f if for every $x, y \in X$, and any rectifiable curve γ connecting x to y we have

$$|f(x) - f(y)| \leq \int_{\gamma} g \, ds.$$

This terminology is due to Heinonen and Koskela [8].

We define the localized Hardy–Littlewood maximal functions with exponent $p \in [1, \infty)$ as

$$\mathcal{M}_{p,s}f(x) = \sup_{r \in (0,s]} \left(\int_{B(x,r)} f^p \, d\mu \right)^{\frac{1}{p}},$$

which makes sense for any non-negative measurable f . The non-localized version is simply

$$\mathcal{M}_p f(x) = \sup_{0 < r} \left(\int_{B(x,r)} f^p \, d\mu \right)^{\frac{1}{p}}.$$

If $p = 1$ we will drop the subscript. Finally, if $A \subset X$, then we denote by 1_A the characteristic function, or indicator function, of the set A .

We have the following weak L^1 -distributional inequality. Its proof is contained in [16].

Theorem 2.2. (Maximal function estimate) *Let (X, d, μ) be a D -measure doubling metric measure space and $s > 0$ and $B(x, r) \subset X$ arbitrary, then for any $p \in [1, \infty)$ and any non-negative $f \in L^1$ and $\lambda > 0$ we have*

$$\mu(\{\mathcal{M}_{p,s}f > \lambda\} \cap B(x, r)) \leq D^3 \frac{\|f 1_{B(x,r+s)}\|_{L^p}^p}{\lambda^p}.$$

We also need a different type of Maximal function estimate, whose proof is similar to the previous theorem, but with an additional observation. It is a “multi-scale” version of the previous inequality.

Lemma 2.3. (Max-max estimate) *Let (X, d, μ) be D -measure doubling and $r, s > 0$, $x \in X$ arbitrary. If f is any non-negative measurable function and $E_{\lambda,p,s} = \{z | \mathcal{M}_{p,s}f(z) > \lambda\}$, then we have*

$$(2.4) \quad \mathcal{M}_r 1_{E_{\lambda,p,s}}(x) \leq \frac{D^4 (\mathcal{M}_{p,s+r}f(x))^p}{\lambda^p}.$$

Proof. Fix x and $t \in (0, r)$ be arbitrary. Without loss of generality, assume $\infty > \lambda^p > D^4 (\mathcal{M}_{p,s+r} f(x))^p$. Were this to fail, the estimate would become trivial (as the left hand side is bounded by 1). We will estimate for the ball $B(x, t) \subset X$

$$\int_{B(x,t)} 1_{E_{\lambda,p,s}} d\mu = \frac{\mu(E_{\lambda,p,s} \cap B(x, t))}{\mu(B(x, t))},$$

by the right hand side. The proof then follows by taking the supremum over $t \in (0, r]$.

Consider $A = B(x, t) \cap E_{\lambda,p,s}$. For every $z \in A$ there exists a ball $B(z, r_z)$ such that $r_z \in (0, s]$ and

$$\int_{B(z,r_z)} f^p d\mu > \lambda^p.$$

There are two cases. Either, for every $z \in A$ we have $r_z < t$, or there exists some such that $r_z \geq t$. If the latter case holds, then

$$\int_{B(x,r_z+t)} f^p \geq \frac{1}{D^2} \int_{B(z,r_z)} f^p d\mu \geq \frac{\lambda^p}{D^2}.$$

But, from $r_z + t \leq s + r$ we would get $(\mathcal{M}_{p,s+r} f)^p \geq \frac{\lambda^p}{D^2}$, which gives a contradiction. So, for every z we have $r_z < t$.

Using the 5-covering Lemma (see [16]), we obtain a collection of balls $\mathcal{B} = \{B(z_i, r_i)\}$ such that $B(z_i, r_i)$ are disjoint, so that $B(z_i, 5r_i)$ cover the set A , $z_i \in B(x, t)$, $r_i \in (0, \min(s, t)]$ and

$$\int_{B(z_i,r_i)} f^p d\mu > \lambda^p.$$

Then, we get

$$\begin{aligned} \frac{\mu(E_{\lambda,p,s} \cap B(x, t))}{\mu(B(x, t))} &\leq \frac{\sum_{B \in \mathcal{B}} \mu(5B)}{\mu(B(x, t))} \leq D^3 \frac{\sum_{B \in \mathcal{B}} \mu(B)}{\mu(B(x, t))} \\ &\leq \frac{D^4}{\lambda^p} \frac{\sum_{B \in \mathcal{B}} \int_B f^p d\mu}{\mu(B(x, t + \min(s, t)))} \leq \frac{D^4}{\lambda^p} \frac{\int_{B(x, t + \min(s, t))} f^p d\mu}{\mu(B(x, t + \min(s, t)))} \\ &\leq \frac{D^4}{\lambda^p} \int_{B(x, t + \min(s, t))} f^p d\mu \leq \frac{D^4 (\mathcal{M}_{p,s+r} f(x))^p}{\lambda^p}. \quad \square \end{aligned}$$

3. Proof of self-improvement

We will use the following definition of A_p -connectivity.

Definition 3.1. Let $C, C_A > 0, p \geq 1$. We say that a metric measure space (X, d, μ) is A_p -connected (with constants (C, C_A)) if for every $x, y \in X$ with $d(x, y) = r > 0$, and every lower semi-continuous and non-negative $g: X \rightarrow [0, \infty)$, there exists a $L > 0$ and a Lipschitz curve $\gamma: [0, L] \rightarrow X$ such that

- (1) $\gamma(0) = x$,
- (2) $\gamma(L) = y$,
- (3) $\text{Len}(\gamma) \leq Cr$ and
- (4)

$$(3.2) \quad \int_{\gamma} g \leq C_A r (\mathcal{M}_{p,Cr} g(x) + \mathcal{M}_{p,Cr} g(y)).$$

We choose the term A_p -connected to draw an analogy to the definition of A_p -weights. Recall, that the class of A_p -weights is defined by $\mu \in A_p(\lambda)$, where λ is Lebesgue measure on $\mathbf{R}^n$ and $d\mu = \omega d\lambda$, if one of the following equivalent conditions holds.

- (1) Maximal function bound: There is a constant $C > 0$ such that for every $f \in L^p(\mu)$ we have

$$(3.3) \quad \left(\int (\mathcal{M}f)^p d\mu \right)^{\frac{1}{p}} \leq C \left(\int f^p d\mu \right)^{\frac{1}{p}}.$$

- (2) Integral bound: $\mu = \omega\lambda$, where ω, ω^{1-p} are locally integrable and there is a $C > 0$ such that for every ball $B = B(x, r)$

$$(3.4) \quad \left(\int_B \omega d\lambda \right) \left(\int_B \omega^{1-p} d\lambda \right)^{\frac{1}{p-1}} \leq C.$$

- (3) Average bound: For some $C > 0$ and for any f locally integrable and any ball $B = B(x, r)$

$$(3.5) \quad \int_B f d\lambda \leq C \left(\frac{1}{\mu(B)} \int_B f^p d\mu \right)^{\frac{1}{p}}.$$

Further, all of these imply that a version of quantitative absolute continuity holds. By this, we mean that there is a constant $C > 0$ such that for all $B(x, r)$ and all $E \subset B(x, r)$ we have

$$(3.6) \quad \frac{\lambda(E)}{\lambda(B(x, r))} \leq C \left(\frac{\mu(E)}{\mu(B(x, r))} \right)^{\frac{1}{p}}.$$

It is subtle, that this quantitative absolute continuity is *not* equivalent to being an A_p -weight. In fact, by work in [14, 15] the condition (3.6) characterizes so called $A_{p,1}$ -weights. While the A_p -conditions characterize boundedness of the Hardy–Littlewood maximal function $\mathcal{M}$ from L^p to L^p , the $A_{p,1}$ -condition characterizes boundedness from $L^p \rightarrow L^{p,\infty}$. It is known, that $A_p \subset A_{p,1}$ strictly. Further, the $A_{p,1}$ -condition does not improve to $A_{q,1}$ for any $q < p$.

Our definition of A_p -connected is analogous to the average bound (3.5). Namely, replace λ by $\mathcal{H}^1|_\gamma$ and the right-hand side by a maximal function bound. The measure $\mathcal{H}_\gamma^1$ is the 1-dimensional Hausdorff measure on the image of γ . The formal difference is that the condition of A_p -connectivity additionally presumed the *existence* of some curve γ such that the estimate holds. In a sense, the A_p -connectivity corresponds to being an “ A_p -weight” with respect to one-dimensional Hausdorff measure on some curve.

The condition (3.6) is somewhat similar to the notion of fine connectivity in [5]. It would correspond to restricting functions g in the definition of A_p -connectivity, with characteristic functions $g = 1_E$. However, we do not need to use that definition here.

At the heart of our proof of self-improvement is the characterization of Poincaré inequalities in terms of A_p -connectivity. We first need some elementary lemmas.

Lemma 3.7. *Let E be a Borel set, $p \in [1, \infty)$ and $s > 0$ a fixed scale parameter. Then, for every $\epsilon \in (0, 1)$ there exists an open set O such that $E \setminus \{x\} \subset O$, $x \notin O$ and such that*

$$\mathcal{M}_s 1_O(x) \leq \mathcal{M}_s 1_E(x) + \epsilon,$$

and

$$\mathcal{M}_{p,s}1_{O \setminus E}(x) \leq \epsilon.$$

Proof. Fix $\epsilon > 0$. By regularity of measure, for each $n \in \mathbf{Z}$ we can find open sets $O_{\epsilon,n}$ such that $E \cap (B(x, 2^{1-n}s) \setminus B(x, 2^{-n-1}s)) \subset O_{\epsilon,n}$, $O_{\epsilon,n} \subset B(x, 2^{2-n}s) \setminus B(x, 2^{-n-2}s)$ and $\mu(O_{\epsilon,n} \setminus E \cap B(x, 2^{2-n}s)) \leq \epsilon^p 4^{-n-1} \mu(B(x, 2^{-n-2}s))$. Define $O = \bigcup_{n=0}^{\infty} O_{\epsilon,n}$. It is clear that $x \notin O$. Now, clearly

$$E \cap B(x, s) \setminus \{x\} = \bigcup_{n \in \mathbf{Z}} E \cap (B(x, 2^{1-n}s) \setminus B(x, 2^{-n-1}s)) \subset \bigcup_{n \in \mathbf{Z}} O_{\epsilon,n} = O.$$

Also, for any $t \in (0, s]$ we have

$$\int_{B(x,t)} 1_{O \setminus E} d\mu \leq \frac{1}{\mu(B(x,t))} \sum_{n \in \mathbf{Z}, 2^{-n-2}s < t} \mu(O_{\epsilon,n} \setminus E \cap B(x, 2^{2-n}s)) \leq \epsilon^p,$$

which gives the second estimate in the statement of the Lemma. Similarly, the first statement follows from the following estimate.

$$\begin{aligned} \int_{B(x,t)} 1_O d\mu &\leq \int_{B(x,t)} 1_E + 1_{O \setminus E} d\mu \\ &= \mathcal{M}_s 1_E(x) + \frac{1}{\mu(B(x,t))} \sum_{n \in \mathbf{Z}, 2^{-n-2}s < t} \mu(O_{\epsilon,n} \setminus E \cap B(x, 2^{2-n}s)) \\ &\leq \mathcal{M}_s 1_E(x) + \epsilon^p \leq \mathcal{M}_s 1_E(x) + \epsilon. \end{aligned} \quad \square$$

Lemma 3.8. *Let g be a non-negative Borel function such that $\mathcal{M}_{p,s}g(x) < \infty$. Then, for every $\epsilon \in (0, 1)$ there exists an lower semi-continuous function $\bar{g}(y)$ such that $\bar{g}(y) \geq g(y)$ for all $y \neq x$ such that*

$$\mathcal{M}_{p,s}\bar{g}(x) \leq \mathcal{M}_{p,s}g(x) + \epsilon.$$

Proof. Let $E_{k,\epsilon} = \{y \in X \setminus \{x\} | g(y) > \frac{k\epsilon}{2}\}$ for $k \geq 0$. Clearly, for $y \neq x$ we have

$$g(y) \geq \sum_{k=0}^{\infty} \frac{\epsilon}{2} 1_{E_{k,\epsilon}} - \frac{\epsilon}{2}.$$

By Lemma 3.7 we have sets $O_{k,\epsilon}$ such that $E_{k,\epsilon} \subset O_{k,\epsilon}$ and

$$\mathcal{M}_{p,s}1_{O_{k,\epsilon} \setminus E_{k,\epsilon}}(x) \leq \epsilon 2^{-k-2}.$$

Finally, define

$$g_{\epsilon} = \sum_{k=0}^{\infty} \frac{\epsilon}{2} 1_{O_{k,\epsilon}}.$$

Now, it is easy to obtain that $g_{\epsilon} \geq g$, except possibly at x . Finally, we also have for all $y \neq x$

$$0 < g_{\epsilon}(y) - g(y) \leq \frac{\epsilon}{2} + \sum_{k=0}^{\infty} \frac{\epsilon}{2} 1_{O_{k,\epsilon} \setminus E_{k,\epsilon}}(y).$$

Thus, from the triangle inequality, we can derive

$$\begin{aligned} \mathcal{M}_{p,s}g_{\epsilon} &\leq \mathcal{M}_{p,s}g + \mathcal{M}_{p,s}(g_{\epsilon} - g) \\ &\leq \mathcal{M}_{p,s}g + \frac{\epsilon}{2} + \sum_{k=0}^{\infty} \mathcal{M}_{p,s}1_{O_{k,\epsilon} \setminus E_{k,\epsilon}} \leq \mathcal{M}_{p,s}g + \epsilon. \end{aligned} \quad \square$$

Remark 3.9. If we have a finite set of points $z_1, \dots, z_n$, then we can choose the lower-semi-continuous approximant $\bar{g}$ so that $\bar{g}(y) \geq g(y)$ for all $y \neq z_1, \dots, z_n$ and $\mathcal{M}_{p,s}\bar{g}(z_i) \leq \mathcal{M}_{p,s}g(z_i) + \epsilon$ for all $i = 1, \dots, n$. Namely, apply the lemma to give functions $\bar{g}_{z_i}$ that satisfy the conclusion for $x = z_i$, and define $\bar{g} = \min_{i=1, \dots, n} \bar{g}_{z_i}$. Similarly, for Lemma 3.7, we can ensure $\mathcal{M}_s 1_O(x) \leq \mathcal{M}_s 1_E(x) + \epsilon$ and $z_i \notin O$ simultaneously for a finite set of points $x = z_1, \dots, z_n$ by considering the intersection of open sets O_{z_i} satisfying the conclusion for individual $x = z_i$.

Proof of Theorem 1.5. That $PI_p \Leftrightarrow PtPI_p$ follows from a classical result, which is presented for example in [8, Lemma 5.15] combined with [11, Theorem 2]¹.

Next, we show that $A_p C \Rightarrow PtPI_p$. Let g be a measurable upper gradient of a continuous function f . Then, using Lemma 3.8 and Remark 3.9 we can find a $\bar{g}_\epsilon$ which is lower semi-continuous, $\bar{g}_\epsilon \geq g$ (except possibly at x, y) and

$$\lim_{\epsilon \rightarrow 0} \mathcal{M}_{p,Cr}\bar{g}_\epsilon(x) = \mathcal{M}_{p,Cr}g(x), \lim_{\epsilon \rightarrow 0} \mathcal{M}_{p,Cr}\bar{g}_\epsilon(y) = \mathcal{M}_{p,Cr}g(y).$$

Then, for any rectifiable curve γ parametrized by length, connecting a pair of points $x, y \in X$, we have

$$|f(x) - f(y)| \leq \int_\gamma g \, ds \leq \int_\gamma \bar{g}_\epsilon \, ds.$$

So, infimizing over curves γ gives

$$|f(x) - f(y)| \leq C_{Ad}(x, y) \left(\mathcal{M}_{p,Cr}\bar{g}_\epsilon(x) + \mathcal{M}_{p,Cr}\bar{g}_\epsilon(y) \right),$$

and then letting ϵ tend to zero gives the desired conclusion.

It remains to show that $PtPI_p$ and PI_p imply $A_p C$. Assume that (X, d, μ) satisfies a $(1, p)$ -Poincaré inequality and $PtPI_p$, and let g be an arbitrary non-negative lower semi-continuous function such that g^p is locally integrable and fix $x, y \in X$. To fix constants, assume the Poincaré inequality in the form

$$\int_B |f - f_B| \, d\mu \leq C_{PIr} \left(\int_{CB} (\text{Lip } f)^p \, d\mu \right)^{\frac{1}{p}}$$

and the second condition as

$$(3.10) \quad |f(x) - f(y)| \leq C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g_f(x) + \mathcal{M}_{p,Cr}g_f(y)),$$

if g_f is an upper gradient for f .

We will construct γ such that

$$\text{Len}(\gamma) \leq 5C_{PPId}(x, y)$$

and

$$\int_\gamma g \, ds \leq 4C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)).$$

Next, define for every $N > 0$ and $\epsilon > 0$ a function $g_{N,\epsilon} = \min(g + \epsilon, N)$. Then

$$(3.11) \quad \mathcal{M}_{p,Cr}g_{N,\epsilon}(x) + \mathcal{M}_{p,Cr}g_{N,\epsilon}(y) \leq \mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y) + 2\epsilon.$$

Now, define for $z \in X$ the set $\Gamma_{x,z}$ as the set of all rectifiable curves starting at x and ending at z . Further, define a function by

$$(3.12) \quad \mathcal{F}_{N,\epsilon}(z) = \inf_{\gamma \in \Gamma_{x,z}} \int_\gamma g_{N,\epsilon} \, ds.$$

¹Keith's result is also needed, since Heinonen and Koskela [8] use a slightly different definition of a Poincaré inequality.

This function is bounded and continuous, since PI-spaces are L -quasiconvex for some $L = L(C_{PI}, D)$ (see e.g. [2, Theorem 4.32], or [4, Appendix]). It is also easy to see that $g_{N,\epsilon}$ is an upper gradient for $\mathcal{F}_{N,\epsilon}$. Next, by the $PtPI_p$ -condition we have

$$(3.13) \quad |\mathcal{F}_{N,\epsilon}(y) - \mathcal{F}_{N,\epsilon}(x)| \leq C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)) + 2C_{PPId}(x, y)\epsilon.$$

Thus, there is a curve $\gamma_{N,\epsilon}$ such that $\gamma_{N,\epsilon}$ connects x to y and

$$\int_{\gamma_{N,\epsilon}} g_{N,\epsilon} ds \leq C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)) + 3C_{PPId}(x, y)\epsilon.$$

Assume now $\epsilon > \mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)$ is arbitrary. Then since $g_{N,\epsilon} \geq \epsilon$, we get

$$(3.14) \quad \begin{aligned} \epsilon \text{Len}(\gamma_{N,\epsilon}) &\leq \int_{\gamma_{N,\epsilon}} g_{N,\epsilon} ds \\ &\leq 2C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)) + 3C_{PPId}(x, y)\epsilon \\ &\leq 5C_{PPId}(x, y)\epsilon. \end{aligned}$$

Thus $\text{Len}(\gamma_{N,\epsilon}) \leq 5C_{PPId}(x, y)$. Assume that $\gamma_{N,\epsilon}$ are parametrized by length. Then, they are 1-Lipschitz and the properness of X allows us to apply Arzela–Ascoli, and to extract a subsequential limit curve γ_ϵ connecting x to y . Up to reindexing, we can assume that the curve is the limit of the original sequence. Then for every N , using lower semi-continuity of curve integrals and the lower semi-continuity of $g_{N,\epsilon}$ (see [11, Proposition 4], we get

$$\begin{aligned} \int_{\gamma_\epsilon} g_{N,\epsilon} ds &\leq \liminf_{M \rightarrow \infty} \int_{\gamma_{M,\epsilon}} g_{N,\epsilon} ds \leq \liminf_{M \rightarrow \infty} \int_{\gamma_{M,\epsilon}} g_{M,\epsilon} \\ &\leq C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)) + 3C_{PPId}(x, y)\epsilon. \end{aligned}$$

Now, letting $N \rightarrow \infty$ and using monotone convergence, we get

$$(3.15) \quad \int_{\gamma_\epsilon} g ds \leq C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)) + 3C_{PPId}(x, y)\epsilon.$$

Let $\epsilon \rightarrow (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y))$. Then, by using Arzela–Ascoli again we obtain a sub-sequential limit γ of γ_ϵ . Finally, using lower semi-continuity of g and the lower semi-continuity of curve integrals we get $\text{Len}(\gamma) \leq 5C_{PPId}(x, y)$ and the desired estimate

$$\int_{\gamma} g ds \leq 4C_{PPId}(x, y) (\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)).$$

We remark, that this final limiting process is only necessary if $\mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y) = 0$. Otherwise, we could just set $\epsilon = \mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y)$. $\square$

Next, we present our proof of Keith–Zhong self-improvement. Some notation and ideas are similar to [15], where the authors show general self-improvement phenomena for maximal-function estimates. There, a crucial role is played by a sub-multiplicative function α . For us, the relevant quantity is the following.

Let $x, y \in X$ be given and denote $r = d(x, y)$. Define with

$$(3.16) \quad \mathcal{E}_{x,y,\tau,C}^p := \{g: X \rightarrow [0, 1] \mid g \text{ lower semi-continuous } \mathcal{M}_{p,Cr}g(x) + \mathcal{M}_{p,Cr}g(y) < \tau\}$$

the set of admissible obstacle functions. Denote by $\Gamma_{x,y}^C$ the set of rectifiable curves γ parametrized by length on the interval $[0, \text{Len}(\gamma)]$ such that $\gamma(0) = x, \gamma(\text{Len}(\gamma)) = y$

and $\text{Len}(\gamma) \leq Cd(x, y)$. Then define

$$(3.17) \quad \alpha^p(C, \tau) := \sup_{x, y \in X} \sup_{g \in \mathcal{E}_{x, y, \tau, C}^p} \inf_{\gamma \in \Gamma_{x, y}^C} \frac{1}{d(x, y)} \int_{\gamma} g \, ds.$$

In a sense, $\alpha^p(C, \tau)$ measures how well a function g with “small” size can block curves, for the worst scale $d(x, y)$ and worst pair of points $x, y \in X$. The additional constraint on g to have values in $[0, 1]$ is used to ensure that α^p is bounded. Namely, if X is L -quasiconvex with $L \leq C$, then for every $p \in [1, \infty)$

$$\alpha^p(C, \tau) \leq L,$$

for all $\tau \in [0, 1]$. In this case, one can estimate the infimum from above by an arbitrary curve γ connecting x, y with length $\text{Len}(\gamma) \leq Ld(x, y)$, and obtain

$$\frac{1}{d(x, y)} \int_{\gamma} g \, ds \leq \frac{Ld(x, y)}{d(x, y)} \leq L.$$

While initially non-intuitive, this expression is a way of condensing the A_p -connectivity property.

Lemma 3.18. *Let $p \in [1, \infty)$. The space X is A_p -connected with constants (C, C_A) if and only if*

$$\alpha^p(C, \tau) \leq C_A \tau,$$

for all $\tau \in [0, 1]$.

Proof. If X is A_p -connected, then for any $g \in \mathcal{E}_{x, y, \tau, C}^p$, we have

$$\inf_{\gamma \in \Gamma_{x, y}^C} \int_{\gamma} g \, ds \leq C_A \tau d(x, y),$$

since $\mathcal{M}_{p, Cr}g(x) + \mathcal{M}_{p, Cr}g(y) < \tau$ by assumption. Now, dividing both sides by $d(x, y)$ and taking a supremum over $g \in \mathcal{E}_{x, y, \tau, C}^p$ and x, y gives the desired inequality for α^p .

The converse direction is somewhat more involved, as the g in the A_p -connectivity condition need not be bounded. This can be resolved with a limiting argument which uses the completeness of X . Now, to verify A_p -connectivity, we need to fix arbitrary $x, y \in X$ and a lower semi-continuous non-negative g and find a curve $\gamma \in \Gamma_{x, y}^C$ with

$$\int_{\gamma} g \, ds \leq C_A r (\mathcal{M}_{p, Cr}g(x) + \mathcal{M}_{p, Cr}g(y)).$$

First, let $g_N = \min\{g, N\}$, and $\overline{g_N} = g_N/(2N)$. Both of these functions are lower semi-continuous and $\overline{g_N}(z) \in [0, 1]$ for every $z \in X$. Let

$$\tau_N = (\mathcal{M}_{p, Cr}g_N(x) + \mathcal{M}_{p, Cr}g_N(y)) / (2N).$$

Since g_N converges to g and is a monotone sequence, it is not hard to see that $\lim_{N \rightarrow \infty} 2N\tau_N = \mathcal{M}_{p, Cr}g(x) + \mathcal{M}_{p, Cr}g(y)$.

Now, by linearity it is easy to see that

$$\mathcal{M}_{p, Cr}\overline{g_N}(x) + \mathcal{M}_{p, Cr}\overline{g_N}(y) = \tau_N \in [0, 1].$$

Thus, $\overline{g_N} \in \mathcal{E}_{x, y, \tau_N, C}^p$. Then, from the definition of $\alpha^p(C, \tau_N)$, and the estimate for α^p , we find for every $\epsilon > 0$ a curve $\gamma_{\epsilon, N} \in \Gamma_{x, y}^C$ such that

$$\int_{\gamma_{\epsilon, N}} \overline{g_N} \, ds \leq (\alpha^p(C, \tau_N) + \epsilon/N)d(x, y) \leq (C_A \tau_N + \epsilon/N)d(x, y).$$

Multiplying both sides by N , we obtain

$$\int_{\gamma_{\epsilon,N}} \min\{g, N\} ds \leq (2C_A \tau_N N + \epsilon) d(x, y).$$

Since $\gamma_{\epsilon,N} \in \Gamma_{x,y}^C$, and due to Arzela–Ascoli the set $\Gamma_{x,y}^C$ is a compact family of curves (with respect to uniform convergence), we can find a subsequential limit γ_ϵ of $\gamma_{\epsilon,N}$ as $N \rightarrow \infty$. To simplify notation, reindex so that this is the original sequence.

Then, using monotone convergence and the lower semi-continuity of curve integrals.

$$\begin{aligned} \int_{\gamma_\epsilon} g ds &\leq \lim_{M \rightarrow \infty} \int_{\gamma_\epsilon} g_M ds \leq \lim_{M \rightarrow \infty} \liminf_{N \rightarrow \infty} \int_{\gamma_{\epsilon,N}} \min\{g, M\} ds \\ &\leq \lim_{M \rightarrow \infty} \liminf_{N \rightarrow \infty} \int_{\gamma_{\epsilon,N}} \min\{g, N\} ds \\ &\leq \lim_{M \rightarrow \infty} \liminf_{N \rightarrow \infty} (2C_A \tau_N N + \epsilon) d(x, y) \\ &\leq (C_A (\mathcal{M}_{p,Cr} g(x) + \mathcal{M}_{p,Cr} g(y)) + \epsilon) d(x, y). \end{aligned}$$

Finally, letting ϵ tend to zero and using the lower semi-continuity of curve integrals again, we obtain a limit curve γ of some subsequence of γ_ϵ such that

$$\int_{\gamma} g ds \leq C_A d(x, y) (\mathcal{M}_{p,Cr} g(x) + \mathcal{M}_{p,Cr} g(y)),$$

as required. $\square$

There is a simple sub-linear estimate for α^p .

Lemma 3.19. *Let $K \geq 1$ and $p \in [1, \infty)$. Then*

$$\alpha^p(C, K\tau) \leq K\alpha^p(C, \tau).$$

Proof. Let $g \in \mathcal{E}_{x,y,K\tau,C}^p$, then $\bar{g}_K = g/K \in \mathcal{E}_{x,y,\tau,C}^p$. Note, that if $K \geq 1$, then $\bar{g}$ still has values in $[0, 1]$. In particular, for any x, y and γ , we have

$$\frac{1}{d(x, y)} \int_{\gamma} g ds = K \frac{1}{d(x, y)} \int_{\gamma} \bar{g}_K ds,$$

and so

$$\begin{aligned} \alpha^p(C, K\tau) &= \sup_{x,y} \sup_{g \in \mathcal{E}_{x,y,K\tau,C}^p} \inf_{\gamma \in \Gamma_{x,y}^C} \frac{1}{d(x, y)} \int_{\gamma} g ds \\ &= K \sup_{x,y} \sup_{g \in \mathcal{E}_{x,y,K\tau,C}^p} \inf_{\gamma \in \Gamma_{x,y}^C} \frac{1}{d(x, y)} \int_{\gamma} \bar{g}_K ds \\ &\leq K \sup_{x,y} \sup_{h \in \mathcal{E}_{x,y,\tau,C}^p} \inf_{\gamma \in \Gamma_{x,y}^C} \frac{1}{d(x, y)} \int_{\gamma} h ds = K\alpha^p(C, \tau). \end{aligned} \quad \square$$

Since in the following proof we are using A_p connectivity to prove A_q -connectivity, we will explicate their connectivity constants with an additional subscript. That is A_p connectivity will be assumed to hold with constants $(C, C_{A,p})$, and we will prove A_q -connectivity with different constants $(L, C_{A,q})$ and with $L \geq C$.

Theorem 3.20. *Assume $p > 1$. If (X, d, μ) is D -doubling and A_p -connected (with constants $C, C_{A,p}$), then there is a $\epsilon(D, p, C_{A,p})$ such that X is also A_q -connected for all $p - \epsilon(D, p, C_{A,p}) < q < p$ with constants depending on $C, C_{A,p}, p$ and q .*

Proof. Recall the definitions of α^q in (3.17). As discussed above, it is sufficient to show that there is an $\epsilon(D, p, C_{A,p})$, such that if $p - \epsilon(D, p, C_{A,p}) < q < p$, then the space is A_q -connected. Fix this $\epsilon(D, p, C_{A,p})$ to be determined, and any such exponent q . The A_q connectivity can be reduced by Lemma 3.18 to showing that there are some constants L and $C_{A,q}$ such that for all $\tau \in [0, 1]$

$$\alpha^q(L, \tau) \leq C_{A,q}\tau.$$

This estimate is shown by proving that for $\delta \in (0, 1)$ there are some $k \in \mathbf{N}$, $M \geq 2$, $S \geq 1$ and for all $L \geq \frac{C}{1-\delta}$ and $\tau \in [0, 1]$ we have

$$(3.21) \quad \alpha^q(L, \tau) \leq S\tau + \delta \max_{i=1, \dots, k} M^{-i} \alpha^q(L, M^i \tau).$$

If we have this estimate, then Lemma 3.19 gives

$$(3.22) \quad \alpha^q(L, \tau) \leq S\tau + \delta \alpha^q(L, \tau),$$

and so

$$\alpha^q(L, \tau) \leq \frac{S}{1-\delta} \tau,$$

which is the estimate we desire with $C_{A,q} = \frac{S}{1-\delta}$. Note, we used the fact that α^q is bounded since X is C -quasiconvex and $L \geq C$. This follows from A_p -connectivity since $\Gamma_{x,y}^C$ is not empty for any $x, y \in X$, as otherwise $\alpha^p(C, \tau)$ would not be bounded. Thus, it remains to prove (3.21).

Before we prove this, we wish to give some intuition. A way to think of this estimate (3.21) is that, we would really like to prove for some $\Lambda, L > 1$, that $\alpha^q(L, \tau) \leq S\tau + \delta \Lambda^{-1} \alpha^q(L, \Lambda\tau)$. Here $\alpha^q(L, \tau)$ corresponds to an estimate “at level” τ , and we can estimate it from above by a small constant δ times a term at level $\Lambda\tau$ and a term of the desired form. This corresponds to an iteration at the level of curves, as we will soon see, since an initial curve γ is constructed to almost avoid points where $\mathcal{M}_{q,Ld(x,y)}g(z) > \Lambda\tau$. It can not fully avoid this set, but replacing the portions in this set, if it is sufficiently small, gives a contribution of the form $\delta \Lambda^{-1} \alpha^q(L, \Lambda\tau)$.

However, we do not know how to choose Λ *a priori* in a way independent of $g \in \mathcal{E}_{x,y,\tau,L}^q$. The proof instead shows that we can always find one “level” $\Lambda\tau = M^{i_0}\tau$ for some $i_0 = 1, \dots, k$ where a desired quantity is sufficiently small (compared to δ). This leads to the less intuitive estimate (3.21) involving the maximum. However, this suffices for our purposes. We note, that the curves used for α^q have length $Ld(x, y)$ due to the fact that the iteration necessarily increases the lengths slightly. However, $L = \frac{C}{1-\delta}$ will suffice for our purposes. It is chosen so that $C + \delta L = L$.

Next, fix any $M \geq 2$ and any $\delta \in (0, 1)$. We will also fix $k \geq 1$ to be determined later and $L = \frac{C}{1-\delta}$. In order to estimate $\alpha^q(L, \tau)$ we are taking a supremum over pairs of points and functions. Thus, let $x, y \in X$, $\tau \in [0, 1]$ and let

$$(3.23) \quad g \in \mathcal{E}_{x,y,\tau,L}^q$$

be arbitrary. Denote $d(x, y) = r$. By adding a small constant to g we can assume $(\mathcal{M}_{q,Lr}g(x))^q + (\mathcal{M}_{q,Lr}g(y))^q > 0$. First define $F_i = \{z | \mathcal{M}_{q,\delta Lr}g > M^i\tau/2\}$, for which Lemma 2.3 gives for $z = x, y$

$$\mathcal{M}_{Cr}1_{F_i}(z) \leq \frac{2^q D^4 (\mathcal{M}_{q,\delta Lr+Cr}g(z))^q}{M^{iq}\tau^q} \leq \frac{2^p D^4 (\mathcal{M}_{q,Lr}g(z))^q}{M^{iq}\tau^q}.$$

From Estimates (3.23), (3.16) and the definition of F_i , it follows that $x, y \notin F_i$. We need to enlarge these sets slightly to be open. Using Lemma 3.7 and Remark 3.9 we

can find open sets E'_i such that $F_i \subset E'_i$, $x, y \notin E'_i$, such that for $z = x, y$

$$\mathcal{M}_{Cr} 1_{E'_i \setminus F_i}(z) \leq \frac{(\mathcal{M}_{q,Lr} g(x))^q + (\mathcal{M}_{q,Lr} g(y))^q}{k M^{iq} \tau^q}$$

for $i = 1, \dots, k$. Now, define $E_i := \bigcup_{j \geq i} E'_j$. We have

$$\mathcal{M}_{Cr} 1_{E_i \setminus F_i}(z) \leq \sum_{j=1}^i \mathcal{M}_{Cr} 1_{E'_j \setminus F_j}(z) \leq \frac{(\mathcal{M}_{q,Lr} g(x))^q + (\mathcal{M}_{q,Lr} g(y))^q}{M^{iq} \tau^q},$$

and thus

$$\begin{aligned} \mathcal{M}_{Cr} 1_{E_i}(z) &\leq \mathcal{M}_{Cr} 1_{F_i}(z) + \frac{(\mathcal{M}_{q,Lr} g(x))^q + (\mathcal{M}_{q,Lr} g(y))^q}{M^{iq} \tau^q} \\ (3.24) \quad &\leq 2^{p+1} D^4 \frac{(\mathcal{M}_{q,Lr} g(x))^q + (\mathcal{M}_{q,Lr} g(y))^q}{M^{iq} \tau^q}. \end{aligned}$$

Next, define the function

$$h := \frac{1}{k} \sum_{i=1}^k M^i 1_{E_i}.$$

Since E_i are open the function is lower semi-continuous.

Recall that $E_i \subset E_j$ for $i > j$. Then for $x \in E_l \setminus E_{l+1}$, where $l = 1, \dots, k-1$, or $x \in E_l$ if $l = k$, we have

$$h^p(x) = \frac{1}{k^p} \left(\sum_{i=1}^k M^i 1_{E_i}(x) \right)^p \leq \frac{1}{k^p} \left(\sum_{i=1}^l M^i \right)^p \leq \frac{2^p M^{lp}}{k^p} 1_{E_l}(x).$$

Thus, it is easy to see,

$$(3.25) \quad h^p \leq \frac{2^p}{k^p} \sum_{i=1}^k M^{ip} 1_{E_i}.$$

Now, take an arbitrary $0 < s < Cr$ and compute with $z = x, y$.

$$\begin{aligned} \int_{B(z,s)} h^p d\mu &\stackrel{(3.25)}{<} \frac{2^p}{k^p} \sum_{i=1}^k M^{ip} \frac{\mu(E_i \cap B(z,s))}{\mu(B(z,s))} \\ &\stackrel{(3.24)}{\leq} \frac{2^{p+1}}{k^p} \sum_{i=1}^k 2^{p+1} D^4 M^{ip} \frac{(\mathcal{M}_{q,Lr} g(x))^q + (\mathcal{M}_{q,Lr} g(y))^q}{M^{iq} \tau^q} \\ &\leq \frac{2^{2p+2} D^4 M^{k(p-q)}}{k^{p-1} \tau^q} \frac{(\mathcal{M}_{q,Lr} g(x))^q + (\mathcal{M}_{q,Lr} g(y))^q}{\tau^q} \\ &\stackrel{(3.23), (3.16)}{\leq} \frac{2^{2p+3} D^4 M^{k(p-q)}}{k^{p-1}}. \end{aligned}$$

So, we get $\mathcal{M}_{p,Cr} h(x) + \mathcal{M}_{p,Cr} h(y) < \frac{4(2^{2p+3} D^4 M^{k(p-q)})^{\frac{1}{p}}}{k^{\frac{p-1}{p}}}$. Define

$$\Delta := \frac{4C_{A,p}(2^{2p+3} D^4 M^{k(p-q)})^{\frac{1}{p}}}{k^{\frac{p-1}{p}}}.$$

By A_p -connectivity, there is a curve γ such that

$$(3.26) \quad \text{Len}(\gamma) \leq Cr,$$

and

$$\int_{\gamma} h \, ds < \Delta r.$$

Since an the minimum of a set of numbers is a lower bound for its mean, there must be some index i_0 such that

$$(3.27) \quad \int_{\gamma} 1_{E_{i_0}} M^{i_0} \, ds \leq \frac{1}{k} \sum_{i=1}^k \int_{\gamma} 1_{E_i} M^i \, ds = \int_{\gamma} h \, ds < \Delta r.$$

Now, we can fix our choices of k and $\epsilon(D, p, C_{A,p})$. Choose k so large that $\frac{4C_{A,p}(2^{2p+3}D^4)^{\frac{1}{p}}}{k^{\frac{p-1}{p}}} < \frac{\delta}{2}$. and $\epsilon(D, p, C_{A,p})$ so small that

$$M^k \frac{\epsilon(D, p, C_{A,p})}{p} \leq 2.$$

Then, since $p - q < \epsilon(D, p, C_{A,p})$ by assumption,

$$\Delta = \frac{4C_{A,p}(2^{2p+3}D^4 M^{k(p-q)})^{\frac{1}{p}}}{k^{\frac{p-1}{p}}} < M^k \frac{\epsilon(C, C_{A,p,p})}{p} \frac{\delta}{2} \leq \delta.$$

Finally, we obtain from this and estimate (3.27) that

$$\int_{\gamma} 1_{E_{i_0}} \, ds < \delta M^{-i_0} r.$$

Parametrize γ by unit speed on the interval $[0, \text{Len}(\gamma)]$ to be a 1-Lipschitz curve. Since E_{i_0} is open, so is $U = \gamma^{-1}(E_{i_0})$. Clearly $|U| < \delta M^{-i_0} r$. Note $0, \text{Len}(\gamma) \notin U$ since $\gamma(0) = x \notin E_{i_0}$ and $\gamma(\text{Len}(\gamma)) = y \notin E_{i_0}$. Define $K = [0, \text{Len}(\gamma)] \setminus U$. Then, clearly $0, \text{Len}(\gamma) \in K$. Also, from (3.26)

$$(3.28) \quad |K| \leq \text{Len}(\gamma) \leq Cr.$$

By our definition of K we have $|[0, \text{Len}(K)] \setminus K| < \delta M^{-i_0} r$. We will now redefine γ on the small set $[0, \text{Len}(K)] \setminus K$.

We can express the open set $[0, \text{Len}(\gamma)] \setminus K$ as a countable union of its components, i.e. $[0, \text{Len}(\gamma)] \setminus K = \bigcup_{j \in J} (a_j, b_j)$ for some countable (possibly finite) index set J . These intervals (a_j, b_j) are also referred to as the gaps of γ . Define $d_j := d(\gamma(b_j), \gamma(a_j))$. By construction and since γ is 1-Lipschitz, we have

$$(3.29) \quad \sum_j d_j \leq \sum_{j \in J} |b_j - a_j| \leq |[0, \text{Len}(\gamma)] \setminus K| < \delta M^{-i_0} r \leq \delta r.$$

In particular, since γ is parametrized by length, $d_j \leq |b_j - a_j|$ and $d_j \leq \delta r$.

For each gap (a_j, b_j) , since $\gamma(K) \cap E_{i_0} = \emptyset$ and $a_j, b_j \in K$, we have $\gamma(a_j), \gamma(b_j) \notin E_{i_0}$, and moreover $\gamma(a_j), \gamma(b_j) \notin F_{i_0}$ (even in the possible cases $\gamma(a_j), \gamma(b_j) = x, y$). Also, $Ld_j \leq \delta Lr$, so by the definition of F_{i_0} we obtain

$$\mathcal{M}_{q, Ld_j} g(\gamma(a_j)) + \mathcal{M}_{q, Ld_j} g(\gamma(b_j)) \leq \mathcal{M}_{q, \delta Lr} g(\gamma(a_j)) + \mathcal{M}_{q, \delta Lr} g(\gamma(b_j)) \leq M^{i_0} \tau.$$

Thus, by the definition of $\alpha^q(L, M^{i_0} \tau)$, there exists curves γ_j connecting $\gamma(a_j)$ to $\gamma(b_j)$ of length at most Ld_j with

$$(3.30) \quad \int_{\gamma_j} g \, ds \leq d_j \alpha^q(L, M^{i_0} \tau),$$

and $\text{Len}(\gamma_j) \leq Ld_j \leq L|b_j - a_j|$. We can reparametrize these curves as L -Lipschitz maps $\gamma_j: [a_j, b_j] \rightarrow X$.

Now, define $\gamma': [0, \text{Len}(\gamma)] \rightarrow X$ by $\gamma'(t) = \gamma(t)$, when $t \in K$, and $\gamma'(t) = \gamma_j(t)$, when $t \in (a_j, b_j)$. Clearly γ' is L -Lipschitz. Since $\gamma'|_K = \gamma$, we obtain that

$$\text{Len}(\gamma') \leq \text{Len}(\gamma) + \sum_{j \in J} \text{Len}(\gamma_j) \leq Cr + \sum_{j \in J} Ld_j \leq (C + \delta L)d(x, y) = Ld(x, y).$$

So we get $\gamma' \in \Gamma_{x,y}^L$. Note, for $t \in K$ we have $\gamma(t) \notin F_{i_0}$ and thus since g is lower semi-continuous $g(\gamma(t)) \leq \mathcal{M}_{q,\delta Lr} g(\gamma(t)) \leq \tau M^{i_0} \leq M^k \tau$.

Further,

$$\begin{aligned} \inf_{\sigma \in \Gamma_{x,y}^L} \frac{1}{d(x, y)} \int_{\sigma} g \, ds &\leq \frac{1}{d(x, y)} \int_{\gamma'} g \, ds \\ &= \frac{1}{d(x, y)} \int_K g(\gamma(t)) \, dt + \frac{1}{d(x, y)} \sum_{j \in J} \int_{\gamma_j} g \, ds \\ &\stackrel{(3.28)}{=} CM^k \tau + \frac{1}{d(x, y)} \sum_j \int_{\gamma_j} g \, ds \\ &\stackrel{(3.30)}{\leq} CM^k \tau + \frac{1}{d(x, y)} \sum_j d_j \alpha^q(L, M^{i_0} \tau) \\ &\stackrel{(3.29)}{\leq} CM^k \tau + \delta M^{-i_0} \alpha^q(L, M^{i_0} \tau) \\ &\leq CM^k \tau + \delta \max_{i=1, \dots, k} M^{-i} \alpha^q(L, M^i \tau). \end{aligned}$$

The right hand side now no longer involves x, y or g . Taking suprema over all functions $g \in \mathcal{E}_{x,y,\tau,L}^q$ and all pairs $x, y \in X$ gives

$$\alpha^q(L, \tau) \leq CM^k \tau + \delta \max_{i=1, \dots, k} M^{-i} \alpha^q(L, M^i \tau).$$

This gives the desired estimate with $S = CM^k \tau$.

Recall, we required the estimates

$$\frac{4C_{A,p}(2^{2p+3}D^4)^{\frac{1}{p}}}{k^{\frac{p-1}{p}}} < \frac{\delta}{2},$$

and

$$M^k \frac{\epsilon(D,p,C_{A,p})}{p} \leq 2.$$

These can be obtained by setting

$$k := \frac{(2^{6p+3}C_{A,p}^p D^4)^{\frac{1}{p-1}}}{\delta^{\frac{p}{p-1}}},$$

and

$$\epsilon(D, p, C_{A,p}) := \frac{p}{\log_2(M)k}.$$

A more detailed analysis will follow after the proof. $\square$

We can present the proof of Keith–Zhong self-improvement using this result.

Proof of Theorem 1.4. From Theorem 1.5 we obtain that X is A_p -connected with constants (C, C_A) . From the previous theorem we see that X is also A_q -connected for all $p - \epsilon(D, p, C_A) < q < p$. Again, applying Theorem 1.5 we see that A_q -connectivity implies the $(1, q)$ -Poincaré inequality. This completes the proof. Since C_A depends quantitatively on D and the constant C_{PI} , we can express $\epsilon(D, p, C_A)$ in terms of D and C_{PI} . See the discussion following for some more detail. $\square$

Finally, with the above choice of k , we see what the bound for q is. The bounds become a little easier of $M = 2$ and $\delta = \frac{1}{2}$. Then, we obtain a bound for $\epsilon(D, p, C_{A,p})$ of the form

$$(3.31) \quad \epsilon \leq \frac{p}{(2^{7p+3} C_{A,p}^p D^4)^{\frac{1}{p-1}}}.$$

As p further increases, the asymptotic behavior of this expression is $\frac{p}{2^7 C_{A,p}}$. This seemingly loses dependence on the doubling constant D . However, Theorem 1.5 gives that the A_p -connectivity constant $C_{A,p}$ is related to both C_{PI} and D . More precisely, by using similar techniques to [8] and the arguments in Theorem 1.5, we can show that $C_{A,p} \leq 2^6 D^3 C_{PI}$ suffices, which gives the following bound for Theorem 1.4 (when $2^p \geq D^3$)

$$\epsilon \leq \frac{p}{(2^{13p+3} C_{PI}^p D^{3p+4})^{\frac{1}{p-1}}}.$$

3.1. Remarks on localizing the estimates. The same proof as above, with slight additional care, can be applied to the localized version

$$(3.32) \quad \alpha_{r_0}^p(C, \tau) := \sup_{x,y \in X, d(x,y) \leq r_0} \sup_{g \in \mathcal{E}_{x,y,\tau,C}^p} \inf_{\gamma \in \Gamma_{x,y}^C} \frac{1}{d(x,y)} \int_{\gamma} g \, ds.$$

The proof also needs a localized version of Theorem 1.5 and [11, Theorem 2]. Here, one loses at most a factor of 2 in applying the proof in [8]. In order to be slightly more precise we will trace the proof backwards. In proving the $(1, q)$ -Poincaré inequality at scale r_0 , we reduce it to the A_q -connectivity at scale $2r_0$, i.e. for points $d(x, y) \leq 2r_0$, using a local version of Theorem 1.5. To obtain A_q -connectivity from A_p -connectivity at scale $2r_0$, we need to repeat the proof of Theorem 1.4 which applies Lemma 2.3. This requires D -measure doubling up to scale $20Cr_0$ (with the choice of $\delta = \frac{1}{2}$). Finally, to obtain A_p -connectivity at scale $2r_0$ we need D -doubling up to scales $20Cr_0$, and a $(1, p)$ -Poincaré inequality at scale $4r_0$. All the required estimates hold, if we assume

$$r_0 \leq \min \left\{ \frac{r_{PI}}{4}, \frac{r_D}{20C} \right\}.$$

Recall, r_{PI} is the scale at which the $(1, p)$ -Poincaré inequality holds, and r_D is the scale for the doubling property. This gives Theorem 1.6.

References

- [1] AMBROSIO, L., N. GIGLI, and G. SAVARÉ: Gradient flows: in metric spaces and in the space of probability measures. - Lectures Math. ETH Zürich, Springer, 2008.
- [2] BJÖRN, A., and J. BJÖRN: Nonlinear potential theory on metric spaces. - EMS Tracts Math. 17, Eur. Math. Soc., 2011.
- [3] BJÖRN, J., S. BUCKLEY, and S. KEITH: Admissible measures in one dimension. - Proc. Amer. Math. Soc. 134:3, 2006, 703–705.
- [4] CHEEGER, J.: Differentiability of Lipschitz functions on metric measure spaces. - Geom. Funct. Anal. 9:3, 1999, 428–517.
- [5] ERIKSSON-BIQUE, S.: Characterizing spaces satisfying Poincaré inequalities and applications to differentiability. - arXiv:1607.07428, 2016.
- [6] HAJLASZ, P., and P. KOSKELA: Sobolev meets Poincaré. - C. R. Acad. Sci. Paris Sér. I 320, 1995, 1211–1215.
- [7] HEINONEN, J.: Lectures on analysis on metric spaces. - Springer, 2000.

- [8] HEINONEN, J., and P. KOSKELA: Quasiconformal maps in metric spaces with controlled geometry. - *Acta Math.* 181:1, 1998, 1–61.
- [9] HEINONEN, J., P. KOSKELA, and N. SHANMUGALINGAM, and J. T. TYSON: Sobolev spaces on metric measure spaces. - *New Math. Monogr.* 27, Cambridge Univ. Press, 2015.
- [10] HYTÖNEN, T., C. PÉREZ, and E. RELA: Sharp reverse Hölder property for A_∞ weights on spaces of homogeneous type. - *J. Funct. Anal.* 263:12, 2012, 3883–3899.
- [11] KEITH, S.: Modulus and the Poincaré inequality on metric measure spaces. - *Math. Z.* 245:2, 2003, 255–292.
- [12] KEITH, S., and ZHONG, X.: The Poincaré inequality is an open ended condition. - *Ann. of Math.* (2) 167:2, 2008, 575–599.
- [13] KINNUNEN, J., and J. LEHRBÄCK, A. V. VÄHÄKANGAS, and X. ZHONG: Maximal function estimates and self-improvement results for Poincaré inequalities. - *arXiv:1705.05072*, 2017.
- [14] KURTZ, D. S., R. A. HUNT, and H. M. CHUNG: The Hardy–Littlewood maximal function on $L(p, q)$ spaces with weights. - *Indiana Univ. Math. J.* 31:1, 1982, 109–120.
- [15] LERNER, A. K., and C. PÉREZ: A new characterization of the Muckenhoupt A_p weights through an extension of the Lorentz–Shimogaki theorem. - *Indiana Univ. Math. J.* 56:6, 2007, 2697–2722.
- [16] STEIN, E. M.: Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals. - Princeton Univ. Press, 2016.

Received 5 August 2017 • Revision received 10 May 2018 • Accepted 18 September 2018