Cent. Eur. J. Math. • 11(5) • 2013 • 865-875
DOI: 10.2478/s11533-013-0213-0

VERSITA

Central European Journal of **Mathematics**

Best constants for metric space inversion inequalities

Research Article

Stephen M. Buckley[1]*, Safia Hamza[1]†

1 *Department of Mathematics and Statistics, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland*

Received 29 November 2011; accepted 20 June 2012

Abstract: For every metric space (X,d) and origin $o\in X$, we show the inequality $I_o(x,y)\le 2d_o(x,y)$, where $I_o(x,y)=d(x,y)/d(x,o)\,d(y,o)$ is the metric space inversion semimetric, d_o is a metric subordinate to I_o, and $x,y\in X\setminus\{o\}$. The constant 2 is best possible.

MSC: 30F45, 54E25

Keywords: Metric space inversion • Best constant

1. Introduction

Inversion (or reflection) about the unit sphere is a bijection on $\mathbb{R}^n\setminus\{0\}$, so we can pull back Euclidean distance to get a new distance on $\mathbb{R}^n\setminus\{0\}$: $I_0(x,y)=|x-y|/|x||y|$. Inversion has been generalized in [3] to the setting of a metric space (X,d) containing at least two points: for fixed $o\in X$, define

$$I_o(x,y)=\frac{d(x,y)}{d(x,o)\,d(y,o)}\,,\qquad x,y\in X_o,$$

where $X_o=X\setminus\{o\}$. Then I_o is a semimetric on X_o, but not in general a metric. However we can define a related function $d_o\colon X_o\times X_o\to[0,\infty)$ subordinate to I_o, and show that d_o is a metric that is bilipschitz equivalent to I_o. Specifically, it is shown in [3, Lemma 3.2] that

$$\frac14\,I_o(x,y)\le d_o(x,y)\le I_o(x,y)=\frac{d(x,y)}{d(x,o)\,d(y,o)}\,. \tag{1}$$

* *E-mail: stephen.buckley@maths.nuim.ie*
† *E-mail: safia.hamza@maths.nuim.ie*

Inversion has been used as a tool to characterize uniform domains in terms of Gromov hyperbolicity and the quasiconformal structure of the Gromov boundary in [7, Theorem 9.1], thus extending a result for bounded Euclidean uniform domains [1, Theorem 1.11].

Many estimates in [3] depend on the above inequality, but few of the constants are sharp. A natural first question related to sharpness is therefore to investigate the sharpness of the first inequality in (1) for general metric spaces. Note that the second inequality is sharp since $d_o = l_o$ whenever X is a CAT(0) space, as explained in [5] or [2]; see also [6] and [4] for more on Ptolemaic spaces (i.e. spaces in which $d_o = l_o$ for all o).

In this paper, we investigate the first inequality in (1) and prove the following sharp replacement.

Theorem 1.1.

For every metric space (X, d) of cardinality at least 2, every $o \in X$, and every $x, y \in X_o$, we have $l_o(x, y) \leq C d_o(x, y)$ for $C = 2$. However this inequality fails for certain choices of data o, x, y, whenever $C < 2$ and X is a non-Euclidean L^∞ space.

Note that a non-Euclidean L^∞ space is precisely an L^∞ space of dimension at least 2.

After some preliminaries in Section 2, we investigate an easier variant of the problem for ℓ^p and counting measure L^∞ in Section 3, and then we prove the main theorem in Section 4.

2. Notation and preliminaries

Generalities

Throughout the remainder of the paper, (X, d) is a metric space of cardinality at least 2 (sometimes satisfying additional restrictions), and d_o, l_o, X_o are as in the introduction. We denote the norm in any L^p space by $\|\cdot\|_p$. As usual, $1/p$ is taken to mean zero when $p = \infty$, and ℓ^p_n is L^p for the n-point counting space. Everything said about L^p or ℓ^p_n is true for both real and complex versions of these spaces.

Consider the following well-known conditions defining a metric $d\colon X \times X \to \mathbb{R}$, where X is a non-empty set:

(a) d is non-negative, and d is zero along the diagonal (i.e. $d(x, x) = 0$);

(b) d is nonzero off the diagonal (i.e. $d(x, y) > 0$ if $x \neq y$);

(c) d is symmetric;

(d) d satisfies the triangle inequality.

A function $d\colon X \times X \to \mathbb{R}$ is a *pseudometric on X* if it satisfies (a), (c), and (d) above, while it is a *semimetric on X* if it satisfies (a), (b), and (c).

The metric d_0 and related constants

We first define a discrete path P and associated “lengths” $d(P)$ and $l_o(P)$.

Definition 2.1.

A *discrete path $P = (z_i)_{i=0}^n$ from x to y in X_o* is a finite sequence $z_0, \dots, z_n$ in X_o satisfying $z_0 = x$ and $z_n = y$. For any such discrete path, we define the “lengths” $d(P) = \sum_{i=1}^n d(z_{i-1}, z_i)$ and $l_o(P) = \sum_{i=1}^n l_o(z_{i-1}, z_i)$.

Let us recall the definition of d_o from [3].

Definition 2.2.

For $x, y \in X_o$, $d_o(x, y)$ is the infimum of $l_o(P)$ over all discrete paths $P = (z_i)_{i=0}^n$ from x to y in X_o.

The above definition involves the standard construction of a pseudometric from a semimetric: in fact it is clearly the largest pseudometric subordinate to the semimetric. The first inequality in (1) ensures that d_0 is a metric. If we wish to emphasize what space we are working in, we write $d_{o,X}(x,y)$ instead of $d_o(x,y)$.

It is trivial that $d_{o,Y}(x,y) \le d_{o,X}(x,y)$ whenever $x, y \in X \subset Y$ but we do not always get equality. For instance if we define the points $o = (0,0)$, $x = (1,0)$, $y = (0,1)$, and $z = (1,1)$ in ℓ_2^1, and take $X = \{o, x, y\}$ and $Y = X \cup \{z\}$, then it is trivial that $d_{o,X}(x,y) = 2$, whereas $d_{o,Y}(x,y) = d_{o,Y}(x,z) + d_{o,Y}(z,y) = 1/2 + 1/2 = 1$. This example motivates the following definition of a variant $\widehat{d}_o$ of d_o.

Definition 2.3.

For $x, y \in X_o$, $\widehat{d}_o(x,y)$ is the infimum of the distances $d_{o,Y}(x,y)$ over all metric spaces (Y, d_Y) containing (an isometric copy of) X.

In the previous example, $d_{o,X}(x,y) = I_o(x,y) = 2$ but $\widehat{d}_o(x,y) \le 1$. Using (1), it follows easily that $\widehat{d}_o$ is also a metric on X_o satisfying $\widehat{d}_o \ge I_o/4$.

By contrast with $d_o(x,y)$, the semimetric quantity $I_o(x,y)$ depends only on $d(x,y)$, $d(x,o)$, and $d(y,o)$, so it is unchanged if we replace X by a space in which X is isometrically embedded. It is clear that d_o and I_o coincide infinitesimally, and that they have length element $ds(z)/d^2(z,o)$ at $z \in X_o$, where ds denotes the d-length element. Thus the d_o-length of a path γ in X_o is

$$\mathrm{len}_o(\gamma) = \int_\gamma \frac{ds(z)}{d^2(z,o)}\,.$$

We now define the main constant $C_{\mathrm{inv}}(X)$ that interests us in this paper, and related constants $c_{\mathrm{inv}}(X)$ and $\widehat{C}_{\mathrm{inv}}(X)$.

Definition 2.4.

We denote by $C_{\mathrm{inv}}(X)$ the smallest constant $C \ge 1$ such that $I_o(x,y) \le C\,d_o(x,y)$ for all $o \in X$, $x, y \in X_o$. We denote by $c_{\mathrm{inv}}(X)$ the smallest constant $C \ge 1$ such that $I_o(x,y) \le C(I_o(x,z) + I_o(z,y))$ for all $o \in X$, $x, y, z \in X_o$. We denote by $\widehat{C}_{\mathrm{inv}}(X)$ the smallest constant $C \ge 1$ such that $I_o(x,y) \le C\,\widehat{d}_o(x,y)$ for all $o \in X$, $x, y \in X_o$.

Let us now list a few basic facts about these constants, with justification for those facts that are non-trivial.

Fact 2.5.

$1 \le c_{\mathrm{inv}}(X) \le C_{\mathrm{inv}}(X) \le \widehat{C}_{\mathrm{inv}}(X)$.

Fact 2.6.

It can occur that $C_{\mathrm{inv}}(X) < \widehat{C}_{\mathrm{inv}}(X)$ (e.g. the example before Definition 2.3) or that $c_{\mathrm{inv}}(X) < C_{\mathrm{inv}}(X)$ (e.g. Example 2.10 below).

Fact 2.7.

If X is isometrically embedded in Y, then $C(X) \le C(Y)$, where $C(\cdot)$ denotes $C_{\mathrm{inv}}(\cdot)$, $c_{\mathrm{inv}}(\cdot)$, or $\widehat{C}_{\mathrm{inv}}(\cdot)$.

[3, Lemma 3.2] says that $\widehat{C}_{\mathrm{inv}}(X) \le 4$. Our Theorem 1.1 says that $C_{\mathrm{inv}}(X) \le 2$ and, since this is true for all spaces, we also have $\widehat{C}_{\mathrm{inv}}(X) \le 2$. Theorem 1.1 also says that $C_{\mathrm{inv}}(X) = 2$ if X is a non-Euclidean L^∞ space. In fact, we will see that $c_{\mathrm{inv}}(X) = 2$ in every non-Euclidean L^∞ space.

Elementary estimates and examples

If for some $o \in X$, $x, y, z \in X_o$, we have $d(x,z) = t\,d(x,y)$, $d(z,y) = (1-t)\,d(x,y)$, and $d(z,o) = (1-t)\,d(x,o) + t\,d(y,o)$, then

$$I_o(x,z) + I_o(z,y) = d(x,y)\left(\frac{t}{d(x,o)\,d(z,o)} + \frac{1-t}{d(z,o)\,d(y,o)}\right) = d(x,y)\left(\frac{t\,d(y,o) + (1-t)\,d(x,o)}{d(x,o)\,d(z,o)\,d(y,o)}\right) = I_o(x,y).$$

Replacing the last equation above by an inequality, we immediately deduce the remaining parts of the following useful observation.

Observation 2.8.

Suppose $o \in X$, $x, y, z \in X_o$, with $d(x,z) = t\,d(x,y)$ and $d(z,y) = (1-t)\,d(x,y)$ for some $0 < t < 1$. Then $I_o(x,z) + I_o(z,y)$ is greater than, equal to, or less than $I_o(x,y)$ depending on whether $d(z,o)$ is less than, equal to, or greater than $(1-t)\,d(x,o) + t\,d(y,o)$, respectively.

Applying this observation to points on a path, we get the following result.

Observation 2.9.

Suppose $\gamma\colon [0,L] \to X_o$ is a d-geodesic segment from x to y parametrized by d-arclength, $x, y \in X_o$. Then $\mathrm{len}_o(\gamma) \le I_o(x,y)$ if $t \mapsto d(\gamma(t))$ is a concave function, while $\mathrm{len}_o(\gamma) \ge I_o(x,y)$ if $t \mapsto d(\gamma(t))$ is convex.

With the above observations in hand, it is not hard to give an example of a space X such that $c_{\mathrm{inv}}(X) < C_{\mathrm{inv}}(X)$.

Example 2.10.

Let X be the subset of ℓ_2^∞ given by $X = \{o, x, y, z_1, z_2\}$ where $o = (0,0)$, $x = (1,1)$, $y = (-1,1)$, $z_1 = (1/2, 3/2)$, $z_2 = (-1/2, 3/2)$. Then $I_o(x,y) = 2/(1)^2 = 2$, while

$$I_o(x,z_i) + I_o(z_i,y) = \frac{1/2}{(1)(3/2)} + \frac{3/2}{(3/2)(1)} = \frac{4}{3},$$

for $i = 1, 2$. Thus $c_{\mathrm{inv}}(X) \le 2/(4/3) = 3/2$. In fact $c_{\mathrm{inv}}(X) = 3/2$. To see this, we need to do a similar calculation for all other pairs of distinct points $u, v \in X_o$: in fact, by symmetry it suffices to consider the pairs $\{x, z_1\}$, $\{x, z_2\}$, and $\{z_1, z_2\}$. In most cases, adding an intermediate point $w \in X_o$ gives an I_o sum at least as large as $I_o(u,v)$, and so $d_o(u,v) = I_o(u,v)$. The only exception is that $I_o(x,z_2) = 1$, while $I_o(x,z_1) + I_o(z_1,z_2) = 1/3 + 4/9 = 7/9$, but this value is large enough to allow us to deduce that $c_{\mathrm{inv}}(X) = 3/2$. However $C_{\mathrm{inv}}(X) \ge 2/(10/9) = 9/5 > c_{\mathrm{inv}}(X)$, because

$$I_o(x,z_1) + I_o(z_1,z_2) + I_o(z_2,y) = 2\,\frac{1/2}{(1)(3/2)} + \frac{1}{(3/2)^2} = \frac{10}{9}.$$

In the above example, one intermediate point was insufficient to obtain $d_o(x,y)$. With a little extra effort, we now show that no finite collection of points may be sufficient to obtain $d_o(x,y)$.

Example 2.11.

Let X consist of the interval $X_o = [0,1] \subset \mathbb{R}$ together with a single extra point o. Let

$$d(s,t) = \begin{cases} |s-t|, & s, t \in [0,1], \\ 2 - t^2/2, & s = o, \quad t \in [0,1], \\ 0, & s = t = o. \end{cases}$$

A straightforward case analysis shows that d is a metric: the only case that is not completely trivial is the inequality $d(o,t) \le d(o,s) + d(s,t)$ for $s,t \in [0,1]$, which can be rewritten as

$$s^2 \le t^2 + 2|s-t|, \qquad 0 \le s,t \le 1,$$

and this is easily established.

The key features of X_o are that it consists entirely of a d-geodesic segment $[0,1]$, and that the distance function $d(o,\cdot)$ is strictly concave on $[0,1]$. Thus by Observation 2.8, adding an extra intermediate point to a discrete path (z_i) from $x=0$ to $y=1$ always decreases the corresponding I_o sum, and so $d_o(x,y)$ is strictly smaller than $I_o(P)$ for any discrete path P from x to y. It readily follows that $[0,1]$ is also a d_o-geodesic segment and

$$d_o(x,y) = \int_0^1 \frac{dt}{(2-t^2/2)^2} = \frac{1}{6} + \frac{1}{4}\tanh^{-1}\frac{1}{2} \approx 0.304.$$

By comparison, note that $I_o(x,y) = 1/(2(3/2)) = 1/3$.

3. The constant $c_{\mathrm{inv}}(\ell^p)$, $1 \le p \le \infty$

In this section, we prove the following rather easy but crucial lemma.

Lemma 3.1.
If $X = L^\infty(S)$, where S is a counting measure space with at least two points, then $c_{\mathrm{inv}}(X) = 2$.

The first step in proving Lemma 3.1 is to compute $c_{\mathrm{inv}}(\ell_2^\infty)$. Since essentially the same method yields a formula for $c_{\mathrm{inv}}(\ell_2^p)$, $1 \le p \le \infty$, we compute all of these constants.

Proposition 3.2.
For $1 \le p \le \infty$, $c_{\mathrm{inv}}(\ell_2^p)$ equals $c_p = 2^{|1-2/p|}$.

Proof. We first show that $I_o(x,y) \le c_p(I_o(x,z) + I_o(z,y))$ for all $x,y,z \ne o$. Multiplying this inequality across by $\|x-o\|_p\|y-o\|_p\|z-o\|_p$, the desired inequality is seen to be equivalent to

$$\|x-y\|_p\|z-o\|_p \le c_p\big(\|x-z\|_p\|y-o\|_p + \|y-z\|_p\|x-o\|_p\big).$$

Since inversion gives a metric on the deleted Euclidean plane, this inequality holds when $p=2$ (and $C_2 = 1$). Using this ℓ^2 inequality and the following well-known estimates (which can be deduced from the inequalities of Hölder and Minkowski):

$$\begin{aligned} \|\cdot\|_2 \le \|\cdot\|_p &\le 2^{1/p-1/2}\,\|\cdot\|_2\,, && 1 \le p \le 2,\\ 2^{1/p-1/2}\,\|\cdot\|_2 \le \|\cdot\|_p &\le \|\cdot\|_2\,, && 2 \le p \le \infty, \end{aligned}$$

it is straightforward to deduce the result. For instance, when $p \ge 2$,

$$\begin{aligned} \|x-y\|_p\|o-z\|_p \le \|x-y\|_2\|o-z\|_2 &\le \|x-z\|_2\|y-o\|_2 + \|y-z\|_2\|x-o\|_2 \\ &\le 2^{1-2/p}\big(\|x-z\|_p\|y-o\|_p + \|y-z\|_p\|x-o\|_p\big), \end{aligned}$$

where we used the above estimates repeatedly in both inequalities comparing ℓ^p and ℓ^2 quantities.

To finish the proof, we show that

$$\|x-y\|_p\|o-z\|_p = c_p\big(\|x-z\|_p\|y-o\|_p + \|y-z\|_p\|x-o\|_p\big)$$

for some choice of distinct points $x, y, z, o \in \ell_2^p$. For $1 \le p \le 2$, take $o = (0,0)$, $x = (1,0)$, $y = (0,1)$, and $z = (1,1)$ so that

$$\|x-y\|_p \cdot \|o-z\|_p = 2^{1/p} \cdot 2^{1/p} = 2^{2/p},$$

while

$$\|x-z\|_p \cdot \|y-o\|_p + \|y-z\|_p \cdot \|x-o\|_p = 1\cdot 1 + 1 \cdot 1 = 2,$$

giving the required equality. For the case $p \ge 2$, we instead use the points $o = (0,0)$, $x = (-1,1)$, $y = (1,1)$, and $z = (0,2)$. □

For general n, we get the following estimates.

Proposition 3.3.

For $1 \le p \le \infty$ and $n \ge 2$, we have $2^{|1-2/p|} \le c_{\rm inv}(\ell_n^p) \le n^{|1-2/p|}$ and $C_{\rm inv}(\ell_n^p) \le n^{|3/2-3/p|}$.

Proof. The proof of the upper bound in Proposition 3.2 is easily adjusted to give a proof of the upper bound for $c_{\rm inv}(\ell_n^p)$: the only difference is that constants of comparison between ℓ^p and ℓ^2 now involve the factor $n^{1/2-1/p}$ rather than $2^{1/2-1/p}$. The upper bound for $C_{\rm inv}(\ell_n^p)$ is similar, except that we need to compare ℓ^p norms with ℓ^2 norms for three vectors in each term rather than just two. Lastly, the lower bound for $c_{\rm inv}(\ell_n^p)$ (and so for $C_{\rm inv}(\ell_n^p)$) follows immediately from Proposition 3.2 since ℓ_2^p is isometrically embedded in ℓ_n^p. □

The estimate $c_{\rm inv}(\ell_n^\infty) \le n$ in Proposition 3.3 is not sharp when $n > 2$. The sharp result is of course Lemma 3.1, which we now prove.

Proof of Lemma 3.1. Since ℓ_2^∞ is isometrically embedded in X, we deduce from Proposition 3.2 that $c_{\rm inv}(X) \ge 2$. Conversely, we need to show that

$$\|x-y\|_\infty\|o-z\|_\infty \le 2\big(\|x-z\|_\infty\|y-o\|_\infty + \|y-z\|_\infty\|x-o\|_\infty\big)$$

for all $o \in X$ and $x, y, z \in X_o$. Let $\epsilon > 0$ be arbitrary, and let us choose $a, b \in S$ so that $(1+\epsilon)|x(a) - y(a)| \ge \|x-y\|$ and $(1+\epsilon)|o(b) - z(b)| \ge \|o-z\|$. Define the functions $x', y', z', o' \in X$ to have the same values as x, y, z, o, respectively, at the two points a and b, and to equal 0 elsewhere. Thus these new functions lie in an isometric copy of the L^∞ plane and so using Proposition 3.3 we get that

$$\begin{aligned}(1+\epsilon)^{-2}\|x-y\|_\infty\|o-z\|_\infty \le \|x'-y'\|_\infty\|o'-z'\|_\infty &\le 2\big(\|x'-z'\|_\infty\|y'-o'\|_\infty + \|y'-z'\|_\infty\|x'-o'\|_\infty\big)\\ &\le 2\big(\|x-z\|_\infty\|y-o\|_\infty + \|y-z\|_\infty\|x-o\|_\infty\big).\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, the result follows. □

Remark 3.4.

The assumption in Lemma 3.1 that the measure on S is counting measure could be eliminated at the expense of a slightly more technical proof. However the given version is sufficient for the proof of Theorem 1.1, which implies such an improved version of Lemma 3.1 as a special case.

Remark 3.5.

We do have an explicit formula for $c_{\rm inv}(\ell_n^p)$, $n > 2$, or for $C_{\rm inv}(\ell_n^p)$, $n > 1$. The only cases in which we can give explicit values are when $p \in \{1, \infty\}$, in which cases both constants equal 2, as follows rather easily from Proposition 3.2, the isometric embedding of ℓ_2^p in ℓ_n^p, and Theorem 1.1.

4. The constants $C_{\mathrm{inv}}(X)$ and $c_{\mathrm{inv}}(X)$ for general metric spaces

Before investigating $C_{\mathrm{inv}}(X)$ and $c_{\mathrm{inv}}(X)$ for general spaces, we first need some notation. We denote by $[x,y]_o$ any d-geodesic segment from x to y in X_o, meaning a path from x to y whose d-length equals $d(x,y)$. We call $[x,y]_o$ a *doubly geodesic segment* if its d_o-length equals $d_o(x,y)$. We write $G(x,y) = (d(x,y) - |d(x,o) - d(y,o)|)/2$ whenever $x, y \in X_o$. Thus $G(u,v) \geq 0$, and $G(u,v)$ is just the *Gromov product* of o and whichever of x, y is further from o, with the third point being the base for the Gromov product. Trivially $G(\cdot,\cdot)$ is non-negative and symmetric. We have the following useful lemma that applies when $G(x,y) = 0$.

Lemma 4.1.
Suppose that for some $x, y \in X_o$, there exists a d-geodesic segment $[x,y]_o$ and $G(x,y) = 0$. Then $d_o(x,y) = I_o(x,y)$, and $[x,y]_o$ is a doubly geodesic segment.

Proof. Without loss of generality, we assume that $d(x,o) \leq d(y,o)$. Note that $G(x,y) = 0$ is just another way of writing $d(y,o) = d(o,x) + d(x,y)$. It follows that there is an isometric embedding R from $[x,y]_o \cup \{o\}$ to the Euclidean half-line $[0,\infty)$ with $R(o) = 0$. Since the inversion semimetric I_o for $o = 0$ is a geodesic metric on $(0,\infty)$, we deduce that $\mathrm{len}_o([x,y]_o) = I_o(x,y)$.

Fixing an arbitrary pair $x, y \in X_o$, it remains to prove that $[x,y]_o$ is a d_o-geodesic. For this it suffices to show that if $z \in X_o \setminus \{x,y\}$, then $S_x = I_o(x,z) + I_o(z,y)$ is at least as large as $I_o(x,y)$. Writing S_z as an expression involving d-distances, we see that there are three quantities that vary as z varies: $d(x,z)$, $d(z,y)$, and $d(z,o)$. Furthermore, S_z is decreased whenever we decrease either of the first two of these quantities or increase the third, if the others are kept fixed. In view of the previous paragraph, we note that $S_w = I_o(x,y)$ whenever $w \in [x,y]_o$.

Consider first the case where $d(z,o) < d(x,o)$. Since $d(x,x) < d(z,x)$ and $d(x,y) = d(y,o) - d(x,o) < d(z,y)$, it is clear that $I_o(x,y) = S_x < S_z$. Consider next the case $d(x,o) \leq d(z,o) \leq d(y,o)$, and let $w \in [x,y]_o$ be such that $d(w,o) = d(z,o)$. By the triangle inequality we see that $d(x,w) = d(w,o) - d(x,o) \leq d(x,z)$, $d(w,y) = d(y,o) - d(w,o) \leq d(z,y)$, and so $I_o(x,y) = S_w \leq S_z$.

Finally, suppose $d(z,o) > d(y,o)$. We write $d_x = d(x,o)$, $d_y = d(y,o)$, $d_z = d(z,o)$, $\alpha = d(x,z)$, $\beta = d(z,y)$, $\delta = d(x,y) = d_y - d_x$, $\gamma = (\alpha + \beta - \delta)/2 > 0$. Then $d_z \leq \min(d_x + \alpha, d_y + \beta)$. We have already fixed x, y but if we allow z to vary while fixing γ, we see that the upper bound on d_z is minimized by choosing α, β so that $d_x + \alpha = d_y + \beta$, which in turn implies that $\beta = \gamma$, and so $\alpha = \gamma + \delta$. Thus

$$S_z \geq \frac{\gamma+\delta}{d_x(d_y+\gamma)} + \frac{\gamma}{d_y(d_y+\gamma)} = \frac{(\gamma+\delta)d_y + \gamma d_x}{d_x d_y (d_y+\gamma)} = \frac{\delta(d_y+\gamma) + 2\gamma d_x}{d_x d_y(d_y+\gamma)} > \frac{\delta}{d_x d_y} = I_o(x,y),$$

as required. □

We are now ready to state a theorem that implies the first statement of Theorem 1.1. Note that the second statement in Theorem 1.1 already follows from Lemma 3.1 since all non-Euclidean L^∞ spaces contain an isometric copy of ℓ_2^∞.

Theorem 4.2.
If (X,d) is a metric space of cardinality at least 2, then $\widehat{C}_{\mathrm{inv}}(X) \leq 2$.

Proof. We will prove that $C_{\mathrm{inv}}(X) \leq 2$ for all bounded spaces (X,d). This readily implies the same inequality for all metric spaces, and hence that $\widehat{C}_{\mathrm{inv}}(X) \leq 2$, as required.

Since (X,d) is a bounded metric space, we can define an isometric embedding I of X into $Y = L^\infty(X,\mu)$, where μ is counting measure, by letting $I(x) = i_x$, where $i_x(u) = d(x,u)$, $u \in X$. By Fact 2.7, $C_{\mathrm{inv}}(X) = C_{\mathrm{inv}}(I(X)) \leq C_{\mathrm{inv}}(Y)$, so it suffices to prove the result when $X = L^\infty(S)$ for some counting measure space S. We assume that X has this form from now on. By translation invariance of L^∞, we may assume that o is the origin 0. We need to prove that $I_0(x,y) \leq 2I_0(P)$ for every discrete path P from x to y in X_0. Certainly this holds if we prove the same inequality for every discrete path P

from x to y in A_0, where (A, d) is some augmented metric space that contains (an isometric copy of) (X, d), and we will pass to such a superspace without further comment during the proof.

We split the rest of the proof into parts for clarity. In Part I, we reduce to considering a class of nice discrete paths. In Part II, we show that any such nice discrete path can be replaced by a (continuous) path, and hence by a discrete path with only three points. The result then follows from Lemma 3.1.

Part I: Reduction to a nicer discrete path Q

We reduce the task to considering only discrete paths $P = (z_i)_{i=1}^n$ such that $G(x_{i-1}, x_i) = 0$ for each $1 \leq i \leq n$. The idea is to insert extra points into P to get a refinement Q for which this is true, and such that $l_0(Q) \leq l_0(P)$ and $d(Q) = d(P)$.

This is often, but not always, possible for points in $X = L^\infty(S)$. However we will show that it is true in general if we make use of a suitable isometric embedding J of X into an augmented L^∞ space A, and then refine (the isometric copy of) $J(P)$ in A. It suffices to take $A = L^\infty(S')$, where $S' = S \cup \{s\}$, $s \notin S$, and counting measure is again attached to S'; we denote the metric in A also by d. We define $J\colon X \to A$, $Ju = u'$, where $u'(a) = u(a)$, $a \in S$, and $u'(s) = d(u, 0) = \|u\|_\infty$. It is clear that J is an isometric embedding.

Given a discrete path P from x to y in X_0, we say that Q is a *refinement* of P in A_0 if Q is a discrete path in A_0 from x' to y' obtained by inserting zero or more additional intermediate points between elements in the discrete path P' identified with P under J. We claim that every discrete path in X_0 from x to y has a refinement in A_0 with the properties that $d(Q) = d(P)$, $l_0(Q) \leq l_0(P)$, and $G(z_{i-1}, z_i) = 0$ for every pair of adjacent points in Q. To prove our claim, it suffices to show that if $u, v \in X_0$ with $G(u, v) > 0$, then there exists $w' \in A_0$ with $d(u, v) = d(u', w') + d(w', z')$, $l_0(u, v) > l_0(u', w') + l_0(w', v')$ and $G(u', w') = G(w', v') = 0$. Assuming without loss of generality that $d(u, o) \leq d(v, o)$, we will prove that it suffices to define $w' \in A$ by the equations

$$\begin{aligned} w'(a) &= v'(a) + G(u, v)\,\mathrm{sgn}(u'(a) - v'(a)), \qquad a \in S, \\ w'(s) &= v'(s) + G(u, v), \end{aligned}$$

where $\mathrm{sgn}\colon \mathbb{R} \to \{-1, 0, 1\}$ is the usual sign function on the real line.

If $a \in S$, then $|w'(a) - v'(a)|$ equals either 0 or $G(u, v)$, depending on whether or not $v(a) = u(a)$. Since $|w'(s) - v'(s)| = G(u, v)$, we see that $d(w', v') = G(u, v)$. Also

$$|v(a) - u(a)| - G(u, v) \leq |w'(a) - u'(a)| \leq d(u, v) - G(u, v), \qquad a \in S.$$

Note that the above lower bound is obvious, while the obvious upper bound is

$$\max\big(d(u, v) - G(u, v), G(u, v)\big),$$

which equals $d(u, v) - G(u, v)$, as is clear from the definition of G. Since also

$$|w'(s) - u'(s)| = d(v, 0) + G(u, v) - d(u, 0) = d(u, v) - G(u, v),$$

we deduce that $d(u', w') = d(u, v) - G(u, v)$. Moreover, since $d(w', 0') = d(v, 0) + G(u, v)$, it follows that

$$\begin{aligned} 2G(u', w') &= d(u', w') + d(u', 0) - d(w', 0) = (d(u, v) - G(u, v)) + d(u, 0) - (d(v, 0) + G(u, v)) = 0, \\ 2G(w', v') &= d(w', v') + d(v', 0) - d(w', 0) = G(u, v) + d(v, 0) - (d(v, 0) + G(u, v)) = 0, \end{aligned}$$

and $d(u', w') + d(w', v') = d(u, v)$. Lastly, the fact that $d(w', 0) > d(v', 0) \geq d(u', 0)$ implies that $l_0(u, v) > l_0(u', w') + l_0(w', v')$ by Observation 2.8.

Part II: Reduction to a 3-point discrete path (via a continuous path)

Fixing two points $u', v' \in A_0$ satisfying $G(u', v') = 0$ and $d(u', 0) \le d(v', 0)$, and writing $L = d(u', v')$, we let $\gamma\colon S[0, L] \to A_0$ be the line segment path parameterized by arclength from u' to v'. Then $d(\gamma(t), 0) = d(u', 0) + t$ for $0 \le t \le L$, and so by Lemma 4.1, $l_0(u', v') = \mathrm{len}_0(\gamma)$.

It follows that for the nicer discrete path Q constructed in Part I, $l_0(Q) = \mathrm{len}_0(\lambda)$, where λ is the (continuous) polygonal path from x' to y' in A_0 obtained by "joining the dots" in Q. Thus our task has now been reduced to proving that $l_0(x', y') \le 2\,\mathrm{len}_0(\lambda)$ for every (continuous) path from x' to y'. Without loss of generality, we assume that $\lambda\colon [0, M] \to A_0$ is parametrized by arclength and that $d(x', 0) \le d(y', 0)$, so that $\mathrm{len}_0(\lambda) = \int_0^M D^{-2}(t)\,dt$, where $D(t) = d(\gamma(t), 0)$. To minimize $\mathrm{len}_0(\lambda)$ among all paths in A_0 from x' to y' of length M, we need to maximize $D(t)$. The triangle inequality gives two constraints: $D(t) \le d(x', 0) + t$ and $D(t) \le d(y', 0) + M - t$. Of these two constraints, the former is stronger when $0 \le t \le M - G_M$, where $G_M = (M - d(y', 0) + d(x', 0))/2 \ge 0$, and the latter is stronger when $M - G_M \le t \le M$.

Let $z \in A_0$ be defined by the equations

$$\begin{aligned} z'(a) &= y'(a) + G_M \,\mathrm{sgn}(x'(a) - y'(a)), \qquad a \in S,\\ z'(s) &= y'(s) + G_M. \end{aligned}$$

If $a \in S$, then $|z'(a) - y'(a)|$ equals either 0 or G_M, depending on whether or not $y(a) = x(a)$. Since also $|z'(s) - y'(s)| = G_M$, we see that $d(z', y') = G_M$. Similarly,

$$\begin{aligned} |y(a) - x(a)| - G_M \le |z'(a) - x'(a)| &\le d(x', y') - G_M \le M - G_M, \qquad a \in S,\\ |z'(s) - x'(s)| &= d(y', 0') + G_M - d(x', 0') = M - G_M, \end{aligned}$$

and we readily deduce that $d(x', z') = M - G_M$. Moreover, since $d(z', 0') = d(y', 0') + G_M$, it follows that

$$\begin{aligned} 2G(x', z') &= d(x', z') + d(x', 0) - d(z', 0) = (M - G_M) + d(x', 0) - (d(y', 0) + G_M) = 0,\\ 2G(z', y') &= d(z', y') + d(y', 0) - d(z', 0) = G_M + d(y', 0) - (d(y', 0) + G_M) = 0, \end{aligned}$$

and $d(x', z') + d(z', y') = d(x', y')$.

It follows that the path ν consisting of two line segments, one from x' to z' and the second from z' to y', maximizes $D(t)$ for all $0 \le t \le M$ and, again using Lemma 4.1, we see that $\mathrm{len}_0(\nu) = l_0(R)$, where R is the 3-point discrete path (x', z', y'). Thus we have reduced the task to showing that $l_0(x', y') \le 2(l_0(x', z') + l_0(z', y'))$, and this inequality is already given by Lemma 3.1. □

We have shown that $c_{\mathrm{inv}}(X) \le C_{\mathrm{inv}}(X) \le \widehat{C}_{\mathrm{inv}}(X) \le 2$ for all metric spaces. We finish by discussing conditions under which these constants equal 2. Let $x', y' \in X_o$ be as in the proof of Theorem 4.2, with $d(x', o) \le d(y', o)$. We saw that the infimum of $l_0(Q)$ among all discrete paths from x' to y' in A_0 is the same as its infimum over all those special discrete paths $Q = (x', z', y')$ where $G(x, z') = G(z', y') = 0$ and $d(z', 0) = d(y', 0) + G_M$, where $G_M = (M - d(y', 0) + d(x', 0))/2$ and $M \ge d(x, y)$. It readily follows from Observation 2.9 that

$$l_0(x', z') = \mathrm{len}_0([x', z']) = \int_{d(x',0)}^{d(y',0)+G_M} t^{-2}\,dt, \qquad l_0(z', y') = \mathrm{len}_0([z', y']) = \int_{d(y',0)}^{d(y',0)+G_M} t^{-2}\,dt.$$

Thus for fixed x', y', $l_0(Q)$ is minimized uniquely by minimizing G_M. Taking $M = d(x', y')$, G_M becomes $G(x', y')$.

Writing $\delta_x = d(x', 0)$, $\delta_y = d(y', 0)$, $\delta = d(x', y')$, and $\gamma = G(x', y') = (\delta - \delta_y + \delta_x)/2$, for this minimizing Q, we see that

$$l_0(Q) = \frac{\delta - \gamma}{\delta_x(\delta_y + \gamma)} + \frac{\gamma}{(\delta_y + \gamma)\delta_y} = \frac{\delta(\delta_x + \delta_y) + (\delta_y - \delta_x)^2}{\delta_x\delta_y(\delta + \delta_x + \delta_y)}.$$

Since a general metric space can be isometrically embedded in L^∞, and then isometrically embedded in an L^∞ space of one extra dimension as in the proof of Theorem 4.2, the above calculations yield the following corollary of the proof of Theorem 4.2.

Corollary 4.3.
If (X,d) is a metric space of cardinality at least 2, and $o\in X$, then

$$\widehat{d}_o(x,y)=\frac{\delta(\delta_x+\delta_y)+(\delta_y-\delta_x)^2}{\delta_x\delta_y(\delta+\delta_x+\delta_y)},$$

where $\delta_x=d(x,o)$, $\delta_y=d(y,o)$, and $\delta=d(x,y)$.

We know that $\widehat{d}_o(x,y)/I_o(x,y)\ge 1/2$, and we now determine when equality holds in this inequality. Assume without loss of generality that $0<\delta_x\le\delta_y$. Corollary 4.3 implies that

$$\frac{\widehat{d}_o(x,y)}{I_o(x,y)}=\frac{\delta(\delta_x+\delta_y)+(\delta_y-\delta_x)^2}{\delta(\delta+\delta_x+\delta_y)}=\frac{b(a+1)+(1-a)^2}{b(b+a+1)},$$

where $a=\delta_x/\delta_y$ and $b=\delta/\delta_y$ are real numbers satisfying $0<a\le 1$ and $1-a\le b\le 1+a$. Thus the task of locating all points where the minimum is achieved is reduced to calculating where this last real-valued expression equals 1/2 on the set

$$R=\{(a,b)\in\mathbb{R}^2 : 0<a\le 1,\ 1-a\le b\le 1+a\}.$$

By splitting

$$f(a,b)=\frac{b(a+1)+(1-a)^2}{b(b+a+1)}=\frac{a+1}{b+a+1}+\frac{(1-a)^2}{b(b+a+1)},$$

we see that this expression is strictly decreasing in b. Thus $f(a,b)$ is minimal for fixed a if and only if b is maximal, i.e. $b=1+a$. Then

$$f(a,1+a)=\frac{1}{2}+\frac{(1-a)^2}{2(1+a)^2},$$

and it is clear that the unique minimum is achieved when $a=1$.

Thus we conclude as before that $I_o(x,y)\le 2\widehat{d}_o(x,y)$, but this time with some extra information: we get equality if and only if x,y are such that $d(x,y)=2d(x,o)=2d(y,o)$. In view of the fact that f is continuous and has a unique minimum point on

$$R'=\{(a,b)\in\mathbb{R}^2 : 1/2\le a\le 1,\ 1\le b\le 1+a\},$$

we readily deduce the following result.

Theorem 4.4.
If (X,d) is a metric space of cardinality at least 2, then $\widehat{C}_{\rm inv}(X)=2$ if and only if for every $\epsilon>0$, there exist $o\in X$ and $x,y\in X_o$ such that $d(x,y)/2d(x,o)>1-\epsilon$ and $d(x,y)/2d(y,o)>1-\epsilon$.

Certainly, if X is a normed vector space, then taking $y=2o-x$, we have $d(x,y)=\|x-y\|=2\|x-o\|=2\|y-o\|$, so $\widehat{C}_{\rm inv}(X)=2$ in all such cases.

Clearly, calculating $\widehat{C}_{\rm inv}(X)$ is often now quite easy. By contrast, calculating $C_{\rm inv}(X)$ is typically much more difficult, but the above calculations do provide us with some insight. In particular, it is clear that $c_{\rm inv}(X)=C_{\rm inv}(X)=2$ if for every $\epsilon>0$ there exists a set of points $o\in X$, $x,y,z\in X_o$, such that the five numbers $d(y,o)/d(x,o)$, $d(x,y)/2d(x,o)$, $d(x,z)/d(x,o)$, $d(z,y)/d(x,o)$, and $d(z,o)/2d(x,o)$ all lie in the interval $[1-\epsilon,1+\epsilon]$.

Simple examples of spaces satisfying $c_{\rm inv}(X)=C_{\rm inv}(X)=2$ include any L^1 or L^∞ space of dimension more than 1, since such spaces include (isometric copies of) ℓ_2^1 or ℓ_2^∞, respectively, and we readily find such a configuration of points in those spaces, even for $\epsilon=0$: it suffices to take the configurations of points given in the second half of Proposition 3.2.

References

[1] Bonk M., Heinonen J., Koskela P., Uniformizing Gromov Hyperbolic Spaces, Astérisque, 207, Société Mathématique de France, Paris, 2001
[2] Buckley S.M., Falk K., Wraith D.J., Ptolemaic spaces and CAT(0), Glasg. Math. J., 2009, 51(2), 301–314
[3] Buckley S.M., Herron D.A., Xie X., Metric space inversions, quasihyperbolic distance, and uniform spaces, Indiana Univ. Math. J., 2008, 57(2), 837–890
[4] Buckley S.M., Wraith D.J., McDougall J., On Ptolemaic metric simplicial complexes, Math. Proc. Cambridge Philos. Soc., 2010, 149(1), 93–104
[5] Foertsch T., Lytchak A., Schroeder V., Nonpositive curvature and the Ptolemy inequality, Int. Math. Res. Not. IMRN, 2007, 22, # rnm100
[6] Foertsch T., Schroeder V., Hyperbolicity, CAT(−1)-spaces and the Ptolemy inequality, Math. Ann., 2011, 350(2), 339–356
[7] Herron D., Shanmugalingam N., Xie X., Uniformity from Gromov hyperbolicity, Illinois J. Math., 2008, 52(4), 1065–1109