Probability Measures on Metric Spaces of Nonpositive Curvature

Karl-Theodor Sturm

Abstract. We present an introduction to metric spaces of nonpositive curvature ("NPC spaces") and a discussion of barycenters of probability measures on such spaces. In our introduction to NPC spaces, we will concentrate on analytic and stochastic aspects of nonpositive curvature. Among others, we present two different characterizations of nonpositive curvature in terms of variance inequalities for probabilities on metric spaces as well as a characterization in terms of a weighted quadruple inequality. The latter generalizes Reshetnyak’s quadruple inequality which played a major role in many developments during the last decade. For Riemannian manifolds, nonpositive curvature will also be characterized by the existence of a contracting barycenter map on the space of probability measures on the respective manifold.

In our discussion of barycenters of probability measures on NPC spaces, basic results like the Law of Large Numbers, Jensen’s inequality and the $L^1$-contraction property will be derived and many examples will be presented. We identify barycenters for images, products, $L^2$-spaces, Hilbert spaces, trees and manifolds. Also some results on convex means will be included.

Contents

1. Geodesic Spaces
2. Global NPC Spaces
3. Examples of Global NPC Spaces
4. Barycenters
5. Identification of Barycenters
6. Jensen’s Inequality and $L^1$ Contraction Property
7. Convex Means
8. Local NPC Spaces
References

1991 Mathematics Subject Classification. Primary 60B, 58E, 28C; Secondary 58C, 52A, 49C.

Key words and phrases. Barycenter, mean value, expectation, center of mass, nonpositive curvature, NPC space, Hadamard space, Alexandrov space, quadruple inequality, variance inequality, Jensen inequality.

The author gratefully acknowledges support by the SFB 611 at the University of Bonn.
1. Geodesic Spaces

A curve in a metric space \((N, d)\) is a continuous map \(\gamma : I \to N\) where \(I \subset \mathbb{R}\) is some interval. Its length \(d(\gamma)\) is defined as the supremum of \(\sum_{k=1}^{n} d(\gamma_{t_k}, \gamma_{t_{k-1}})\) where \(t_0 \leq t_1 \leq \ldots \leq t_n\) and \(t_0, \ldots, t_n \in I\). A curve is called geodesic iff \(d(\gamma_r, \gamma_t) = d(\gamma_s, \gamma_t) + d(\gamma_s, \gamma_r)\) for all \(r, s, t \in I\) with \(r < s < t\) or, equivalently, iff \(d(\gamma_s, \gamma_t) = d(\gamma_{s, t}]\) for all \(s, t \in I\) with \(s < t\). Note that geodesics in the sense of Riemannian geometry are only required to minimize locally the length (i.e. the above holds true only if \(|s - t|\) is sufficiently small) whereas geodesics in our sense are always globally minimizing the length.

We say that a curve \(\gamma : [a, b] \to N\) connects the points \(x, y \in N\) iff \(\gamma_a = x\) and \(\gamma_b = y\). Obviously, this implies \(d(\gamma) \geq d(x, y)\).

**Definition 1.1.** A metric space is called a length space (or inner metric space) iff for all \(x, y \in N\)

\[
d(x, y) = \inf_{\gamma} L_d(\gamma),
\]

where the infimum is taken over all curves which connect \(x\) and \(y\). It is called a geodesic space iff each pair of points \(x, y \in N\) is connected by a curve \(\gamma\) of length \(L_d(\gamma) = d(x, y)\). This curve is not required to be unique.

**Proposition 1.2.** A complete metric space \((N, d)\) is a geodesic space if and only if \(\forall x_0, x_1 \in N : \exists z \in N:\)

\[
d(x_0, z) = d(x_1, z) = \frac{1}{2} d(x_0, x_1).
\]

Any point \(z \in N\) with the above properties will be called midpoint of \(x_0\) and \(x_1\).

**Proof.** It only remains to prove the "if"-implication. Given \(x_0, x_1 \in N\), we firstly obtain their midpoint \(x_{1/2} \in N\). Then the points \(x_{1/4}\) and \(x_{3/4}\) are obtained as midpoints of \(x_0\) and \(x_{1/2}\) or \(x_{1/2}\) and \(x_1\) respectively. By this procedure, we obtain the points \(x_t\) for all dyadic \(t \in [0, 1]\) and obviously \(d(x_r, x_s) = d(x_r, x_s) + d(x_s, x_t)\) for all dyadic \(0 \leq r < s < t \leq 1\). By completeness of \(N\), it yields the existence of \(x_t \in N\) for all \(t \in [0, 1]\) such that \(x : [0, 1] \to N\) is a geodesic.

**Remark 1.3.** A characterization (in terms of "approximative midpoints") similar to Proposition 1.2 holds true for length spaces:

A complete metric space \((N, d)\) is a length space (or geodesic space) if and only if for all \(x_0, x_1 \in N\) and \(\epsilon > 0\) (or for \(\epsilon = 0\), resp.) there exists \(y \in N\) such that

\[
d^2(x_0, y) + d^2(x_1, y) \leq \frac{1}{2} d^2(x_0, x_1) + \epsilon.
\]

Let \((N, d)\) be a geodesic space.

**Definition 1.4.** A set \(N_0 \subset N\) is called convex iff \(\gamma([0, 1]) \subset N_0\) for each geodesic \(\gamma : [0, 1] \to N\) with \(\gamma_0, \gamma_1 \in N_0\). A function \(\varphi : N \to \mathbb{R}\) is called convex iff the function \(\varphi \circ \gamma : [0, 1] \to \mathbb{R}\) is convex for each geodesic \(\gamma : [0, 1] \to N\), i.e. iff \(\forall t \in [0, 1]\)

\[
\varphi(\gamma_t) \leq (1 - t)\varphi(\gamma_0) + t\varphi(\gamma_1).
\]

**Proposition 1.5.** For \(\varphi : N \to \mathbb{R}\) define its epigraph

\[
N_\varphi = \{(x, r) \in N \times \mathbb{R} : \varphi(x) \leq r\} \subset N \times \mathbb{R}.
\]

Then
PROBABILITY MEASURES ON METRIC SPACES

(i) \( \varphi \) is convex if and only if \( N_\varphi \) is convex.

(ii) \( \varphi \) is lower semicontinuous if and only if \( N_\varphi \) is closed.

PROOF. (i) \( N_\varphi \) will be regarded as a subset of the space \( \hat{N} = N \times \mathbb{R} \) with the metric \( d((x, r), (y, s)) = (d^2(x, y) + |r - s|^2)^{1/2} \). Hence, \( \hat{\gamma} : [0, 1] \to \hat{N} \) is a geodesic if and only if \( \hat{\gamma}(t) = (\gamma(t), c_0 + c_1 t) \) with a geodesic \( \gamma : [0, 1] \to N \) and \( c_0, c_1 \in \mathbb{R} \).

Now let \( \hat{\gamma} \) be a geodesic with \( \hat{\gamma}(0), \hat{\gamma}(1) \in N_\varphi \), that is, with \( \varphi \circ \gamma(0) \leq c_0 \) and \( \varphi \circ \gamma(1) \leq c_0 + c_1 \). Convexity of \( \varphi : N \to \mathbb{R} \) implies convexity of \( \varphi \circ \gamma : [0, 1] \to \mathbb{R} \) and this in turn
\[
\varphi(\gamma(t)) \leq c_0 + c_1 t
\]
or, in other words, \( \varphi(t) \in N_\varphi \). This proves the convexity on \( N_\varphi \).

Now, conversely, assume that \( N_\varphi \) is convex. Let \( \gamma : [0, 1] \to N \) be any geodesic. Choose \( c_0 = \varphi \circ \gamma(0), c_1 = \varphi \circ \gamma(1) - \varphi \circ \gamma(0) \) and \( \hat{\gamma}(t) := (\gamma(t), c_0 + c_1 t) \). Then \( \hat{\gamma}(0), \hat{\gamma}(1) \in N_\varphi \) and thus also \( \hat{\gamma} \in N_\varphi \). The latter states that
\[
\varphi \circ \gamma(t) \leq c_0 + c_1 t = (1 - t)\varphi \circ \gamma(0) + t\varphi \circ \gamma(1)
\]
for all \( t \in [0, 1] \). That is, \( \varphi \circ \gamma : [0, 1] \to \mathbb{R} \) is convex for each geodesic \( \gamma : [0, 1] \to N \) and thus \( \varphi : N \to \mathbb{R} \) is convex.

(ii) \( N_\varphi \) is closed \( \iff \) \( \{x_n \to x, r_k \to r \Rightarrow \varphi(x) \leq r \} \iff \varphi \) is lower semicontinuous. \( \square \)

DEFINITION 1.6. A function \( \varphi : N \to \mathbb{R} \) is called \textit{uniformly convex} iff there exists a strictly increasing function \( \eta : \mathbb{R}_+ \to \mathbb{R}_+ \) such that for any geodesic \( \gamma : [0, 1] \to N \)
\[
\varphi(\gamma(t)) \leq \frac{1}{2}(\varphi(\gamma(0)) + \varphi(\gamma(1))) - \eta(d(\gamma(0), \gamma(1))).
\]
\( \varphi \) is called \textit{strictly convex} iff for any geodesic \( \gamma : [0, 1] \to N \) with \( \gamma_0 \neq \gamma_1 \)
\[
\varphi(\gamma_{1/2}) < \frac{1}{2}(\varphi(\gamma_0) + \varphi(\gamma_1)).
\]

PROPOSITION 1.7. Let \( \varphi : N \to \mathbb{R} \) be a uniformly convex, lower semicontinuous function on a complete geodesic space \( (N, d) \). Then there exists a unique minimizer \( z \in N \), i.e. a unique point \( z \in N \) with \( \varphi(z) = \inf_{w \in N} \varphi(w) \). We write
\[
z = \arg \min_{w \in N} \varphi(w).
\]

PROOF. (i) Existence: Let \( z_n \) be a sequence of points in \( N \) with \( \lim \varphi(z_n) = \inf \varphi(z) =: \alpha \) and let \( z_{n,k} \) the midpoint between \( z_n \) and \( z_k \). Then for \( n, k \to \infty \)
\[
\alpha \leq \varphi(z_{n,k}) \leq \frac{1}{2}(\alpha + \frac{1}{2}\varphi(z_n) + \frac{1}{2}\varphi(z_k)) - \eta(d(z_n, z_k)).
\]
Consequently, \( d(z_n, z_k) \to 0 \) for \( n, k \to \infty \), i.e. \( (z_n)_n \) is a Cauchy sequence and therefore there exists \( z^* = \lim_{n \to \infty} z_n \in N \) since \( N \) is complete. Moreover, \( \varphi(z^*) = \inf \varphi(z) \) by lower semicontinuity of \( \varphi \).

(ii) Uniqueness: Assume \( \varphi(z_0) = \varphi(z_1) = \inf \varphi(z) = \alpha \) and \( z_0 \neq z_1 \). For the midpoint \( z_{1/2} \) between \( z_0 \) and \( z_1 \) we get the contradiction \( \alpha \leq \varphi(z_{1/2}) < \frac{1}{2}\alpha + \frac{1}{2}(\alpha) \). \( \square \)
Remark 1.8. For the uniqueness of the minimizer it suffices to require that $\varphi$ is strictly convex.

If $N$ is compact then for the existence of the minimizer it suffices to require that $\varphi$ is convex and lower semicontinuous.

Definition 1.9. A geodesic space $(N,d)$ is called doubly convex iff the function $d : (x,y) \mapsto d(x,y)$ is convex on $N \times N$ or, in other words, iff the function $t \mapsto d(\gamma_t, \eta_t)$ is convex for each pair of geodesics $\gamma, \eta : [0,1] \to N$.

It is called strictly doubly convex iff $d$ is strictly convex on $N \times N$.

Remark 1.10. (i) In a doubly convex geodesic space, any two of its points are joined by a unique geodesic and this geodesics depends continuously on its endpoints.

(ii) If a geodesic space is locally doubly convex and simply connected then it is doubly convex [Gr87], [EF01].

Let $(M,\mathcal{M})$ be a measurable space and $(N,d)$ be a metric space. A map $f : M \to N$ is called measurable iff it is measurable with respect to the given $\sigma$-field $\mathcal{M}$ on $M$ and the Borel $\sigma$-field $\mathcal{B}(N)$ on $N$, i.e. iff $f^{-1}(N') \in \mathcal{M}$ for all $N' \in \mathcal{B}(N)$. It is well known that for the latter it suffices that $f^{-1}(N') \in \mathcal{M}$ for all open $N' \subset N$.

A map $f : M \to N$ is called elementary measurable iff it is measurable and has finite range. In other words, iff there exists a decomposition of $M$ into finitely many disjoint sets $M_i \in \mathcal{M}$ such that $f$ is constant on each of the $M_i$.

It is called strongly measurable iff it is the (pointwise) limit of a sequence of elementary measurable maps or, equivalently, iff it is measurable and has separable range $f(M)$. See [St02].

2. Global NPC Spaces

We present an introduction to metric spaces of nonpositive curvature ("NPC spaces") with emphasis on analytic and stochastic aspects of nonpositive curvature. For instance, we do not deal with triangle or angle comparison but use the explicit estimates for the distance function. Also we do not introduce the tangent cone or the space of directions.

For the many and deep geometric aspects we refer to the huge literature on NPC spaces. The whole development started with the investigations of A. D. Alexandrov [Al51] and Yu. G. Reshetnyak [Re68] and was strongly influenced by the work of M. Gromov [Gr81]. Recently, there appeared various monographs devoted exclusively to NPC spaces: [Ba95], [Jo97] and [BH99]. Also the monographs [BBI01], [BGS85] and [EF01] contain much material on this subject. Moreover, we recommend the articles [AB90], [KS93] and [Jo94].

Definition 2.1. A metric space $(N,d)$ is called global NPC space if it is complete and if for each pair of points $x_0, x_1 \in N$ there exits a point $y \in N$ with the property that for all points $z \in N$:

$$d^2(z,y) \leq \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1).$$

Here "NPC" stands for "nonpositive curvature". Global NPC spaces are also called Hadamard spaces.
Property (2.1) (or the equivalent property (2.3) below) is called the NPC inequality. The defining property (2.1) of global NPC spaces can be weakened.

**Remark 2.2.** A complete metric space \((N,d)\) is a global NPC space if and only if for all \(x_0, x_1 \in N\) and \(\epsilon > 0\) there exists \(y \in N\) such that for all \(z \in N\)

\[
d^2(z,y) \leq \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1) + \epsilon.
\]

**Proof.** Given \(x_0, x_1 \in N\) denote for each \(\epsilon > 0\) the point \(y \in N\) with the required property by \(y_\epsilon\). Choosing \(z = x_0\) or \(z = x_1\) yields \(\frac{1}{2}d^2(y_\epsilon,x_0) + \frac{1}{2}d^2(y_\epsilon,x_1) \leq \frac{1}{4}d^2(x_0,x_1) + \epsilon\) and thus \(d^2(y_\epsilon,y) \leq \frac{1}{4}d^2(y_\epsilon,x_0) + \frac{1}{4}d^2(y_\epsilon,x_1) - \frac{1}{4}d^2(x_0,x_1) + \delta \leq \delta + \epsilon\) for each \(\delta > 0\). That is, \((y_\epsilon)_{\epsilon>0}\) is a Cauchy family and there exists a unique \(y = \lim_{\epsilon \to 0} y_\epsilon \in N\). Obviously, this \(y\) satisfies \(d^2(z,y) \leq \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1)\) for all \(z \in N\) which proves the claim. \(\Box\)

**Proposition 2.3.** If \((N,d)\) is a global NPC space then it is a geodesic space. Even more, for any pair of points \(x_0, x_1 \in N\) there exists a unique geodesic \(x : [0,1] \to N\) connecting them. For \(t \in [0,1]\) the intermediate points \(x_t\) depend continuously on the endpoints \(x_0, x_1\). Finally, for any \(z \in N\)

\[
d^2(z, x_t) \leq (1-t)d^2(z, x_0) + td^2(z, x_1) - t(1-t)d^2(x_0, x_1).
\]

**Figure 1.** NPC inequality

**Proof.** (i) Choosing \(z = x_0\) or \(z = x_1\) in (2.1) yields \(d(x_0, x_{1/2}) \leq \frac{1}{2}d(x_0, x_1)\) and \(d(x_2, x_1) \leq \frac{1}{2}d(x_0, x_1)\). Hence, \(x_2\) is a midpoint and (by Proposition 1.2) \((N,d)\) is a geodesic space. Choosing \(z\) to be any other midpoint of \(x_0\) and \(x_1\) yields \(d(z, x_{1/2}) = 0\). That is, midpoints are unique and thus also geodesics are unique.

(ii) Given any geodesic \(x : [0,1] \to N\) it suffices to prove (2.3) for all dyadic \(t \in [0,1]\). It obviously holds for \(t = 0\) and \(t = 1\). Assume that it holds for all \(t = k2^{-n}\) with \(k = 0,1,\ldots,2^n\) We want to prove that then (2.3) also holds for all \(t = k2^{-(n+1)}\) for all \(t = k2^{-n}\) with \(k = 0,1,\ldots,2^{n+1}\). For even \(k\) this is just the assumption. Fix \(t = k2^{-(n+1)}\) with an odd \(k\) and put \(\Delta t = 2^{-(n+1)}\). Then by (2.1)

\[
d^2(z,x_{1/2}) \leq \frac{1}{2}d^2(z,x_{1-\Delta t}) + \frac{1}{2}d^2(z,x_{1+\Delta t}) - \frac{1}{4}d^2(x_{1-\Delta t},x_{1+\Delta t})
\]

and by (2.3) (for multiples of \(2^{-n}\))

\[
d^2(z,x_{t\pm\Delta t}) \leq (1-t\mp\Delta t)d^2(z,x_0) + (t\pm\Delta t)d^2(z,x_1) - (1-t\mp\Delta t)(t\pm\Delta t)d^2(x_0,x_1).
\]
Thus
\[ d^2(z, x_1) \leq (1 - t)d^2(z, x_0) + td^2(z, x_1) \]
\[- \left[ (\Delta t)^2 - \frac{1}{2}(1 - t - \Delta t)(t + \Delta t) - \frac{1}{2}(1 - t + \Delta t)(t - \Delta t) \right] d^2(x_0, x_1) \]
\[ = (1 - t)d^2(z, x_0) + td^2(z, x_1) - t(1 - t)d^2(x_0, x_1). \]

(iii) Now let \( x, y : [0, 1] \rightarrow N \) be two geodesics. Then applying (2.3) twice yields
\[ d^2(x_1, y_t) \leq (1 - t)d^2(x_0, y_t) + td^2(x_1, y_t) - t(1 - t)d^2(x_0, x_1) \]
\[ \leq (1 - t)^2d^2(x_0, y_0) + t^2d^2(x_1, y_1) \]
\[ + t(1 - t) \left[ d^2(x_0, y_1) + d^2(x_1, y_0) - d^2(x_0, x_1) - d^2(y_0, y_1) \right]. \]

Obviously, the right hand side converges to 0 if \( y_0 \rightarrow x_0 \) and \( y_1 \rightarrow x_1 \), and thus \( y_t \rightarrow x_t \), that is \( x_t \) depends continuously on \( x_0 \) and \( x_1 \). \( \square \)

**Proposition 2.4** (Reshetnyak’s Quadruple Comparison). For all \( x_1, x_2, x_3, x_4 \in N \) in a global NPC space \((N, d)\)
\[ d^2(x_1, x_3) + d^2(x_2, x_4) \leq d^2(x_2, x_3) + d^2(x_4, x_1) + 2d(x_1, x_2) \cdot d(x_3, x_4), \]
in particular,
\[ d^2(x_1, x_3) + d^2(x_2, x_4) \leq d^2(x_2, x_3) + d^2(x_4, x_1) + d^2(x_1, x_2) + d^2(x_3, x_4). \]

![Quadruple comparison](image)

**Figure 2. Quadruple comparison**

**Proof.** This is an immediate consequence of the more general Theorem 4.9 which will be proven below. Indeed, property (iv) of that result with \( s = t \) yields
\[ d^2(x_1, x_3) + d^2(x_2, x_4) \leq \frac{t}{1 - t}d^2(x_1, x_2) + d^2(x_2, x_3) + \frac{1 - t}{t}d^2(x_3, x_4) + d^2(x_4, x_1). \]
Choosing \( t \) optimal proves the claim. \( \square \)

The original proof of Reshetnyak is based on the fact that the quadruple can be embedded into a twodimensional Euclidean space in such a way that the side lengths are preserved but the diagonals expand. Another proof which avoids embedding can be found in \([Jo97]\).
Inequality (2.4) is equivalent to the following (which might seem to be stronger):

\[
x_1, x_2, x_3, x_4 \in N, \forall r \in [0, 1]:
\]

\[
d^2(x_1, x_3) + d^2(x_2, x_4) \leq d^2(x_2, x_3) + d^2(x_4, x_1) + d^2(x_1, x_2) + d^2(x_3, x_4)
- r [d(x_1, x_2) - d(x_3, x_4)]^2 - (1 - r) [d(x_2, x_3) - d(x_4, x_1)]^2.
\]

Indeed, for \( r = 1 \) this is just (2.4), and for \( r = 0 \) it is (2.4) with a cyclic permutation of \( x_1, \ldots, x_4 \). The general case is a linear combination of these two cases.

**Corollary 2.5 (Geodesic Comparison).** Let \((N, d)\) be a global NPC space, \(\gamma, \eta : [0, 1] \to N\) be geodesics and \(t \in [0, 1]\). Then

(2.5) \[ d^2(\gamma_t, \eta_t) \leq (1 - t)d^2(\gamma_0, \eta_0) + td^2(\gamma_1, \eta_1) - t(1 - t)[d(\gamma_0, \gamma_1) - d(\eta_0, \eta_1)]^2 \]

and

(2.6) \[ d(\gamma_t, \eta_t) \leq (1 - t)d(\gamma_0, \eta_0) + td(\gamma_1, \eta_1). \]

**Figure 3.** Geodesic comparison

**Proof.** In part (iii) of the proof of Proposition 2.3 we have already deduced that

\[
d^2(\gamma_t, \eta_t) - (1 - t)^2d^2(\gamma_0, \eta_0) - t^2d^2(\gamma_1, \eta_1)
\leq t(1 - t)[d^2(\gamma_0, \eta_1) + d^2(\gamma_1, \eta_0) - d^2(\gamma_0, \gamma_1) - d^2(\eta_0, \eta_1)].
\]

By quadruple comparison, the right hand side is

\[
\leq t(1 - t)\left[d^2(\gamma_0, \eta_0) + d^2(\gamma_1, \eta_1)
- r \left(d(\gamma_0, \eta_0) - d(\gamma_1, \eta_1))^2 - (1 - r) \left(d(\gamma_0, \gamma_1) - d(\eta_0, \eta_1))^2\right]\right]
\]

for each \( r \in [0, 1] \). For \( r = 0 \) this yields (2.4) and for \( r = 1 \) it yields

\[
d^2(\gamma_t, \eta_t) \leq (1 - t)d^2(\gamma_0, \eta_0) + td^2(\gamma_1, \eta_1) - t(1 - t)[d(\gamma_0, \eta_0) - d(\gamma_0, \eta_1)]^2
\]

\[
= [(1 - t)d(\gamma_0, \eta_0) + td(\gamma_1, \eta_1)]^2.
\]

□
In other words, (2.6) states that $d$ is doubly convex, i.e. $(x, y) \mapsto d(x, y)$ is a
convex function on $N \times N$. Obvious consequences are:

(i) For each $z \in N$ the function $x \mapsto d(x, z)$ is convex; in particular, all balls
$B_r(z) \subset N$ are convex.

(ii) Geodesics depend continuously on their endpoints in the following quanti-
tative way:
\[
d_\infty(\eta, \gamma) = \sup \{d(\eta_0, \gamma_0), d(\eta_1, \gamma_1)\},
\]
where for any curves $\eta, \gamma : [0, 1] \to N$ we put $d_\infty(\eta, \gamma) := \sup\{d(\eta_t, \gamma_t) : t \in [0, 1]\}$.

(iii) $N$ is contractible and, in particular, simply connected.

**Proposition 2.6** (Convex Projection). (i) For each convex closed set $K \subset N$
in a global NPC space $(N, d)$ there exists a unique map $\pi_K : N \to K$ ("projection
onto $K$") with
\[
d(\pi_K(z), z) = \inf_{w \in K} d(w, z) \quad (\forall z \in N)
\]
(ii) $\pi_K$ is "orthogonal":
\[
d^2(z, w) \geq d^2(z, \pi_K(z)) + d^2(\pi_K(z), w) \quad (\forall z \in N, \ w \in K)
\]
(iii) $\pi_K$ is a contraction:
\[
d(\pi_K(z), \pi_K(w)) \leq d(z, w) \quad (\forall z, w \in N)
\]

![Figure 4. Convex projection](image)

**Proof.** (i) Fix $z \in N$ and a closed convex set $K \subset N$. Then $K$ is a global
NPC space and the function $\varphi : K \to R, x \mapsto d^2(x, z)$ in continuous and uniformly
convex on $K$. Hence by Proposition 1.2.11 there exists a unique minimizer in $K$.

(ii) Let $t \mapsto w_t$ be the geodesic joining $w_0 := \pi_K(z)$ and $w_1 := w$. Then
$w_t \in K$ for all $t \in [0, 1]$ by convexity and closedness of $K$. Hence, by the NPC
inequality
\[
d^2(\pi_K(z), z) \leq d^2(w_t, z) \leq (1 - t)d^2(\pi_K(z), z) + td^2(w, z) - t(1 - t)d^2(\pi_K(z), w)
\]
and therefore
\[
d^2(\pi_K(z), z) + (1 - t)d^2(\pi_K(z), w) \leq d^2(w, z).
\]
(iii) Put \( z' = \pi_K(z), w' = \pi_K(w) \). Then (ii) and quadruple comparison imply

\[
d^2(z, w) + d^2(w, w') + d^2(w', z') + d^2(z', w) \geq d^2(z, w') + d^2(z', w) \geq d^2(z, z') + d^2(w, w') + 2d^2(w', z') \text{ which yields the claim.}
\]

The important fact here is the existence of a unique projection without assuming any kind of compactness of \( K \).

**Remark 2.7.** (i) Given any subset \( A \subset N \) in a global NPC space \((N, d)\), there exists a unique smallest convex set \( C(A) \) containing \( A \), called convex hull of \( A \). It can be constructed as \( C(A) = \bigcup_{n=0}^{\infty} A_n \) where \( A_0 := A \) and for \( n \in \mathbb{N} \), the set \( A_n \) consists of all points in \( N \) which lie on geodesics which start and end in \( A_{n-1} \).

(ii) Given any bounded subset \( A \subset N \) in a global NPC space \((N, d)\) there exists a unique closed ball of minimal radius which contains \( A \). In other words, there exists a unique point \( x \in N \) (the circumcenter of \( A \)) such that

\[
r(x, A) = \inf_{z \in N} r(z, A)
\]

where \( r(z, A) := \sup_{y \in A} d(z, y) \). This is an immediate consequence of Proposition 1.7 since the function \( z \mapsto r^2(z, A) \) is uniformly convex.

3. Examples of Global NPC Spaces

Our main examples for global NPC spaces are manifolds, trees and Hilbert spaces. Further examples are cones, buildings and surfaces of revolution. New global NPC spaces can be built out of given global NPC spaces as subsets, images, gluings, products or \( L^2 \)-spaces.

**Proposition 3.1** (Manifolds). Let \((N, g)\) be a Riemannian manifold and let \( d \) be its Riemannian distance. Then \((N, d)\) is a global NPC space if and only if it is complete, simply connected and of nonpositive (sectional) curvature.

Besides manifolds, the most important examples of NPC spaces are trees, in particular, spiders.

**Example 3.2** (Spiders). Let \( K \) be an arbitrary set and for each \( i \in K \) let \( N_i = \{(i, r) : r \in \mathbb{R}_+\} \) be a copy of \( \mathbb{R}_+ \) (equipped with the usual metric). Define the spider over \( K \) or \( K \)-spider \((N, d)\) by gluing together all these spaces \( N_i, i \in K \), at their origins, i.e.

\[
N = \{(i, r) : i \in K, r \in \mathbb{R}_+\}/\sim \text{ where } (i, 0) \sim (j, 0) \text{ (\forall i, j)}
\]

and

\[
d(((i, r), (j, s))) = \begin{cases} |r - s|, & \text{if } i = j \\ |r| + |s|, & \text{else.} \end{cases}
\]

The rays \( N_i \) can be regarded as closed subsets of \( N \). Any two rays \( N_i \) and \( N_j \) with \( i \neq j \) intersect at the origin \( o := (i, 0) = (j, 0) \) of \( N \).

The \( K \)-spider \( N \) depends (upto isometry) only on the cardinality of \( K \). If \( K = \{1, \ldots, k\} \) for some \( k \in \mathbb{N} \) then it is called \( k \)-spider. It can be realized as a subset of the complex plane

\[
\{r \cdot \exp\left(\frac{l}{k} 2\pi i\right) \in \mathbb{C} : r \in \mathbb{R}_+, l \in \{1, \ldots, k\}\},
\]

however, equipped with a non-Euclidean metric. If \( k = 1 \) or \( = 2 \) then it is isometric to \( \mathbb{R}_+ \) or \( \mathbb{R} \), resp. The 3-spider is also called tripod.
Example 3.3 (Booklet). Gluing together $k$ copies of halfspaces $\mathbb{R}_+ \times \mathbb{R}^{n-1}$ along the set $\{0\} \times \mathbb{R}^{n-1}$ yields a booklet $N$. In other words, $N = N_0 \times \mathbb{R}^{n-1}$ where $N_0$ is the $k$-spider.

Proposition 3.4 (Trees). Each metric tree is a global NPC space.

Proof. We have to prove the NPC inequality (2.1) for each triple of points $x_0, x_1, z \in N$. Without restriction, we may replace $N$ by the convex hull of these three points which is isometric to the convex hull of three points in the tripod. That is, without restriction $N$ is the tripod.

Firstly, consider the case where $x_0, x_1, z$ lie on one geodesic $\gamma : I \to N$. Then $\gamma$ is an isometry between $I \subset \mathbb{R}$ and $\gamma(I) \subset N$. Since $I$ is globally NPC, so is $\gamma(I)$. Actually, $I$ and thus $\gamma(I)$ are even "flat", i.e.

$$d^2(z, x_{1/2}) = \frac{1}{2}d^2(z, x_0) + \frac{1}{2}d^2(z, x_1) - \frac{1}{4}d^2(x_0, x_1)$$

for all $x_0, x_1, z \in \gamma(I)$ and with $x_{1/2}$ being the midpoint of $x_0, x_1$.

Secondly, consider the non-degenerate case $x_0 = (i, r), x_1 = (j, s)$ and $z = (k, t)$ with $r \cdot s \cdot t > 0$ and different $i, j, k \in \{1, 2, 3\}$. Assume without restriction $r \geq s$ and put $z' = (j, t)$. Then $x_0, x_{1/2} \in N_i$ and $z' \in N_j$. The points $x_0, x_1, z'$ lie on one geodesic. Hence, by the previous considerations

$$d^2(z', x_{1/2}) = \frac{1}{2}d^2(z', x_0) + \frac{1}{2}d^2(z', x_1) - \frac{1}{4}d^2(x_0, x_1).$$

Moreover, $d(x_0, z) = d(x_0, z')$ and $d(x_{1/2}, z) = d(x_{1/2}, z')$ whereas $d(x_1, z) \geq d(x_1, z')$. Hence, finally,

$$d^2(z, x_{1/2}) = d^2(z', x_{1/2}) = \frac{1}{2}d^2(z', x_0) + \frac{1}{2}d^2(z', x_1) - \frac{1}{4}d^2(x_0, x_1) \leq \frac{1}{2}d^2(z, x_0) + \frac{1}{2}d^2(z, x_1) - \frac{1}{4}d^2(x_0, x_1).$$

□

Proposition 3.5 (Hilbert spaces).

(i) Each Hilbert space is a global NPC space.

(ii) A Banach space is a global NPC space if and only if it is a Hilbert space.

(iii) A metric space is (derived from) a Hilbert space if and only if it is a nonempty, geodesically complete global NPC space with curvature $\geq 0$.

One possible (of many equivalent) definitions for the latter is to require that in (2.1) also the reverse inequality holds true.
Choosing a subset $K$ with weights yields a closed convex set $\psi$ if and only if it is closed and convex.

Then $\psi$ is a "parallelogram inequality". Replacing $x_0$ and $x_1$ in this inequality by $x_0 + x_1$ and $x_0 - x_1$, resp., yields the opposite inequality and thus proves the parallelogram equality.

(iii) The "only if"-implication is easy. For the "if"-implication fix an arbitrary point $o \in N$. Then for each $x \in N$ there exists a unique geodesic $x : \mathbb{R} \to N$ with $x_0 = o$ and $x_1 = x$. Using these geodesics we define a scalar multiplication by $\lambda \cdot x := x_\lambda$ ($\forall \lambda \in \mathbb{R}, x \in N$), an addition by $x + y := \text{midpoint of } 2 \cdot x$ and $2 \cdot y$ ($\forall x, y \in N$), and an inner product by $(x, y) := \frac{1}{2}(d^2(x, y) - d^2(o, x) - d^2(o, y))$ ($\forall x, y \in N$).

For details, see [KS93].

**Lemma 3.6 (Subsets).** A subset $N_0 \subset N$ of a global NPC space $N$ is a global NPC space if and only if it is closed and convex.

**Lemma 3.7 (Images).** Let $\psi : N \to N'$ be an isometry between metric spaces $(N, d)$ and $(N', d')$. Then $(N, d)$ is globally NPC if and only if $(N', d')$ is.

**Example 3.8.** Let $\psi : N \to \mathbb{R}$ be a bijective map from an arbitrary set $N$ onto a closed interval $\psi(N) \subset \mathbb{R}$ and define

$$d(x, y) := |\psi(x) - \psi(y)| \quad (\forall x, y \in N).$$

Then $(N, d)$ is a global NPC space since $\psi$ defines an isometry between $N$ and the closed convex set $\psi(N) \subset \mathbb{R}$.

**Lemma 3.9 (Products).** The direct product of metric spaces $(N_i, d_i)$, $i = 1, ..., k$ is the metric space $(N, d)$ defined by

$$N = \bigotimes_{i=1}^{k} N_i, \quad d(x, y) = \left( \sum_{i=1}^{k} d_i(x_i, y_i)^2 \right)^{1/2}.$$

It is a global NPC space if all factors are global NPC spaces.

More generally, given an arbitrary family of pointed metric spaces $(N_i, d_i, o_i)$, $i \in K$ with weights $m_i > 0$, the weighted $l^2$-product is the metric space $(N, d)$ defined by

$$N = \left\{ x \in \bigotimes_{i \in K} N_i : \sum_{i \in K} m_i \cdot d_i(x_i, o_i)^2 < \infty \right\}, \quad d(x, y) = \left[ \sum_{i \in K} m_i \cdot d_i(x_i, y_i)^2 \right]^{1/2}.$$

It is a global NPC space if all factors are global NPC spaces.

Here a pointed metric space is a triple $(N, d, o)$ consisting of a metric space $(N, d)$ and a specified ("base") point $o \in N$.

**Proof.** One easily verifies that $(N, d)$ is a complete metric space. Given $x(0)$ and $x(1) \in N$, we define a curve $x : t \mapsto x(t) \in N$ by $x(t) := (x_i(t))_{i \in K}$ where $x_i : t \mapsto x_i(t) \in N_i$ is the unique geodesic in $N_i$ connecting $x_i(0)$ and $x_i(1)$ (for each $i \in K$). Obviously (2.1) carries over from $(N_i, d_i)$ to $(N, d)$.
The results are well-known, see e.g. \cite{KS93, Jo94, Jo97}. For the convenience of the reader we present the main arguments.

(i) Is obvious.

(ii) The NPC inequality
\[ d_2^2(g, f_t) \leq (1 - t)d_2^2(g, f_0) + td_2^2(g, f_1) - (1 - t)td_2^2(f_0, f_1) \]
for fixed versions of the maps \( g, f_0, f_1 \in L^2(M, N, h) \) follows from integrating (w.r.t. \( m(dx) \)) the NPC inequality
\[ d^2(g(x), f_t(x)) \leq (1 - t)d^2(g(x), f_0(x)) + td^2(g(x), f_1(x)) - (1 - t)td^2(f_0(x), f_1(x)) \]
for the points \( g(x), f_0(x), f_1(x) \in N \).

For each \( x \in M \) let \( t \mapsto f_t(x) \) be the (unique) geodesic connecting \( f_0(x), f_1(x) \in N \). It depends continuously on the endpoints \( f_0(x), f_1(x) \in N \) and thus in a measurable way on \( x \in M \). Moreover, \( d(f_s(x), f_t(x)) = |t - s| \cdot d(f_0(x), f_1(x)) \). Integrating this with respect to \( m(dx) \) yields \( d_2(f_s, f_t) = |t - s| \cdot d_2(f_0, f_1) < \infty \). Hence, \( f_t \in L^2(M, N, h) \) and \( t \mapsto f_t \) is a geodesic in \( L^2(M, N, h) \) connecting \( f_0 \) and \( f_1 \).

(iii) If \( m \) is finite then \( N \) can be isometrically embedded into \( L^2(M, N) \) by identifying points \( z \in N \) with constant maps \( f \equiv z \in L^2(M, N) \).

PROPOSITION 3.11 (Limits). The (pointed or unpointed) Gromov-Hausdorff limit \((N, d)\) of a converging sequence of global NPC spaces \((N_k, d_k), k \in \mathbb{N}\), is again a global NPC space.

PROOF. \cite[Corollary II.3.10.]{BH99} \( \square \)

THEOREM 3.12 (Reshetnyak’s Gluing Theorem). Let \((N_i, d_i), i \in K\), be an arbitrary family of global NPC spaces and for each \( i \in K \), let \( C_i \) be a closed convex subset of \( N_i \). Assume that all these sets \( C_i \) are isometric and fix isometries \( \psi_{ij} : C_i \to C_j, i, j \in K \). Define a metric space \( N \) by gluing together the spaces \( N_i, i \in K \) along the isometries \( \psi_{ij}, i, j \in K \). That is,
\[ N = \bigsqcup_{i \in K} N_i / \sim \quad \text{with } x \sim y \text{ iff } \psi_{ij}(x) = y \text{ for some } i, j \in K. \]
Then \((N, d)\) is a global NPC space.
Each of the original spaces \((N_i, d_i)\) is isometrically embedded into \((N, d)\) as a closed convex subset. The intersection of (the embedding of) two different spaces \((N_i, d_i)\) and \((N_j, d_j)\) coincides with the common embedding of all the sets \(C_k, k \in K\).

**Proof.** [BH99, Thm. II.11.3] (or in a slightly more restrictive form [BBI01, Thm. 9.1.21]).

**Example 3.13 (Cones).** (i) The cone \((N, d)\) over a length space \((M, \rho)\) is globally NPC if and only if \((M, \rho)\) has globally curvature \(\leq 1\).

(ii) The cone over a circle of length \(L\) is globally NPC if and only if \(L \geq 2\pi\).

**Example 3.14 (Warped Products).** Let \((N_1, d_1)\) and \((N_2, d_2)\) be global NPC spaces and \(\varphi : N_1 \to \mathbb{R}\) be a convex and continuous function. Then the warped product \(N_1 \times \varphi N_2\) is a global NPC space.

**Proof.** [AB98].

**Example 3.15 (Buildings).** Each Euclidean Bruhat-Tits Building is a global NPC space. See [BH99] for details.

### 4. Barycenters

Let \((N, d)\) be a complete metric space and let \(\mathcal{P}(N)\) denote the set of all probability measures \(p\) on \((N, \mathcal{B}(N))\) with separable support \(\text{supp}(p) \subset N\). For \(1 \leq \theta < \infty\), \(\mathcal{P}_\theta(N)\) will denote the set of all \(p \in \mathcal{P}(N)\) with \(\int d^\theta(x, y)p(dy) < \infty\) for some (hence all) \(x \in N\), and \(\mathcal{P}_\infty(N)\) will denote the set of all \(p \in \mathcal{P}(N)\) with bounded support. Finally, we denote by \(\mathcal{P}_0(N)\) the set of all \(p \in \mathcal{P}(N)\) of the form \(p = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}\) with suitable \(x_i \in N\). Here and henceforth, \(\delta_x : A \mapsto 1_A(x)\) denotes the Dirac measure in the point \(x \in N\). Obviously, \(\mathcal{P}_0(N) \subset \mathcal{P}_\infty(N) \subset \mathcal{P}_\theta(N) \subset \mathcal{P}^1(N)\).

For \(q \in \mathcal{P}(N)\) the number \(\text{var}(q) := \inf_{z \in N} \int_N d^2(z, x)q(dx)\) is called variance of \(q\). Of course, \(q \in \mathcal{P}^2(N)\) if and only if \(\text{var}(q) < \infty\).

Given \(p, q \in \mathcal{P}(N)\) we say that \(\mu \in \mathcal{P}(N^2)\) is a coupling of \(p\) and \(q\) iff its marginals are \(p\) and \(q\), that is, \(\forall A \in \mathcal{B}(N)\)

\[
\mu(A \times N) = p(A) \quad \text{and} \quad \mu(N \times A) = q(A).
\]

One such coupling \(\mu\) is the product measure \(p \otimes q\).

**Definition 4.1.** We define the \((L^1-)\) Wasserstein distance or Kantorovich-Rubinstein distance \(d^W\) on \(\mathcal{P}^1(N)\) by

\[
d^W(p, q) = \inf \left\{ \int_N \int_N d(x, y)\mu(dx dy) : \mu \in \mathcal{P}(N^2) \text{ is coupling of } p \text{ and } q \right\}.
\]

Now let \((\Omega, \mathcal{A}, \mathbb{P})\) be an arbitrary probability space and \(X : \Omega \to N\) a strongly measurable map. It defines a probability measure \(X_\ast \mathbb{P} \in \mathcal{P}(N)\), called push forward (or image) measure of \(\mathbb{P}\) under \(X\), by

\[
X_\ast \mathbb{P}(A) := \mathbb{P}(X^{-1}(A)) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) \quad (\forall A \in \mathcal{B}(N)),
\]
Let $\mathbb{P}_X \in \mathcal{P}^0(N)$ if and only if $X \in L^0(\Omega, N)$, and the variance of $X$ is

$$V(X) := \inf_{z \in N} \mathbb{E} d^2(z, X) = \text{var}(\mathbb{P}_X).$$

Moreover,

$$d^W(p, q) = \inf d(X, Y),$$

where the infimum is over all probability spaces $(\Omega, \mathcal{A}, \mathbb{P})$ and all strongly measurable maps $X : \Omega \to N$ and $Y : \Omega \to N$ with distributions $\mathbb{P}_X = p$ and $\mathbb{P}_Y = q$.

**Remark 4.2.** $d^W$ is a complete metric on $\mathcal{P}^1(N)$. The space $\mathcal{P}_0(N)$ (and hence each space $\mathcal{P}^0(N)$) is dense in $\mathcal{P}^1(N)$. Cf. [Du89] or [RR98].

**Proposition 4.3.** Let $(N, d)$ be a global NPC space and fix $y \in N$. For each $q \in \mathcal{P}^1(N)$ there exists a unique point $z \in N$ which minimizes the uniformly convex, continuous function $z \mapsto \int_N [d^2(z, x) - d^2(y, x)] q(dx)$. This point is independent of $y$; it is called $q$-barycenter (or, more precisely, $d^2$-barycenter) of $q$ and denoted by

$$b(q) = \arg\min_{z \in N} \int_N [d^2(z, x) - d^2(y, x)] q(dx).$$

If $q \in \mathcal{P}^2(N)$ then $b(q) = \arg\min_{z \in N} \int_N d^2(z, x) q(dx)$.

**Proof.** Let $F_y(z) = \int [d^2(z, x) - d^2(y, x)] q(dx)$. Then $F_y(z) - F_y'(z) = \int [d^2(y', x) - d^2(y, x)] q(dx)$ is independent of $z$. Moreover, $|F_y(z)| < \infty$ since

$$|F_y(z)| = \left| \int_N [d(z, x) - d(y, x)] \cdot [d(z, x) + d(y, x)] q(dx) \right|$$

$$\leq d(z, y) \cdot \left[ \int_N d(z, x) q(dx) + \int_N d(y, x) q(dx) \right].$$

The uniform convexity of $z \mapsto d^2(z, x)$ as stated in Proposition 2.3 implies that $z \mapsto F_y(z)$ is uniformly convex: For any two points $z_0, z_1 \in N$ let $t \mapsto z_t$ denote the joining geodesic. Application of (2.3) gives

$$F_y(z_t) = \int [d^2(z_t, x) - d^2(y, x)] q(dx)$$

$$\leq (1 - t) \int [d^2(z_0, x) - d^2(y, x)] q(dx) + t \int [d^2(z_1, x) - d^2(y, x)] q(dx)$$

$$- t(1 - t) d^2(z_0, z_1)$$

$$= (1 - t) F_y(z_0) + t F_y(z_1) - t(1 - t) d^2(z_0, z_1).$$

Moreover, continuity of $z \mapsto F_y(z)$ is obvious from $|F_y(z) - F_y(z')| \leq \int_N |d^2(z, x) - d^2(z', x)| q(dx)$. According to Proposition 1.7, uniform convexity and lower semi-continuity of $F_y$ implies existence and uniqueness of a minimizer. \qed
PROPOSITION 4.4 (Variance Inequality). Let \((N, d)\) be a global NPC space. For any probability measure \(q \in \mathcal{P}^1(N)\) and for all \(z \in N\):

\[
(4.2) \quad \int_N [d^2(z, x) - d^2(b(q), x)] q(dx) \geq d^2(z, b(q)).
\]

PROOF. Given \(q\) and \(z\), apply the estimate from the previous proof with \(z_1 := z, z_0 := b(q)\) and \(y := b(q)\). The fact that \(b(q)\) is minimizer yields

\[
0 \leq F(z_t) \leq 0 + t \cdot F(z) - t(1 - t)d^2(z, b(q)).
\]

That is, for all \(t > 0\)

\[
\int_N [d^2(z, x) - d^2(b(q), x)] q(dx) \geq (1 - t)d^2(z, b(q)).
\]

For \(t \to 0\) this yields the claim. \(\square\)

REMARK 4.5. Geodesical completeness and global curvature bounds \(-\kappa^2 \leq \text{curv}(N, d) \leq 0\) imply the following reverse variance inequality: For each \(q \in \mathcal{P}^2(N)\) and for each \(z \in N\)

\[
\int [d^2(z, x) - d^2(z, b(q)) - d^2(b(q), x)] q(dx) \leq \frac{2\kappa^2}{3} \int [d^4(z, b(q)) + d^4(b(q), x)] q(dx).
\]

See \[St03\].

For \(X \in L^1(\Omega, N)\) we define its expectation by

\[
\mathbb{E} X := \arg\min_{z \in N} \mathbb{E} [d^2(z, X) - d^2(y, X)] = b(F_X).
\]

That is, \(\mathbb{E} X\) is the unique minimizer of the function \(z \mapsto \mathbb{E} [d^2(z, X) - d^2(y, X)] = \int_N [d^2(z, x) - d^2(y, x)] P_X(dx)\) on \(N\) (for each fixed \(y \in N\)). Analysts might prefer to write \(\int_N X dP\) instead of \(\mathbb{E} X\) and to call it integral of \(X\) against \(P\).

The above definition immediately implies the following transformation rule: If \(\varphi : \Omega \to \Omega'\) is a measurable map into another measurable space \((\Omega', \mathcal{A}')\) and if \(Y : \Omega' \to N\) is strongly measurable and integrable then

\[
\int_{\Omega} Y(\varphi) dP \, = \, b(F_{Y \circ \varphi}) \, = \, \int_{\Omega'} Y dP_{\varphi}.
\]

Now let us restrict to \(X \in L^2(\Omega, N)\). Then \(\mathbb{E} X\) is the unique minimizer of \(z \mapsto \mathbb{E} d^2(z, X)\).

Identifying points in \(N\) with constant maps in \(L^2(\Omega, N)\), the map \(L^2(\Omega, N) \to N, X \mapsto \mathbb{E} X\) can also be regarded as the convex projection (in the sense of Proposition 2.6) from the global NPC space \(L^2(\Omega, N)\) onto the closed convex subset of constant maps. Proposition 2.6(ii) yields another proof of the variance inequality:

\[
\mathbb{E} d^2(z, X) \geq d^2(z, \mathbb{E} X) + \mathbb{E} d^2(\mathbb{E} X, X)
\]

for all \(X \in L^2(\Omega, N)\) and \(z \in N\). In the classical case \(N = \mathbb{R}\), the corresponding equality should be well known after the first lessons in probability theory.

Our approach to barycenters, integrals and expectations is based on the classical point of view of \[G1809\]. He defined the expectation of a random variable (in Euclidean space) to be the uniquely determined point which minimizes the \(L^2\)-distance ("Methode der kleinsten Quadrate").
In the context of metric spaces, this point of view was successfully used by [Ca28], [Fr48], [Ka77], and many others, under the name of barycenter, center of mass or center of gravity.

Iterations of barycenters on Riemannian manifolds were used by [Ke90], [EM91] and [Pi94]. [Jo94] applied these concepts on global NPC spaces.

Another natural way to define the ”expectation” $E Y$ of a random variable $Y$ is to use (generalizations of) the law of large numbers. This requires to give a meaning to $\frac{1}{n} \sum_{i=1}^{n} Y_i$. Our definition below only uses the fact that any two points in $N$ are joined by unique geodesics. Our law of large numbers for global NPC spaces gives convergence towards the expectation defined as minimizer of the $L^2$ distance.

**Definition 4.6.** Given any sequence $(y_i)_{i \in \mathbb{N}}$ of points in $N$ we define a new sequence $(s_n)_{n \in \mathbb{N}}$ of points $s_n \in N$ by induction on $n$ as follows:

\[ s_1 := y_1 \quad \text{and} \quad s_n := \left( 1 - \frac{1}{n} \right) s_{n-1} + \frac{1}{n} y_n, \]

where the RHS should denote the point $\gamma_{1/n}$ on the geodesic $\gamma : [0, 1] \to N$ connecting $\gamma_0 = s_{n-1}$ and $\gamma_1 = y_n$. The point $s_n$ will be denoted by $\frac{1}{n} \sum_{i=1,\ldots,n} y_i$ and called *inductive mean value* of $y_1, \ldots, y_n$.

Note that in general the point $\frac{1}{n} \sum_{i=1,\ldots,n} y_i$ will strongly depend on permutations of the $y_i$.

**Theorem 4.7 (Law of Large Numbers).** Let $(Y_i)_{i \in \mathbb{N}}$ be a sequence of independent, identically distributed random variables $Y_i \in L^2(\Omega, N)$ on a probability space $(\Omega, \mathcal{A}, P)$ with values in a global NPC space $(N, d)$. Then

\[ \frac{1}{n} \sum_{i=1,\ldots,n} Y_i \to E Y_1 \quad \text{for} \ n \to \infty \]

in $L^2(\Omega, N)$ and in probability ("weak law of large numbers").

If moreover $Y_i \in L^\infty(\Omega, N)$ then for $P$-almost every $\omega \in \Omega$

\[ \frac{1}{n} \sum_{i=1,\ldots,n} Y_i(\omega) \to E Y_1 \quad \text{for} \ n \to \infty \]

("strong law of large numbers").

**Remark 4.8.** (i) In strong contrast to the linear case, the inductive mean value $S_n = \frac{1}{n} \sum_{i=1,\ldots,n} Y_i$ will in general strongly depend on permutations of the iid variables $Y_i, i = 1, \ldots, n$. The distribution $P_{S_n}$ is of course invariant under such permutations. But even $E S_n$ in general depends on $n \in \mathbb{N}$. The law of large numbers only yields that $E S_1 = \lim_{n \to \infty} E S_n$.

(ii) It might seem more natural to define the mean value of the random variables $Y_1, \ldots, Y_n$ as the barycenter of these points, more precisely, as the barycenter of the uniform distribution on these points, i.e.

\[ \mathbb{S}_n(\omega) := b \left( \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i(\omega)} \right). \]
In this case we also obtain a law of large numbers. Indeed, it is much easier to derive (and it holds for more or less arbitrary choices of \( b(.) \)), see Proposition 6.6. However, it is also of much less interest: we will obtain convergence of \( \bar{S}_n(\omega) \) towards \( b(\mathbb{P}_{Y_1}) \), the barycenter of the distribution of \( Y_1 \), but to define \( \bar{S}_n \) we already have to know how \( b(.) \) acts on discrete uniform distributions.

(iii) Of course there are many other ways to define a mean value \( \tilde{S}_n \) of the random variables \( Y_1, \ldots, Y_n \) which do not depend on the a priori knowledge of \( b(.) \). And indeed for many of these choices one can prove that \( \tilde{S}_n \) converges almost surely to a point \( \tilde{b} \) which only depends on the distribution of \( Y_1 \). For instance, define \( S_n, 1 := Y_n \) and recursively \( S_{n,k+1} \) to be the midpoint of \( S_{2n-1,k} \) and \( S_{2n,k} \). Then \( \tilde{S}_n(\omega) := S_{1,n}(\omega) \) converges for a.e. \( \omega \) as \( n \to \infty \) towards a point \( \tilde{b} = b(\mathbb{P}_{Y_1}) \). (Note that in the flat case, \( \tilde{S}_n = 2^{-n} \sum_{i=1}^{2^n} Y_i \).) Another example is given by the mean value in the sense of \([ESH99]\) which will be described in Remark 6.4.

However, no choice of \( \tilde{S}_n \) other than \( S_n \) is known to the author where one obtains convergence towards a point which can be characterized "extrinsically", like in our case as the minimizer of the function \( z \mapsto \mathbb{E} d^2(z, Y_1) \).

**Proof.** (a) Our first claim is that \( \forall n \in \mathbb{N} : \)

\[
\mathbb{E} d^2(\mathbb{E} Y_1, S_n) \leq \frac{1}{n} \mathbb{V}(Y_1)
\]

This is obviously true for \( n = 1 \). We will prove it for all \( n \in \mathbb{N} \) by induction. Assuming that it holds for \( n \) we conclude (using inequalities (2.3) and (4.2) from Proposition 2.3 and Corollary 4.4)

\[
\mathbb{E} d^2(\mathbb{E} Y_1, S_{n+1})
\]

\[
= \mathbb{E} d^2\left( \mathbb{E} Y_1, \frac{n}{n+1} S_n + \frac{1}{n+1} Y_{n+1} \right)
\]

\[
\overset{(2.3)}{\leq} \frac{n}{n+1} \mathbb{E} d^2(\mathbb{E} Y_1, S_n) + \frac{1}{n+1} \mathbb{E} d^2(\mathbb{E} Y_1, Y_{n+1}) - \frac{n}{(n+1)^2} \mathbb{E} d^2(S_n, Y_{n+1})
\]

\[
\overset{(4.2)}{\leq} \frac{n}{n+1} \mathbb{E} d^2(\mathbb{E} Y_1, S_n) + \frac{1}{n+1} \mathbb{E} d^2(\mathbb{E} Y_1, Y_{n+1})
\]

\[
- \frac{n}{(n+1)^2} \left[ \mathbb{E} d^2(\mathbb{E} Y_{n+1}, S_n) + \mathbb{E} d^2(\mathbb{E} Y_{n+1}, Y_{n+1}) \right]
\]

\[
= \left( \frac{n}{n+1} \right)^2 \mathbb{E} d^2(\mathbb{E} Y_1, S_n) + \frac{1}{(n+1)^2} \mathbb{V}(Y_1)
\]

\[
\leq \frac{1}{n+1} \mathbb{V}(Y_1).
\]

This proves the first claim. And of course it also proves the \( L^2 \)-convergence as well as the weak law of large numbers

\[
S_n \to \mathbb{E} Y_1 \quad \text{in probability}
\]

as \( n \to \infty \), i.e. for all \( \varepsilon > 0 \)

\[
\mathbb{P}(d(S_n, \mathbb{E} Y_1) > \varepsilon) \to 0
\]

as \( n \to \infty \).

(b) Our second claim is that

\[
S_{n^2} \to \mathbb{E} Y_1
\]
a.s. for \( n \to \infty \). Indeed, by (a)
\[
\sum_{n=1}^{\infty} \mathbb{P}(d(S_n, Y_1) > \varepsilon) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 N} \mathbb{E}d^2(S_n, Y_1) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 n^2} \mathbb{V}(Y_1) < \infty.
\]
Due to the Borel-Cantelli lemma, this implies the second claim.

Now assume that \( Y_1 \in L^\infty(\Omega, N) \), say \( d(Y_1, z) \leq R \) a.s. for some \( z \in N \) and some \( R \in \mathbb{R} \). Then by convexity \( d(S_n, z) \leq R \) a.s. for all \( n \in \mathbb{N} \) and
\[
d(S_n, S_{n+1}) \leq \frac{1}{n+1} d(S_n, Y_{n+1}) \leq \frac{2}{n+1} R
\]
a.s. Therefore, for all \( k, n \in \mathbb{N} \) with \( n^2 \leq k < (n+1)^2 \)
\[
d(S_k, S_n) \leq \left( \frac{1}{n^2 + 1} + \frac{1}{n^2 + 2} + \ldots + \frac{1}{k} \right) 2R \leq \frac{k - n^2}{n^2} 2R \leq \frac{4}{n} R
\]
a.s. Together with the second claim, this proves the strong law of large numbers. \( \square \)

Finally, we will give various characterizations of nonpositive curvature in terms of properties of probability measures on the spaces. For instance, the validity of a variance inequality turns out to characterize NPC spaces. Similarly, an inequality between two kind of variances as well as a weighted quadruple inequality.

**Theorem 4.9.** Let \((N, d)\) be a complete metric space. Then the following properties are equivalent:

(i) \((N, d)\) is a global NPC space

(ii) For any probability measure \( q \in \mathcal{P}^2(N) \) there exists a point \( z_q \in N \) such that for all \( z \in N \)
\[
\int_N d^2(z, x)q(dx) \geq d^2(z, z_q) + \int_N d^2(z_q, x)q(dx).
\]

(iii) For any probability measure \( q \in \mathcal{P}(N) \)
\[
\text{var}(q) \leq \frac{1}{2} \int_N \int_N d^2(x, y)q(dx)q(dy).
\]

(iv) \((N, d)\) is a length space with the property that for any \( x_1, x_2, x_3, x_4 \in N \) and \( s, t \in [0, 1] \)
\[
s(1-s)d^2(x_1, x_3) + t(1-t)d^2(x_2, x_4)
\leq
std^2(x_1, x_2) + (1-s)td^2(x_2, x_3) + (1-t)(1-s)d^2(x_3, x_4) + s(1-t)d^2(x_4, x_1).
\]

The proof will show that in (iii) it suffices to consider probability measures \( q \) which are supported by four points and in (iv) it suffices to consider \( t = \frac{1}{2} \).

**Proof.** "(i) \(\Rightarrow\) (ii)" : Corollary 4.4

"(ii) \(\Rightarrow\) (i)" : Given points \( \gamma_0, \gamma_1 \in N \) and \( t \in [0, 1] \), choose the probability measure \( q = (1-t)\delta_{\gamma_0} + t\delta_{\gamma_1} \) and denote the point \( z_q \) by \( \gamma_t \). Then (ii) implies for all \( z \in N \)
\[
(1-t)d^2(z, \gamma_0) + td^2(z, \gamma_1) \geq (1-t)d^2(\gamma_t, \gamma_0) + td^2(\gamma_t, \gamma_1) + d^2(\gamma_t, z)
\geq (1-t)d^2(\gamma_0, \gamma_1) + d^2(\gamma_t, z),
\]
where the last inequality is a simple consequence of the triangle inequality. This proves (i).
Then for each $N$ in $\mathbb{R}$.

Since this holds for any $\gamma$ where again the last inequality is a simple consequence of the triangle inequality. According to Remark 1.3, this already implies that $(\gamma, \gamma)$ yields $(\gamma, \gamma)$. To see the second claim, choose $\epsilon > 0$.

Choose the probability measure $q = \frac{1}{2}d_{\gamma_0} + \frac{1}{2}d_{\gamma_1}$. Then (ii) implies that there exists a point $z \in N$ with

$$d^2(z, \gamma_0) + d^2(z, \gamma_1) \leq \frac{1}{2}d^2(\gamma_0, \gamma_1) + \epsilon.$$ 

According to Remark 1.3, this already implies that $(N, d)$ is a length space. To see the second claim, choose $q = \frac{1}{2}[s\delta_{x_1} + t\delta_{x_2} + (1-s)\delta_{x_3} + (1-t)\delta_{x_4}]$. Then for each $\epsilon > 0$, (iii) implies that for suitable $z \in N$

$$\frac{1}{8}[st^2(x_1, x_2) + (1-s)t^2(x_2, x_3) + (1-s)(1-t)d^2(x_3, x_4)$$

$$+ s(1-t)d^2(x_4, x_1) + s(1-s)d^2(x_1, x_3) + t(1-t)d^2(x_2, x_4)] + \epsilon$$

$$\geq \frac{1}{4}[sd^2(z, x_1) + td^2(z, x_2) + (1-s)d^2(z, x_3) + (1-t)d^2(z, x_4)]$$

$$\geq \frac{1}{4}[s(1-s)d^2(x_1, x_3) + t(1-t)d^2(x_2, x_4)],$$

where again the last inequality is a simple consequence of the triangle inequality. Since this holds for any $\epsilon > 0$ it proves the claim.

"(iv) $\Longrightarrow$ (i)"; The fact that $(N, d)$ is a length space implies that, given $\gamma_0, \gamma_1 \in N$ and $s > 0$, there exists $y \in N$ such that

$$d^2(\gamma_0, y) + d^2(\gamma_1, y) \leq \frac{1}{2}d^2(\gamma_0, \gamma_1) + s^2.$$ 

For arbitrary $z \in N$, apply (iv) to $x_1 = z, x_2 = \gamma_1, x_3 = y, x_4 = \gamma_0$ and $t = \frac{1}{2}$. It yields

$$s(1-s)d^2(z, \gamma_1)$$

$$\leq \frac{s}{2}d^2(z, \gamma_1) + \frac{s}{2}d^2(z, \gamma_0) + \frac{1}{2}d^2(\gamma_1, \gamma_0) + \frac{1-s}{2}d^2(y, \gamma_1) + \frac{1-s}{2}d^2(y, \gamma_0) - \frac{1}{4}d^2(\gamma_0, \gamma_1)$$

$$\leq \frac{s}{2}d^2(z, \gamma_1) + \frac{s}{2}d^2(z, \gamma_0) - \frac{s}{4}d^2(\gamma_0, \gamma_1) + \frac{s}{2}(1-s).$$

\[\text{Figure 6. Weighted quadruple comparison}\]
Dividing by $s$ and then letting $s \to 0$ this (together with Remark 2.2) yields the claim. \hfill \Box

5. Identification of Barycenters

**Lemma 5.1 (Images).** Let $\psi : N \to N'$ be an isometry between global NPC spaces $(N,d)$ and $(N',d')$. Then for each $q \in \mathcal{P}^1(N)$, the push forward $\psi_* q$ belongs to $\mathcal{P}^1(N')$ and

$$\psi_* : \mathcal{P}^1(N) \to \mathcal{P}^1(N')$$

is an isometry between the spaces of probability measures equipped with the Wasserstein distances. Moreover,

$$b(\psi_* q) = \psi(b(q)).$$

This can be interpreted as equivariance of $b$ with respect to isometries.

**Example 5.2.** Let $\gamma : I \to N$ be a geodesic and let $q \in \mathcal{P}^1(N)$ with $\text{supp}(q) \subset \gamma(I)$. Then

$$b(q) = \gamma(b(p)) = \gamma \left( \int_N \gamma^{-1}(y) q(dy) \right),$$

where $b(p) = \int_N x p(dx) = \int_N \gamma^{-1}(y) q(dy)$ is the barycenter (= usual mean value) of the probability measure $p = (\gamma^{-1})_* q$ on $I \subset \mathbb{R}$.

**Proof.** Each geodesic $\gamma$ is an isometry between $I \subset \mathbb{R}$ and $\gamma(I) \subset N$. Hence, applying the previous result to $\psi = \gamma^{-1}$ yields the claim. \hfill \Box

**Example 5.3.** Let $\psi : N \to \mathbb{R}$ be a bijective map from an arbitrary set $N$ onto a closed interval $\psi(N) \subset \mathbb{R}$ and define a metric $d$ on $N$ by $d(x,y) := |\psi(x) - \psi(y)|$. Then for each $q \in \mathcal{P}^1(N)$

$$b(q) = \psi^{-1} \left( \int_{\mathbb{R}} \psi(x) q(dx) \right).$$

For an $L^1$-random variable $X : \Omega \to N$, the above means that $\mathbb{E} X = \psi^{-1} \mathbb{E} \psi(X)$ or, more suggestively,

$$\psi(\mathbb{E} X) = \mathbb{E} \psi(X).$$

In typical applications, $N$ itself is a closed interval $J \subset \mathbb{R}$ and $\psi$ is a strictly increasing and continuous function. Probabilists then might call it scale function.

**Proof.** By the above Lemma $b(q) = \psi^{-1}(b(q'))$ where $q' = \psi_* q \in \mathcal{P}^1(\mathbb{R})$ and thus $b(q') = \int_{\mathbb{R}} \psi(x) q(dx)$. \hfill \Box

**Proposition 5.4 (Hilbert spaces).** If $N$ is a Hilbert space then for each $q \in \mathcal{P}^1(N)$

$$b(q) = \int_N x q(dx)$$

in the sense that

$$\langle b(q), y \rangle = \int_N \langle x, y \rangle q(dx) \quad (\forall y \in N).$$
Note that this identity is true only for probability measures. Namely, let \( m \) be a measure on \((N, \mathcal{B}(N))\) with \( 0 < m(N) < \infty \). Then the barycenter \( b(m) \) of \( m \) can be defined as before by

\[
b(m) = \frac{1}{m(N)} \int_N x \, m(dx).
\]

**Proof.** Recall that \( b(q) \) is the unique minimizer of \( F : z \mapsto \int_N [d^2(z, x) - d^2(0, x)] q(dx) = \int_N ||z - x||^2 - ||x||^2 q(dx) \). Hence, \( z = b(q) \) if and only if

\[
\frac{d}{d\epsilon} F(z + \epsilon y)|_{\epsilon=0} = 2 \int_N \langle y, z - x \rangle q(dx) \uparrow = 0
\]

for all \( y \in N \).

Recall that every separable Hilbert space is either isomorphic to some Euclidean space \( \mathbb{R}^k \) or to the space \( l^2 \). In other words, it is isomorphic to \( \bigotimes_{i \in K} \mathbb{R} \) with a finite or countable set \( K \). By the preceding \( b(q) = (b(q_i))_{i \in K} \) with \( b(q_i) = \int_N x_i q_i(dx) = \int_N \sum_{i=1}^N x_i q_i(dx) \) where \( x_i \) and \( q_i \) denote the projection of \( x \) and \( q \), resp., onto the \( i \)-th factor of \( N \).

Let \( N = \bigotimes_{i=1}^k N_i \) with global NPC spaces \( N_1, \ldots, N_k \) and let \( q = q_1 \otimes \ldots \otimes q_k \) with \( q_i \in \mathcal{P}^1(N_i) \) for each \( i = 1, \ldots, k \). Then our next result states that \( b(q) = (b(q_1), \ldots, b(q_k)) \). Actually, the latter holds true not only for product measures but for arbitrary \( q \in \mathcal{P}^1(N) \) if we define \( q_i \in \mathcal{P}^1(N_i) \), for \( i = 1, \ldots, k \), to be the \( i \)-th marginals of \( q \) or projections of \( q \) onto the \( i \)-th factor, that is,

\[
q_i(A) = q(N_1 \times \ldots \times N_{i-1} \times A \times N_{i+1} \times \ldots \times N_k)
\]

for all \( A \in \mathcal{B}(N_i) \). And it holds true not only for a finite number of factors \( N_i \) but for an arbitrary family.

**Proposition 5.5** (Products). *Given an arbitrary set \( K \), let

\[
N = \left\{ x \in \bigotimes_{i \in K} N_i : \sum_{i \in K} m_i \cdot d^2_i(x_i, o_i) < \infty \right\}
\]

with global NPC spaces \( (N_i, d_i) \), weights \( m_i \in [0, \infty] \) and base points \( o_i \in N_i, i \in K \). Let \( q \in \mathcal{P}^2(N) \) with marginals \( q_i \in \mathcal{P}^2(N_i) \). Then \( b(q) = (b(q_i))_{i \in K} \).

**Proof.** Put \( y = (b(q_i))_{i \in K} \). Then by definition, \( b(q) \) is the minimizer of

\[
z \mapsto \int_N [d^2(z, x) - d^2(y, x)] q(dx) = \int_N \sum_{i \in K} m_i \cdot [d^2_i(z_i, x_i) - d^2_i(b(q_i), x_i)] q(dx)
\]

\[
= \sum_{i \in K} m_i \cdot \int_{N_i} [d^2_i(z_i, x_i) - d^2_i(b(q_i), x_i)] q_i(dx_i)
\]

\[
\geq \sum_{i \in K} m_i \cdot d^2_i(z_i, b(q_i))
\]

\[
= d^2(z, y),
\]

where the inequality follows from the variance inequality for each \( q_i \). Hence, \( b(q) = y \). \( \square \)
Proposition 5.6 \((L^2\)-spaces). Let \((M, M, m)\) be a measure space, \((N, d)\) a global NPC space and \(\hat{N} = L^2(M, N, h)\) for some \(h: M \to N\), equipped with the \(L^2\)-distance \(\hat{d}\). If \(\hat{q} \in \mathcal{P}^2(\hat{N})\) with marginals \(q(x, .) \in \mathcal{P}^2(N)\) then \(b(\hat{q}) : x \mapsto b(q(x, .))\).

Proof. Assume without restriction that \(d\) and \(\hat{d}\) are bounded. Recall that \(q(x, .) \in \mathcal{P}(N)\), \(x \in M\), are called marginals of \(\hat{q}\) \(\in \mathcal{P}(\hat{N})\) iff for each \(x \in M\) and each bounded measurable \(u : N \to \mathbb{R}\)
\[
\int_{\hat{N}} u(g(x)) \hat{q}(dg) = \int_{N} u(z) q(x, dz).
\]

Choose \(h \in \hat{N}\) with \(h(x) = b(q(x, .))\). Then \(b(\hat{q})\) is the minimizer of
\[
f \mapsto \int_{\hat{N}} [\hat{d}^2(f, g) - \hat{d}^2(h, g)] \hat{q}(dg)
\]
\[
= \int_{\hat{N}} \left( \int_{M} \left( d^2(f, g(x)) - d^2(h(x), g(x)) \right) m(dx) \right) \hat{q}(dg)
\]
\[
= \int_{M} \int_{\hat{N}} \left( d^2(f, g(x)) - d^2(h(x), g(x)) \right) \hat{q}(dg) m(dx)
\]
\[
= \int_{M} \int_{\hat{N}} \left( d^2(f, z) - d^2(h(x), z) \right) q(x, dz) m(dx)
\]
\[
\geq \int_{M} d^2(f, h(x)) m(dx) = \hat{d}^2(f, h),
\]
where we have used the variance inequality for each \(q(x, .)\). \(\square\)

Before studying arbitrary trees, we will have a look on spiders. Let \(K\) be an arbitrary set and \(N\) be the corresponding \(K\)-spider. Given \(q \in \mathcal{P}^1(N)\) we define numbers
\[
r_i(q) := \int_{N_i} d(o, x) q(dx), \quad b_i(q) := r_i(q) - \sum_{j \neq i} r_j(q)
\]
for \(i \in K\). (The point \(b_i(q)\) is the usual mean value of the image of \(q\) on \(\mathbb{R}\) if \(N_i\) is identified with \(\mathbb{R}_+\) and all the other \(N_j\) are glued together and identified with \(\mathbb{R}_-\).

Note that \(b_i(q) > 0\) for at most one \(i \in K\).

Proposition 5.7 (Spiders). If \(b_i(q) > 0\) for some \(i \in K\) then \(b(q) = (i, b_i(q))\).
Otherwise, \(b(q) = o\).

Proof. Fix \(q\) and \(i\). If \(b(q) = (i, r_0)\) for some \(r_0 > 0\) then \(r \mapsto F(r)\), where
\[
F(r) := \int_{N} d^2((i, r), x) q(dx) = \int_{N_i} (r - d(o, x))^2 q(dx) + \sum_{j \neq i} \int_{N_j} (r + d(o, x))^2 q(dx),
\]
attains its minimum on \([0, \infty]\) in \(r = r_0\). The latter implies
\[
0 = \frac{1}{2} F'(r_0) = \int_{N_i} (r_0 - d(o, x)) q(dx) + \sum_{j \neq i} \int_{N_j} (r_0 + d(o, x)) q(dx)
\]
\[
= r_0 - r_i(q) + \sum_{j \neq i} r_j(q) = r_0 - b_i(q)
\]
and thus \(r_0 = b_i(q)\). Similarly, \(b(q) = o\) implies \(F'(0) \geq 0\) and thus \(0 \geq b_i(q)\). \(\square\)
The barycentric midpoint (for \(i \lambda\) numbers \(z\))

Then by the same arguments as before we conclude

\[ b(p) = b(\overline{p}). \]

Indeed, with the notations from above,

\[ r_i(\overline{p}) = \int_{N_i} d(o,x) \overline{p}(dx) = \lambda_i \cdot d(o,b(p_i)) = \lambda_i \int_{N_i} d(o,x) p_i(dx) = r_i(p) \]

for each \(i\) and hence the claim follows. Here the crucial point is that each \(p_i\) is supported by a flat space \(N_i\).

(ii) In general, given \(p_1, \ldots, p_k \in P^1(N)\) on some metric space \((N,d)\) and numbers \(\lambda_1, \ldots, \lambda_k \in \mathbb{R}_+\) with \(\sum_{i=1}^k \lambda_i = 1\) one might ask whether the barycenter of \(p := \sum_{i=1}^k \lambda_i \cdot p_i\) coincides with the barycenter of \(\overline{p} = \sum_{i=1}^k \lambda_i \cdot \delta_{b(p_i)}\) where each \(p_i\) is replaced by a Dirac mass with the same barycenter.

Or (more or less equivalently) whether the barycentric midpoint of points \(x_1, \ldots, x_{kn} \in N\) coincides with the barycentric midpoint of \(z_1, \ldots, z_k\) where \(z_i\) (for \(i = 1, \ldots, k\)) is the barycentric midpoint of \(x_{(i-1)n+1}, \ldots, x_{in}\).

Even for \(n = k = 2\) this is not true in general. For instance, let \((N,d)\) be the tripod and let \(x_i = (i,1)\) for \(i = 1, 2, 3\) and \(x_4 = o\). Then obviously (e.g. by symmetry arguments) the barycentric midpoint of \(x_1, \ldots, x_4\) is \(o\). Also the (barycentric) midpoint \(z_1\) of \(x_1\) and \(x_2\) is \(o\) whereas the (barycentric) midpoint \(z_2\) of \(x_3\) and \(x_4\) is \((3, \frac{1}{2})\). Hence, the midpoint of \(z_1\) and \(z_2\) is \((3, \frac{1}{2}) \neq o\).

A noteworthy exception is the above Remark (i) and the Law of Large Numbers.

Now let \((N,d)\) be a discrete metric graph and fix \(z \in N\). The set \(N \setminus \{z\}\) decomposes into a (finite or infinite) disjoint family \(K_z\) of connected components \(K_{z,i}, i \in K_z\), and for sufficiently small \(\epsilon > 0\) each \(K_{z,i} \cap B_r(z) \subset N\) is isometric to the interval \([0, \epsilon] \subset \mathbb{R}\). (In other words, \(B_r(z) \subset N\) is isometric to the \(\epsilon\)-ball in the \(K_z\)-spider.) For each \(i \in K_z\) define

\[ r_{z,i}(q) := \int_{N_{z,i}} d(z,x) q(dx), \quad b_{z,i}(q) := r_{z,i}(q) - \sum_{j \in K_z, j \neq i} r_{z,j}(q). \]

Then by the same arguments as before we conclude

**Proposition 5.9.** (i) For each \(z \in N\)

\[ b(q) = z \iff b_{z,i}(q) \leq 0 \quad (\forall i \in K_z). \]

(ii) If \(z\) lies on an edge then \(K_z = \{1, 2\}\) and the previous condition simplifies to

\[ b(q) = z \iff b_{z,1}(q) = 0. \]

These results can be used to identify barycenters on higher dimensional buildings. For instance, let \((N,d) = N_0 \times \mathbb{R}^{n-1}\) be a booklet where \(N_0\) is a \(k\)-spider. Then for each \(q \in P^1(N)\) the barycenter is given by \(b(q) = (b(q_0), b(q_1))\) where \(q_0\) and \(q_1\) are the projections of \(q\) onto \(N_0\) and \(\mathbb{R}^{n-1}\), resp.

Finally, let us consider the classical case of Riemannian manifolds. A simple variational argument yields
Proposition 5.10. Let \((N,q)\) be a complete, simply connected Riemannian manifold with nonpositive curvature and let \(q \in P^1(N)\). Then \(z = b(q)\) if and only if
\[
\int_N d(z, x) \grad_1 d(z, x) q(dx) = 0,
\]
where \(\grad_1 d(z, x) \in T_z N\) is the gradient of \(d(z, x)\) with respect to the first variable.

6. Jensen’s Inequality and \(L^1\) Contraction Property

Throughout this section \((N,d)\) will always be a global NPC space.

Proposition 6.1. If a probability measure \(q \in P^1(N)\) is supported by a convex closed set \(K \subset N\) then its barycenter \(b(q)\) lies in \(K\). In particular, if \(\text{supp}(q) \subset B_r(x)\) then \(b(q) \in B_r(x)\).

Proof. Assume \(b(q) \notin K\). Then by Proposition 2.6
\[
\int [d^2(b(q), x) - d^2(y, x)] q(dx) \geq \int [d^2(\pi_K(b(q)), x) - d^2(y, x)] q(dx)
\]
which contradicts the minimizing property of \(b(q)\). \(\square\)

Theorem 6.2 (Jensen’s inequality). For any lower semicontinuous convex function \(\varphi : N \to \mathbb{R}\) and any \(q \in P^1(N)\)
\[
\varphi(b(q)) \leq \int_N \varphi(x) q(dx),
\]
provided the RHS is well-defined.

Let us mention that the above RHS is well-defined if either \(\int \varphi^+ dq < \infty\) or \(\int \varphi^- dq < \infty\). In particular, it is well-defined if \(\varphi\) is Lipschitz continuous.

If \(\int \varphi dq\) is well-defined then in Jensen’s inequality we may assume without restriction that \(\varphi\) is bounded from below and \(\int |\varphi| dq < \infty\). Indeed, the assumption implies that \(\int \varphi dq = \lim_{k \to \infty} \int \varphi_k dq\) with \(\varphi_k := \varphi \vee (-k)\) being bounded from below and convex. Moreover, \(\int \varphi^+ dq = \infty\) would imply \(\int \varphi dq = \infty\) in which case Jensen’s inequality is trivially true.

We will present two entirely different, elementary proofs.

First Proof following [EF01]. Given \(\varphi\) and \(q\) as above, let \(\tilde{N} = N \times \mathbb{R}\) and \(N_{\varphi} = \{(x,t) \in \tilde{N} : \varphi(x) \leq t\}\) which is a closed convex subset of the global NPC space \(\tilde{N}\).

Put \(\tilde{\varphi} : N \to \tilde{N}, x \mapsto (x,\varphi(x))\) and let \(\tilde{q} = q \circ \tilde{\varphi}^{-1}\) be the image of the probability measure \(q\) under the map \(\tilde{\varphi}\).

Without restriction, we may assume \(\int_{\tilde{N}} |\varphi(x)| q(dx) < \infty\). Then \(\tilde{\varphi} \in P^1(\tilde{N})\) since for \(\tilde{z} = (z,t) \in \tilde{N}\)
\[
\int_{\tilde{N}} d(\tilde{z}, \tilde{x}) \tilde{q}(d\tilde{x}) \leq \int_N [d(z, x) + |t - \varphi(x)|] q(dx) < \infty.
\]
According to Proposition 5.5
\[ b(\hat{q}) = \left( b(q), \int_N \varphi(x)q(dx) \right). \]
Moreover, supp(\(\hat{q}\)) \(\in N\); hence, by Proposition 6.1 \(b(\hat{q}) \in N\). That is, \(\varphi(b(q)) \leq \int_N \varphi(x)q(dx)\).

**Second Proof.** Now for simplicity assume \(q \in P_2(N)\) and \(\int_N \varphi^2(x)q(dx) < \infty\). The general case follows by an approximation argument. Choose a probability space \((\Omega, \mathcal{A}, \mathbb{P})\) and an iid sequence \((Y_i)\) of random variables \(Y_i : \Omega \to N\) with distribution \(P_{Y_i} = q\). Put
\[ Z_i = \varphi(Y_i), \quad S_n := \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{and} \quad T_n := \frac{1}{n} \sum_{i=1}^n Z_i. \]
Then by the weak law of large numbers (for \(N\)-valued and for \(\mathbb{R}\)-valued random variables, resp.)
\[ S_n \to EY_1 = b(q), \quad T_n \to E\varphi(Y_1) = \int \varphi dq \]
in probability. Moreover, we claim that
\[ \varphi(S_n) \leq T_n. \]
Indeed, this is true for \(n = 1\) and follows for general \(n\) by induction:
\[ \varphi(S_{n+1}) = \varphi\left( \frac{n}{n+1}S_n + \frac{1}{n+1}Y_{n+1} \right) \]
\[ \leq \frac{n}{n+1} \varphi(S_n) + \frac{1}{n+1} \varphi(Y_{n+1}) \]
\[ \leq \frac{n}{n+1} T_n + \frac{1}{n+1} Z_{n+1} = T_{n+1}, \]
where we only used the convexity of \(\varphi\) along geodesics. Hence, by lower semicontinuity of \(\varphi\)
\[ \varphi(b(q)) \leq \liminf_{n \to \infty} \varphi(S_n) \leq \liminf_{n \to \infty} T_n = \int \varphi dq. \]

**Theorem 6.3 (Fundamental Contraction Property).** \(\forall p, q \in P^1(N)\)
\[ (6.1) \quad d(b(p), b(q)) \leq d^W(p, q). \]

**Proof.** Given \(p, q \in P^1(N)\) consider \(\mu \in P^1(N^2)\) with marginals \(p\) and \(q\). Then \(b(\mu) = (b(p), b(q))\). Thus Jensen’s inequality with the convex function \(d : N^2 \to \mathbb{R}\) yields
\[ d(b(p), b(q)) = d(b(\mu)) \leq \int_{N^2} d(x)\mu(dx). \]
Therefore, \(d(b(p), b(q)) \leq d^W(p, q)\).

**Remark 6.4.** Now let \((N, d)\) be an arbitrary complete metric space. A contracting barycenter map is a map \(\beta : P^1(N) \to N\) such that
\[\beta(\delta_x) = x \text{ for all } x \in N;\]
\[d(\beta(p), \beta(q)) \leq d^W(p, q) \text{ for all } p, q \in \mathcal{P}^1(N).\]

(i) If there exists a contracting barycenter map on \((N, d)\) then \((N, d)\) is a geodesic space: For each pair of points \(x_0, x_1 \in N\) we can define one geodesic \(t \mapsto x_t\) connecting \(x_0\) and \(x_1\) by
\[x_t := \beta((1-t)\delta_{x_0} + t\delta_{x_1}).\]

Given any for points \(x_0, x_1, y_0, y_1 \in N\), the function \(t \mapsto d(x_t, y_t)\) is convex. In particular, the geodesic \(t \mapsto x_t\) depends continuously on \(x_0\) and \(x_1\). However, it is not necessarily the only geodesic connecting \(x_0\) and \(x_1\).

If geodesics in \(N\) are unique then the existence of a contracting barycenter map implies that \(d : N \times N \to \mathbb{R}\) is convex. Thus \(N\) has (globally) "nonpositive curvature" in the sense of Busemann.

(ii) A complete, simply connected Riemannian manifold \((N, d)\) admits a contracting barycenter map \(\beta\) if and only if it has nonpositive sectional curvature. Indeed, if \((N, d)\) admits a contracting barycenter map then so does \((N_0, d)\) for each closed convex \(N_0 \subset N\). Choosing \(N_0\) sufficiently small, geodesics in \(N_0\) are unique and thus \(t \mapsto d(\gamma_t, \eta_t)\) is convex for any pair of geodesics \(\gamma\) and \(\eta\) in \(N_0\). This implies that \(N\) has nonpositive curvature.

Conversely, if \(N\) has nonpositive curvature then it admits a barycenter contraction by the next Remark (iii).

(iii) If \((N, d)\) is a global NPC space then the \(d^2\)-barycenter \(b\) defines such a contracting barycenter map. We emphasize, however, that for a given global NPC space there may exist contracting barycenter maps \(\beta : \mathcal{P}^1(N) \to N\) different from \(b\).

(iv) Each contracting barycenter map \(\beta\) on a complete metric space \((N, d)\) gives rise to a whole family of contracting barycenter maps \(\beta_n, n \in \mathbb{N}\) (which in general do not coincide with \(\beta\)).

More precisely: Let \((N, d)\) be a complete metric space with a contracting barycenter map \(\beta : \mathcal{P}^1(N) \to N\) and let \(\Phi : N \times N \to N\) be the midpoint map induced by \(\beta\), i.e. \(\Phi(x, y) = \beta(\frac{1}{2}\delta_x + \frac{1}{2}\delta_y)\). Define a map \(\Xi : \mathcal{P}^1(N) \to \mathcal{P}^1(N)\) by
\[\Xi(q) := \Phi_* (q \otimes q).\]

Then \(\Xi\) is a contraction with respect to \(d^W\). Thus for each \(n \in \mathbb{N}\)
\[\beta_n(q) := \beta(\Xi^n(q))\]
defines a contracting barycenter map \(\beta_n : \mathcal{P}^1(N) \to N\).

**Example 6.5 (Barycenter Map of Es-Sahib & Heinich).** Let \((N, d)\) be a locally compact, global NPC space. Then one can define recursively for each \(n \in \mathbb{N}\) a unique map \(\beta_n : N^n \to N\) satisfying
\[\beta_n(x_1, \ldots, x_1) = x_1\]
\[d(\beta_n(x_1, \ldots, x_n), \beta_n(y_1, \ldots, y_n)) \leq \frac{1}{n} \sum_{i=1}^n d(x_i, y_i)\]
\[\beta_n(x_1, \ldots, x_n) = \beta_n(x_1, \ldots, \bar{x}_n)\text{ where } \bar{x}_i := \beta_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).\]

This map is symmetric (= invariant under permutation of coordinates) and satisfies \(d(z, \beta_n(x_1, \ldots, x_n)) \leq \frac{1}{n} \sum_{i=1}^n d(z, x_i)\) for all \(z \in N\).

Given any \(p \in \mathcal{P}^1(N)\) let \((Y_i)\) be an independent sequence of maps \(Y_i : \Omega \to N\) on some probability space \((\Omega, \mathcal{A}, \mathbb{P})\) with distribution \(\mathbb{P}_{Y_i} = p\) and define \(\tilde{S}_n(\omega) :=\)
Then there exists a point $\beta(p) \in N$ such that
\[ \tilde{S}_n(\omega) \to \beta(p) \]
for $\mathbb{P}$-a.e. $\omega$ as $n \to \infty$. The map $\beta : \mathcal{P}^1(N) \to N$ is easily seen to be a contracting barycenter map. [Actually, here the assumption of nonpositive curvature in the sense of Definition 2.1 (= in the sense of Alexandrov) may be replaced by the weaker condition of nonpositive curvature in the sense of Busemann; local compactness, however, seems to be essential.] Note, however, that in general, $\beta \left( \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \right) \neq \beta_n(x_1, \ldots, x_n).

Moreover, we emphasize that our definition of barycenter $b(p)$, expectation $\mathbb{E}Y$ and mean value $\frac{1}{n} \sum_{i=1}^n Y_i$ are different from the ones used by [ESH99]. Their law of large numbers proves convergence to a point, which may be different from our expectation. For instance, let $(N, \mathbb{P})$ be the tripod and let $\mathbb{P}_Y = \frac{1}{2} \delta_{(1,1)} + \frac{1}{4} \delta_{(2,1)} + \frac{1}{4} \delta_{(3,1)}$. Then our expectation $\mathbb{E}Y$ will be the origin $o = (1,0)$, whereas an easy calculation shows that the expectation in the sense of [ESH99] is the point $(1,1/6)$.

**Proposition 6.6 (Empirical Law of Large Numbers).** Let $(N, d)$ be a complete metric space with a contracting barycenter map $\beta : \mathcal{P}^1(N) \to N$ and fix $p \in \mathcal{P}^\infty(N)$. Moreover, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(X_i)_{i \in \mathbb{N}}$ be an independent sequence of measurable maps $X_i : \Omega \to N$ with identical distribution $\mathbb{P}_{X_i} = p$. Define the "barycentric mean value" $s_n : \Omega \to N$ by $s_n(\omega) := \beta \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \right)$. Then for $\mathbb{P}$-almost every $\omega \in \Omega$
\[ s_n(\omega) \to \beta(p) \quad \text{as } n \to \infty. \]

**Proof.** Given an iid sequence $(X_i)_{i \in \mathbb{N}}$ as above, we define its empirical distribution $p_n$ as usual: $\forall \omega \in \Omega:$
\[ p_n(\omega, \cdot) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)} \in \mathcal{P}^\infty(N). \]

A theorem of Varadarajan (which is stated in [Du89], Theorem 11.4.1) for probability measures on complete separable metric spaces and which easily extends to probability measures with separable supports on complete metric spaces) states that $p_n(\omega, \cdot) \to p$ weakly for $\mathbb{P}$-almost every $\omega \in \Omega$. Hence,
\[ d^W(p_n(\omega, \cdot), p) \to 0 \quad \text{for } \mathbb{P}\text{-almost every } \omega \in \Omega \]
and by assumption on the barycenter map $\beta$, the latter implies
\[ d(\beta(p_n(\omega, \cdot)), \beta(p)) \to 0 \quad \text{for } \mathbb{P}\text{-almost every } \omega \in \Omega. \]

This proves the claim since by definition $s_n(\omega) = \beta(p_n(\omega, \cdot))$. \qed

I learned the above Empirical Law of Large Numbers and its proof from Heinrich von Weizsäcker (private communication) but it can also be found in [ESH99].
7. Convex Means

**Definition 7.1.** Given a probability measure \( p \in \mathcal{P}^1(N) \) we say that a point \( z \in N \) is a convex mean of \( q \) iff for all convex, Lipschitz continuous \( \varphi : N \to \mathbb{R} \):

\[
\varphi(z) \leq \int \varphi \, dp.
\]

(7.1)

The set of all convex means of \( p \) is denoted by \( C(p) \).

Obviously, \( C(p) \) is a closed convex set and contains the barycenter of \( p \).

**Remark 7.2.** (i) Jensen’s inequality can be used to characterize convex functions. Namely, a lower semicontinuous function \( \varphi : N \to \mathbb{R} \) is convex if and only if

\[
\int_{N} \varphi(x)q(dx) \geq \varphi(b(q)) \quad \text{for any } q \in \mathcal{P}^1(N) \quad \text{(or equivalently, for any } q \in \mathcal{P}^\infty(N) \).
\]

It remains to prove the "if"-implication. Let a function \( \varphi \) as above be given. Choose a geodesic \( \gamma \) and \( t \in [0,1] \). Define a probability measure on \( N \) by \( q = (1-t)\delta_{\gamma_0} + t\delta_{\gamma_1} \). Obviously, \( q \in \mathcal{P}^\infty(N) \) and \( b(q) = \gamma_t \). Hence, \( \varphi(\gamma_t) = \varphi(b(q)) \leq \int_{N} \varphi(x)q(dx) = (1-t)\varphi(\gamma_0) + t\varphi(\gamma_1) \). That is, \( \varphi \) is convex.

(ii) Jensen's inequality, however, can not be used in general to characterize barycenters. In other words, \( C(p) \neq \{b(p)\} \) in general. See Example 7.5 below.

(iii) [EM91] define the barycenter of \( q \in \mathcal{P}(N) \) to be the set \( C_*(q) \) of all \( x \in N \) such that \( \varphi(x) \leq \int_{N} \varphi(y)q(dy) \) for all bounded continuous convex functions \( \varphi : N \to \mathbb{R} \). A related point of view was used by [Do49] and [He91] based on the set \( C^*(q) \) of all \( x \in N \) such that \( d(z,x) \leq \int_{N} d(z,y)q(dy) \) for all \( z \in N \).

Note that on bounded global NPC spaces, the functions \( x \mapsto d(z,x) \) are bounded and convex. Hence, \( C_*(q) \subset C(q) \subset C^*(q) \).

**Proposition 7.3.** If \((N,d)\) is a Hilbert space then \( C(q) = \{b(q)\} \) for each \( q \in \mathcal{P}^1(N) \).

**Proof.** Given \( z \in C(q) \), define a convex, Lipschitz continuous function \( \varphi \) on \( N \) by \( \varphi(x) := \langle x, z-b(q) \rangle \). By assumption and Proposition 5.4

\[
\langle z, z-b(q) \rangle = \varphi(z) \leq \int \varphi \, dq = \langle b(q), z-b(q) \rangle.
\]

This implies \( z = b(q) \). \( \square \)

Hence, for Hilbert spaces our definition of barycenters, expectations and integrals coincides with any other of the usual definitions (e.g. Bochner integral).

Now let us consider trees. For convenience, we start with spiders. Recall the notations from the previous section.

**Proposition 7.4.** Let \((N,d)\) be the spider over some set \( K \) and \( q \in \mathcal{P}^1(N) \) then

\[
\kern1cm C(q) \subset \{(i,s) \in K \times \mathbb{R}_+ : b_i(q) \leq s \leq r_i(q)\}.
\]

In particular,

\[
b(q) \in N_i \setminus \{o\} \implies C(q) \subset N_i \setminus B_{b_i(q)}(o)
\]

and

\[
b(q) = o \iff o \in C(q).
\]
PROOF. Fix \( i \in K \) and \( z = (i, s) \in C(q) \cap N_i \) and define convex, Lipschitz continuous functions \( \varphi \) and \( \psi \) on \( N \) by

\[
\varphi(x) := 1_{N_i}(x) \cdot d(o, x), \quad \psi(x) := (1_{N \setminus N_i}(x) - 1_{N_i}(x)) \cdot d(o, x).
\]

By assumption

\[
s = \varphi(z) \leq \int \varphi \, dq = r_i(q)
\]
and

\[
-s = \psi(z) \leq \int \psi \, dq = -b_i(q).
\]

\[\square\]

EXAMPLE 7.5. Let \( (N, d) \) be the tripod and \( q = \frac{1}{3} \sum_{i=1}^3 \delta_{(i,1)} \). Then \( b(q) = o \) and \( C(q) = \overline{B}_{1/3}(o) \).

Finally, let \( (N, d) \) be a discrete metric tree. For \( z \in N \) let \( K_z, N_{z,i}, r_{z,i}(q) \) and \( b_{z,i}(q) \) be as in Proposition 5.9. Moreover, put \( N_{z,i}^* := \emptyset \) if \( N_{z,i} \) is isometric to an interval. Otherwise, let \( z_i \) be the branch point in \( N_{z,i} \) which is closest to \( z \) and let \( N_{z,i}^* \) be the set of points in \( N_{z,i} \) which are joined with \( z \) by geodesics through \( z_i \).

PROPOSITION 7.6. For \( q \in \mathcal{P}^1(N) \) and \( z \in C(q) \) let \( i \in K_z \) such that \( b(q) \in N_{z,i} \cup \{z\} \). Then

\[
d(z, b(q)) \leq \int_{N_{z,i}} d(z_i, x) \, q(dx).
\]

In particular, \( b(q) = z \) if there is no branch point in \( N_{z,i} \).

Moreover, if \( q \in \mathcal{P}^\theta(N) \) for some \( \theta > 1 \) and \( \rho(z) := \inf_v d(v, z) \) where the infimum is over all branchpoints \( v \in N \) for which \( b(q) \) lies on the geodesic from \( z \) to \( v \), then

\[
d(z, b(q)) \leq \frac{1}{\rho(z)^{\theta-1}} \int_N d^\theta(z, x) \, q(dx).
\]

PROOF. A slight generalization of Proposition 5.9 yields that

\[
b_{z,i}(q) = -d(z, b(q))
\]
whenever \( b(q) \in N_{z,i} \cup \{z\} \). Now fix \( z \) and \( i \) and define a convex function \( \varphi \) on \( N \) by

\[
\varphi(x) = 1_{N \setminus N_{z,i}^*}(x) \cdot d(x, z_i) - d(z_i, z)
\]
if \( N_{z,i}^* \neq \emptyset \) and \( \varphi(x) = \left(1_{N \setminus N_{z,i}^*}(x) - 1_{N_{z,i}^*}(x)\right) \cdot d(x, z) \) else. Assuming that \( z \in C(q) \) yields

\[
0 = \varphi(z) \leq \int \varphi \, dq = b_{z,i}(q) + \int_{N_{z,i}^*} d(z_i, x) \, q(dx).
\]

This proves the first claim. For the second claim, note that

\[
\int_{N_{z,i}^*} d(z_i, x) \, q(dx) \leq \int_{N_{z,i}^*} d(z, x) \, q(dx) \leq \left(\int_N d^\theta(z, x) \, q(dx)\right)^{1/\theta} \cdot q\left(N_{z,i}^*\right)^{1-1/\theta}
\]
and \( q\left(N_{z,i}^*\right) \cdot \rho^\theta(z) \leq \int_N d^\theta(z, x) \, q(dx) \). \[\square\]
Proposition 7.7. Let \((N, d)\) be (derived from) a smooth, simply connected, complete Riemannian manifold of nonpositive curvature and let \(N'\) be an open subset of \(N\) with lower bounded curvature.

Then \(\forall \delta > 0 : \exists r > 0 : \forall q \in P^2(N), \forall z \in C(q) \text{ with } B_r(z) \subset N' : \)

\[
d(z, b(q)) \leq \delta \cdot \int \frac{d^2(z, x)q(dx)}{r} + 4 \cdot \delta \cdot q(N \setminus B_r(z))
\]

Proof. [St03]

8. Local NPC Spaces

This section is devoted to the study of integrals and/or expectations for maps with values in local NPC spaces. A metric space \((N, d)\) is called local NPC space iff it is a complete length space of local curvature \(\leq 0\). In other words, iff it is a complete length space where each point \(x \in N\) has a neighborhood \(N' \subset N\) which is a global NPC space.

Note that completeness is really essential in the definition of local NPC spaces. For instance, the subspace \(N = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0 \text{ or } x_2 > 0\}\) with the induced length metric obviously has locally curvature \(\leq 0\) but neither \(N\) nor its completion (= closure in \(\mathbb{R}^2\)) is a local NPC space.

Example 8.1. Let \((N, d)\) be a global NPC space and \(G\) be a subgroup of the isometry group of \((N, d)\), that is, \(G\) is a group of isometries \(\eta : N \to N\). Assume that \(G\) acts properly discontinuous on \(N\), i.e. each point \(z \in N\) has a neighborhood \(N_0 \subset N\) such that \(N_0 \cap \eta N_0 = \emptyset\) for all \(\eta \in G \setminus \{1\}\). Then \(N/G\) is a local NPC space.

Proposition 8.2. (i) A local NPC space \((N, d)\) is path connected, locally path connected and locally simply connected. Its universal cover \((\tilde{N}, \tilde{d})\) is a global NPC space. A metric space \((N, d)\) is global NPC if and only if it is local NPC and simply connected.

(ii) The fundamental group \(\pi_1(N)\) of \((N, d)\) is canonically isomorphic to the group of covering transformations \(G_N\) which acts properly discontinuously on \(\tilde{N}\). The space \(N\) can be identified with \(\tilde{N}/G_N\). Then the covering map \(\varphi_N : \tilde{N} \to N\) is given by \(\varphi_N(x) = G_N x\).

The group \(G_N\) is a subgroup of the isometry group of \(\tilde{N}\).

Proof. For the first assertion, see e.g. [Jo97, Corollary 2.3.2]. For the topological results, we refer to any textbook on algebraic topology, e.g. [GH81].

Example 8.3 (Surfaces of Revolution). Let \(\varphi : I \to \mathbb{R}_+\) be a continuous function defined on an interval \(I \subset \mathbb{R}\) and define its surface of revolution by

\[
N := \{(x, y, z) \in \mathbb{R}^3 : \sqrt{x^2 + y^2} = \varphi(z), z \in I\}.
\]

The Euclidean distance on \(\mathbb{R}^3\) induces a geodesic metric \(d\) on \(N\) which allows to identify \(N\) with the warped product \(I \times_{\varphi} S^1\). \((N, d)\) is a local NPC space if and only if \(\varphi\) is convex.

Example 8.4 (Graphs). Each discrete metric graph is a local NPC space.
EXAMPLE 8.5 (Polyhedra). A two-dimensional polyhedral space is a local NPC space if and only if it has no vertex whose link contains a subspace isometric to a circle of length $\leq 2\pi$.

There is no canonical way to define barycenters on local NPC spaces. However, it is possible to define expectations or integrals of continuous maps $f : M \to N$.

For this purpose, let $M$ be a connected and locally connected topological space with a fixed point $m \in M$ and let $\mathcal{P}(M)$ denote the set of all probability measures on $M$ equipped with its Borel $\sigma$-field. Let $(N,d)$ be a local NPC space with universal cover $(\tilde{N},\tilde{d})$ and covering map $\varphi_N : \tilde{N} \to N$.

PROPOSITION 8.6. (i) For every continuous map $f : M \to N$ and every $\tilde{o} \in \varphi_N^{-1}(f(m))$ there exists a unique continuous map $\tilde{f} : M \to \tilde{N}$ with

$$\varphi_N \circ \tilde{f} = f$$

and $\tilde{f}(\tilde{m}) = \tilde{o}$ ("$\tilde{f}$ is the lifting of $f$""). $\tilde{f}$ has separable range if and only if $f$ has so.

If we choose another point $\tilde{o}' \in \varphi_N^{-1}(f(m))$ then $\tilde{f}' = \eta \tilde{f}$ with suitable $\eta \in \pi_1(N,f(m))$.

(ii) For each $p \in \mathcal{P}(M)$ and each $\theta \geq 1$, the number

$$\inf_{\tilde{z} \in \tilde{N}} \int_M d^\theta(\tilde{z},\tilde{f}(x)) p(dx)$$

does not depend on the choice of $\tilde{o}$. If this number is finite and if $f$ has separable range, then we say that $f \in \tilde{L}^\theta(M,N,p)$.

(iii) For each continuous $f \in \tilde{L}^1(M,N,p)$ we define the integral or expectation

$$\int_M f dp := \varphi_N \left( \tilde{b}(\tilde{f}, p) \right)$$

where $\tilde{f}$ is the lifting of $f$ for some choice of $\tilde{o} \in \varphi_N^{-1}(f(m))$ and where $\tilde{b}$ denotes the $d^2$-barycenter map on $(\tilde{N},\tilde{d})$.

The point $\int_M f dp$ does not depend on the choice of $\tilde{o}$. Indeed, for each $\eta \in \pi_1(N,f(m))$

$$\tilde{b} \left( (\eta \tilde{f}), p \right) = \eta \left( \tilde{b}(\tilde{f}, p) \right).$$
(iv) Given \( p, q \in \mathcal{P}(M) \) and continuous \( f \in \tilde{L}^1(M, N, p), g \in \tilde{L}^1(M, N, q) \) then
\[
\inf \int_M d(\tilde{f} dp, \int_M g dq) \leq \inf \int_M \tilde{a}(\tilde{f}(x), \tilde{g}(x)) \mu(dx, dy)
\]
where the inf is over all couplings \( \mu \) of \( p \) and \( q \).

(v) Let \( M' \) be another connected and simply connected topological space and let \( \Psi : M' \to M \) be a continuous map. Then the following transformation rule holds true:
\[
\int_{M'} f(\Psi) dp = \int_M f dp_{\Psi}
\]
where \( p_{\Psi} = \Psi_{*}p \) is the push forward of \( p \) under \( \Psi \).

(vi) If in the above considerations, \( N \) is already simply connected then the integral \( \int_M f dp \) as defined above will coincide with the usual definition by means of barycenters on \( N \) (as given in section 4 and used throughout this chapter), i.e.
\[
\int_M f dp = b(f, p).
\]

For the transformation rule in (v) it is crucial that \( M \) and \( M' \) are simply connected.

**Proof.** (i) can be found in any textbook in topology. (ii) follows from
\[
\inf \int_M d^g(\tilde{z}, \tilde{f})(x)) p(dx) = \inf \int_M d^g(\eta \tilde{z}, \eta \tilde{f})(x)) p(dx)
\]
and (iii) for continuous \( f \in \tilde{L}^2(M, N, p) \) and \( \eta \in \pi_1(N, f(m)) \) from
\[
\hat{b}(\eta \tilde{f}) = \arg\min \int_M d^g(\tilde{z}, \eta \tilde{f})(x)) p(dx) = \eta \left( \arg\min \int_M d^g(\tilde{z}, \tilde{f})(x)) p(dx) \right) = \eta \left( \hat{b}(\tilde{f}) \right).
\]

For (iv) observe that for each coupling \( \mu \) of \( p \) and \( q \)
\[
d\left(\int_M f dp, \int_M g dq\right) \leq \tilde{a}(\hat{b}(f, p), \hat{b}(f, p)) \leq \tilde{a} \left( \int_M \hat{f} dp, \int_M \hat{f} dp \right) \leq \inf \int_M \tilde{a}(\hat{f}(x), \hat{g}(x)) \mu(dx, dy).
\]

Using (iv) one can extend the previous results from \( \tilde{L}^2(M, N, p) \) to \( \tilde{L}^1(M, N, p) \).

(v) follows from
\[
\int_{M'} f(\Psi) dp = \varphi_N \left( \hat{b}(f(\Psi), p) \right) = \varphi_N \left( \hat{b}(f, p_{\Psi}) \right) = \int_M f dp_{\Psi}.
\]

(vi) is obvious. \( \square \)

**Example 8.7.** Let \( N \) be a local NPC space, \( \tilde{N} \) its universal cover and \( o \in N \), \( \tilde{o} \in \tilde{N} \) fixed points with \( \varphi_N \tilde{o} = o \). Let \( \Omega = C_o(\mathbb{R}_+, N) \) be the space of continuous maps \( \omega : \mathbb{R}_+ \to N \) with \( \omega_0 = o \) and equipped with the topology of local uniform convergence. Let \( A \) be its Borel \( \sigma \)-field, \( \mathbb{P} \) be some probability measure on \( (\Omega, A) \) and \( X : \Omega \to N \) the continuous map defined by
\[
X(\omega) = \omega(t)
\]
for some $t \in \mathbb{R}_+$. Then according to the previous Theorem
\[ \mathbb{E}X := \int_{\Omega} X d\mathbb{P} := \phi_N \left( b(\mathbb{P}_{\bar{X}}) \right) \]
is well defined.

Actually, in this case, it can be seen more directly. Let $\bar{\Omega} = C_0(\mathbb{R}_+, \hat{N})$ be the space of continuous maps $\bar{\omega} : \mathbb{R}_+ \to \hat{N}$ with $\bar{\omega}_0 = \bar{a}$. Then for each $\omega \in \Omega$ there exists a unique $\bar{\omega} \in \bar{\Omega}$ with $\phi_N \circ \bar{\omega} = \omega$ ("lifting of $\omega$") and vice versa ("projection of $\bar{\omega}$"), i.e. $\phi_N$ induces a bijective map
\[ \phi_N : \bar{\Omega} \to \Omega. \]
This lifting induces a unique probability measure $\bar{\mathbb{P}} := (\phi_N^{-1})_*\mathbb{P}$ on $(\bar{\Omega}, \bar{A})$ and a continuous map $\bar{X} : \bar{\Omega} \to \hat{N}$ such that $\phi_N \circ \bar{X}(\bar{\omega}) = X(\omega)$. Then
\[ \mathbb{E}X = \phi_N \left( b(\bar{\mathbb{P}}_{\bar{X}}) \right) = \phi_N \left( \int_{\bar{\Omega}} \bar{X} d\bar{\mathbb{P}} \right) = \phi_N \left( \mathbb{E}\bar{X} \right). \]

Acknowledgement. I am grateful to Lutz Mattner from whom I learned the idea to minimize $\int [d^2(\cdot, x) - d^2(x, y)] q(dx)$ [instead of $\int d^2(\cdot, x) q(dx)$] which allows to define directly $b(q)$ for $q \in \mathcal{P}^1(N)$.

References

[A151] A. D. Alexandrov (1951): A theorem on triangles in a metric space and some applications, Trudy Math. Inst. Steklov 38, 5-23. (Russian; translated into German and combined with more material in [A157a])


INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT BONN
WEGELERSTRASSE 6
53115 BONN
GERMANY
E-mail address: sturm@uni-bonn.de