



**HAL**  
open science

# Contextual distances, a mathematical framework for geographical distances

Benoît Kloeckner, Alain L'Hostis, Thomas Richard

► **To cite this version:**

Benoît Kloeckner, Alain L'Hostis, Thomas Richard. Contextual distances, a mathematical framework for geographical distances. 2019. hal-02147263v1

**HAL Id: hal-02147263**

**<https://hal.science/hal-02147263v1>**

Preprint submitted on 4 Jun 2019 (v1), last revised 24 Sep 2020 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Contextual distances, a mathematical framework for geographical distances

Benoît R. Kloeckner<sup>1</sup>, Alain L’Hostis<sup>2</sup>, and Thomas Richard<sup>1</sup>

<sup>1</sup>Université Paris-Est, Laboratoire d’Analyse et de Mathématiques  
Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-94010, Créteil,  
France

<sup>2</sup>LVMT, UMR-T 9403, Ecole des Ponts, IFSTTAR, UPEM,  
Champs-sur-Marne, France

June 4, 2019

## Abstract

Our goal is to establish a mathematical framework suitable for the description of geographical distances related to movement in space and produced by transport means. We propose to root the idea of effort – or disutility – minimisation into the definition of geographical distance and its mathematical formalisation. This objective is non-trivial to achieve because of the issue of possible triangle inequality violation due mainly to the need for *break* in any movement. This issue translates into additivity problems that may disturb the order of proximities in geographical spaces, and undermine their basic geometrical properties. In order to address this issue, we introduce the concept of *contextual distances* that, in parallel to pure geometric movement, considers a context where resources used to move are accounted for, that it be time, money or other relevant quantities.

We show that contextual distances follow some of the properties of metrics. In particular contextual distance respects the triangle inequality. This result paves the way for its use within the context of spatial analysis in geography.

# 1 Introduction

Distance is a central concept for both geography and geometry. In geography, our focus, distance is generated by the necessary separation of geographic entities: all geographic objects can't occupy the same place (Isnard et al. 1981). If the “central problem of geography is to place interacting objects as near to each other as possible when the definition of distance is chosen which minimizes movement” (Bunge 1962, p. 200), then distance definition is a critical task for the geographer. Three modes of management of geographic distances, at least since the neolithic period, have been observed: co-presence, mobility and telecommunication (Lévy 2009). Hence transport is essential for understanding geographical distances.

Our intention is to establish a mathematical framework suitable for the description of geographical distances related to movement in space and produced by transport means.

A related problem is the cartography of time-distances (Shepard 1962; Marchand 1973; Kruskal and Wish 1978; Hyman and Mayhew 2004; Axhausen et al. 2008; Shimizu and Inoue 2009; Ficzer, Ulmann, and Török 2014; Dusek and Szalkai 2017). Contributions in this field develop practical solutions that minimize or control the *stress* of the cartographic representation, but do not introduce new mathematical definitions of distances. Other efforts of developing mathematical description of geographical spaces, like for instance the mathematical framework developed by (Harvey J Miller 2005) for time geography, involve specific sub-domains of geography and do not engage a new definition of distance.

The set of approaches the closest to our intention have concentrated on the development of a mathematical distance framework suitable for the computation of travel costs, in the context of spatial economy (Smith 1989; Huriot, Smith, and Thisse 1989). Our own focus, on the geometry of geographical distances, differs in the sense that it cares more for geographical and cartographic issues while also considering spatial economy costs.

In dealing with movement in geographical space we need to consider metrics but also networks. Hence our starting point is a description of path.

## 2 Path optimization and metrics

Let us first introduce a suitably general formal framework, borrowed from graph theory, to construct a classical distance from travel costs, to be later expanded to take more information into account. We shall rely on “paths” and cost optimization, much as in (Huriot, Smith, and Thisse 1989), but we shall give a more abstract

representation of these paths in order for this model to better fit several common situations to be found in geographical spaces. First, this model shall be usable to describe both continuous spaces and discrete spaces; e.g. road trips can be described by continuous paths on the physical space, or by discrete paths on the graph defined by the road network. Second, this model shall be suitable to include plane flights, for which continuous paths are ill-suited: at any time of the flight, the path takes as value a mid-air point that is *not* a viable location: one only travels *through* such a point, which is never a destination. Not including such points in the space forbids the use of continuous paths, while adding them creates artificial locations that are the starting or ending point of no trip.

## 2.1 Networks and cost functions

Let  $X$  be a set, called the *space*. An element in  $X$  shall be called a *location*, and in the modeling process one should only include in  $X$  those places where people could go to or get from; in particular,  $X$  might not represent the whole of the “physical space” that is being modeled.

A *network* on  $X$  is a set  $N$  together with two functions  $s, e : N \rightarrow X$ ; elements of  $N$  are called *arcs*, and given an arc  $\alpha$  the locations  $s(\alpha)$  and  $e(\alpha)$  are called its starting point and endpoint respectively. In the modeling process, one should include in  $N$  either all possible travels, or a set of elementary travels that will be enough to recover all of them by concatenation (see next paragraph). It is possible to define several arcs that correspond to the same displacement in the physical space, in order to account for other differences (e.g. traveling along a given fixed road by car, or on foot, or by bike could be modeled by defining three distinct arcs).

Given a network  $N$  on a space  $X$ , a *path*  $\gamma$  from  $p \in X$  to  $q \in X$  is a finite word with letters in  $N$  (i.e. an ordered tuple of arbitrary length, written  $\gamma = \alpha_1 \alpha_2 \cdots \alpha_k$  for some  $k \in \mathbb{N}$  and some  $\alpha_i \in N$ ), such that  $s(\alpha_1) = p$ ,  $e(\alpha_k) = q$  and  $e(\alpha_i) = s(\alpha_{i+1})$  for all  $i \in \{1, 2, \dots, k-1\}$ . We also include as paths the trivial paths  $\emptyset_p$  for each  $p \in X$ , which start and end at  $p$  and are empty words (this necessitate to include the point  $p$  in the data describing the path, which is unnecessary for non-empty paths). Given two paths  $\gamma = \alpha_1 \dots \alpha_k$  and  $\eta = \beta_1 \dots \beta_\ell$  such that  $e(\alpha_k) = s(\beta_1)$ , we define their *concatenation* as the path  $\gamma * \eta := \alpha_1 \dots \alpha_k \beta_1 \dots \beta_\ell$ . A path is thus a concatenation of arcs, and represents a travel possibly made of a combination of elementary travels. The set of paths shall be denoted by  $N^*$ , and we shall write  $\gamma : p \rightarrow q$  to express that  $\gamma$  is a path from  $p$  to  $q$ . Note that despite the notation,  $N^*$  depends on  $s, e$  as much as it depends on  $N$ .

A *length function* on a network  $N$  is simply a function  $\ell : N \rightarrow (0, +\infty)$  (we

could have included the value 0, but this would have made the introduction of an additional adjective necessary). We immediately extend the function  $\ell$  to a function defined on  $N^*$ , still denoted by  $\ell$ , by

$$\ell(\alpha_1 \dots \alpha_k) = \sum_{i=1}^k \ell(\alpha_i) \tag{1}$$

with the usual convention that an empty sum is zero, i.e.  $\ell(\emptyset_p) = 0$  for all  $p \in X$ . (Another equivalent way to formalize the same framework would be to consider a function  $\ell : N^* \rightarrow [0, +\infty)$  asked to be additive, i.e.  $\ell(\gamma * \eta) = \ell(\gamma) + \ell(\eta)$  for all paths  $\gamma, \eta$ , and positive on non-trivial paths). The *distance* on  $X$  defined by  $(N, \ell)$  is then set as

$$d(p, q) = \inf_{\gamma: p \rightarrow q} \ell(\gamma).$$

Note that this is not necessarily a metric unless we make additional assumptions; it can even be infinite, whenever there are no paths from  $p$  to  $q$ .

The term *length* function should be taken with a grain of salt:  $\ell$  could be about any effort or disutility function (length, travel time, cost, etc.) In some cases, it could be argued that disutility should be allowed to be sub-additive; the reader can check that most of the properties below would not be altered by this relaxation, with the exception of Symmetry in 2.5 that would need an additional assumption on the cost function. In any case, sub-additive disutility can always be modelled in the present framework by adding an arc for each travel combination that allows for some savings.

## 2.2 When is the distance induced by a length function, a metric?

We will now explore assumptions that ensure the various axioms defining a metric. The most troublesome is distinguishability, more precisely that different points should be at positive distance one from another. The following example shows what could go wrong.

**Example 2.1.** Take  $X = \{p, q\}$ ,  $N = \{\alpha_i : i \geq 1\}$  with  $s(\alpha_i) = p$ ,  $e(\alpha_i) = q$ , and  $\ell(\alpha_i) = 1/i$  (we could enlarge  $N$  to get symmetry and finiteness of  $N$  but this would mostly obscure the point). Then we have an infinite sequence of paths (each consisting of only one arc,  $\alpha_i$ ) from  $p$  to  $q$ , of length  $1/i$ . Since this goes to zero as  $i$  goes to infinity, we get  $d(p_1, p_2) = 0$ .

**Example 2.2.** Let us give a second example that feels less ad hoc. Take  $X = [0, 1]$ ,  $N = \{\alpha_p^q : (p, q) \in [0, 1] \times [0, 1]\}$  with  $s(\alpha_p^q) = p$ ,  $e(\alpha_p^q) = q$  (we have one arc for

each possible pair of starting and ending points) and  $\ell(\alpha_p^q) = |p - q|^2$ . Then for each  $n \in \mathbb{N}$ , the path

$$\gamma_n := \alpha_0^{1/n} \alpha_{1/n}^{2/n} \cdots \alpha_{(n-1)/n}^1$$

goes from 0 to 1 and has length  $\ell(\gamma_n) = n \times (1/n)^2 = 1/n$ . It follows  $d(0, 1) = 0$ .

(Note that in both examples, the distance can only be approximated by the length of path, there are no given path linking the given points whose length is their distance. This phenomenon can occur even in cases where all axioms of a metric hold true.)

This leads us to the following definition:

**Definition 2.3.** We say that the length function is *non-degenerate* when there exist a function  $m : \{(p, q) : p, q \in X, p \neq q\} \rightarrow (0, +\infty)$  such that for all  $p \neq q \in X$  and all  $\gamma : p \rightarrow q$  we have  $\ell(\gamma) > m(p, q)$ .

This is somewhat trivial as it is tuned to exactly ensure distinguishability, but is still an operational definition as it leads one to find a lower bounding function  $b$ , which in most cases of interest shall be easy to either find or at least prove into existence. There are two particularly simple cases ensuring non-degeneracy.

**Proposition 2.4.** *Let  $X$  be a space and  $N$  be a network on  $X$ .*

- i. If for all  $p, q \in X$  the set  $\{\gamma : p \rightarrow q\}$  of paths from  $p$  to  $q$  is finite, then every length function is non-degenerate.*
- ii. Let  $\ell$  be a length function such that for some  $\varepsilon > 0$ , for all arc  $\alpha$  it holds  $\ell(\alpha) \geq \varepsilon$ . Then  $\ell$  is non-degenerate.*

*Proof.* Note that a length function  $\ell$  is assumed to be positive on each arc, and is thus positive on each non-trivial path. When  $\{\gamma : p \rightarrow q\}$  is finite, one can take  $m(p, q) = \min_{\gamma: p \rightarrow q} \ell(\gamma)$  which is positive as soon as  $p \neq q$ .

The second case actually assumes  $\inf_N \ell \geq \varepsilon > 0$ , which implies that each non-trivial path has length at least  $\varepsilon$ : one can then take  $m(p, q) \equiv \varepsilon$ .  $\square$

**Theorem 2.5.** *Let  $X$  be a space endowed with a network  $N$  and a length function  $\ell$ , and denote by  $d$  the corresponding distance.*

- i. (Finitess) For all  $p, q \in X$ , we have  $d(p, q) < +\infty$  if and only if there exists  $\gamma : p \rightarrow q$ .*
- ii. (Symmetry) If for all arc  $\alpha$  there exist an arc  $\alpha'$  such that  $s(\alpha') = e(\alpha)$ ,  $e(\alpha') = s(\alpha)$  and  $\ell(\alpha') = \ell(\alpha)$ , then  $d(p, q) = d(q, p)$  for all  $p, q \in X$ .*

- iii. (Triangular inequality)  $d(p, r) \leq d(p, q) + d(q, r)$  for all  $p, q, r \in X$ .
- iv. (Distinguishability) *The property  $d(p, q) = 0$  if and only if  $p = q$  holds if and only if  $\ell$  is non-degenerate.*

In particular, assuming that every pair of locations is linked by a path, that each arc has a reverse arc of the same length (case of a symmetric network), and that the length function is non-degenerate ensures that  $d$  is a metric. While among the three axioms defining a metric, the triangular inequality is probably the one most discussed, we see as in (Huriot, Smith, and Thisse 1989) that it is the one that needs the less hypotheses: it follows from the optimization of paths.

*Proof.* The first point follows from the fact that  $\ell$  does not take the value  $\infty$ , so that  $d(p, q) = +\infty$  if and only if the infimum defining it is over the empty set.

To prove the second point we consider  $p, q \in X$  and prove  $d(q, p) \leq d(p, q)$ ; equality follows by exchanging the roles of  $p$  and  $q$ . If  $d(p, q) = +\infty$ , this is obvious. Otherwise, given any  $\varepsilon > 0$  there exist a path  $\gamma = \alpha_1 \alpha_2 \cdots \alpha_k$  from  $p$  to  $q$  with  $\ell(\gamma) \leq d(p, q) + \varepsilon$ . Then, using the notation  $\alpha'_i$  for the reverse arc of  $\alpha_i$  provided by the hypothesis,  $\gamma' := \alpha'_k \alpha'_{k-1} \cdots \alpha'_1$  is a path from  $q$  to  $p$  and

$$\ell(\gamma') = \sum_{i=1}^k \ell(\alpha'_{k-i}) = \sum_{i=1}^k \ell(\alpha_{k_i}) = \ell(\gamma) \leq d(p, q) + \varepsilon.$$

Letting  $\varepsilon$  go to zero, it follows  $d(q, p) \leq d(p, q)$ .

To prove the third point, let  $p, q, r \in X$ . If there is no path from  $p$  to  $q$  or no path from  $q$  to  $r$ , then the right-hand side is  $+\infty$  and the inequality is trivially true. Assume otherwise and let  $\varepsilon > 0$  be arbitrary. There exists paths  $\gamma_1 : p \rightarrow q$  and  $\gamma_2 : q \rightarrow r$  such that  $\ell(\gamma_1) \leq d(p, q) + \varepsilon$  and  $\ell(\gamma_2) \leq d(q, r) + \varepsilon$ . Then  $\gamma_1 * \gamma_2 : p \rightarrow r$  so that

$$d(p, r) \leq \ell(\gamma_1 * \gamma_2) = \ell(\gamma_1) + \ell(\gamma_2) \leq d(p, q) + d(q, r) + 2\varepsilon.$$

Letting  $\varepsilon$  go to zero, we get the triangular inequality.

Concerning the fourth and last point, note that we always have  $d(p, p) = 0$  since the empty path  $\emptyset_p$  has by definition length 0. Assume now that  $\ell$  is non-degenerate and let  $m$  a function as given in Definition 2.3; then whenever  $p \neq q \in X$ , for all  $\gamma : p \rightarrow q$  we have  $\ell(\gamma) \geq m(p, q)$ , so that  $d(p, q) \geq m(p, q) > 0$ . Conversely, if  $p \neq q \implies d(p, q) > 0$ , then letting  $m(p, q) = d(p, q)$  yields non-degeneracy of  $\ell$ .  $\square$

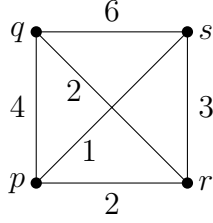


Figure 1: Four cities linked by six roads with time measurements from (Haggett 2001, p. 341)

### 2.3 Examples

Let us give a few examples showing how the above framework can be used to model various geographically relevant situations.

**Example 2.6.** Let us start with a discrete example. We choose the infamous example of Haggett (Haggett 2001, p. 341):  $X = \{p, q, r, s\}$  is a set of four cities, linked two by two by six roads that can be traveled in times given in figure 1. These connections can be modeled in our framework by a network of twelve arcs  $N = \{\alpha_{ij} : i, j \in X\}$  (including both directions for each road) with  $s(\alpha_{ij}) = i$  and  $e(\alpha_{ij}) = j$ ; and by the length function (expressed in the unit of one hour)

$$\begin{aligned} \ell(\alpha_{pq}) = \ell(\alpha_{qp}) = 4 & & \ell(\alpha_{pr}) = \ell(\alpha_{rp}) = 2 & & \ell(\alpha_{ps}) = \ell(\alpha_{sp}) = 1 \\ \ell(\alpha_{qr}) = \ell(\alpha_{rq}) = 2 & & \ell(\alpha_{qs}) = \ell(\alpha_{sq}) = 6 & & \ell(\alpha_{rs}) = \ell(\alpha_{sr}) = 3. \end{aligned}$$

Up to now, the model fits exactly Haggett example. However, considering paths we see that what we call the distance becomes quite different; in particular, the road  $\alpha_{qs}$  becomes irrelevant as the paths  $\gamma_1 := \alpha_{qr}\alpha_{rs}$  and  $\gamma_2 := \alpha_{qp}\alpha_{ps}$  each have cost 5, less than the cost 6 of the direct arc  $\alpha_{qs}$ . This translates the fact that a traveler *can* go from  $q$  to  $s$  in only five hours, by avoiding the direct road. As underlined in (L’Hostis 2016b), optimization ensures the validity of the triangular inequality.

More generally, any graph with positive labels on its edges can be translated into the present framework, and the distance is the usual path-minimizing distance (actually, this case is a very classical framework in graph theory).

**Example 2.7.** Let us now show how the above framework can be used to model a continuous space. Let  $X$  be a domain of the plane  $\mathbb{R}^2$  (endowed with its canonical scalar product), meant to represent a small enough region that the curvature of Earth can be neglected. Assume that we are to model e.g. bird or foot travel and



that the region is homogeneous, without obstacles or roads facilitating certain travels compared to other.

Then it makes sense to consider as arcs all continuously differentiable curves  $\alpha : [T_0, T_1] \rightarrow X$  (where  $T_0 < T_1 \in \mathbb{R}$ ), with  $s(\alpha) = \alpha(T_0)$  and  $e(\alpha) = \alpha(T_1)$ , and to take the classical length function

$$\ell(\alpha) := \int_{T_0}^{T_1} \|\alpha'(t)\| dt$$

Then paths can be identified with continuous, piecewise continuously differentiable curves in an obvious way. Note that if we preferred to have “length” be physically homogeneous to a time, we could have written

$$\ell(\alpha) := \int_{T_0}^{T_1} R \|\alpha'(t)\| dt$$

where  $R$  is the inverse of the speed of the modeled traveler. This number can be factored into a global scale for our model, and we disregard this consideration for now.

If  $X$  is convex, then the distance obtained by optimizing the length of curves coincides with the Euclidean metric:  $d(p, q) = \|p - q\|$ , and the unique shortest path from  $p$  to  $q$  is the line segment between these points (parameterized in any one-to-one way).

Note that convexity of  $X$  is necessary for this statement to hold. Consider for example an annulus  $X = \bar{B}(0, 3) \setminus B(0, 1)$  where  $B(p, r)$  (respectively  $\bar{B}(p, r)$ ) denotes the open (respectively closed) Euclidean disk of center  $p$  and radius  $r$ . Then the points  $p, q$  of coordinates  $(0, -2)$  and  $(0, 2)$  respectively cannot be joined by a line segment, as paths are constrained to stay in  $X$ . In this case  $d(p, q) > \|q - p\| = 4$  (in general, the same phenomenon occurs for each pair of point defining a line segment that is not contained in  $X$ ). There are exactly two shortest paths from  $p$  to  $q$ , one avoiding the inner disk from the left, the other avoiding it from the right. If we had considered  $X = \bar{B}(0, 3) \setminus \bar{B}(0, 1)$ , then the distance would have been the same, but there would not exist any shortest path (as in Example 2.2 from  $p$  to  $q$ , their distance being only arbitrarily well approximated by lengths of paths).

Even without convexity, it is well-known and easy to check that  $d$  is a metric if and only if  $X$  is connected by continuous, piecewise continuously differentiable paths (otherwise, only the finiteness fails). In particular, non-degeneracy is easily proven by taking  $b(p, q) = \|q - p\|$ .

**Example 2.8.** More general situations can be modeled in a similar way than in Example 2.7, for example by introducing inhomogeneity in the formula of the cost

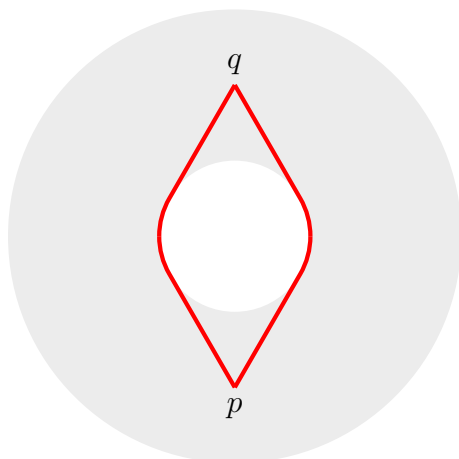


Figure 2: Distances on an annulus

of an arc

$$\ell(\alpha) = \int_{T_0}^{T_1} R(\alpha_t) \|\alpha'(t)\| dt$$

where  $R : X \rightarrow (0, +\infty)$  represent the roughness of the terrain at each point (in mathematics, this is called a *conformal change of Riemannian metric*). We shall denote by  $d_R$  the induced distance (which is a metric in many cases, e.g. whenever there exist some  $\varepsilon > 0$  such that  $R(p) \geq \varepsilon$ ). Such a situation is easily pictured using shades of gray to represent  $R$ , travel being more difficult in darker regions than in lighter ones. In the domain of geographical cartography, such a representation of distances takes the form of *cost-of-passage surface* or *cost surface* (Collischonn and Pilar 2000; White and Barber 2012). Related representations include Bunge proposing a crumpled space model to describe a marsh area difficult to cross surrounding a road (Bunge 1962, p. 271). Representations with the same aim of describing different geographical time-distances use graphs in two (Plassard and Routhier 1987; Tobler 1997) or three dimensions (Mathis 1990; Mathis, Polombo, and L’Hostis 1993; L’Hostis 2009). One could also model general surfaces with *Riemannian metrics* (i.e. metrics which, at a very local scale, are Euclidean up to order 1 approximations); but note that the marvelous *uniformization theorem* from the early XX<sup>th</sup> century shows that *any* such surface which is homeomorphic to a domain in the plane, can be represented by a distance  $d_R$  obtained as above *isometrically* (i.e. in a way that distances are perfectly preserved). In particular, any part  $Y$  of the earth that is not the whole can be represented in the plane by a map  $\varphi : Y \rightarrow X \subset \mathbb{R}^2$ , and a function  $R$  can be chosen, in a way that  $d_R(\varphi(p), \varphi(q))$  is exactly equal to the shortest path

distance in  $Y$  between  $p$  and  $q$ . For a detailed historical and mathematical account of the uniformization theorem, the reader can consult (Saint-Gervais 2016).

**Example 2.9.** The above examples are all classical in the geometry of metric spaces, and the only advantage of our framework seems to unify the discrete and continuous models. Let us now give an example that mixes both continuous and discrete aspects, and that we claim is quite satisfactorily modeled in our framework.

Let again  $X$  be a domain in  $\mathbb{R}^2$ , convex say, and assume we have two ways to travel in  $X$ : by car, where the constraints of the road network are assumed to be negligible, so that we modeled car travel as in Example 2.7; and by plane, with exactly two airports located at  $a_1, a_2 \in X$ , with a fast two-ways connexion between them.

The network  $N$  shall now be the union of the set  $N_g$  of all continuously differentiable arcs on  $X$  in the one hand, and of the set of the two connexions between the airports  $N_f = \{\alpha_{f:12}, \alpha_{f:21}\}$  ( $g$  stands for “ground” and  $f$  for “flight”). The starting and endpoints of arcs in  $N_g$  are defined as usual, and we set  $s(\alpha_{f:12}) = e(\alpha_{f:21}) = a_1$  and  $s(\alpha_{f:21}) = e(\alpha_{f:12}) = a_2$ .

The length function is given for  $\alpha \in N_g$  by the physically homogeneous formula of Example 2.7

$$\ell(\alpha) := \int_{T_0}^{T_1} R \|\alpha'(t)\| dt$$

where  $R$  is the inverse of the speed of ground travel; and assuming both aerial connexions take the same time  $T_f$ ,  $\ell(\alpha_{f:12}) = \ell(\alpha_{f:21}) = T_f$ .

Now, applying the framework above defines paths that can combine car travels with flights, and enable to construct the underlying optimized distance. Here, provided  $R > 0$  and  $T_f > 0$ , this distance is a metric, and shortest paths are either line segments, or combination of one or two line segments and one flight.

Of course, more complicated situations with more airports, more connexions (not necessarily between all possible pairs of airports), possibly varying ground travel difficulty, can be modeled in the same way. One can also model alternative ground travel means by taking  $N_g = N_{gb} \cup N_{gc}$  where  $N_{gb}$  and  $N_{gc}$  are two disjoint copies of the set of continuously differentiable curves, the first ones corresponding to bike travels and the second ones corresponding to car travel. Then one sets

$$\begin{aligned} \ell(\alpha) &:= \int_{T_0}^{T_1} R_b \|\alpha'(t)\| dt \quad \forall \alpha \in N_{gb} \\ \ell(\alpha) &:= \int_{T_0}^{T_1} R_c \|\alpha'(t)\| dt \quad \forall \alpha \in N_{gc} \end{aligned}$$

where  $R_b$  is the inverse of the biking speed, and  $R_c$  is the inverse of the driving speed. As such, if we make the reasonable assumption  $R_b > R_c$ , the introduction of biking does not change the distance as it is a slower mean of transport than car. But if one turns  $R_b$  and  $R_c$  into functions, with  $R_b(p) < R_c(p)$  when  $p$  lies in some regions, or if one changes the formula of the cost function to take into account the economical and ecological cost of gas, then this enriched model becomes relevant.

Note that compared to (Smith 1989), the present approach does notably not include axiom N3 (subpath closure) of Smith’s definition 3.1: arcs need not be restrictable into subarcs, and this better represent flights or any other travels that cannot in practice be decomposed. On the other hand, Smith’s path networks are particular cases of our definition of networks.

### 3 Contextual networks and metrics

We shall now enrich the above notions of space, network, and length by adding some contextual information. To motivate and explain the need for this enrichment, let us discuss how the notion of *break* has been considered incorrectly as an impediment to the universal validity of the triangular inequality (L’Hostis 2016a; L’Hostis 2016b). *Breaks* in itineraries are necessary to relaunch movement, so they do not entail the idea of sub-optimality in distances that the triangle inequality violation suggest. In particular (Huriot, Smith, and Thisse 1989, p. 313) admit that their proposed *minimum cost distance* may violate the triangle inequality. While they do not give an explicit example, it is plausible that what they have in mind is close to example 3.1 below.

Regarding this situation, two ways forward can be envisaged:

- i.* developing a semi-metric structure that assumes a violation of triangle inequality; these structures will be based on a relaxation of the triangle inequality replaced by a less demanding inequality (Wallace Alvin Wilson 1931)
- ii.* developing a framework where triangle inequality of metrics is always respected

The first approach has been indicated by (Huriot, Smith, and Thisse 1989). The renunciation to triangular inequality entails losing major geometrical properties, the inability to use the spatial analytical techniques of Geography (Ahmed and Harvey J. Miller 2007), and entering abstract spaces (W. A. Wilson 1932, p. 517). The second approach allows to remain in a metric domain concerning geometry, and allows to consider that optimisation is included in the idea of distance. For these reasons we chose the second approach.

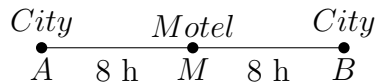


Figure 3: Two cities, a midway motel and the length of two paths  $AM$  and  $MB$

As extensively shown in (L’Hostis 2016b), several authors attach the idea of a human as an effort minimizer (Zipf 1949) to cost and “not just to distances” (Montello 1991, p. 113). We defend here an opposite view and propose to root the idea of optimisation into the definition of geographical distance and its mathematical formalisation. We refute the idea that geographical distance measurement could eventually be suboptimal. Departing from the point of view of (Huriot, Smith, and Thisse 1989, p. 313), we claim that as soon as distances are well-defined through an optimization process, the triangle inequality is satisfied. Apparent violations of the triangular inequality are either a lack of optimization, as in the “two errors” debunked in (L’Hostis 2016b), or a lack of well-definiteness as we will explain below. The reinterpretation of the following simple example described in (L’Hostis 2016b) is at the core of our approach.

**Example 3.1.** Assume, as on figure 3, two cities  $A, B$  are connected by a single road with a motel  $M$  in between; to travel by car between  $A$  and  $M$ , or between  $M$  and  $B$  takes 8 hours. Assume further that one is not able or allowed to drive for more than 8 hours in a row, and that after 8 hours of driving an 8 hours rest is needed. It has been argued that this is a counter-example to the triangular inequality: the distance (in travel time) from  $A$  to  $C$  would be 24 hours (twice 8 hours of driving and 8 hours of rest in between), larger than the sum of the distance from  $A$  to  $M$  and the distance from  $M$  to  $B$  (both 8 hours). It has been proposed to relax the triangular inequality in such a case (L’Hostis 2016b), in a way that could be modeled by replacing in the above framework the additivity of cost (1) by some other rule to determine  $\ell(\gamma_1 * \gamma_2)$ .

We argue that the problem here does not lie in the triangular inequality, but in the fact that distances (travel times) are not well-defined out of *context*. Geometry alone cannot explain the form of these path and of the corresponding metric spaces. In particular, it is not true in general that one standing at  $A$  can be at  $M$  8 hours later. Imagine the motel manager has to go to  $A$  fetch some commodities: neglecting the time involved in logistics, after her 8 hours drive to  $A$  she cannot drive back straight away. She thus stands in  $A$  with no way to be at  $M$  until 16 hours later. The travel time between any two points in fact depends on whether the traveler is rested or tired: there is a *context* to be taken into account. Modelling this is the goal of the

framework we are about to develop.

### 3.1 Contextual networks

Let  $X$  be a set called the *space*, and  $C$  be another set called the *context set*. Elements of  $X$  are called *locations*, while elements of  $C$  are called *contexts*. As before,  $X$  should contain only those points of the physical space that are meaningful locations;  $C$  should be made rich enough to model all elements of context that are relevant to travel cost. An element  $(p, c)$  of  $X \times C$  is called a *state*: it compounds the data of the location and the context.

Contextual networks will be defined exactly as networks, but on the set of states. We still repeat the definitions as the different roles played by  $X$  and  $C$  will have an importance in the modeling process and in the interpretation. The case of a singleton  $C = \{c_0\}$  will correspond to the above framework, while in example 3.1 we could take  $C = \{r, t\}$  ( $r$  for “rested”,  $t$  for “tired”).  $C$  can be a product space, to take into account several variables (gas gauge, accumulated fatigue, available visas, etc.)

A *contextual network* on  $X$  with context set  $C$  is a network  $N$  on the states set  $X \times C$ , i.e.  $N$  is a set endowed with two functions:  $s, e : N \rightarrow X \times C$ . Elements of  $N$  are still called *arcs*: given an arc  $\alpha$  we call  $s(\alpha)$  and  $e(\alpha)$  its *starting state* and *endstate*; we write  $s_X, e_X : N \rightarrow X$  and  $s_C, e_C : N \rightarrow C$  the functions defined by

$$s(\alpha) = (s_X(\alpha), s_C(\alpha)) \quad \text{and} \quad e(\alpha) = (e_X(\alpha), e_C(\alpha)) \quad \forall \alpha \in N.$$

As before, several arcs can be used to describe travels corresponding to the same displacement in the physical space; actually, most of the time a lot of different arcs will be needed for each movement in the physical space to take into account the starting context. Even for a given movement and a given starting context, several arcs can be used, e.g. one for fast driving and another for slow driving (with different ending contexts, notably in term of gas consumption and fatigue).

Given a contextual network  $N$  on a space  $X$  with context  $C$ , a *path*  $\gamma$  from  $(p, c) \in X \times C$  to  $(q, b) \in X \times C$  is a finite word  $\alpha_1 \alpha_2 \dots \alpha_k$  with letters in  $N$  such that  $s(\alpha_1) = (p, c)$ ,  $e(\alpha_k) = (q, b)$  and  $e(\alpha_i) = s(\alpha_{i+1})$  for each  $i \in \{1, 2, \dots, k-1\}$ . We write  $\gamma : (p, c) \rightarrow (q, b)$  to express that  $\gamma$  is some path from  $(p, c)$  to  $(q, b)$ , but we may want to forget some information by speaking of a path from  $p$  to  $q$ , writing  $\gamma : p \rightarrow q$ ; or of a path from  $(p, c)$  to  $q$  by writing  $\gamma : (p, c) \rightarrow q$ . The set of paths is again denoted by  $N^*$ , and includes a trivial path  $\varnothing_{p,c} : (p, c) \rightarrow (p, c)$  for each state  $(p, c)$ . Two paths  $\gamma : (p, c) \rightarrow (q, b)$  and  $\eta : (p', c') \rightarrow (q', b')$  are *chainable* if  $(p', c') = (q, b)$ , and the concatenation of two chainable paths is a path denoted by  $\gamma * \eta$ .

A *length function* on a contextual network  $N$  is a function  $\ell : N \rightarrow (0, +\infty)$  which we extend to a function on  $N^*$ , still denoted by  $\ell$ , by

$$\ell(\alpha_1 \cdots \alpha_k) = \sum_{i=1}^k \ell(\alpha_i)$$

with the same convention  $\ell(\emptyset_{p,c}) = 0$ .

Up to this point, there is *no* conceptual difference between the first framework with a simple network and the present contextual network. It is only when considering distances that the role of locations will be prominent compared to contexts: more often than not, distances between locations will be of primary interest, rather than distances between states.

### 3.2 Contextual distances

Given a space  $X$ , a context set  $C$ , a contextual network  $N$  and a length function  $\ell$ , we can construct several distances by optimization, depending on how the context is taken into account. Let the *semi-specific*, *specific*, *minimal* and *maximal contextual distances* be defined respectively by

$$\begin{aligned} d_c(p, q) &= \inf_{\gamma: (p,c) \rightarrow q} \ell(\gamma), \\ d_{c,b}(p, q) &= \inf_{\gamma: (p,c) \rightarrow (q,b)} \ell(\gamma), \\ d_{\min}(p, q) &= \inf_{\gamma: p \rightarrow q} \ell(\gamma), \\ d_{\max}(p, q) &= \sup_{c \in C} d_c(p, q). \end{aligned}$$

The *semi-specific* distance is lowest value achieved or approximated by the length of a path  $\gamma$  from  $p$  with a given starting context state  $c$ . In this definition the starting state is fully specified, while only the arriving *location* is constrained (this definition can also be inverted to care for situations where we want to fix the ending state, e.g. a time of arrival). This first distance will often be the object of interest: it disregards the context in which the traveler arrives at its destination, and only takes into account the context he or she starts in.

The *specific* distance is formally the same object as the non-contextual distance introduced above, with  $X \times C$  as space. In particular, it satisfies the triangular inequality in the sense that for all  $(p, c), (q, b), (r, a) \in X \times C$ ,

$$d_{c,a}(p, r) \leq d_{c,b}(p, q) + d_{b,a}(q, r).$$

It refers to a path  $\gamma$  that modifies the position in  $X$  from  $p$  to  $q$ , and at the same time, modifies the context in  $C$  from  $c$  to  $b$ .

In the motel example 3.1, what had originally been considered as the distance between two locations is in our framework their *minimal* distance: it corresponds to a best-case scenario and will in general not satisfy the triangle inequality.

The *maximal* distance corresponds to the worst-case scenario, and will always satisfy the triangle inequality, as we will show below.

**Theorem 3.2.** *For all contextual data  $(X, C, N, \ell)$  the following hold.*

- i. (Finiteness) *Given  $p, q \in X$  and  $c, b \in C$ , we have:*
  - (a)  $d_c(p, q) < +\infty$  if and only if there is a path  $\gamma : (p, c) \rightarrow q$ ,
  - (b)  $d_{c,b}(p, q) < +\infty$  if and only if there is a path  $\gamma : (p, c) \rightarrow (q, b)$ ,
  - (c)  $d_{\min}(p, q) < +\infty$  if and only if there is a path  $\gamma : p \rightarrow q$ ,
  - (d)  $d_{\max}(p, q) < +\infty$  if and only if there is a number  $A$  such that for all  $c \in C$ , there is a path  $\gamma : (p, c) \rightarrow q$  with  $\ell(\gamma) \leq A$ . In particular, if  $C$  is finite, then  $d_{\max}(p, q) < +\infty$  if and only if there is a path  $\gamma : p \rightarrow q$ .
- ii. (Symmetry) *If for all arc  $\alpha$ , there exist an arc  $\alpha'$  such that  $s_X(\alpha') = e_X(\alpha)$ ,  $e_X(\alpha') = s_X(\alpha)$  and  $\ell(\alpha') = \ell(\alpha)$ , then:  $d_{\min}(p, q) = d_{\min}(q, p)$  for all  $p, q \in X$ .*
- iii. (Triangle inequality) *for all  $p, q, r \in X$  and all  $c, b, a \in C$  we have*
  - (a)  $d_c(p, r) \leq d_c(p, q) + d_{\max}(q, r)$ ,
  - (b)  $d_{c,a}(p, r) \leq d_{c,b}(p, q) + d_{b,a}(q, r)$ ,
  - (c)  $d_{\max}(p, r) \leq d_{\max}(p, q) + d_{\max}(q, r)$ .
- iv. (Distinguishability) *If  $\ell$  is non-degenerate (note that Definition 2.3 applies word-for-word to contextual length functions), then for all  $p, q \in X$  and all  $c, b \in C$  we have*

$$d_c(p, q) = 0 \Leftrightarrow d_{\min}(p, q) = 0 \Leftrightarrow d_{\max}(p, q) = 0 \Leftrightarrow p = q.$$

Note that symmetry is not relevant for the semi-specific distance, and is covered by Theorem 2.5 for the specific distance. One can state conditions ensuring symmetry of the maximal distance, but they seem either not general enough or not simple enough to warrant a statement rather than a case-by-case analysis.



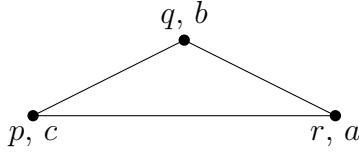


Figure 4: A triangle in the state space: each vertex is a state, i.e. a pair (location, context), and each edge is a path.

*Proof.* The four finiteness assertions are direct consequences of the hypothesis that length functions take finite values on arcs. The symmetry assertion for  $d_{\min}$  has *mutatis mutandis* the same proof as the symmetry in Theorem 2.5.

Let us now prove the triangle inequalities; we fix  $\varepsilon > 0$  and in each case assume the contextual distances in the right-hand side are finite, as otherwise the statement is vacuously true.

There is a path  $\gamma : (p, c) \rightarrow q$  such that  $\ell(\gamma) \leq d_c(p, q) + \varepsilon$ ; let  $b$  be the context at its endpoint. By definition  $d_b(q, r) \leq d_{\max}(q, r)$ , so there is a path  $\eta : (q, b) \rightarrow r$  such that  $\ell(\eta) \leq d_b(q, r) + \varepsilon \leq d_{\max}(q, r) + \varepsilon$ . Now  $\gamma * \eta : (p, c) \rightarrow r$  has length at most  $d_c(p, q) + d_{\max}(q, r) + 2\varepsilon$ , so that letting  $\varepsilon$  go to zero we get  $d_c(p, r) \leq d_c(p, q) + d_{\max}(q, r)$ . By taking the supremum over  $c \in C$ , we get precisely  $d_{\max}(p, r) \leq d_{\max}(p, q) + d_{\max}(q, r)$ . The case of the specific distance follows from Theorem 2.5, seeing the contextual network as a non-contextual network on  $X \times C$  (see Figure 4).

Concerning distinguishability, simply observe that  $p = q$  implies each of the contextual distances to be zero, while  $p \neq q$  implies each of them to be at least  $m(p, q) > 0$ .  $\square$

### 3.3 Examples

Let us now see how much flexibility we gain from this framework in a bunch of examples. First, given a space  $X$ , a network  $N$  and a length function  $\ell$ , one can take  $C = \{0\}$  and get in the obvious way contextual network and cost function. We thus do not lose any generality in adding context. We are of course more interested in less trivial examples.

**Example 3.3.** Let us revisit the motel example 3.1. There are several ways to model this example in our framework, we propose the following:

- $X = \{A, B, M\}$ ,  $A, B$  representing the two cities and  $M$  the motel in between them,

- $C = \{r, t\}$  where  $r$  is for “rested” and  $t$  for “tired”,
- $N = \{\alpha_{IJ}: I \neq J \in X\} \cup \{\rho_I: I \in X\}$ , with  $s(\alpha_{IJ}) = (I, r)$ ,  $e(\alpha_{IJ}) = (J, t)$ ,  $s(\rho_I) = (I, t)$  and  $e(\rho_I) = (I, r)$ ; the  $\alpha_{IJ}$  correspond to driving from  $I$  to  $J$ , which needs one to be rested and provokes fatigue, while  $\rho_I$  consists in resting at  $I$ ,
- $\ell(\alpha_{IJ}) = \ell(\rho_I) = 8$  (in hours) for all  $I, J$ .

Now we have  $d_{\min}(A, M) = d_r(A, M) = 8$  and  $d_{\min}(M, B) = d_r(M, B) = 8$ , but  $d_{\min}(A, B) = d_r(A, B) = 24$ : the shortest path  $A \rightarrow B$  starts from a rested context and is  $\alpha_{AM}\rho_M\alpha_{MB}$ . We see that neither  $d_{\min}$  nor  $d_r$  satisfy the triangle inequality. But for example  $d_{r,r}(A, B) = 32 = d_{r,r}(A, M) + d_{r,r}(M, B)$ , and we can check more generally that  $d_{r,r}$  is a metric. We also have  $d_{\max}(A, B) = 32 = d_{\max}(A, M) + d_{\max}(M, B)$ : the worst case in each travel is when one is tired, implying an 8 hours rest before driving (and an 8 hours rest between the two segments when one goes from  $A$  to  $B$ ).

**Example 3.4.** Assume that we are given a space  $X$  of cities, a (non-contextual) network of roads  $N$  on  $X$  (with  $s, e$  as starting and endpoint functions) and a length function  $\ell$  describing the duration of driving each road, in hours. Assume we want to take into account a legislation that imposes a resting time of duration  $P > 0$  after driving for a time  $D > 0$ . This situation can be observed in the freight road transport with driving rules for lorry drivers (Chapelon 2006).  $D$  is the amount of fatigue in hours of driving that makes a break necessary or mandatory; in the case of freight road transport this level is fixed by professional rules, while in the domain of individual road transport this level are indicated by good practice.

If we assume that a driver can rest anywhere, even along a road, then instead of having *driving* arcs and *resting* arcs as in example 3.3, it is simpler to construct the following contextual data (all contextual objects will be denoted with a  $\tilde{\phantom{x}}$  to distinguish them from the original ones).

- The space is still  $X$ , and the set of context is  $\tilde{C} = [0, D]$  representing the fatigue in hours of driving since the last rest.
- The contextual network is  $\tilde{N} = \{\tilde{\alpha}_c: \alpha \in N, c \in \tilde{C}\}$ , with starting and endpoint maps defined by  $\tilde{s}(\tilde{\alpha}_c) = (s(\alpha), c)$  and  $\tilde{e}(\tilde{\alpha}_c) = (e(\alpha), c + \ell(\alpha) \bmod D)$  where  $x \bmod D$  means the representative in  $(0, D]$  of the class of  $x$  modulo  $D$ , i.e.  $x \bmod D := x - kD$  whenever  $k$  is the integer with  $kD < x \leq (k + 1)D$ . We say that  $k$  is the *number of breaks* in  $\tilde{\alpha}_c$ . Note that  $c + \ell(\alpha) > 0$  since by

assumption  $\ell(\alpha) > 0$ , and we take the representative in  $(0, D]$  rather than  $[0, D)$  because when one arrives at destination with fatigue  $D$ , resting is usually not considered part of the travel. This contextual distance is relevant to describe time-distance of travel. Nevertheless, from the description of trips generated with this framework, it is quite easy to infer a cost function if one wants to take into account the arriving context.

- The contextual length function is given by  $\tilde{\ell}(\tilde{\alpha}_c) = \ell(\alpha) + kP$  where  $k$  is the number of breaks.

Note that in this case, we can compute contextual distances from the non-contextual distance: for all  $p \neq q \in X$ ,  $\tilde{d}_c(p, q) = d(p, q) + kP$  where  $k$  is the number of breaks, i.e. the integer such that  $d(p, q) - kD \in (0, D]$ . We remark that here, whenever  $d$  is symmetric, the maximal and minimal contextual distances are both symmetric too.

**Example 3.5.** Assume again that we are given a space  $X$  of cities (and possibly other locations such as motels), a (non-contextual) network of roads  $N$  on  $X$  (with  $s, e$  as starting and endpoint functions) and a length function  $\ell$  describing the duration of driving each road, in hours, and that again we want to take into account a legislation that imposes a resting time of duration  $P > 0$  after driving for a time  $D > 0$ . However, assume that rest is only possible at a location, not in between; or worse, resting is only possible at *some* locations: we denote by  $R \subset X$  the set of locations where one can rest. This situation is closer to the observed practices of lorry driving especially on motorways where stop is only allowed in dedicated service areas. Then we shall model this situation as in the simple motel example above.

- The space is still  $X$ , and the set of context is  $\tilde{C} = [0, D]$  representing the fatigue in hours of driving since the last rest.
- The contextual network is  $\tilde{N} = \{\tilde{\alpha}_c : \alpha \in N, c \in C \mid c + \ell(\alpha) \leq D\} \cup \{\rho_{p,c} : p \in R, c \in C\}$ , with starting and endpoint maps defined by  $\tilde{s}(\tilde{\alpha}_c) = (s(\alpha), c)$ ,  $\tilde{e}(\tilde{\alpha}_c) = (e(\alpha), c + \ell(\alpha))$  and  $\tilde{s}(\rho_{p,c}) = (p, c)$ ,  $\tilde{e}(\rho_{p,c}) = (p, 0)$ .
- The contextual length function is given by  $\tilde{\ell}(\tilde{\alpha}_c) = \ell(\alpha)$  and  $\tilde{\ell}(\rho_{p,c}) = P$ .

The computation of contextual distances now needs the full contextual data: we cannot recover it from the non-contextual distance only. Indeed, what was an optimal path in the non-contextual case can now be unavailable (e.g. if there was an arc with length greater than  $D$ ); or it could become non-optimal e.g. because of the presence of pairs of consecutive arcs both of length slightly above  $D/2$ , making it mandatory to rest in between them while some other path has slightly greater driving time,

but optimizes rests to happen near the limit  $D$ . Note that in some cases we can get non-symmetric contextual distances. For example, service areas that are only accessible from one side of a motorway can break the symmetry. Symmetry almost never exists in geographical spaces. Moreover, this framework can be extended to for instance caring simultaneously for breaks linked to resting of the driver, and for breaks linked to reloading the fuel consumed by vehicles.

The location of break points, or stations, is tightly related to the spatial patterns of trips and represents a highly relevant feature of geographical spaces. The caravanserais, playing the role of halting places, where the major infrastructure of the Silk Road (Williams 2014). The need for the first break for Parisian travellers to the south of France has fuelled the economic specialisation of Burgundy in the touristic sector around hotels and restaurants that shaped the reputation of the region (Bavoux 2009). The availability of – and the market served by – petrol stations on motorways is directly linked to the consumption of fuel by vehicles on the trunk lines. These examples show that a suitable framework caring for rest and energy reloading for movement sustaining, analysed with the contextual distance, is of significant value for spatial analysis.

**Example 3.6.** This next example models travel using scheduled trips between various stations of a public transportation system, such as flights between airports, or trains or buses between stations. Our space  $X$  will consist of finitely many stations, while our context set  $C$  will be a subset of  $\mathbb{R}$ , considered as a time variable. Our network  $N$  will contain two kinds of arcs:  $N = T \cup W$  where

- each scheduled trip will correspond to an arc  $f \in T$  such that  $s_X(f)$  and  $e_X(f)$  are the departure and arrival stations of the flight service while  $s_C(f)$  and  $e_C(f)$  are the departure and arrival time of the trip, note that  $s_C(f) < e_C(f)$ .
- for every station  $x \in X$  and each couple  $(t_1, t_2) \in \mathbb{R}^2$  such that  $t_1 \leq t_2$ , there is a path  $w_{x,t_1,t_2}$  such that  $s(w_{x,t_1,t_2}) = (x, t_1)$  and  $e(w_{x,t_1,t_2}) = (x, t_2)$  which models waiting at station  $x$  from time  $t_1$  to  $t_2$ . We will denote by  $W$  the set of paths coming from waiting time.

As in the previous case, non symmetry already arises from the network itself : our definition of  $N$  ensures that every path  $\gamma : (x_1, t_1) \rightarrow (x_2, t_2)$  satisfies  $t_1 < t_2$ . Hence for every two airports  $x_1$  and  $x_2$  and any two times  $t_1$  and  $t_2$  such that  $t_1 > t_2$ , the specific distance  $d_{t_1,t_2}(x_1, x_2)$  will be infinite for every choice of length function  $\ell$  whereas  $d_{t_2,t_1}(x_2, x_1)$  can be finite. Hence except in very degenerate cases the specific distance will not be symmetric. The semi-specific distance need not be symmetric, since the flights from  $A$  to  $B$  and  $B$  to  $A$  need not be synchronized. The maximal and

minimal distances need not be symmetric either, since the waiting time at connections can be different for the return trip.

One length function of interest in this setting is the duration of the trip  $\ell$  defined for  $\alpha \in N$  by  $\ell(\alpha) = e_C(\alpha) - s_C(\alpha)$ ; we immediately get the same expression  $\ell(\gamma) = e_C(\gamma) - s_C(\gamma)$  for all  $\gamma \in N^*$ . In this example the four contextual distances have different meanings :

- the semi-specific distance  $d_{t_1}(x_1, x_2)$  will be  $T - t_1$  where  $T$  is the infimum of the possible arrival times of a trip from  $x_1$  to  $x_2$  which leaves  $x_1$  after time  $t_1$  (if no such trip exists,  $T = +\infty$ ).
- the specific distance  $d_{t_1, t_2}(x_1, x_2)$  will be equal to  $t_2 - t_1$  if there is a trip from  $x_1$  to  $x_2$  which takes place during the time interval  $[t_1, t_2]$  and to  $+\infty$  otherwise.
- the minimal distance  $d_{\min}(x_1, x_2)$  will be equal to the infimum of the duration of a trip from  $x_1$  to  $x_2$ .
- the maximal distance will have properties which depend a lot on the travel network : in particular if for every flight  $f \in F$ ,  $s_C(f) \geq 0$  (no trip starts before time 0), taking  $C = \mathbb{R}$  would yield  $d_{\max}(x_1, x_2) = +\infty$  whenever  $x_1$  and  $x_2$  are different. Here the modeling choices are important to obtain relevant distances.

Let us be more specific and consider 3 airports  $x$ ,  $y$  and  $z$  such that :

- there is a flight from  $x$  to  $y$  taking off at 10 and landing at 11.
- there is a flight from  $y$  to  $z$  taking off at 12 and landing at 13.
- there is a flight from  $x$  to  $z$  taking off at 12 and landing at 14.

Here the space  $X$  is  $\{x, y, z\}$ , the context is  $C = [9, 15]$  considered as a time variable and the network  $N$  is built from the timetable as described above. In this example :

$$d_{\min}(x, z) = 2, \quad d_{\max}(x, z) = +\infty, \quad d_t(x, z) = \begin{cases} 13 - t & \text{if } t \leq 10 \\ 14 - t & \text{if } 10 < t \leq 12 \\ +\infty & \text{if } 12 < t. \end{cases}$$

The appearance of infinite distance should not be treated as a weakness of the example : it carries information. The infiniteness of  $d_t(x, z)$  for  $t > 12$  shows that  $z$  is not reachable if the trip starts later than time  $t = 12$ .

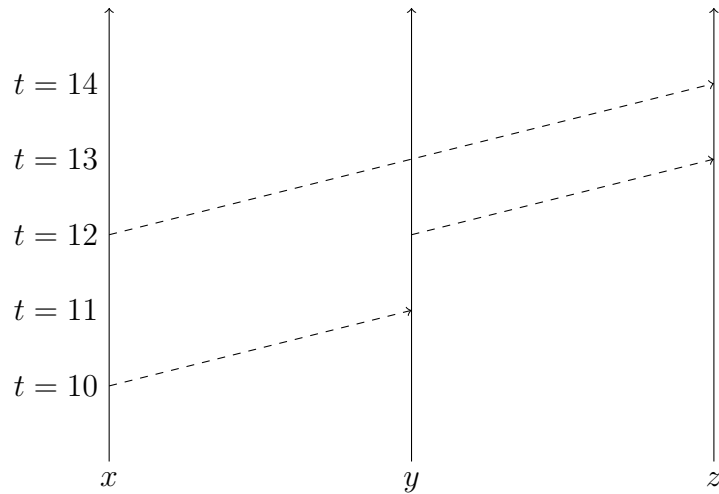


Figure 5: Representation of the contextual network from example 3.6. The space  $X$  is represented horizontally while the context is represented vertically. Any subsegment of the vertical solid arrows is a valid path in the network, while the dashed path cannot be subdivided.

The computation of  $d_{t_1}(x_1, x_2)$  for a real public transportation network, defined as time-dependant networks, is an important problem in practice and a very active field of research, see (Müller-Hannemann et al. 2007) for timetable algorithms and (Hall 1986; Delling et al. 2009; Bast et al. 2016) for routing algorithms in time-dependant graphs. The semi-specific distance formula is relevant for the issue of a given starting time, but can also be inverted to deal with the problem of a given desired time at destination. These two cases cover the main problems posed by time-dependant graphs. Note that two approaches can be found in this field: in the time extended approach the problem of computing  $d_{t_1}(x_1, x_2)$  is tackled by choosing shortest path in the static graph whose vertices are couples  $(x_i, t_i)$ , while in the time dependent approach the underlying graph has fixed vertices  $x$  but dynamic edges which appear and disappear as time progresses according to available trips. In our framework they corresponds to considering either the distance on the space  $X \times \mathbb{R}$  of all states or the semi-specific contextual distance on the space  $X$ .

The length function we have presented here only takes into account the travel time, but the flexibility of our framework allows to take into account various criteria. One can for instance take as a length function the total price of the trip by assigning to each arcs in  $W$  a length of 0 and each arc in  $T$  the price of the ticket. One can also penalize connections by assigning to each possible trip  $t \in T$  a length

$\ell(t) = e_C(t) - s_C(t) + p$  where  $p$  is a fixed penalty.

## 4 Conclusion

A large part of our effort lies in the introduction of the description of actual trips into the definition of distances. In our view, geographical distance is not exclusively a matter of geometry, it is also related to physical movement in geographical space.

We propose to root the idea of effort minimisation into the definition of geographical distance and its mathematical formalisation. This objective is non-trivial to achieve notably because of the issue of possible triangle inequality violation due to the need for *break* in any movement. This issue translates into additivity problems that may disturb the order of proximities in geographical spaces, and undermine their basic geometrical properties. In order to address this issue, we introduce the concept of *contextual distances* that, in parallel to pure geometric movement, considers a context where resources used for – or related to – movement, are accounted for, that it be time, money or other relevant quantities.

We show that contextual distances follow many properties of metrics, in particular they respect the triangle inequality. This result paves the way for its use within the context of spatial analysis in geography.

Future work could examine the observation of triangle inequality on empirical datasets. Time, kilometres and cost distances by various transport modes could be measured, and then tested regarding triangle inequality. The investigation could consider these datasets as contextual distances, adjusting contextual and non contextual parameters, and test whether this mathematical framework allows to better understand the geometry of geographic spaces. In this direction, our affirmation that the triangle inequality is always satisfied opens more questions than it closes: first, in some spaces the triangle inequality is almost an equality for most triple of points while in other spaces a much stronger inequality is almost always true; this could be used as a geometric property classifying geographical spaces. Are there geographical spaces which seem very different from a geographic perspective but which share such geometric properties? On the contrary, are there geographical spaces that are thought of as close one to each other, but in fact have very different geometric behavior in this respect?

Second, one could consider other inequalities, possibly involving more than three points. One could for example classify spaces by the constraints its geometry entails for the 6-tuple

$$\left(d(p, q), d(p, r), d(p, s), d(q, r), d(q, s), d(r, s)\right)$$

when  $(p, q, r, s)$  runs over all 4-tuples of point in the given space. Searching various sets of geometric inequalities that, when verified by a given space, ensures the planar representation of this space up to some reasonable error would also be an interesting endeavour: it would inform map designer by telling them when no good map can exist, or rather how bad they must be, in a variety of contexts far beyond the mapping of Earth with its geometric distance. These findings could prove valuable in the domain of the representation of time-distances, where stress control is a key issue.

## 5 Acknowledgements

The authors thank T. Smith for encouraging them to pursue the idea of introducing a formula of distance that would care for the triangle inequality property.

This research is conducted under the umbrella of the project “Voyageur virtuel” funded by Ifsttar and Université Paris-Est.

## References

- [1] Nobbir Ahmed and Harvey J. Miller. “Time–space transformations of geographic space for exploring, analyzing and visualizing transportation systems”. In: *Journal of Transport Geography* 15.1 (2007), pp. 2–17. ISSN: 0966-6923. DOI: [10.1016/j.jtrangeo.2005.11.004](https://doi.org/10.1016/j.jtrangeo.2005.11.004).
- [2] Kay W. Axhausen et al. “Constructing Time-Scaled Maps: Switzerland from 1950 to 2000”. In: *Transport Reviews* 28.3 (2008), pp. 391–413. DOI: [10.1080/01441640701747451](https://doi.org/10.1080/01441640701747451).
- [3] Hannah Bast et al. “Route Planning in Transportation Networks”. en. In: *Algorithm Engineering: Selected Results and Surveys*. Ed. by Lasse Kliemann and Peter Sanders. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 19–80. ISBN: 978-3-319-49487-6. DOI: [10.1007/978-3-319-49487-6\\_2](https://doi.org/10.1007/978-3-319-49487-6_2).
- [4] Jean-Jacques Bavoux. “Structuration du territoire bourguignon : de l’île à l’isthme”. fr. In: *Revue Géographique de l’Est* 46.1-2 (2009). ISSN: 035-3213. URL: <http://rge.revues.org/1195#tocto2n7>.
- [5] William Bunge. *Theoretical geography*. seconde éd. augmentée 1966. Lund: Gleerup, 1962.



- [6] Laurent Chapelon. “L’accessibilité, marqueur des inégalités de rayonnement des villes portuaires en Europe”. fr. In: *Cybergeog : European Journal of Geography* (2006). ISSN: 1278-3366. DOI: [10.4000/cybergeog.2463](https://doi.org/10.4000/cybergeog.2463).
- [7] Walter Collischonn and Jorge Victor Pilar. “A direction dependent least-cost-path algorithm for roads and canals”. In: *International Journal of Geographical Information Science* 14.4 (2000), pp. 397–406. DOI: [10.1080/13658810050024304](https://doi.org/10.1080/13658810050024304).
- [8] Daniel Delling et al. “Engineering route planning algorithms”. In: *Algorithmics of large and complex networks*. Springer, 2009, pp. 117–139. ISBN: 978-3-642-02094-0. URL: [https://doi.org/10.1007/978-3-642-02094-0\\_7](https://doi.org/10.1007/978-3-642-02094-0_7).
- [9] Tamás Dusek and Gábor Szalkai. “Time- spaces in Hungary”. In: *Raportul spațiu-timp în Ungaria*. 16.2 (2017), pp. 161–170. ISSN: 15831523. DOI: [10.5775/fg.2017.009.d](https://doi.org/10.5775/fg.2017.009.d).
- [10] Péter Ficzer, Zita Ulmann, and Ádám Török. “Time–space analysis of transport system using different mapping methods”. In: *Transport* ahead-of-print (2014), pp. 1–7. DOI: [10.3846/16484142.2014.916747](https://doi.org/10.3846/16484142.2014.916747).
- [11] Peter Haggett. *Geography, a global synthesis*. Harlow: Prentice Hall, 2001.
- [12] Randolph W. Hall. “The fastest path through a network with random time-dependent travel times”. In: *Transportation science* 20.3 (1986), pp. 182–188. DOI: [10.1287/trsc.20.3.182](https://doi.org/10.1287/trsc.20.3.182).
- [13] Jean-Marie Huriot, Tony E. Smith, and Jacques François Thisse. “Minimum-Cost Distances in Spatial Analysis”. en. In: *Geographical Analysis* 21.4 (1989), pp. 294–315. ISSN: 1538-4632. DOI: [10.1111/j.1538-4632.1989.tb00898.x](https://doi.org/10.1111/j.1538-4632.1989.tb00898.x).
- [14] G.M. Hyman and L. Mayhew. “Advances in travel geometry and urban modelling”. In: *GeoJournal* 59 (2004), pp. 191–207. DOI: [10.1023/B:GEJO.0000026689.48422.96](https://doi.org/10.1023/B:GEJO.0000026689.48422.96).
- [15] Hildebert Isnard et al. *Problématiques de la géographie*. français. SUP, Le géographe. Paris: Presses universitaires de France, 1981. ISBN: 2-13-036800-X.
- [16] J. Kruskal and M. Wish. *Multidimensional scaling*. Beverly Hills/London: Sage Publications, 1978.
- [17] Jacques Lévy. “Entre contact et écart: la distance au coeur de la réflexion”. In: *Atala* 12 (2009), pp. 175–185.
- [18] Alain L’Hostis. “Detour and break optimising distance, a new perspective on transport and urbanism”. In: *Environment and Planning B: Planning and Design* 44.3 (2016), pp. 441–463. DOI: [10.1177/0265813516638849](https://doi.org/10.1177/0265813516638849).

- [19] Alain L’Hostis. “Misunderstanding geographical distances: two errors and an issue in the interpretation of violations of the triangle inequality”. In: *Cybergeo : European Journal of Geography* 793 (2016). DOI: [10.4000/cybergeo.27810](https://doi.org/10.4000/cybergeo.27810).
- [20] Alain L’Hostis. “The shrivelled USA: representing time-space in the context of metropolitanization and the development of high-speed transport”. In: *Journal of Transport Geography* 17.6 (2009), pp. 433–439. ISSN: 0966-6923. DOI: [10.1016/j.jtrangeo.2009.04.005](https://doi.org/10.1016/j.jtrangeo.2009.04.005).
- [21] Bernard Marchand. “Deformation of a transportation surface”. In: *Annals of the Association of American Geographers* 63 (1973), pp. 507–521. DOI: [10.1111/j.1467-8306.1973.tb00944.x](https://doi.org/10.1111/j.1467-8306.1973.tb00944.x).
- [22] Philippe Mathis. “Espace et graphe, le p-graphe t-modal 1-planaire”. FR. In: Chamonix, 1990, p. 10. URL: <https://hal.archives-ouvertes.fr/hal-02070946/>.
- [23] Philippe Mathis, Nadine Polombo, and Alain L’Hostis. “Les Grandes vitesses”. In: *Circuler demain. Monde en cours, Prospective et territoires*. La Tour d’Aigues: DATAR Editions de l’Aube, 1993, pp. 129–142.
- [24] Harvey J Miller. “A Measurement Theory for Time Geography”. en. In: *Geographical Analysis* 37.1 (2005), pp. 17–45. ISSN: 1538-4632. DOI: [10.1111/j.1538-4632.2005.00575.x](https://doi.org/10.1111/j.1538-4632.2005.00575.x).
- [25] Daniel R. Montello. “The measurement of cognitive distance: Methods and construct validity”. In: *Journal of Environmental Psychology* 11.2 (1991), pp. 101–122. DOI: [10.1016/S0272-4944\(05\)80071-4](https://doi.org/10.1016/S0272-4944(05)80071-4).
- [26] Matthias Müller-Hannemann et al. “Timetable Information: Models and Algorithms”. en. In: *Algorithmic Methods for Railway Optimization*. Ed. by Frank Geraets et al. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2007, pp. 67–90. ISBN: 978-3-540-74247-0.
- [27] François Plassard and Jean-Louis Routhier. *Sémiologie graphique et évaluation*. Lyon: A.R.T.U.R., 1987.
- [28] Henri Paul de Saint-Gervais. *Uniformization of Riemann Surfaces*. 2016. ISBN: 978-3-03719-145-3. DOI: [10.4171/145](https://doi.org/10.4171/145).
- [29] R.N. Shepard. “The Analysis of proximities: multidimensional scaling with an unknown distance function”. In: *Psychometrika* 27.2 (1962), pp. 125–140. DOI: [10.1007/BF02289630](https://doi.org/10.1007/BF02289630).

- [30] Eihan Shimizu and Ryo Inoue. “A new algorithm for distance cartogram construction”. In: *International Journal of Geographical Information Science* 23.11 (2009), pp. 1453–1470. ISSN: 13658816. DOI: [10.1080/13658810802186882](https://doi.org/10.1080/13658810802186882).
- [31] Tony E. Smith. “Shortest-Path Distances: An Axiomatic Approach”. In: *Geographical analysis* 21.1 (1989), pp. 1–31. DOI: [10.1111/j.1538-4632.1989.tb00874.x](https://doi.org/10.1111/j.1538-4632.1989.tb00874.x).
- [32] Waldo Rudolph Tobler. “Visualizing the impact of transportation on spatial relations”. In: *Western Regional Science Association meeting*. Hawaii, 1997, p. 7.
- [33] Devin A. White and Sarah B. Barber. “Geospatial modeling of pedestrian transportation networks: a case study from precolumbian Oaxaca, Mexico”. In: *Journal of archaeological science* 39.8 (2012), pp. 2684–2696. DOI: [10.1016/j.jas.2012.04.017](https://doi.org/10.1016/j.jas.2012.04.017).
- [34] Tim Williams. *The Silk Roads: An ICOMOS Thematic Study*. EN. Monographie. UNESCO WHC, 2014, p. 152. URL: [http://www.icomos.org/images/mediatheque/ICOMOS\\_WHThematicStudy\\_SilkRoads\\_final\\_lv\\_201406.pdf](http://www.icomos.org/images/mediatheque/ICOMOS_WHThematicStudy_SilkRoads_final_lv_201406.pdf).
- [35] W. A. Wilson. “A Relation Between Metric and Euclidean Spaces”. In: *American Journal of Mathematics* 54.3 (1932), pp. 505–517. ISSN: 0002-9327. DOI: [10.2307/2370894](https://doi.org/10.2307/2370894).
- [36] Wallace Alvin Wilson. “On Semi-Metric Spaces”. In: *American Journal of Mathematics* 53.2 (1931), pp. 361–373. ISSN: 0002-9327. DOI: [10.2307/2370790](https://doi.org/10.2307/2370790).
- [37] George Kingsley Zipf. *Human Behaviour and the Principle of Least-Effort*. Cambridge, USA: Addison-Wesley, 1949. URL: <http://www.citeulike.org/group/2050/article/1283136>.