Shrinking Complexity:
Some Heuristics for Contractible Spaces

Eduardo Francisco Rêgo

Abstract

This paper deals with the existence, or absence, of certain "simple" contractions of a contractible space; the main concept involved is that of a *coalescent contraction*: when the tracks of any two points meet, at time t_0, they remain together thereafter. Examples of contractible spaces where no coalescent contractions exist are simplicial complexes, like the *dunce hat*, with no *free* simplices. We introduce two notions for measuring the complexity of a homotopy: its *volume* and its *energy*. We relate the problem of the existence of a coalescent contraction with ways of minimalizing the volume and energy, in a suitable space of contractions. The main results and problems - considered as a start for a research project, and for some of which we advance supporting heuristics based on the analysis of the three dimensional cases - are about the existence of coalescent contractions in compact, triangulable, contractible manifolds.

1 Introduction

(...) Mathematics rigorously treated from [the] point of view [of] deducing theorems exclusively by means of introspective construction, is called intuitionistic mathematics...

(...) I hope I have made clear that intuitionism on the one hand subtilizes logic, on the other hand denounces logic as a source of truth. Further that intuitionistic mathematics is inner architecture, and that research in foundations of mathematics is inner inquiry with revealing and liberating consequences, also in non-mathematical domains of thought.

(Brouwer, L. E. J.,[1])

This paper is about contractible spaces and our main purpose is to search for *simple* contractions. Of course, we have to establish meanings to the word 'simple', in this context. One such meaning is given by the notion of *coalescence*: we say that a homotopy, $H : X \times I \longrightarrow Y$, is *coalescent* if

$$H(a, t_0) = H(b, t_0) \ \Rightarrow H(a, t) = H(b, t) \ \forall t \geq t_0$$

that is, when the tracks of any two points, $a, b \in X$ meet, at time t_0, they remain together thereafter; other meanings are relative to two notions we introduce, in section 4, as a way of measuring the complexity of a homotopy: its *volume* and its *energy*.

In section 2 we revisit the *dunce hat*: this is the simplest example of a contractible finite simplicial complex which is non-collapsible, [2]. It is not collapsible, simply because there is no free simplex to start a collapse. The easiest proof that it is contractible goes like this: one readily sees that the dunce hat can be embedded in euclidian space, $\mathbb{E}^3$, where it is a (strong) deformation retract of a neighbourhood (a regular neighbourhood) homeomorphic to the 3-ball,

B^3, and is therefore contractible [1]. While it is reasonably easy to visualize, in $\mathbb{E}^3$, the embedding of the dunce hat, its thickening to a 3-ball and the corresponding deformation, it is far more difficult to visualize an actual contraction of the dunce hat, a least until one gets used to that mental exercise. What we shall do, in section 2, is to describe, in intrinsic terms, a contraction of the dunce hat; our intention is that the picture thus formed will serve as a role model, to be contrasted with the heuristic arguments that lie ahead.

In section 3 we prove that the dunce hat doesn't have any coalescent contractions. This comes as a consequence of the dunce hat not having *free* simplices, which also prevents it from being collapsible; of course, for a collapsible space, K, we can associate to each collapse, $K \searrow *$, a coalescent contraction. The arguments used in this section can be generally described as variations on Brower's fixed point theorem and degrees of maps between spheres; for simplicity, we focus on the dunce hat, but the arguments can be generalized for other finite simplicial complexes with no free simplices.

In section 4 we introduce the notions of *volume* and *energy* of a homotopy. These are intended to be measures of the waste, that is unnecessary moves, in the timely process of a space contracting in itself: in very simple heuristic terms, the volume measures the *folding* of the space and the energy measures the *wandering* of the point paths. Given a contractible, finite simplicial complex, K, we will think mainly in terms of simplicial contractions, $H : (K \times I)^{sub} \longrightarrow K^{sub}$, the upper-scripts denoting simplicial subdivisions, including the baricentric ones, denote by (r) [1], and their limits in the space $\mathcal{C}$ of contractions endowed with the *supreme metric.* Volume and energy are lower semi-continuous functions in that space, therefore the existence of contractions with minimal energy and volume follows when we restrict to a suitable compact subspace, $\mathcal{G}$.

In section 5 we consider the case when the contractible finite complex K is a *manifold*: we sketch a general argument, some details of which are essentially heuristic - since we only work them in dimension three - to provide evidence and support to the conjecture that coalescent contractions exist in this case when a certain geometric condition is assumed - *convexity at the boundary.* The idea is to consider homotopies with minimal energy and volume and argue that, for those, the nature of a manifold, endowed with a metric which is *locally geometric* - in the sense that geodesics intersect transversely - doesn't allow non coalescence to occur: although the spaces $\mathcal{G}$ are quite restricted in the space of all possible contractions, they still allow enough local flexibility for the arguments to be carried through. This is intended to start a research project on the subject of coalescence; We end the paper with some general remarks on the relations between, coalescence, collapsibility and compactness.

This is the second paper on a series started with [6]. In there, a longer list of references can be found, as well as a short introduction to contractible spaces and the Poincaré conjecture and its history, which are relevant to the issues dealt with in the present paper. As a matter of fact, the notions of volume and energy we are using, were inspired by the mechanical picture underlying the 'heuristic tour' we took in that paper, and actually can be used to give mathematical content to some of the heuristic views in that tour; but the present paper is totally independent and can be read without any reference to that previous one; it is also, we hope, much less heuristic.

[1] For the background notions just invoked (simplicial complexes, collapsible complex, regular neighbourhoods, etc.), see [3, Chapter 2] for a quick introduction and [4], [5] for deeper treatments.

2 From inside the dunce hat I: contraction

Recall that the dunce hat, D, is the quotient space obtained from a triangle (2-simplex) by identifying the three oriented edges where the orientations of two of the edges are pointing to the same vertex. See Figure 1, bellow.

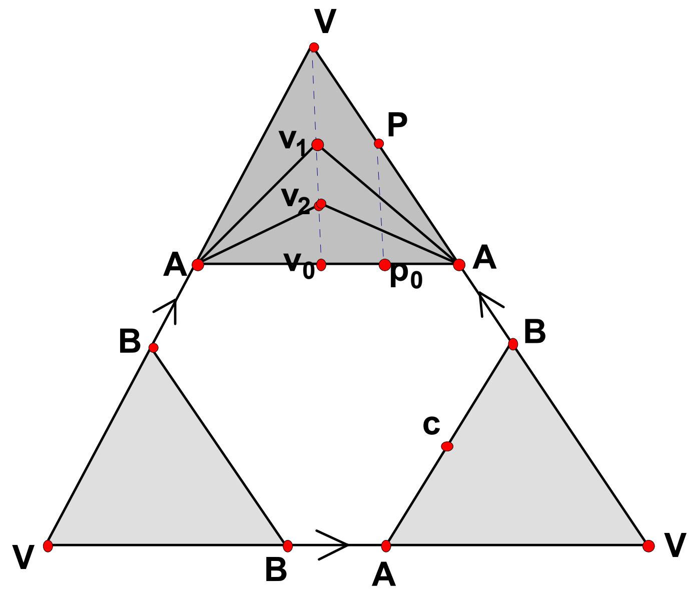

Figure 1

The three vertices of the triangle are all identified to the same point V of D, which we call the *vertex* of the dunce hat: as we shall see, this point plays an essential part in several properties of D. Represented in shade is a typical neighbourhood of V: it is easy to see that it is homeomorphic to a disc - the bottom right triangle AVC - with two cones glued along generators, VA and VB: see Figure 2 for a picture of an embedding of that neighbourhood in three space.

We start now, describing a contraction of the dunce hat. We look first at the cone corresponding to the top triangle in Figure 1. We divide it into three regions, depicted in Figures 1 and 2, as follows. Consider the perpendicular segment from V to the foot V_0, and divide it in thirds by the points V_1, V_2. Consider next the two pairs of lines from A to V_1 and V_2: each gives a loop based in A and going through V_i. The perpendicular from a generic point P, in the generator AV, to its foot P_0, gets divided into thirds by the intersections with those loops, P_1 and P_2: one such perpendicular is represented in Figure 2.

We construct the contraction of D as a sequence of several homotopy movements which we describe in order.

The first movement fixes all points outside the cone, and also those on the base of the cone, that is on loop AV_0A, as well as on the generator AV and moves just those points in the interior of the top triangle, in Figure 1. The tracks of the points in the interior of the triangle follow the perpendicular segments: for each perpendicular segment PP_0, the homotopy fixes the end points, shrinks the segment P_2P to the end point P and stretches the segment P_0P_2 to a homeomorphism into P_0P. At the end of the movement, for any point V' in the perpendicular segment V_2V each of the two arcs from A to V' is sent homeomorphically to AV: so that the region between the loops AV_0A and AV_2A is stretched to give the original cone by identification of the two arcs AV_2 to AV, while the complementary region of the cone closes down to AV - as a fan made of those loops $AV'A$.

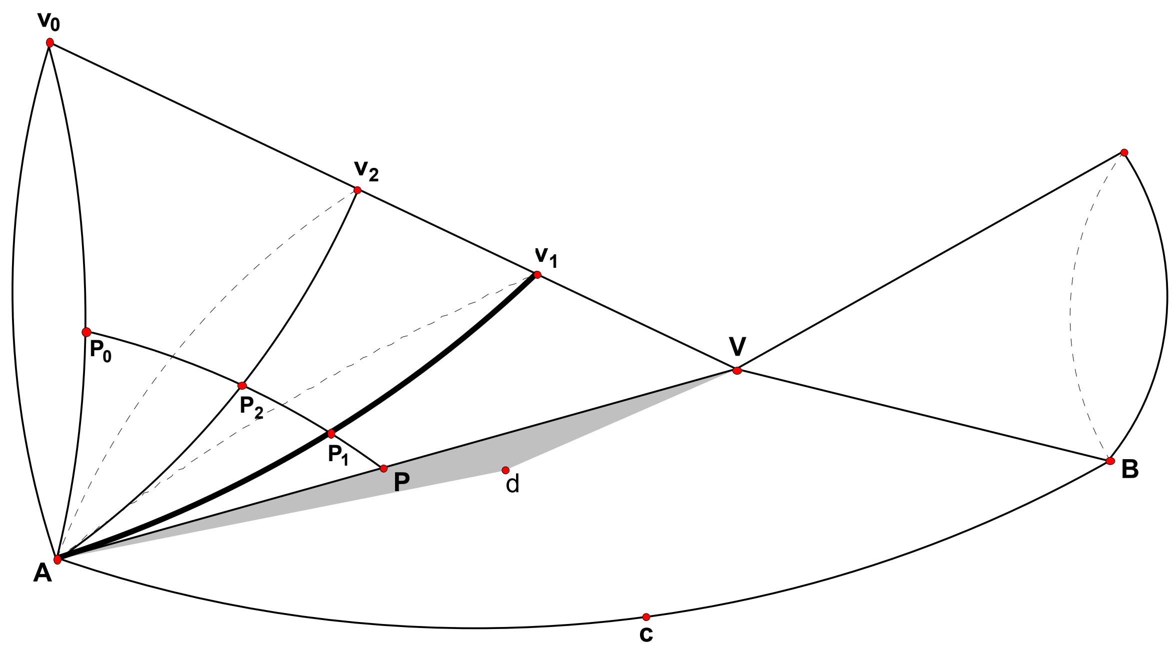

Figure 2

The second part of the homotopy keeps moving just those points interior to the "fan region" just mentioned, and the movement takes place entirely within the segment AV, where they all lie by the end of the first movement. Recall that for a generic $P \in AV$, the "perpendicular segment" P_2P was squeezed to P; now, the points P_1 move along AV, from P to A and correspondingly the two segments PP_1 and P_2P_1 expand again, keeping the other end points P and P_2 fixed at P, to fill the segment PA homeomorphically. Note that at the end of this process, the loop AV_1A is shrunk to the point A. In Figure 3 bellow, we try to give a visual impression of the combination of the first two homotopy movements, just described. But note that it is not a realistic picture: the two arrows to the right, pointing downward, suggest rightly the first movement, that of closing the fan and the third arrow, pointing to the left, suggests the second movement, with the thick loop AV_1A shrinking in the direction of point A; but in this respect it is a fake picture, since the second movement happens all inside the segment VA: lets say it's an "infinitesimal vision" from inside AV, allowing us to keep seeing all the elements of the closed fan - like the points P_2 and P or the segments P_2P_1 and P_2P_1in the picture - set apart ! Anyway, if the reader doesn't find this picture helpful he/she may simply ignore it.

The third part of the homotopy starts by pushing the segment AV, keeping the end points fixed, slightly inside the disc that corresponds to the bottom right triangle AVB, sending it to an inside arc AdV, as represented in Figure 2: by 'pushing' we mean to perform a strong deformation retraction of the disc through the shaded area. In the process, we must drag along all the points from the interior of the top most region of the triangle - the part of the fan above loop AV_1A - that by the end of the first two movements were all laid inside the segment AV: the important point now is that since the loop AV_1A was shrunk to the point A that now remains fixed, the dragging gives a continuous function!

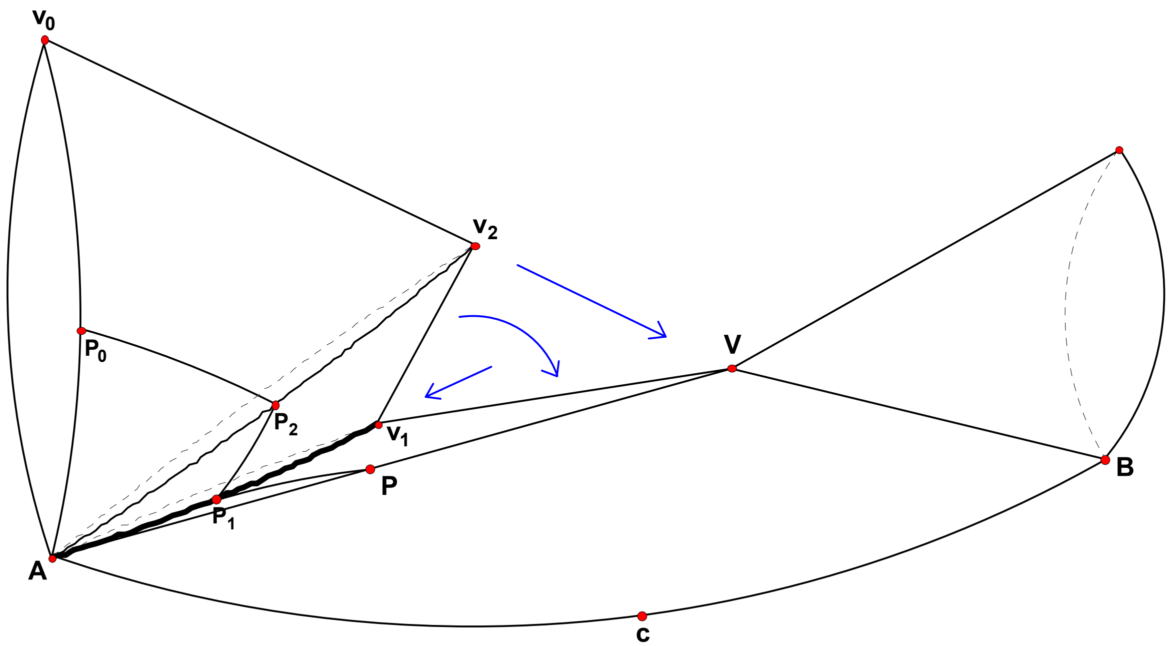

Figure 3

This is the stage where the image of the dunce-hat by the homotopy is no longer surjective. Figure 4 represents the state of affairs at this point.

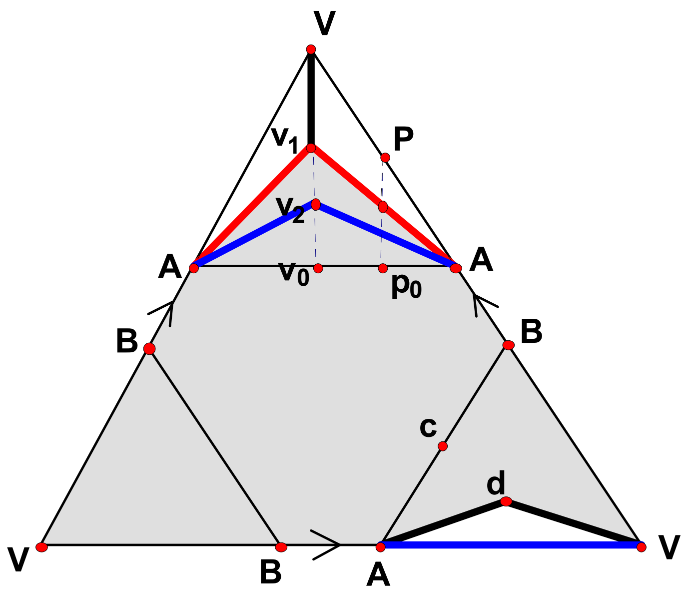

Figure 4

The image of the dunce-hat D, after the small deformation retraction of disc AVB from AV across the disc ΔAVd, is the image after the first two movements of the shaded region plus the segment AV which remains the image, after those previous movements, of the upper region enclosed by the loop AV_2A - recall that each one of the arcs AV_2 was mapped homeomorphically into AV - and the loop AV_1A which was shrunk to A; the image of the top region is now the arc AdV, with the segment V_1V - represented thick - mapped homeomorphically from A to V.

The interior of the triangle ΔAdV is what is missing from the image of D at this point.

The fourth part of the contraction of D is the continuation of the previous small deformation retraction all the way across the interior of the shaded region, from the arc AdV to the complementary part of the boundary of that region, as suggested in Figure 5 bellow, composed with the final map of the first two movements. At the end of this deformation, the image of D will just be the image of segment VA plus the image of arc AV_1A by that map. That map was the identity on VA and on arc AV_1A was the constant map to A. Therefore the image of D at the end of this fourth part will be an arc, namely VA.

Since an arc is a contractible space, the fifth and last part of the contraction of D will be a contraction of the arc VA: note that if we choose this final contraction to be a strong deformation retraction to the end point V, this point - the vertex of D - will have remained fixed through out all the contraction of D.

This finishes the description of a contraction of the dunce hat, as seen from the inside.

We should note that many points that meet at a certain stage of the contraction - in the terminology of the introduction, they *coalesce* - separate again at later stages. As we shall see in the next section, that phenomenon can't be avoided in any contraction of D.

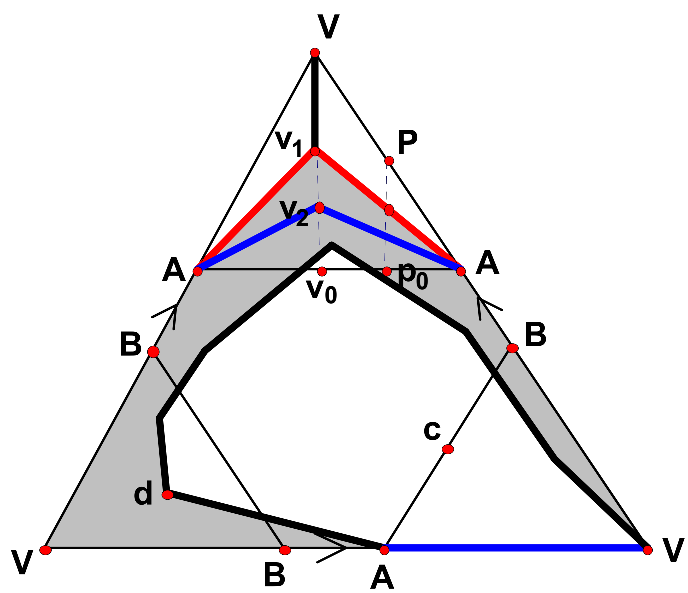

Figure 5

Note: In the first two parts of the contraction we only moved points in the cone with generator AV, until we were able to dig a hole near the vertex V, by pushing along AV; note that we can take the cone - equivalently the generator AV - to be arbitrarily small. If we consider an *infinitesimal* cone (more formally, the uniform limit of a sequence of homotopies defined, as the one before, on a sequence of cones converging to vertex V...) the hole near V, that is non-surjectiveness, appears right from the start of the homotopy: $\forall \varepsilon > 0,\ \exists t < \varepsilon\ : H_t(D) \neq D$, where $H_t, t \in [0,1]$ denotes the contraction. That is not always the case for contractible spaces; the reader may convince himself of that fact after solving the following exercise on another famous contractible non-collapsible space, Bing's house with two rooms (Figure 6)

The house is built in the following way. Start with a closed rectangular box, and add a slab in the middle; then add two more boxes, as shown in Figure 6, one from the mid slab to the top and another one from the mid slab to the bottom; remove then the base and top from each of these two chimneys. You got a house with two separated rooms: you can get into the bottom room through the top entrance and into the top room from the bottom entrance (see[7]). Finally add two walls, one in each room as shown shaded in Figure 6. You can easily see that a regular neighbourhood of this object in $\mathbb{R}^3$ is a 3-ball, therefore it is contractible; but it is not collapsible since there are no free simplices.

Exercise 1 *Describe a contraction of Bing's house: the essential point is the art of digging a hole somewhere, as we did for the dunce-hat...*

***Hint**: having experienced with the dunce hat D before, what you should expect now is to be able to dig a hole near one of the "special" corners of the house where, as in the vertex V of D, there is some complexity in the way the several walls meet at that point (intuitively, and as we shall see later, one can not do such a thing as digging a hole near a manifold point); but, after some trials, you will soon find that you can not act locally near just one such point, instead you will have to work in a larger region with several of them.*

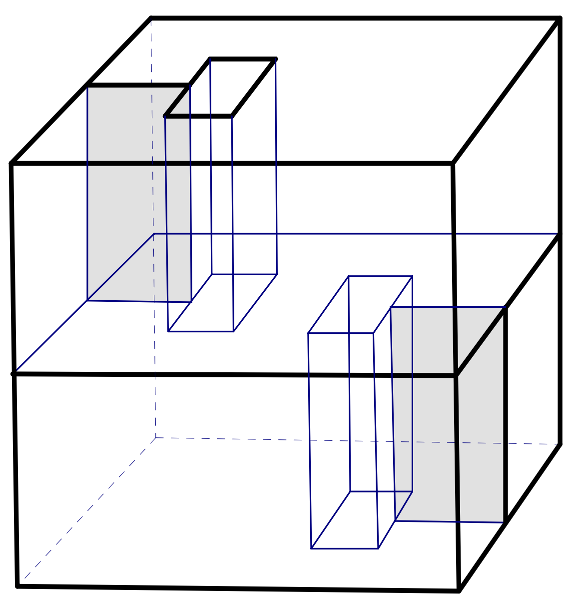

Figure 6

3 From inside the dunce hat II: coalescence

We have already said, in the introduction, what it means for a homotopy $H : X \times I \longrightarrow Y$ to be *coalescent*:

$$H(a, t_0) = H(b, t_0) \Rightarrow H(a, t) = H(b, t), \forall t \geq t_0$$

that is, if the the tracks of two points, a and b, meet at time t_0, they remain together thereafter.

What we shall establish, in this section, is the impossibility of coalescent contractions of the dunce-hat; the arguments apply equally well to other contractible spaces, for instance Bing's house. Before proceeding, we want to draw attention to a specially nice feature of coalescent contractions.

From now on we will assume, unless otherwise stated, that all spaces are finite simplicial complexes, although most of the arguments can be carried over to more general classes of spaces, especially compact metric spaces.

Let Y be a contractible space and $H : Y \times I \longrightarrow Y$ a contraction that is coalescent. Let, for each $t \in I$, Y_t be the image of Y at time t: $Y_t = H_t(Y) = H(Y \times t)$. Since H is continuous, each Y_t is a compact subspace of Y, homeomorphic to the identification space Y/R where R is the relation of coalescence at time t, $aRb \Leftrightarrow H_t(a) = H_t(b)$. The nice feature of coalescent contractions we were alluding to is that given any Y_{t_0}, what we see during the contraction, from time t_0 until the end, is also a contraction of subspace Y_{t_0} in the ambient space Y. This is an immediate consequence of transgression for *identification spaces* that we recall:

Theorem 2 *Let $f : X \longrightarrow Z$ be a continuous map and $p : X \longrightarrow Y$ an identification map such that $h = fp^{-1}$ is well defined (that is, f is constant on each fibre $p^{-1}(y)$).Then:*

a) *$h : Y \longrightarrow Z$ is a continuous map.*

b) *In the particular case when $Y = X_f = X/K_f$ with K_f the relation induced by f, defined by $aK_f b \Leftrightarrow f(a) = f(b)$, h is injective. Therefore each continuous map, f, factors through the composition of a continuous surjection and a continuous injection.*

c) *If, furthermore, f is surjective, then $h = fp^{-1} : X_f \longrightarrow Z$ is a continuous bijection and is a homeomorphism, $h : X_f \cong Z$, if and only if f is an identification map.*

Proof. *See, for instance, [8, Chapter VI].* ■

We can now give the precise formulation of that nice feature we described above:

Remark 3 *Let Y_{t_0} be the image of Y at time t_0, and let $p : Y \times [t_0, 1] \longrightarrow Y_{t_0} \times [t_0, 1]$ be defined by $p(y, t) = (H_{t_0} y, t)$. Since we are assuming that H is* coalescent, *$J = p^{-1}H$ is single-valued, thus defines a function $J : Y_{t_0} \times [t_0, 1] \longrightarrow Y$. Because $Y \times [t_0, 1]$ is compact, p is an identification map and so, by a) of the previous theorem, it is continuous. Furthermore, it's clear that for each $y \in Y_{t_0}$, $J(y, t_0) = y$, that is J_{t_0} is the identity map on Y_{t_0}, and since H_1 is constant we have that J_1 is also constant and therefore J realizes a contraction of Y_{t_0} in the ambient space Y. Clearly, J is also coalescent.*

Let $H : Y \times I \longrightarrow Y$ be a contraction. Define the *opening time*, t_{op}, by $t_{op} = \sup \{t : Y_s = Y \ , \ \forall s \leq t\}$. The opening time is the time when space Y first opens, in the sense that we start seeing non-surjectiveness. It is an easy exercise to show that $Y_{t_{op}} = Y$.

After our construction, in the previous section, of a contraction for the dunce-hat D, we explained in a note how it could be modified, through a limiting process, to yield a contraction H such that $\forall\varepsilon > 0,\ \exists t < \varepsilon\ : H_t(D) \neq D$. This means that $t_{op} = 0$, the opening time is 0. The next result shows that if there is a coalescent contraction, there is a contraction with opening time 0.

Theorem 4 *Let $H : Y \times I \longrightarrow Y$ be a coalescent contraction. Then there is a (coalescent) contraction K_s, $s \in I$, of Y with opening time $s_{op} = 0$.*
Proof. *Let t_{op} be the opening time of H and consider $Y_{t_{op}}$. As we've just stated above, $Y_{t_{op}} = Y$: with the usual notation for $\varepsilon > 0$ neighbourhoods of subspaces $A \subset Y$, $N_\varepsilon(A) = \{y \in Y : d(y, A) < \varepsilon\}$, where d is a metric for Y and $d(y, A) = \inf\{d(y, a) : a \in A\}$, we have: given an arbitrary Y_{t_0} and neighbourhood $N_\varepsilon(Y_{t_0})$, using the uniform continuity of H, due to the domain being compact, we can get a $\delta > 0$ such that $|t - t_0| < \delta \Rightarrow Y_t \subset N_\varepsilon(Y_{t_0})$; in particular, if $Y_{t_0} \neq Y$, since Y_{t_0} is compact, we have $d(a, Y_{t_0}) > 0$ for $a \in Y - Y_{t_0}$ and so taking $\varepsilon < d(a, Y_{t_0})$ we have that $Y_t \neq Y$ for all t such that $|t - t_0| < \delta$; therefore, by the definition of opening time, it has to be the case that $Y_{t_{op}} = Y$.*

Refer back to the discussion in the remark above: with $p : Y \times [t_{op}, 1] \longrightarrow Y_{t_{op}} \times [t_0, 1] = Y \times [t_{op}, 1]$, we have that $J = Hp^{-1} : Y \times [t_{op}, 1] \longrightarrow Y$ realizes a (coalescent) contraction of Y and clearly, by definition of t_{op} as a supremum, $\forall\varepsilon > 0,\ \exists t < t_{op} + \varepsilon\ : J_t(Y) \neq Y$. The final contraction K_s, $s \in I$, is obtained by an obvious reparametrization of J: $t = s(1 - t_{op}) + t_{op}$; clearly, $s_{op} = 0$. ∎

We are now in the position of establishing the non-existence of coalescent contractions of the dunce-hat D. But we look first at Bing's house; let's denote it B. Let H be a contraction of B and assume that it is coalescent: by the previous theorem we can assume, without loss of generality, that the opening time for H is 0, that is $\forall\varepsilon > 0,\ \exists t < \varepsilon\ : H_t(B) = B_t \neq B$. Let t_n, $n \in \mathbb{N}$, be a sequence convergent to 0, such that $B_{t_n} \neq B$ for all n, and, for each n, choose some point $c_n \in B - B_{t_n}$; without loss of generality, considering a subsequence if necessary, we can assume that c_n converges to a point, say $c_n \longrightarrow c$. So, as soon as the contraction of B gets started, we see points arbitrarily close to c that are missed at arbitrarily early stages.

Assume B is endowed with a triangulation by 2-simplices linearly embedded in 3-space: this means that the corners of the house are vertices of the triangulation and that the edges of the house are contained in the 1-skeleton. Looking at the *star* of point c, $St(c)$, the union of all simplices that contain c, we can assume without loss of generality that the points c_n all belong to the same simplex $\sigma \in St(c)$. It should be clear, looking at B, that we can always get an embedded disc Δ, made up of simplices of $St(c)$ including σ, and with c in its interior. Here interior means interior of the manifold - the points not in boundary circle $\Sigma = \partial\Delta$ - not the interior of Δ as a subspace of B: for those points of B which are not manifold points - there are eight of them at the corners of the two shaded walls in Figure 6 - we have several ways of constructing Δ. We can also assume that, alongside with c, all the c_n are also in the interior of Δ. Now, we just need to look at (the first stages of) the contraction of disc Δ in B, that is, to look at the restriction of H to $\Delta \times I$.

Let us first recall some results about self-maps of n-balls and spheres. Given the n-ball $B^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ and an interior point x, let $r_x : B^n - \{x\} \longrightarrow S^{n-1}$ denote the usual retraction given by *radial projection* from x.

Lemma 5 *Let x_n, $n \geq 1$, be a sequence of interior points of B^n converging to an interior point x_0 and N a closed neighbourhood of S^{n-1} disjoint from the set of points $x_i, i \geq 0$. Then the* *restrictions to* N *of the set of retractions* $\{r_{x_n}\}_{n\geq 0}$ *is* equicontinuous, *that is,*

$$\forall\varepsilon > 0, \exists\delta > 0 : \forall x \in N, \forall n \geq 0, \|r_{x_n}(x) - x\| < \varepsilon$$

Proof. *We leave the proof as an exercise for the reader.* ■

The next result says that continuous self-maps of the sphere which are sufficiently close to the identity are homotopic to the identity. In fact more can be said:

Lemma 6 *Let* $f : S^{n-1} \longrightarrow S^{n-1}$ *be a continuous map such that no point is sent to its antipodal, that is,* $f(x) \neq -x, \forall x \in S^{n-1}$*. Then* f *is homotopic to the identity.*
Proof. *Consider for each point* $x \in S^{n-1}$ *the segment in* B^n *with end points* x *and* $f(x)$*. We leave it for the reader to check that we obtain a homotopy* $H : S^{n-1} \times I \longrightarrow S^{n-1}$ *between the identity and* f *by sliding from* x *to* $f(x)$ *along those segments and compose with the radial projection,* r*, from the origin - more specifically* $H(x,t) = r(x + t(f(x) - x))$ *(the track of each point is the (shortest) arc of great circle that joins it to its image).* ■

Consider again the disc Δ and the contraction H. Through some specific homeomorphism $(\Delta, \Sigma) \longrightarrow (B^2, S^1)$ we get the analogous of the two previous lemmas for the pair (Δ, Σ), namely a set of retractions $r_{c_n} : \Delta - \{c_n\} \longrightarrow \Sigma$, $n \geq 0$ (with $c_0 = c$), which is *equicontinuous* on any closed neighbourhood $N(\Sigma)$ disjoint from the $c_i, i \geq 0$, and a positive constant k such that any continuous map of $f : \Sigma \longrightarrow \Sigma$ with $d(f, id_\Sigma) \leq k$
(d the sup metric) is homotopic to id_Σ. Fix a α-closed neighbourhood of Σ in Δ, $N_\alpha(\Sigma) = \{x \in \Delta : d(x, \Sigma) \leq \alpha\}$ which doesn't contain any of the points c_i: by equicontinuity, there exists $\delta > 0$ such that $\forall x \in N_\alpha(\Sigma), \forall n \geq 0, \|r_{c_n}(x) - x\| < k/2$. Let, for each $\varepsilon > 0$, M_ε denote the ε-neighbourhood of Δ in B: it is clear that for sufficiently small ε there is a (strong deformation) retraction r^ε of M_ε into Δ, such that, for all $y \in M_\varepsilon - \Delta$, $r^\varepsilon(y) \in \Sigma$ and $d(y, r^\varepsilon(y)) \leq \varepsilon$. Choose such an ε, with $\varepsilon \leq \min\{\alpha, \delta, k/2\}$. By uniform continuity of H, there is a time $t_0 > 0$ such that for all $x \in B$ and for all $t \leq t_0$ we have $d(x, H(x,t)) < \varepsilon$; choose a time $t_{c_n} < t_0$, call it s to simplify notation and let f be the restriction of H_s to the disc Δ. Let $g : \Delta \longrightarrow \Sigma$ be defined by $g = r_{c_n} \circ r^\varepsilon \circ f$: by the choice of $s = t_{c_n} < t_0$, $f(\Delta) \subset M_\varepsilon$, so $r^\varepsilon \circ f$ is well defined. Let $x \in \Sigma$: if $f(x) \in M_\varepsilon - \Delta$ we have $g(x) = r^\varepsilon \circ f(x)$ and $d(f(x), r^\varepsilon f(x)) \leq \varepsilon$, so $d(x, g(x)) \leq d(x, f(x)) + d(f(x), r^\varepsilon \circ f(x)) \leq \varepsilon + \varepsilon \leq k$; if $f(x) \in \Delta$, $g(x) = r_{c_n} \circ f(x)$, and since $d(x, f(x)) < \varepsilon \leq \min\{\alpha, \delta, k/2\}$, we have that $f(x) \in N_\alpha$; therefore, since $d(x, f(x)) < \delta \Rightarrow d(f(x), r_{c_n} \circ f(x)) < k/2$ we have $d(x, g(x)) \leq d(x, f(x)) + d(f(x), r_{c_n} \circ f(x)) \leq k/2 + k/2 = k$. In both cases we conclude that $\forall x \in \Sigma$, $d(g(x), x) < k$ and so, by the choice of K above, the restriction of g to Σ is homotopic to the identity and so has *degree* 1; on the other hand, being the restriction of a map defined on the disc Δ it is nulhomotopic and so has *degree* 0, contradiction[2].

We have thus finished a proof that no contraction of space B has opening time 0 - a result we hinted at in the note of the previous section - and as a corollary that no contraction of B is coalescent:

Theorem 7 *No contraction of Bing's house,* B*, has* opening time $t_{op} = 0$*; therefore no contraction of* B *is coalescent.*

[2] The reader who hasn't yet learned about this most important notion of *degree* - and the associated results on maps between spheres - introduced by Brower, is urged to do so: see [8, Chapters XV-XVII] or books on algebraic topology, for instance [3], [9], [10].

We should stress what was the key factor in the previous proof: given any point c where space B "opens up" (at time 0) and whatever simplex, $\sigma \in St(c)$, contains points arbitrarily close to c from the complements of images $H_t(B)$, it is possible to get an embedded disc Δ containing σ and with c an interior point.

Definition 8 *We say a point c in a simplicial complex C has the* star-disc *property if for each simplex $\sigma \in St(c)$ there is an embedded disc Δ that contains σ and with c an interior point (note that, in particular, this property implies that there are no free simplices and so C is non-collapsible)*

It is straightforward - we leave it as an exercise for the reader - to generalize the proof to higher dimensions to get:

Theorem 9 *Let C be a finite, contractible simplicial complex such that every point c has the* star-disc *property (therefore C is non-collapsible)*

Then, every contraction of C has opening time $t_{op} > 0$ and so is non-coalescent.

The dunce-hat D doesn't satisfy the hypothesis of the theorem: there is exactly one point which doesn't have the required property, the vertex V. Referring back to Figures 1 and 2, it is clear that a simplex in $St(V)$ that comes from the right-bottom part, like triangle AVB, can not be extended to an embedded disc with V an interior point. This fits in well with our note in the previous section, where it was observed that our construction of a contraction of D would have opening time 0 if it was done with an *infinitesimal* cone; on the other hand any contraction of D with $t_{op} = 0$ will have to open up at the vertex V since it's easily seen that any point other than V has the star-disc property.

Although the dunce-hat has contractions with opening time 0, it doesn't have any coalescent contractions, but to settle this some further arguments are needed to deal with the special situation of the vertex. We start by establishing a lemma to the effect of getting rid of a certain type of possible wilderness of the track of vertex V under a contraction H of D: for technical reasons, that will become apparent in the course of the final proof, we do not want the track $H(\{V\} \times I)$ to cover the complements of images $H_t(D)$ that arise after opening time.

Lemma 10 *Let H be a (coalescent) contraction of D and p any manifold point of D (that is, p is any point in the interior of the identification triangle in Figure 1). Then there is a (coalescent) contraction, J, such that the track of V under J misses p.*

Proof. *Let $\gamma(t) = H(V,t)$. Let B be a closed neighbourhood of p, homeomorphic to a disc (see Figure 7) and disjoint from V.*

Let u_o be any time such that $\gamma(u_o) = p$,
$t_0 = \inf\{t \in I : \gamma([t, u_0]) \subset B\}$ and $s_0 = \sup\{t \in I : \gamma([u_0, t]) \subset B\}$; this means that in the interval $[t_0, s_0]$ the track of V stays in disc B. Note that Figure 7 is an over simplified picture, with $\gamma([t_0, s_0])$ represented by a very simple polygonal arc: the point is that it could be much complicated, even with $\gamma([t_0, s_0])$ filling the whole of B ! If there is another time u_1 outside this interval such that $\gamma(u_1) = p$, we have another interval, $[t_1, s_1]$, similarly defined and disjoint from the first: it's clear that we can repeat this process until we get a finite number of disjoint closed intervals $[t_i, s_i]\, , i = 0, ..., m$ whose union contains the pre-image $\gamma^{-1}(p)$. For each $i = 0, ..., m$ we modify the homotopy H in each time period $[t_i, s_i]$. Starting with $[t_0, s_0]$: let E be another disc surrounding B as shown in Figure 7; consider an isotopy *of D, $K : D \times I \longrightarrow D$ which is the identity outside E and on E is the identity on the boundary and moves interior points*

so as to push disc B to a final position $k(B)$, $k = K_1$, disjoint from point p; as suggested in Figure 7, we can further assume that the isotopy is also fixed in an neighbourhood of an arc in ∂B that contains the two points $\gamma(t_0) = H_{t_0}(V)$ and $\gamma(s_0) = H_{s_0}(V)$: just expand the shaded portion of the ring, keeping the complementary white part fixed, while pushing disc B along the corresponding boundary arc.

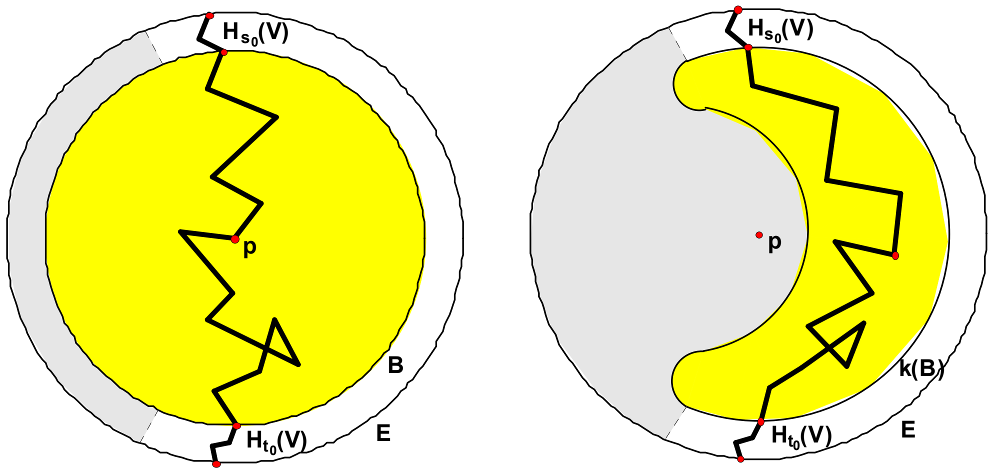

Figure 7

Let $N_\varepsilon(\gamma(t_0))$ and $N_\varepsilon(\gamma(s_0))$ be neighbourhoods of those two points, contained in E and fixed through the isotopy. Let $\delta_0 > 0$ be sufficiently small so that $[t_0 - \delta_0, s_0 + \delta_0]$ is still disjoint from the other intervals $[t_i, s_i]$ and such that $\gamma([t_0 - \delta_0, t_0]) \subset N_\varepsilon(\gamma(t_0))$, $\gamma([s_o, s_0 + \delta_0]) \subset N_\varepsilon(\gamma(s_0))$. Now we modify H in the time period $[t_0 - \delta_0, s_0 + \delta_0]$: consider the reparametrizations $s(t) = (1/\delta_0)\,[t - (t_0 - \delta_0)]$, $t_0 - \delta_0 \le t \le t_0$, and $r(t) = (1/\delta_0)\,[t - s_0]$, $s_0 \le t \le s_0 + \delta_0$, and define $J : D \times I \longrightarrow D$ by

$$
\begin{aligned}
J(x,t) &= H(x,t) \text{ if } t \le t_0 - \delta_0 \\
J(x,t) &= K(H(x,t), s(t)) \text{ if } t_0 - \delta_0 \le t \le t_0 \\
J(x,t) &= k(H(x,t)) \text{ if } t_0 \le t \le s_0 \\
J(x,t) &= K(H(x,t), 1 - r(t)) \text{ if } s_0 \le t \le s_0 + \delta_0 \\
J(x,t) &= H(x,t) \text{ if } s_0 + \delta_0 \le t
\end{aligned}
$$

In short, between times $t_0 - \delta_0$ and t_0 we combine, through a suitably reparametrization, H with the isotopy K, next from t_0 until s_0 we compose H with the final map $k = K_1$ of that isotopy, and finally between times s_0 and $s_0 + \delta_0$ we combine H with the time reversed isotopy. It should be clear that under this new homotopy, J, the track of V misses p during period $[t_0, s_0]$ and was unchanged before or after. We next repeat the process, in turn, for the other time periods $[t_1, s_1]$,$[t_2, s_2]$, ..., $[t_m, s_m]$. The proof ends by noting that, since each changing of H was achieved by combining it with an isotopy, *if H is* coalescent *so J will be.* ■

With the previous lemma in hand, we can now adjust the arguments we used in proving theorem 7 to the exceptional situation of vertex V in D.

Let H be a *coalescent* contraction of D. Choose any *manifold* point p in a complement of an image $D - H_t(D)$, for some $t > t_{op}$ (of course we could assume, without loss of generality, that $t_{op} = 0$ but there is no gain in assuming that). By the previous lemma, there is a coalescent contraction J such that the track of V misses p. By usual reasoning with uniform continuity and compactness we can assume there is a neighbourhood of V, $N_\varepsilon(V)$, such that the track of each of its points also misses p, that is $p \notin J(N_\varepsilon(V) \times I)$. Refer back to Figures 1 and 2 that represent a neighbourhood of V: consider in each of the two cones a closed segment of a generator with end point V and contained in $N_\varepsilon(V)$ and glue to the dunce-hat, by an homeomorphism, a disc E along the union of both segments as represented in Figure 8.

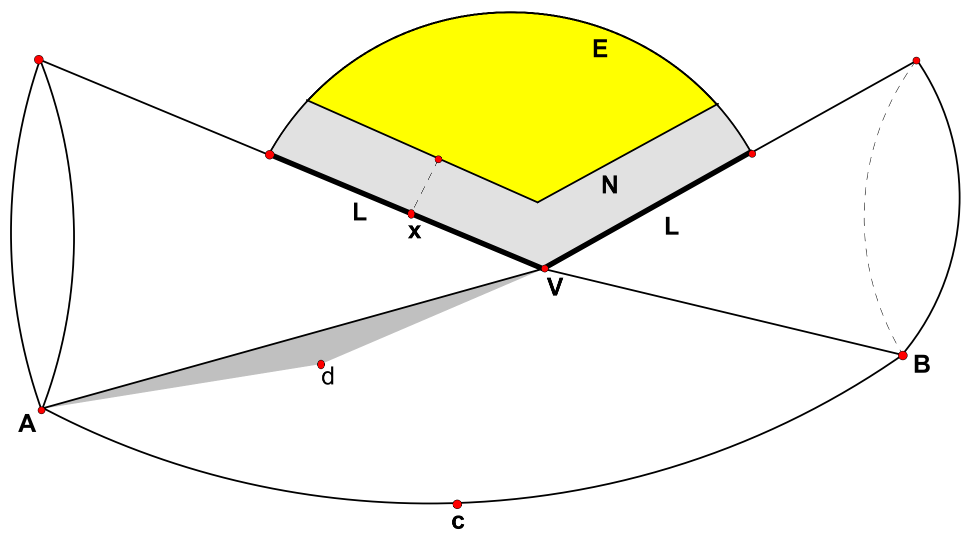

Figure 8

Labelled L, the union of those two segments is represented thick. Denote the resulting space by F.

Remark 11 *Observe that the only points $c \in F$ that do not satisfy the* star-disc *property (Definition 8) are the points on the* free arc *of E, including the two end points where that arc meets L; but even for these extreme points the property only fails for those simplices $\sigma \in St(c)$ in the complement of D.*

F is of course contractible: there is an obvious strong deformation retract of F into D - across the disc E from the free arc to the gluing arc L - and therefore we can follow this first homotopy by a contraction of D. But we are going to construct a different contraction of F. Consider a regular neighbourhood N of L in E, homeomorphic to $L \times I$ (represented shaded gray in Figure 8). We define a contraction $K : F \times I \longrightarrow F$ in two steps, first for $t \leq 1/2$ and second for $1/2 \leq t \leq 1$. In the first part: if $x \in D$ we define $K(x,t) = J(x, 2t)$, if $x \in \overline{E - N}$ we have $K(x,t) = x$ and for those points in the region $N \cong L \times I$ the homotopy stretches each stalk $S_x = \{x\} \times I$, $x \in L$ (one of those is depicted in Figure 8), keeping the top end-point in $\overline{E - N}$ fixed, so as to follow x along its track under J - we leave it as an exercise for the reader to define suitable formulas for this. At the end of this first part, we have $K_{1/2}(F) = E \cup J(L \times I)$: D was contracted to a point, say q, by J (with double speed), and for each $x \in L$ we developed its former track under J as we've just explained. In the second part of the homotopy, for times $1/2 \leq t \leq 1$, we contract $K_{1/2}(F)$ in itself, to the point q, in the obvious way: we first use a strong deformation retraction of $E \cup J(L \times I)$ into $J(L \times I)$, across the disc E, as alluded above; then we contract $J(L \times I)$ to q erasing all the tracks - we also leave the appropriate formulation of this *erasing* to the reader.

It is easy to check that each step of the construction of K preserves the given coalescence of J, so K is also coalescent.

By our assumptions about J, and the fact that $L \subset N_\varepsilon(V)$, the tracks under K of all points $x \in L$ miss p and so $K_{1/2}(F) \neq F$; therefore the opening time for K is strictly less than $1/2$ - recall that at its opening time any homotopy is still surjective. So, by theorem 4, we can assume that $t_{op} = 0$ - starting afresh at opening time and reparametrizing, which we do here by sending $[t_{op}, 1/2]$ to $[0, 1/2]$ and keeping the other times fixed. Now, as in the proof of theorem 7, let t_n, $n \in \mathbb{N}$, be a sequence of times convergent to 0, and $c_n \longrightarrow c$ a sequence of points, all in a simplex $\sigma \in St(c)$, and such that, for each n, $c_n \in F - K_{t_n}(F)$. Since for all $t \leq 1/2$ we have $E \subset K_t(F)$, the simplex σ is not contained in E and therefore - see Remark 11, above - the point c has the star-disc property that allows us to carry on the argument of theorem 7 to arrive at a contradiction.

We have thus proved:

Theorem 12 *Any contraction of the dunce-hat is non-coalescent.*

We end this section with two problems for the reader:

Exercise 13 *Generalize last theorem to arbitrary finite contractible simplicial complexes.*

The troublesome part of the generalization will be to work out appropriate analogues of the previous gluing of disc E: that is, to add appropriate pageants near a point where the star-disc property fails.

The second problem is just a fun-exercise about contractions: it asks about a property, whose existence is not at all revealing of the structure of the space.

Exercise 14 *Let H be a (coalescent) contraction of space Y. Is there always a (coalescent) contraction of Y that is descending in the sense that $t' > t \Rightarrow H_{t'}(Y) \subset H_t(Y)$?*

4 Shrinking complexity I: volume and energy

In our construction, in section 2, of a contraction of the dunce-hat D we needed first to perform some *folding* of that space (recall the visual suggestion in "fake" Figure 3) to be able, in a second movement, to open up the space, thus effectively starting its shrinking; on the other hand, the existence of contractions of D with opening time 0 - recall the note in that section - suggests that we can keep that initial folding phase to a minimum. We also needed, in the course of establishing the non-existence of coalescent contractions of the dunce-hat (theorem 12), a technical result, lemma 10, to deal with the situation where the track of a single point, *wandering* wildly, may cover up the shrinking - for instance, by filling whole manifold regions. This sort of wild wandering is also relevant to the exercise in the end of last section.

In this section we introduce ways of measuring the complexity - or waste - due to those *folding* and *wandering*, and ways of minimizing it.

We will be working mainly in the *piecewise-linear* setting [3], with *simplicial* contractions, $H : (K \times I)^{sub} \longrightarrow K^{sub}$, where the upper-scripts denote simplicial subdivisions (with the usual notation $K^{(r)}$ for the r-th baricentric subdivision), and their limits in the space of contractions endowed with the *supreme metric*. Of course, for simplicial homotopies the tracks of points

[3] See footnote 1 on page 2.

don't suffer from the sort of wild wandering we've just referred, nevertheless they can still be pretty complicated in view of the well known fact - see theorems 16 and 17 bellow - that any continuous map $f : K \longrightarrow L$, between simplicial complexes, can be arbitrarily approximated, in the *supreme metric*, by a simplicial one, $g : K^{(r)} \longrightarrow L^{(n)}$. A question is in order here: why haven't we consider such simplicial approximations in the last section, thus avoiding the need for lemma 10? The answer is that we were dealing with coalescence and so we would need simplicial approximations that preserved coalescence also: and the proofs we know of the existence of simplicial approximations don't provide that.

Let us recall the fundamental facts about approximating continuous maps by simplicial ones (see [3, Chapter 2]).

Given simplicial complexes K and L and a continuous map $f : K \longrightarrow L$, a simplicial map $g : K \longrightarrow L$ is called a *simplicial approximation* to f if for each vertex c of K, $f(St_K(c)) \subset St_L(g(c))$.

Proposition 15 *Let $f : K \longrightarrow L$ be a continuous map between simplicial complexes. Then f is homotopic to any simplicial approximation g; moreover the homotopy can be defined so that it is relative to the subspace of K where f and g coincide.*

The fundamental result is the famous *Simplicial Approximation Theorem*, due to James Alexander [11] and Oscar Veblen [12].

Theorem 16 (Simplicial Approximation Theorem) *Let K and L be simplicial complexes and $f : K \longrightarrow L$ a continuous map. Then there is $r \in \mathbb{N}$ such that $f : K^{(r)} \longrightarrow L$ has a simplicial approximation.*

As a corollary to the theorem we can approximate any continuous map by a simplicial one, less than any given $\varepsilon > 0$. If g is a simplicial approximation to f, since we also have $g(St_K(c)) \subset St_L(g(c))$, for each $x \in St_K(c)$, $d(g(x), f(x)) \leq \delta(St_L(g(c)))$ where δ stands for *diameter*. Therefore all we need to do is to first substitute L by an iterated baricentric subdivision $L^{(n)}$ such that *mesh* $L^{(n)} < \varepsilon$, where *mesh* denotes the supremum of the diameters of all the stars of vertices of L: *mesh* $L = \sup \{St(c) : c \in L^0\}$; then for any simplicial approximation to f, $g : K^{(r)} \longrightarrow L^{(n)}$, we have $d(f(x), g(x)) \leq \varepsilon$, for all $x \in K$.

What we actually need, for our purposes, is the relative version of the previous theorem, due to Zeeman [13].

Theorem 17 (Relative Simplicial Approximation Theorem) *Let $f : K \longrightarrow L$ be a continuous map between simplicial complexes, such that its restriction, $f|_M$, to a subcomplex $M \subset K$, is simplicial. Then there is a natural $r \in \mathbb{N}$ and a simplicial map $g : (K, M)^{(r)} \longrightarrow L$ such that $g = f$ on M, and $g \simeq f$ rel M.*

In the statement of the theorem, $(K, M)^{(r)}$ denotes the (r-th) *baricentric subdivision relative to M*, where relative means: obtained by introducing as new vertices the baricentres of all simplices except those in M. There are examples that show one can not get what would be a full relative version of the simplicial approximation, with g a (true) simplicial approximation to f; nevertheless the proof of the theorem (see [3, page 55]) shows that, starting with sufficiently fine subdivisons $K^{(n)}$ and $L^{(n)}$, we can get g as ε-close to f as we like it.

For a simplicial complex K, $K \times I$ has a natural associated triangulation with vertices the points $K^0 \times \{0, 1\}$, that is the vertices of K at levels 0 and 1, plus another vertex at $(\hat{\sigma}, 1/2)$, where $\hat{\sigma}$ denotes the baricentre, for each simplex $\sigma \in K$. Given a contraction, $h : K \times I \longrightarrow K$, by taking in the previous theorem $K = K \times I$, $L = K$ and $M = K \times \{0, 1\}$, the simplicial map we get, $H : (K \times I, K \times \{0, 1\})^{(r)} \longrightarrow K$, is also a contraction. For simplicity, we shall denote the subdivisions of $K \times I$ relative to the base-and-top, $M = K \times \{0, 1\}$, by $[K \times I]^{(r)}$.

Let us now establish the precise metric settings in which the measures *volume* and *energy* will be defined. Let K be a simplicial complex; consider for each simplex of dimension n, $\sigma \in K^n$ a linear embedding in euclidian n-space, $\mathbb{E}^n$, in such a way that all those embeddings are compatible: for any two simplices with a common face, the restrictions to that face agree (or agree up to some isometric motion of $\mathbb{E}^n$) - a simple way of getting such a compatible set of embeddings is to consider for each simplex of top dimension, say n, a linear embedding to the standard *regular* n-simplex, $\Delta^n \subset \mathbb{E}^n$, and for each simplex of smaller dimension $j < n$, a linear embedding into a j-face of Δ^n. Clearly we can extend, in a natural way, the set of compatible linear embedding to a set of compatible linear embeddings of simplices in $K \times I$. Each simplex in K, or in $K \times I$, has a transport-metric brought by its linear embedding, and all those metrics are compatible. Now, using those transport-metrics, each piecewise-linear path $\gamma : I \longrightarrow N$, where N stands for either K or $K \times I$, has a well defined length, $l(\gamma)$. We take, in both K and in $K \times I$, the *length-metric*, $d(x, y) = \inf_{\gamma : x \to y} l(\gamma)$ (the infimum over all pl-paths joining x to y); obviously, in each simplex the length-metric and the transport-metric coincide; by compactness, that is, one's having only a finite number of simplices, there is at least one path from x to y - a *geodesic* - which minimizes length, and since in each simplex the metric is euclidian, the geodesics are also pl-paths.

Exercise 18 *Prove the existence of geodesics.*

We can now define volume and energy: although these can be defined for general homotopies, we will be talking only of contractions to keep the notation slim. The generalizations to arbitrary homotopies are straightforward. The definition of volume follows.

4.1 Volume

Consider a simplicial contraction $H : [K \times I]^{sub} \longrightarrow K^{sub}$. Let $M = [K \times I]^{sub}$and for each $t \in I$, $M_t = K \times \{t\}$. The idea is to measure, for each $t \in I$, the *area* of $H_t(M) = H(M_t)$: the word *area* comes from the consideration of 2-dimensional cases, like the dunce-hat or Bing's house. To keep the discussion simpler, we will assume that each simplex in K is the face of a top-dimensional simplex, say of dimension m: we say, in these circumstances, that K is a *full* m-complex; then, for all $i, r \in \mathbb{N}$, K^{sub} is a full m-complex and $[K \times I]^{sub}$ is an full $m + 1$-complex. Take all the intersections of M_t with $m + 1$ simplices of M. By the way $K \times I$ is triangulated, we have that for each simplex $\sigma \in M^{\ m+1}$, the intersection $\sigma_t = \sigma \cap M_t$ is just like the intersection of an $m + 1$ simplex in euclidian space $\mathbb{E}^{n+1}$ with an hyperplane: it is either a face of σ, maybe a single vertex - those cases happen for $t = 1/2$ and for other heights of baricentres - or, more generally, the *convex span* of $j \geq m + 1$ points in the 1- skeleton, σ^1, of σ - in this case, it is either an m-face of σ or a properly embedded (affine) m-disc. Since H is linear on σ, the image $H(\sigma \cap M_t)$ is the convex span of $k \leq j$ points in the 1- skeleton of the simplex $\rho = H(\sigma)$; we consider the m-area of that image, $A(H(\sigma_t))$: of course it is non zero only if $k \geq m + 1$. Naturally, the idea now is to sum these areas over all the intersections σ_t of top dimensional simplices with M_t, but avoiding replicating contributions; if two such intersections coincide, say $\sigma_t = \tau_t$, then it is a face common two both σ and τ and it doesn't matter which of the two simplices, $H(\sigma)$ or $H(\tau)$, we choose to take in the sum: $H(\sigma_t) = H(\tau_t)$ has a well defined m-area independently of the simplices where it lies (the metric in K was built from *compatible* transport metrics in the simplices). We thus define the area of H at time t by

$$A_H(t) = \sum_{\sigma_t : \sigma \in M^{m+1}} A(H(\sigma_t))$$

The *volume* of the contraction H is now defined by integration:

$$V(H) = \int_0^1 A_H(t)dt$$

Note that if σ intersects M_t, then, for a sufficiently small $\varepsilon > 0$, it intercepts $M_{t'}$ for all t' belonging to one of the intervals $[t, t+\varepsilon]$, $[t-\varepsilon, t]$, or to both: obviously, whatever is the case, $A(H(\sigma_t))$ varies continuously with t; considering all the $m+1$ simplices that intersect M_t and the combination of their intersections with nearby levels it is not difficult to see that $A_H(t)$ *is a continuous function of* t: this we leave as an exercise for the reader.

Note: the heuristic idea of introducing a notion of volume as a mean to measure the folding of a space during a contraction, could be implemented in a variety of ways; for instance, we could have defined the volume, more simply and directly, to be the sum over all top dimensional simplices $\sigma \in M^{m+1}$ of the m-areas $A(H(\sigma))$, without ever bothering to consider levels and dropping the final integration; but the definition we adopted, has some technical advantages, which will be put to an use in the next section: it can be combined with the *energy* that we will define shortly, when it comes to control the effect of local changes caused by choosing to perform certain movements at earlier or later stages; as we will see next, it has the disadvantage of not being sensible to arbitrary reparametrizations.

Let $\mathcal{C}$ denote the space of all contractions of the space K, endowed with the *supreme metric* and $\mathcal{S}$ the subspace of all simplicial contractions $H : [K \times I]^{sub} \longrightarrow K^{sub}$, for all simplicial subdivisions *sub*. It is easy to see that the volume function $V : \mathcal{S} \longrightarrow [0, +\infty)$ doesn't have a minimum. On the one hand, for all $H \in \mathcal{S}$ we have $V(H) > 0$, since $A_H(t) : I \longrightarrow [0, +\infty)$ is continuous and not constant equal to 0: at $t = 0$ its value is the total area of K: $A_H(0) = A(K) = \sum_{\rho \in K} A(\rho)$. On the other hand, given any $H \in \mathcal{S}$, we can reparametrize it in such a way as to perform the whole contraction in an arbitrarily small initial time interval $[0, \delta]$, $\delta = 1/r^n$, $n \in \mathbb{N}$, and stay constant thereafter: clearly the infimum of the respective volumes when $\delta \longrightarrow 0$ is 0.

The volume function $V : \mathcal{S} \longrightarrow [0, +\infty)$ is not continuous because it is not *upper-semi-continuous*: there are contractions $H \in \mathcal{S}$ such that for all $\varepsilon > 0$ we can find other contractions $J \in \mathcal{S}$ such that $d(H, J) < \varepsilon$ but $V(J) \geq V(H) + c$, c a positive constant. Let's describe an easy example with contractions of the unit interval I: refer to figure 9 bellow.

In this figure there are represented the six 2-simplices of $I \times I$, where we took I with the trivial simplicial decomposition consisting of the two vertices, 0 and 1, and the one edge: those are the triangles defined by the two diagonals plus the segment $t = 1/2$ (remember how one triangulates $K \times I$). The four chained rectangles centred at $(1/2, 1/2)$ are just a schematic representation. Consider a subdivision $[I \times I]^{(r)}$ with $r \geq 2$. Let S be the union of the two vertical edges of $[I \times I]^{(r)}$ with common vertex $(1/2, 1/2)$; it is easy to see they will be the two points

$(1/2, 1/2 \pm 1/2^{r-1}3)$; the first subdivision introduces the two baricentres for the top and bottom triangles, since the medians of a triangle trisect each other we get the points $(1/2, 1/2 \pm 1/3)$; the next subdivisions will introduce the baricentres for the two segments that make up S, thus halving them.

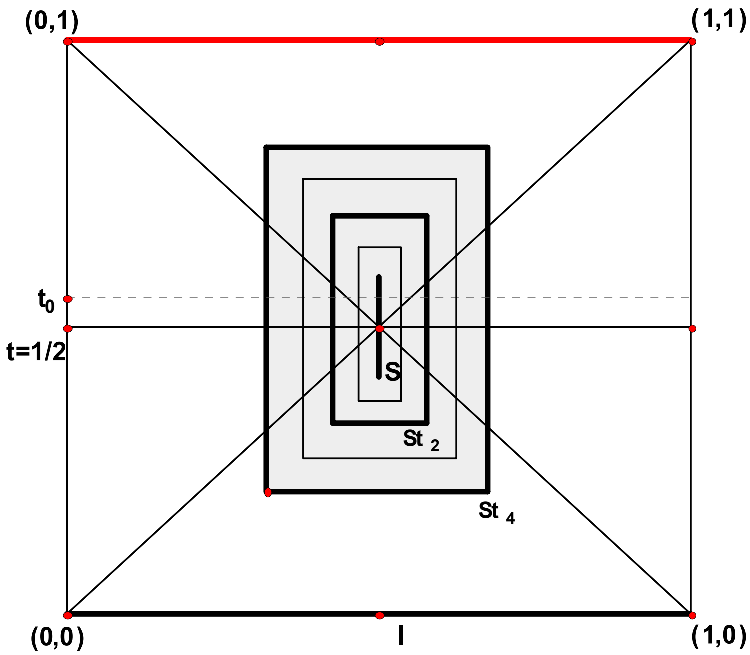

Figure 9

Consider to further subdivisions, $M = [I \times I]^{(r+2)}$ The outer most rectangle represents a neighbourhood N of S, $N = St(S)$, the *star* of S in M. Note that this is just schematic: in reality, stars are not rectangles, sometimes are not even convex; but $St(S)$ is a *regular neighbourhood* and as such a manifold with boundary that collapses to S and therefore, since S is contractible, it will be a disc whose boundary is the link of S in M, $Lk(S)$. Consider a simplicial contraction $H : M \longrightarrow I$ such that $H(St(S)) = 0$; in this very simple case we can do it by hand, specifying the images of the vertices: since there are only the two vertices 0 and 1 in I, there is always compatibility whatever choice we make - we can choose to send $(0, 0)$ to 0 and $(1, 0)$ to 1, thus getting the identity in $I \times \{0\}$, and all the other vertices to 0; we could be more sophisticated: first extend the map from $I \times \{0, 1\} \cup St(S)$ to I which is the identity in $I \times \{0\}$ and is constant 0 in the complement to a continuous map on the whole of $I \times I$, invoking *Tietze's characterization of normality* (the Extension Lemma) (see [8, Chapter VII]), and then use Theorem 17. For each $n \in \mathbb{N}$ consider $M^{(2^n)}$: in this complex let $St_0 = S$, $St_1 = St(S)$, $St_2 = St(St_1)$,..., $St_i = St(St_{i-1})$, ..., $St_{2^n} = St(St_{2^n-1})$, and for each $i = 1, ..., 2^n$, denote the respective link by Lk_i; these (St_i, Lk_i) form a chained sequence of discs, and respective boundary circles, the last one, St_{2^n}, being the original N: figure 9 represents the situation when $n = 2$, with the two rectangles with thicker edges representing the case $n = 1$. Now, construct a simplicial contraction $J_n : M^{(2^n)} \longrightarrow I^{(2^n)}$ as follows; in the closure of $M^{(2^n)} - N$, whose boundary is Lk_{2^n}, J_n coincides with H, in particular it is constant 0 in Lk_{2^n}: since H is linear on each simplex σ, the 2^n baricentric subdivisions of σ naturally match, under the image by H, the 2^n baricentric subdivisions of $H(\sigma)$ in I; in N, J_n is defined by stipulating that the vertices of the links Lk_i with i *even* go to 0 and the vertices on the others go to $1/2^n$, the vertex next to 0 in $I^{(2^n)}$. For any level $I \times t_0$ that intersects S - represented dashed in the figure - the segment has to cross all links in succession - at least once each, but maybe more since the discs St_i are not necessarily convex - until it reaches S, and then cross them over again in reverse order to leave N, correspondingly the image by J_n goes back and forth between 0 and $1/2^n$ at least 2^{n+1} times; therefore, the *area* - which, in the present situation, should be called more appropriately *length* - at time t_0, $A_{J_n}(t_0)$ is greater or equal to 2. So, the volume $V(J_n)$ is greater than twice the length of S, which is $1/2^{r-2}3$. We concluded then that for all $n \in \mathbb{N}$, $V(J_n) > 1/2^{r-1}3$; on the other hand, by the definition of J_n we have $d(H, J_n) \leq 1/2^n$. This completes the example, showing that the volume function is not upper-semi-continuous in the space $\mathcal{S}$.

Although the volume function $V : \mathcal{S} \longrightarrow [0, +\infty)$ is discontinuous, it is *lower-semi-continuous*: given any $H \in \mathcal{S}$ we have $\forall \varepsilon > 0, \exists \delta > 0 : d(H, J) < \delta \Rightarrow V(J) > V(H) - \varepsilon$. This means simply that while the volume can increase suddenly (upper-semi-discontinuity), it can not decrease suddenly.

Consider the area of H at time t, $A_H(t)$. Let σ be a top dimensional simplex that intersects $M_t = K \times \{t\}$, where, as above, $M = [K \times I]^{sub}$. Since we are dealing with the question of lower-semi-continuity, we consider only those simplices that contribute positively to $A_H(t)$, that is such that $A(H(\sigma_t) > 0$; in those cases, $\sigma_t = \sigma \cap M_t$ has to be the convex span of $j \geq m + 1$ points in the 1- skeleton of σ, that is σ_t is an m-face of σ or a properly embedded (affine) m-disc, and furthermore $H(\sigma_t)$, which the convex span of $j \leq k$ points in the 1-skeleton of the m-simplex $\rho = H(\sigma)$ will have to be an m-disc also (possibly the whole of ρ but, apart from this very exception, not properly embedded in ρ): then, because H is linear its restriction to σ_t is a homeomorphism between the two m-discs. As we noted before, for a sufficiently small $\alpha > 0$, we can consider $\sigma_{t'}$ for t' belonging to one of the intervals $[t, t + \alpha]$, $[t - \alpha, t]$, or to both. In the case σ_t is an m-face of σ, σ is all in one side of M_t and we have intersections $\sigma_{t'}$ for t' in just one of those intervals: for t' in the other interval there is another simplex $\mathring{\sigma}$ in the opposite side of σ, with $\mathring{\sigma}_t = \sigma_t$ and for which $\mathring{\sigma}_{t'} \neq \emptyset$; we will not distinguish the two cases and will refer simply to all t' close to t. When σ_t intersects the interior of σ (is a properly embedded m-disc) we have intersections for $t' \in [t - \alpha, t + \alpha]$. In any case the nearby intersections, $\sigma_{t'}$, are all properly embedded m-discs close to σ_t and we have natural homeomorphisms $h_{t'} : \sigma_t \longrightarrow \sigma_{t'}$ such that for all $x \in \sigma_t$ $d(x, h_{t'}(x)) < c\alpha$ for some constant $c \geq 1$. Consider the m-disc $\Delta = H(\sigma_t)$ in the m-simplex $\rho = H(\sigma)$ and let $\Sigma = \dot{\Delta}$ be its boundary-sphere. Recall - from our analysis of Bing's house in section 3 (Lemma 6 and what follows it) - that there is a positive constant k such that any continuous map of $f : \Sigma \longrightarrow \Sigma$ with $d(f, id_\Sigma) \leq k$ (d the sup metric) is homotopic to id_Σ. Let $C(\Sigma)$ be a *collar* of Σ in Δ, that is a closed neighbourhood homeomorphic to $\Sigma \times I$ with $\Sigma \times \{0\} \equiv \Sigma$, and for each $t \in I$, let $C_t(\Sigma)$ be the collar $\Sigma \times [0, t]$. Fix $t_0 \in I$. Consider closed λ-neighbourhoods $N(\Delta, \lambda)$ of Δ in K, and $N_\Delta(\Sigma, \lambda)$ of Σ in Δ, with λ sufficiently small so that $\lambda < k/3$, there is a (strong deformation) retraction $R : N(\Delta, \lambda) \longrightarrow \Delta$ with $d(x, R(x)) < k/3$ and $N_\Delta(\Sigma, \lambda) \subset C_{t_0/2}(\Sigma)$.

See figure 10 bellow. It represents the situation for a 2-dimensional K; the simplex σ is on the top left, with vertices a, b, c, d and σ_t is represented by the (yellow) shaded quadrilateral; $\rho = H(\sigma)$ is the triangle with vertices A, B, C which are the images of the vertices a, b, c, respectively, with the fourth vertex D sent to a. The picture shows three other simplices of K adjacent to ρ. The quadrilateral $\Delta = H(\sigma_t) \subset \rho$ is represented by the four thick edges, forming its boundary Σ, and in its interior, $cl((\Delta - C(\Sigma))$, the closure of the complement of collar $C(\Sigma)$, is the dark shaded quadrilateral. Also represented in lighter shade is the neighbourhood

$N(\Delta, \lambda)$.

We claim there are $\alpha_0 > 0$ and $\varepsilon_0 > 0$ such that for all $t' \in [t, t+\alpha_0] \cup [t-\alpha_0, t]$ and for all $J \in S$ with $d(H, J) < \varepsilon_0$, we have $J(\sigma_{t'}) \supset (\Delta - C_{t_0})$, that is the images of all the discs α_0-close to σ_t by all contractions ε_0-close to H in $\mathcal{S}$, cover the complement in Δ of the collar C_{t_0}. Suppose not: then for all $n \in \mathbb{N}$ there are $t_n = t \pm \alpha_n$, $0 < \alpha_n < 1/n$ and $J_n \in \mathcal{S}$, $d(H, J_n) < 1/n$, such that $J_n(\sigma_{t_n}) \not\supset (\Delta - C_{t_0}(\Sigma))$.

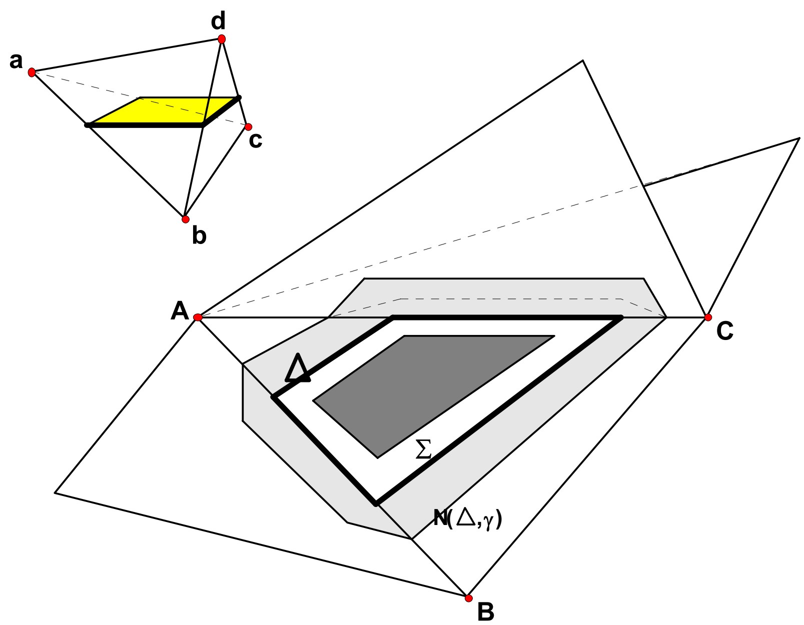

Figure 10

Now the argument is just a clone of the argument previously used in the case of Bing's house to show it had an opening time greater than 0.

Let, for each $n \in \mathbb{N}$, $x_n \in (\Delta - C_{t_0}(\Sigma)) - J_n(\sigma_{t_n})$: by passing to a subsequence and renumbering if necessary, we can assume without loss of generality that $x_n \longrightarrow x$ with x in the closure $cl((\Delta - C_{t_0}(\Sigma))$. As in the case of Bing's house, we can consider retractions $r_n : \Delta - \{x_n\} \longrightarrow \Sigma$, in such a way that their restrictions to $C_{t_0/2}(\Sigma)$ form an *equicontinuous* set of functions (see Lemma 5): therefore, $\exists \delta_0 > 0$ such that for all $n \in \mathbb{N}$ and for all $x, y \in C_{t_0/2}(\Sigma)$, $d(x, y) < \delta_0 \Rightarrow d(r_n(x), r_n(y)) < k/3$. Take $\lambda_0 \leq \delta_0$. Since H is uniformly continuous, $\exists \delta_1 > 0$ such that $\forall x, y \in M, d(x, y) < \delta_1 \Rightarrow d(H(x), H(y)) < \lambda_0/2$. Consider, for each $n \in \mathbb{N}$, the map $g_n : \Delta \longrightarrow K$ defined as $g_n = J_n \circ h_{t_n} \circ H^{-1}$:

$$\Delta \xrightarrow{H^{-1}} \sigma_t \xrightarrow{h_{t_n}} \sigma_{t_n} \xrightarrow{J_n} K$$

Let $N_0 \in \mathbb{N}$ be such that $n \geq N_0 \Rightarrow c(1/n) < \min\{\delta_1, \lambda_0/2\}$; let $y \in \Delta$ arbitrary and $x = H^{-1}(y)$; for $n \geq N_0$ we have that $d(x, h_{t_n}(x)) < c(1/n \leq \delta_1$ so, by the choice of δ_1, $d(H(x), H(h_{t_n}(x))) < \lambda_0/2$; therefore, since $d(J_n, H) < 1/n \leq \lambda_0/2$, we have

$$d(y, g_n(y)) \leq d(y, H(h_{t_n}(x)) + d(H(h_{t_n}(x), J_n(h_{t_n}(x))) < \lambda_0/2 + \lambda_0/2 = \lambda_0$$

So, $g_n(\Delta) \subset N(\Delta, \lambda_0)$ and since J_n misses the point x_n, so does g_n. Let $f_n = R \circ g_n : \Delta \longrightarrow \Delta$; consider an arbitrary $y \in \Sigma$: if $g_n(y)$ is not in Δ, and since by the definition of the neighbourhoods

$N(\Delta, \lambda)$ we have

$$d(x, R(x)) < k/3, d(y, f_n(y)) \leq d(y, g_n(y)) + d(g_n(y), f_n(y)) < \lambda_0 + k/3 < k/3 + k/3 < k;$$

if $g_n(y)$ is in Δ, then $f_n(y) = g_n(y) \in N_\Delta(\Sigma, \lambda_0) \subset C_{t0/2}(\Sigma)$ and so, by the choice of $\lambda_0 \leq \delta_0$, $d(y, r_n(f_n(y)) < k/3$. In conclusion, $j_n = r_n \circ f_n$ is a continuous map $\Delta \longrightarrow \Sigma$ such that for all $y \in \Sigma$, $d(y, j_n(y)) < k$ and so by the choice of k its restriction to Σ is homotopic to id_Σ which is a contradiction. We have thus justified our claiming the existence of α_0 and ε_0; for all t' with $|t - t'| < \alpha_0$ and $J \in \mathcal{S}$ with $d(H, J) < \varepsilon_0$ we have that $J(\sigma_{t'}) \supset (\Delta - C_{t_0}(\Sigma))$ and so $A(H(\sigma_t)) - A(J(\sigma_{t'})) = A(\Delta) - A(J(\sigma_{t'})) \leq A(C_{t_0}(\Sigma))$; let's denote the area of the collar $C_{t_0}(\Sigma)$ by a_0: $a_0 = A(C_{t_0}(\Sigma))$. Note that, since t_0 is arbitrary and clearly $A(C_t(\Sigma)) \longrightarrow 0$ when $t \longrightarrow 0$, we can consider that a_0 is as small as we like. Consider all top dimensional simplices σ that contribute positively to $A_H(t)$: suppose there are $n+1$ such simplices, call them ${}^0\sigma, {}^1\sigma, ..., {}^n\sigma$. For each $i = 0, 1, ..., n$ get positive constants α_i, ε_i and a_i as before and let $\alpha_t = \min_{i=0,...n}\{\alpha_i\}$, $\varepsilon_t = \min_{i=0,...n}\{\varepsilon_i\}$ and $a_t = \sum_{i=0}^n a_i$. If $|t - t'| < \alpha_t$ and $d(H, J) < \varepsilon_t$, summing over all the simplices ${}^i\sigma$ we have

$$\begin{aligned} A_J(t') &\geq \sum_{i=0}^n A(J({}^i\sigma_{t'}) \geq \sum_{i=0}^n A(J({}^i\sigma_{t'}) \geq \sum_{i=0}^n (A(H({}^i\sigma_t)) - a_i) \\ &= \sum_{i=0}^n A(H({}^i\sigma_t)) - \sum_{i=0}^n a_i = A_H(t) - \sum_{i=0}^n a_i = A_H(t) - a_t \end{aligned}$$

where the first inequality may be strict since it possible that $A(J(\sigma_{t'})) > 0$ for other simplices different from the ${}^i\sigma$ - those for which $A(H(\sigma_t)) = 0$ - and a_t can be arbitrarily small. Let $\varepsilon > 0$ and, for each $t \in I$, $a_t \leq \varepsilon$; we've noted before that $A_H(t)$ is continuous, so by the compactness of I it is uniformly continuous, therefore $\exists \lambda > 0$ such that for $|t - t'| < \lambda \Rightarrow A_H(t') \in (A_H(t) - \varepsilon, A_H(t) + \varepsilon)$. For each $t \in I$ consider an interval $[t - \beta_t, t + \beta_t]$ such that $0 < \beta_t < \min\{\lambda, \alpha_t\}$ and consider a finite number of such intervals $[t_0, t_0 + \beta_{t_0}]$, $[t_1 - \beta_{t_1}, t_1 + \beta_{t_1}]$, $[t_2 - \beta_{t_2}, t_2 + \beta_{t_2}]$, ..., $[t_m - \beta_{t_m}, t_m]$ covering I, with $t_0 = 0$ and $t_m = 1$. Take $\delta = \min_{i=0,...,m}\{\varepsilon_{t_i}\}$ and let $d(H, J) < \delta$. Then for an arbitrary $t \in I$ we have that t belongs to one of these intervals, say the one around t_i, and since $d(H; J) < \delta \leq \varepsilon_{t_i}$, we have

$$|t_i - t| < \beta_i \Rightarrow A_J(t) \geq A_H(t_i) - a_{t_i} \geq A_H(t_i) - \varepsilon \geq A_H(t) - 2\varepsilon$$

Integrating we get

$$V(J) = \int_0^1 A_J(t)dt \geq \int_0^1 (A_H(t) - 2\varepsilon)dt = V(H) - 2\varepsilon$$

finishing the proof that V is lower-semicontinuous. We state this fact as a theorem:

Theorem 19 *Let K be a simplicial complex and $\mathcal{S}$ the space of all simplicial contractions $H : [K \times I]^{sub} \longrightarrow K^{sub}$, for all simplicial subdivisions, sub. Then the volume function $V : \mathcal{S} \longrightarrow [0, +\infty)$ is discontinuous and lower-semi-continuous.*

Next we will discuss the notion of energy of a contraction.

4.2 Energy

We start by spelling out some notation. Let $\mathcal{S}$ be the space of simplicial contractions of simplicial complex K, as previously defined, and $H \in \mathcal{S}$, say $H : [K \times I]^{sub} \longrightarrow K^{sub}$. We assume as before that K is *m-full*, so $K \times I$ is $(m+1)$-full. For each $x \in K$, the track $t_H(x) = H(\{x\} \times I)$ is a polygonal path in K which has a well defined *length*, say l_x; let $L_H : K \longrightarrow [0, +\infty)$ be the length function, $L_H(x) = l_x$, which is continuous as we will see. Let $S_x = \{x\} \times I$ be the stalk over x and $N(S_x, \varepsilon)$ a closed neighbourhood of S_x in $M = [K \times I]^{sub}$. Let σ be a top dimensional simplex of M intersecting S_x; $\sigma_x = S_x \cap \sigma$ is either a single vertex of σ or a (vertical) segment, which may or may not intersect the interior of σ (if it doesn't then it is contained in a proper face of σ, but that's not relevant here). We want to control the lengths of all segments $\sigma_{x'}$ for points x' close to x: to do this, we will recall first some facts of affine geometry, leaving the details for the reader.

Let M, N be two hyperplanes in euclidian $(m+1)$-space $\mathbb{E}^{m+1}$, r_x a vertical line, $r_x = x \times \mathbb{E}, x \in \mathbb{E}^n \times \{0\}$, intersecting M and N in two distinct points, s_x be the segment with those end points and $l(s_x)$ its length. Consider a cylindrical neighbourhood $N(r_x, \varepsilon) = D(x, \varepsilon) \times \mathbb{E}$, $D(x, \varepsilon)$ the closed ε-disc in $\mathbb{E}^n$. Clearly, for ε sufficiently small, all lines parallel to r_x that make up $N(r_x, \varepsilon)$ still intersect both hyperplanes in distinct points; consider one such line, $r_{x'}, x' \in D(x, \varepsilon)$ and the corresponding segment $s_{x'}$. Clearly, $l(s_{x'})$ varies continuously with x' and therefore the differences in length to s_x, $l(s_{x'}) - l(s_x)$, attain a maximum and a minimum values, say D_ε and d_ε respectively; it's also clear that D_ε and d_ε tend to 0 when $\varepsilon \longrightarrow 0$. For each x', consider the quadrilateral $Q_{x'}$ that has s_x and $s_{x'}$ as opposite sides; the other transverse sides lie in the hyperplanes M and N (see figure 11).

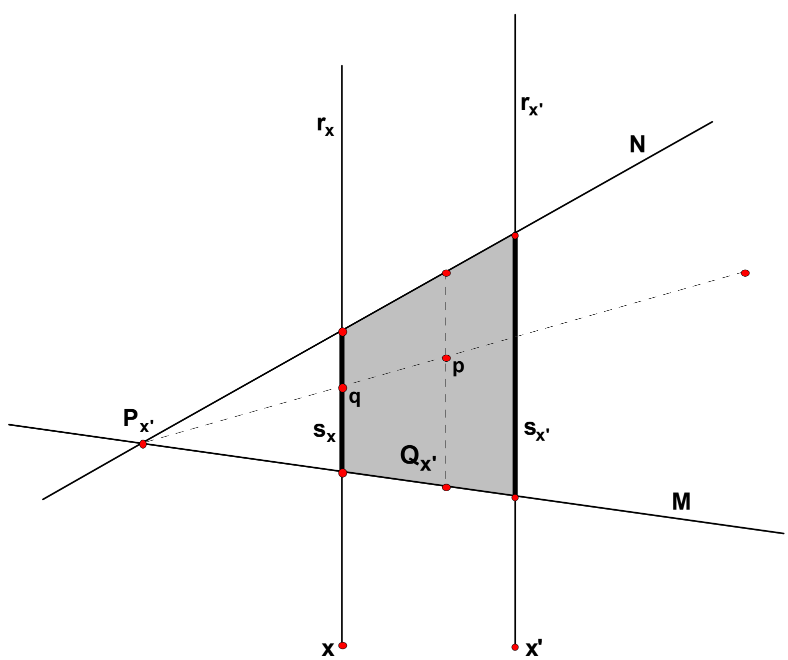

Figure 11

There is a strong deformation retraction $R_{x'}$ of $Q_{x'}$ into the side s_x, such that the track of each point follows the *radial projection* from the point $P_{x'}$ where the two lines containing the transverse sides meet ($P_{x'}$ belongs to the affine subspace $M \cap N$; if those lines are parallel, that is if M and N are parallel hyperplanes, then the projection is horizontal instead) and that sends vertical segments to vertical segments: for an arbitrary point p, if its radial projection into s_x is the point q simply define $R_{x'}$ to be the final map of the homotopy $(p,t) \longrightarrow q + (1-t)(p-q)$. Since the quadrilaterals $Q_{x'}$ and respective radial projection points $P_{x'}$ vary continuously with x', the retractions $R_{x'}$ fit together to provide a global (strong deformation) retraction R of the cylindrical chunk between the two hyperplanes - call this $C(s_x, \varepsilon)$ - into s_x.

Take an $(m+1)$-simplex σ in $\mathbb{E}^{m+1}$and the intersections $\sigma_{x'} = \sigma \cap r_{x'}$ - which, for each $x' \in \pi(N(r_x, \varepsilon) \cap \sigma)$, $\pi : \mathbb{E}^{n+1} \longrightarrow \mathbb{E}^n \times \{0\}$ the vertical projection, coincide with the previous $s_{x'}$ for two hyperplanes containing two m-faces of σ: the previous argument neatly applies to this more general situation to provide us with a strong deformation retraction R of $N(r_x, \varepsilon) \cap \sigma$ into σ_x. And of course we also have a maximum and a minimum values, that tend to 0 when $\varepsilon \longrightarrow 0$, for the differences in length $l(\sigma_{x'}) - l(\sigma_x)$.

For later use, we need to single out an important property of the retraction $R : C(s_x, \varepsilon) \longrightarrow s_x$. Let L be a linear segment embedded in $C(s_x, \varepsilon)$, with end points p_1, p_2; as we shall see shortly, the restriction of R to L is either constant - $R(L)$ is then a single point in s_x - or a homeomorphism into a segment in s_x; if one compares the ratio of the lengths of L and $R(L)$, one sees that although $l(R(L))/l(L)$ can be arbitrarily small - even 0 when $R(L)$ is a point - it is bounded above; this is the property we want to single out and will refer to by saying that R has *bounded magnifying factor*. Let $x_i = \pi(p_i), i = 1, 2$; $\pi(L)$ is a segment in $\mathbb{E}^n \times \{0\}$, degenerate if L is vertical, and for each point $p \in L$, $p \in s_{\pi(p)}$; let, for $i = 0, 1, 2$ the bottom and top end points of s_{x_i} be b_i and a_i, respectively, where we are taking $x_0 = x$ (see figure 12 bellow)

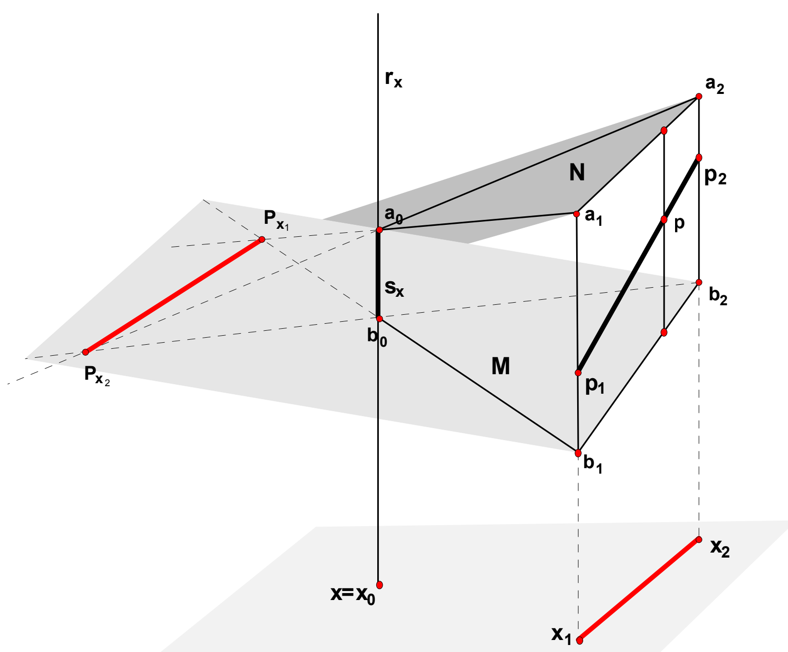

Figure 12

Figure 12 pictures a 3-dimensional situation: here, the (hyper)planes M and N are represented by the two shaded planes, that contain the triangles with vertices b_1, b_2, b_3 and a_1, a_2, a_3,respectively. Note first that the set of radial projection points $P_{x'}$ for all $x' \in \pi(L)$ is also line segment - degenerate if x, x_1, x_2 are collinear, that is if L is contained in the larger of the two quadrilaterals Q_{x_1}, Q_{x_2} (in this situation, one is contained in the other): by the definition of $P_{x'}$, the segment $\overline{P_{x_1}P_{x_2}}$ lies in the intersection of the plane spanned by all the lines $\overleftrightarrow{aa_0}$, $a \in \overline{a_1a_2}$, that is the plane that contains the triangle with vertices a_1, a_2, a_3, with the plane spanned by the lines $\overleftrightarrow{bb_0}$, $b \in \overline{b_1 b_2}$, that is the plane that contains the triangle with vertices b_1, b_2, b_3 - in the 3-dimensional situation of figure 12, these planes agree with N, M but in general they are just 2-dimensional subspaces of these hyperplanes and the intersection line $\overleftrightarrow{P_{x_1} P_{x_2}}$ is just a proper 1- dimensional subspace of the $(m-1)$-subspace $M \cap N$. It's clear that for each $p \in L$, $R(p)$ is the intersection with s_x of the plane, call it H_p, defined by p and the line $\overleftrightarrow{P_{x_1} P_{x_2}}$: so, either L and $\overleftrightarrow{P_{x_1} P_{x_2}}$ are parallel, and then $R(L)$ is just a point in s_x, or R is injective in L and therefore a homeomorphism to $R(L)$.

Note that the length $l(R(L))$ depends only on the *angle* of the planes H_{p_1} and H_{p_2}, call it the angle of L, $\measuredangle(L)$; clearly these angles are bounded above: otherwise, we could get a sequence of segments $L_n, n \in \mathbb{N}$, such that for each n $\measuredangle(L) \geq n$; since for each segment L there is a vertical segment with angle greater or equal to the angle of L, we can assume without loss of generality that all segments are vertical, $L_n = s_{x_n} = \overline{b_n a_n}$; by passing to a subsequence, and by compactness of $C(s_x, \varepsilon)$, we can assume that $a_n \longrightarrow a$: but then, since the angles increase, we would have that $a \in r_x - s_x$, which is a contradiction. If the ratios $l(R(L))/l(L)$ were not bounded, there would be a sequence of segments $L_n = \overline{p_1^n p_2^n}$, $n \in \mathbb{N}$, such that $l(R(L_n))/l(L_n) \geq n$; since, for all segments L, $l(R(L)) \leq l(s_x)$, we would have $l(L_n) \longrightarrow 0$; because the angles $\measuredangle(L_n)$ are bounded and $l(R(L_n))$ are bounded as well by $l(s_x)$, to get $l(R(L_n))/l(L_n) \geq n$ and $l(L_n) \longrightarrow 0$ we would need to have either the distances from the *hinges* of the angles to s_x increasing to ∞, or the segments L_n ever closer to the *hinges* of their angles; but obviously neither of these can happen: the set of all hinges $\overline{P_{x_1} P_{x_2}}$ is a compact set contained in the intersection of the two hyperplanes, $M \cap N$, which is disjoint from $N(r_x, \varepsilon)$ and so from the chunk $C(s_x, \varepsilon) \subset N(r_x, \varepsilon)$. It should also be clear, from the previous considerations, that the closer L is to s_x the closer $l(R(L))$ is to $l(L)$, which means that the (maximum) magnifying factor tends to 1 when $\varepsilon \longrightarrow 0$.

As before, in the more general case of an $(m+1)$-simplex σ in $\mathbb{E}^{m+1}$ and the retraction $R : N(r_x, \varepsilon) \cap \sigma \longrightarrow \sigma_x$, since this retraction agrees in $N(r_x, \varepsilon) \cap \sigma$ with restrictions of the previous retractions $C(s_x, \varepsilon) \longrightarrow s_x = \sigma_x$ for chunks between pairs of hyperplanes containing top faces of σ, it has *bounded magnifying factor* k_σ^ε, with $k_\sigma^\varepsilon \longrightarrow 1$ when $\varepsilon \longrightarrow 0$.

We return now to the analyses of the *length function*, $L_H : K \longrightarrow [0, +\infty)$, started in the first paragraph of this section; recall the notation we introduced there. For an arbitrary $x \in K$, $L_H(x)$ is defined as the length $l_x = l(H(S_x)) = l(t_H(x)$, S_x the stalk over x, $t_H(x)$ the track of x under H; we consider a closed neighbourhood $N(S_x, \varepsilon)$ of S_x in $M = [K \times I]^{sub}$ with ε sufficiently small so that $N(S_x, \varepsilon) \subset St(S_x)$. There is a $\delta > 0$ such that for all $x' \in K$ with $d(x, x') < \delta$, $S_{x'} \subset N(S_x, \varepsilon)$. If we consider all $(m+1)$-simplices σ in $St(S_x)$, that is all those that intersect S_x, then $l_x = \sum_\sigma l(H(\sigma_x)$: it is understood that when $\sigma_x = \sigma'_x$ for distinct simplices σ, σ', that is when σ_x is contained in a proper face common to both σ and σ', in the summation we only consider the contribution from one of the simplices. As we've seen, for each simplex σ, there is a maximum, D_ε^σ, and minimum, d_ε^σ, for the length differences $l(\sigma_{x'}) - l(\sigma_x)$ for all $\sigma_{x'} \subset N(S_x, \varepsilon) \cap \sigma$. Since H is *linear* on each simplex σ, there is in σ a *direction of maximal stretch*, with factor say $k_\sigma > 0$. For any $x' \in K$ with $d(x, x') < \delta$ we have for

$$\begin{aligned} l_{x'} - l_x &= \sum_\sigma l(H(\sigma_{x'}) - \sum_\sigma l(H(\sigma_x) = \sum_\sigma (l(H(\sigma_{x'}) - l(H(\sigma_x)), \text{ that} \\ n(\max_\sigma k_\sigma)(\min_\sigma d_\varepsilon^\sigma) &\leq \sum_\sigma k_\sigma d_\varepsilon^\sigma \leq l_{x'} - l_x \leq \sum_\sigma k_\sigma D_\varepsilon^\sigma \leq n(\max_\sigma k_\sigma)(\max_\sigma D_\varepsilon^\sigma) \end{aligned}$$

where n is the number of simplices σ in the summation. Since, for all σ, D_ε^σ and d_ε^σ tend to zero when $\varepsilon \longrightarrow 0$, we readily conclude that L_H is continuous.

The definition of *energy* $E(H)$, for a contraction $H : M = [K \times I]^{sub} \longrightarrow K$ follows:

$$E(H) = \int_K L_H(x) = \sum_{\sigma \in K^m} \int_\sigma L_H(x)\ dx$$

that is, we integrate the length function over each top-dimensional simplex σ of K, using its transport euclidian structure, and then sum all those integrals.

Analogously to the volume function, the *energy function* $E : \mathcal{S} \longrightarrow [0, +\infty)$ is not continuous but is *lower-semi-continuous*.

Let $H : M = [K \times I]^{sub} \longrightarrow K^{sub}$ be an element of $\mathcal{S}$ and $x \in K^{sub}$. Take an $\varepsilon > 0$; We claim that for all x' sufficiently close to x and for all J sufficiently close to H, $L_J(x') > L_H(x) - \varepsilon$, that is the lengths of the tracks of all points near x, for all contractions sufficiently close to H, don't diminish suddenly. Consider the track $t_H(x) = H(S_x)$ and a closed neighbourhood $N(t_H(x), \lambda)$ with λ sufficiently small so that $N(t_H(x), \lambda)$ is contained in the star of $t_H(x)$ in K^{sub}. Since H is uniformly continuous, there is an $\varepsilon > 0$ such that $d((x, t), (x', t') < \varepsilon \Rightarrow d(H_t(x), H_{t'}(x')) < \lambda/2$, and so, in particular, if we take a closed neighbourhood $N(S_x, \varepsilon)$ of S_x in M, then $H(N(S_x, \varepsilon)) \subset N(t_H(x), \lambda/2)$. We may consider $\varepsilon > 0$ small enough so that $N(S_x, \varepsilon) \subset St(S_x)$ There is also a $\delta > 0$ such that $d(x, x') < \delta \Rightarrow S_{x'} \subset N(S_x, \varepsilon)$. Let $J \in \mathcal{S}$ be such that $d(H, J) < \lambda/2$: then, if $d(x, x') < \delta$, $J(S_{x'}) = t_J(x') \subset N(t_H(x), \lambda)$. Let $\sigma_x^1, \sigma_x^2, ..., \sigma_x^n$ be the segments that make up S_x - in the usual notation σ_x^i is the intersection of S_x with the $(m+1)$-simplex σ^i - ordered from bottom to top, that is, $\sigma_x^i \cap \sigma_x^{i+1}$ is a point $x_i = (x, t_i)$ with $t_i < t_{i+1}$; we make $x_0 = (x, 0)$ and $x_n = (x, 1)$. Then $t_H(x) = H(S_x) = \bigcup_{i=1,...,n} H(\sigma_x^i)$: some of the $H(\sigma_x^i)$ may be vertices - that happens when $H(x_{i-1}) = H(x_i)$ - the others are all the segments that make up the track. Given another stalk $S_{x'}$ with $d(x, x') < \delta$, consider a parallel subdivision of $S_{x'}$ into segments $s_{x'}^1, s_{x'}^2, ..., s_{x'}^n$ with end points $x_i' = (x', t_i)$ - note that we substituted the σ for the s, because these segments are no longer intersections of $S_{x'}$ with simplices of M. By the choice of δ and because $s_{x'}^i$ is parallel to σ_x^i, for each $i = 1, 2, ..., n$ we have $J(s_{x'}^i) \subset N(H(\sigma_x^i), \lambda)$ and $x_i' \in N(H(x_i), \lambda)$ (note that $J(s_{x'}^i)$ is a polygonal line, but not necessarily a simple segment: this happens because in the space $\mathcal{S}$ the subdivisions $[K \times I]^{sub}$, K^{sub} vary and may not be the same for H and J, so what is sent by H into a segment may be sent by J into a polygonal line and vice-versa). There is a path $\gamma^i : I \longrightarrow N(H(x_i), \lambda)$ from x_i' to x_i with length smaller than λ: we modify the path $t_J(x') = J(S_{x'})$ inside each $N(H(x_i), \lambda)$ by considering a small interval time interval, $[t_i - a, t_i + a]$, such that $J(x' \times [t_i - a, t_i + a]) \subset N(H(x_i), \lambda)$ and redefine J in that interval in the following way: in the first quarter J does, by reparametrization, what it did in the first half, in the second quarter follows the path γ^i, in the third quarter follows the inverse path $(\gamma^i)^{-1}$ and in the last quarter does, again by reparametrization, what was previously done in the second half; the new path J' thus obtained sends each x_i' to x_i and clearly $l(J'(S_{x'})) \leq l(J(S_{x'})) + n(2\lambda)$. Next we construct a map of this new polygonal line, $J'(S_{x'})$, into $t_H(x)$: note that we still have, for each $i = 1, ..., n$, $J'(s_{x'}^i) \subset N(H(\sigma_x^i), \lambda)$ with the end points of the segment $H(\sigma_x^i)$ and the end points of the polygonal line $J'(s_{x'}^i)$ coinciding; consider a retraction $R_i : N(H(\sigma_x^i), \lambda) \longrightarrow H(\sigma_x^i)$ such that the restriction to each simplex of $St(H(\sigma_x^i))$ is of the *radial projection type* we treated above, and consider the image $R_i(J'(s_{x'}^i))$: since the end points coincide and $J'(s_{x'}^i)$ is a path, $R_i(J'(s_{x'}^i)) = H(\sigma_x^i)$. As explained before, the restriction of R_i to each simplex has a (maximum) magnifying factor - which tends to 1 when $\lambda \longrightarrow 0$ - and having only a finite number of simplices, we have a maximum magnifying factor for R_i, say k_i: therefore, with $k = \max_{i=1,...,n} k_i$ we have

$$l(H(\sigma^i_x)) \leq k_i l(J'(s^i_{x'})) \leq kl(J'(s^i_{x'})) \leq k(l(J(s^i_{x'}) + 2\lambda)$$

and summing up

$$\begin{aligned} L_H(x) &= \sum_{i=1}^{n} l(H(\sigma^i_x)) \leq \sum_{i=1}^{n} kl(J(s^i_{x'}) + 2\lambda) = 2n\lambda + \sum_{i=1}^{n} kl(J(s^i_{x'})) \\ &\leq 2n\lambda + k\sum_{i=1}^{n} l(J(s^i_{x'})) = 2n\lambda + kL_J(x') \end{aligned}$$

We concluded that $kL_J(x') \geq L_H(x) - 2n\lambda$. Since $k \longrightarrow 1$ when $\lambda \longrightarrow 0$ we have proved our claim that for every $\varepsilon > 0$, there are $\delta > 0$ and $\lambda > 0$ such that $d(x, x') < \delta, d(H, J) < \lambda/2 \Rightarrow L_J(x') > L_H(x) - \varepsilon$.

The proof that the energy function is lower-semi-continuous follows easily from the claim, and is analogous to the correspondent proof for the volume. By the definition of the energy function as a finite sum of integrals over the top-dimensional simplices of K, it's clearly enough to concentrate in one of those integrals, say over the simplex σ. Let $\varepsilon > 0$; since L_H is uniformly continuous, there is a $\delta > 0$ such that $d(x, x') < 0 \Rightarrow |L_H(x) - L_H(x')| < \varepsilon$. Pick for each $x \in K$ two positive constants $\delta_x, \lambda_x > 0$, satisfying the claim for that ε and with $\delta_x \leq \delta$. Choose a finite cover of σ by neighbourhoods $N(x_1, \delta_{x_1}), ..., N(x_n, \delta_{x_n})$ and let $\lambda_\sigma = \min\{\lambda_{x_i}\}_{i=1,...,n}$. Then, for an arbitrary $x' \in \sigma$ and any $J \in \mathcal{S}$, with $d(H, J) < \lambda_\sigma$, if $x' \in N(x_i, \delta_{x_i})$ we have that $L_J(x') > L_H(x_i) - \varepsilon$ and since $|L_H(x_i) - L_H(x')| < \varepsilon$, $L_J(x') > L_H(x') - 2\varepsilon$. If, in σ, $d(H, J) < \lambda_\sigma \Rightarrow L_J > L_H - 2\varepsilon$, integrating we have

$$\int_\sigma L_J(x)\ dx \geq \int_\sigma (L_H(x) - 2\varepsilon)\ dx = \int_\sigma L_H(x)\ dx - 2\varepsilon A(\sigma)$$

where $A(\sigma)$ denotes the "area" of σ. Summing over all simplices, we have that if $d(H, J) < \lambda$, with $\lambda = \min_{\sigma\in K^m}\{\lambda_\sigma\}$

$$\begin{aligned} E(J) &= \int_K L_J(x) = \sum_{\sigma\in K^m} \int_\sigma L_J(x)\ dx \geq \sum_{\sigma\in K^m} \left(\int_\sigma L_H(x)\ dx - 2\varepsilon A(\sigma)\right) \\ &= \sum_{\sigma\in K^m} \int_\sigma L_H(x)\ dx - 2\varepsilon \sum_{\sigma\in K^m} A(\sigma) = E(H) - 2\varepsilon A(K) \end{aligned}$$

$A(K)$ the total area of the complex K, and so we've finished the proof that E is lower-semi-continuous.

It is easy to adapt the example of contractions of the unit interval we used to show that the volume function is discontinuous - represented in figure 9 above - to show that the energy function is discontinuous as well; instead of the vertical segment S, just start with a horizontal one, all the rest being the same: then, for each stalk $t_0 \times I$ that intersects the horizontal segment S, the length $L_{J_n}(t_0)$ has the same value as the area $A_{J_n}(t_0)$ in the original example, and so the argument and calculations are exactly the same.

Like the volume the energy function doesn't have minima. For the volume that fact was easily established by reparametrizing any contraction, speeding it up; that simple procedure doesn't work for the energy since energy is invariant by reparametrizations. But consider the example of the space K represented in figure 13.

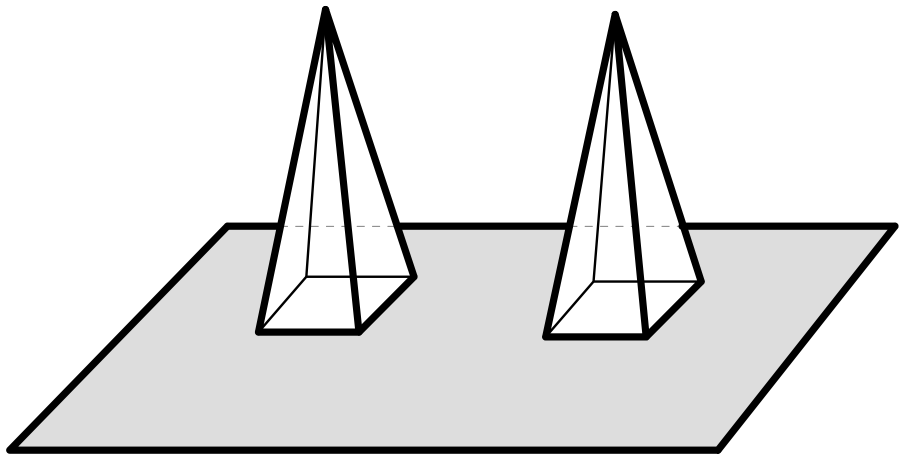

Figure 13

This space, homeomorphic to the 2-disc, is constructed from a square, by removing the interiors of two embedded squares and gluing two pyramids (their lateral faces) to the edges of those squares; the metric is the length metric, induced by the embedding in euclidian 3-space.

Exercise 20 *Show that the energy function has not a minimum in the space $\mathcal{S}$ of contractions associated with the complex K.*

Hint: *note that given a contraction $H : [K \times I]^{sub} \longrightarrow K^{sub}$ of K to an arbitrary point x_0, for any point $x \in K$, the length of its track, l_x, is greater or equal to the length of a* geodesic *from x to x_0 ,which is precisely the distance $d(x, x_0)$ - recall exercise 18 - therefore, $E(H) \geq m = \int_K d(x, x_0)$. Show that one can get contractions of K with energy as close to m as we like but always greater than m (m is an infimum but not a minimum): looking from x_0, there is hiding behind each pyramid a line from the summit to the boundary of the disc, consisting of points each of which has two geodesics to x_0 that go around the pyramid in opposite ways (in case x_0 is the summit of one of the pyramids, then there is just one line behind the other pyramid)*

Remark 21 *The definitions of the volume and energy functions, $V, E : \mathcal{S} \longrightarrow [0, +\infty)$, and the proofs of their* lower-semi-continuity *don't depend on the fact that we were considering simplicial maps between simplicial complexes: in reality, we only used the fact the contractions $H \in \mathcal{S}$ were* (affine) linear on each simplex. *It is natural, and most useful, to relax the previous demands and include in the space $\mathcal{S}$ the larger class of functions which satisfy this weaker condition: it is useful because, as hinted in the introduction, we want to analyse the volume and energy of contractions $H : M = (K \times I)^{sub} \longrightarrow K^{sub}$ through local small changes, and that is achieved by considering - at least locally - sufficiently small subdivisions of M and K^{sub}. The problem is that staying in the simplicial realm would force us to look at the changes of energy and volume in simplices of M other than the local ones we need to consider, perhaps distant ones but whose images by H are also the chosen local simplices of K we are now subdividing. That happens because despite the great advantages of affine coordinates they don't behave well under subdivisions, even baricentric ones. Look at figure 14. We have two 2-simplices of a complex K that are sent to a 1-simplex, the segment $\overline{AB}$ whose baricentre is C: the two linear maps are defined by sending the vertices V_1, V_2 and W_1, W_2 to A and V_3 and W_3 to B. Suppose we change the image of the first simplex* *through a (first) baricentric subdivision, sending* V_1, V_2 *and the baricentre of* $\overline{V_1V_2}$ *to* A *and the other four vertices of the subdivision (including* V_3*) to* C*;*

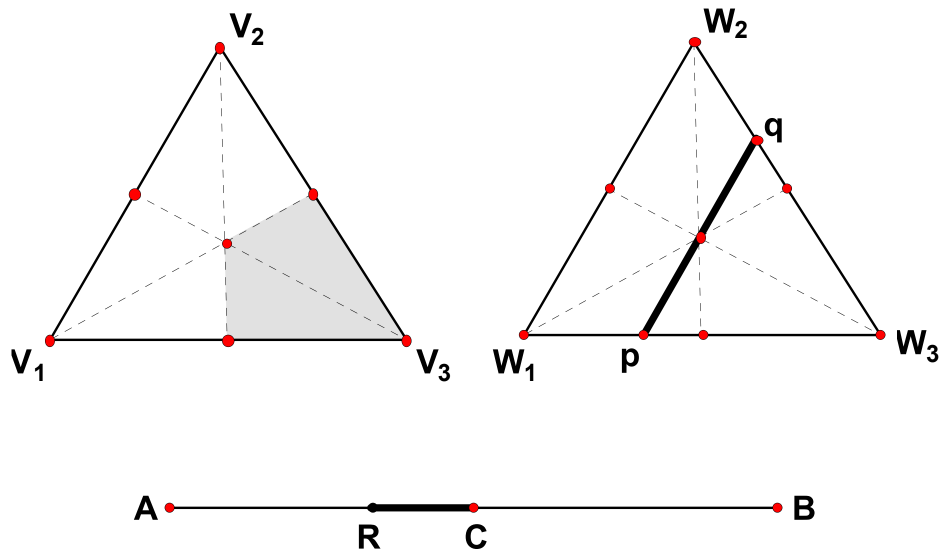

Figure 14

the image is now $\overline{AC}$*: the shaded triangles are squeezed to vertex* C*, the segment* $\overline{V_1V_2}$ *to vertex* A *and the other five edges - and remaining four white simplices - are sent to* $\overline{AC}$*. If, for instance,*$\overline{V_1V_3}$ *was part of the stalk of some point* $x \in K$*, this process would reduce the length* l_x *of the respective track by* $d(A,B)/2$*. In the second 2-simplex consider the segment* $\overline{pq}$ *parallel to* $\overline{W_1W_2}$*. Since the medians of a triangle trisect each other, the segment* $\overline{W_1p}$ *is one third of* $\overline{W_1W_3}$ *and likewise* $\overline{W_2q}$ *is one third of* $\overline{W_2W_3}$*, therefore, by linearity, the whole segment* $\overline{pq}$ *is sent to the point* R *whose distance from* A *is one third of the length of* $\overline{AB}$ *- in particular the baricentre is not sent to* C*. If we consider the first baricentric subdivision and adjust the map on this* 2*-simplex to remain simplicial so as to conform to the new subdivision of* $\overline{AB}$*, it easy to see that the points* p, q *are still sent to the point* R *(*$2/3 \times 1/2 = 1/3$*!) but now the baricentre in the interior of the* 2*-simplex is sent to* C *and so the segment* $\overline{pq}$ *is sent from* R *to* C *and back from* C *to* R*, thus increasing the length of its image: if we think of* $\overline{pq}$ *as part of the stalk of some point, the corresponding track would have its length increased by* $d(A,B)/3$*. That's the sort of problem we want to avoid by our extending space* $\mathcal{S}$ *to include all maps that are* linear on each simplex, but not necessarily simplicial*: in the present example we would not change the map on the second* 2*-simplex, even if it is no longer simplicial relative to the* new *subdivision of* $\overline{AB}$*. In accordance with this, and indeed equivalently, we don't need to assume that the subdivisions* M^{sub}, K^{sub}*, are simplicial, that is, the non-empty intersection of any two simplices is a common face, but simply that they are linear: we can consider linear subdivisions of any simplex without having to extend in a compatible way to neighbouring simplices.*

To sum up our remarks: we assume, from now on, that the space $\mathcal{S}$ *consists of all contractions,* $H : M = (K \times I)^{sub} \longrightarrow K^{sub}$*, where the subdivisions,* $?^{sub}$*, are not always simplicial but consist of - not necessarily coherent - (simplicial) subdivisions of chosen sets of simplices and the maps* H *are* linear on each simplex σ *of the subdivision, where linearity is relative not necessarily to some simplex of* K^{sub}*, but more generally to some simplex,* $\rho \in K$*,* ancestor *of the subdivision and, furthermore, being linear doesn't mean it is simplicial, in the sense that the image of* σ *is not necessarily the whole of* ρ*.*

4.3 Minimizing volume and energy

Next, we will be looking for suitable subspaces of $\mathcal{C}$, the space of contractions of a given simplicial complex K, where the restrictions of the energy and volume functions attain a minimum. The natural thing to do is to look for *compact subspaces of* $\mathcal{C}$ but, as we've already hinted at in the introduction, we want those spaces to be sufficiently rich and flexible to allow us to perform some natural local movements - as those we will consider in section 5.

For technical reasons which related to the previous remark, we will consider decompositions of spaces into *cubes instead of simplices*. The standard n-cube, $\mathbb{I}^n \subset \mathbb{E}^n$, is the the cartesian product of I by itself n times: $\mathbb{I}^n = I \times ... \times I = \{(x_1, ..., x_n) \in \mathbb{E}^n : 0 \leq x_i \leq 1\}$. In $\mathbb{I}^n$ we have 2^n vertices, all the points with coordinates 0 or 1, and $2n$ faces of dimension $n-1$ which are the intersections of $\mathbb{I}^n$ with the $2n$ hyperplanes with equations $x_i = 0$ or $x_i = 1$. For each of these faces there is a natural isometry into the $n-1$ cube, given by the forgetting of the i-coordinate. Then we have faces of dimension $n-2$, given by the intersection of two $n-1$ faces, that is by the conjunction of two of those equations $x_i = 0, 1$ and $x_j = 0, 1$, and more generally faces of all dimensions from $n-1$ down to 1, the dimension of the edges, and 0 the dimension of the vertices. We will consider more generally an n-cube in $\mathbb{E}^n$ as the convex polyhedron determined by the combinatorial equivalent arrangement of $2n$ hyperplanes, with the corresponding sets of faces of the various dimensions.

We will consider our spaces K decomposed into cubes: a *squaring* (or *cubication* to have a word analogous to triangulation) of the space K is a decomposition into *cubes* - that is, into subspaces each of which is homeomorphic to a cube of a certain dimension - such that each two cubes either do not intersect or intersect along a common proper face. We will say that the space K is *squared* or *cubical*. We have for cubical spaces the analogues of star, St, link, Lk, k-skeleton, K^k, subcomplex, $M \leq K$, etc..

Of course, for any cubical complex K, the product $K \times I$ has a very natural squaring, with cubes the products $\omega \times I$ of all the cubes of K with I. As in the simplicial case, we consider in K and $K \times I$ compatible euclidian embeddings of the constituent cubes, locally in each cube the transport metric from the corresponding embedding, and then globally the length-metric. The definition of *full* cubical complex is identical to the previous one.

The two types of decomposition, triangulations and squarings, are naturally intertwined. To any squaring of the space K there is associated its *standard triangulation*, obtained by subdividing each n-cube into n-simplices in the following standard way - in what follows it might be useful to visualize in dimension 3. Let ψ be an n-cube; its *baricentre*, $\hat{\psi}$ is the affine combination of its 2^n vertices, $\sum_{i=1}^{2^n} \lambda_i v_i$, $\sum \lambda_i = 1$, with all coefficients equal to $1/2^n$. Keep in mind that, since the vertices of a cube are not affine independent, the affine coordinates relative to those vertices are not unique: for that reason, and apart the convenience for certain definitions, like the one just given of baricentre, we will not be able to use them as in the simplicial case, for instance to define what will be the equivalent for the squarings of a simplicial map. We use the usual notation for proper faces: if ω is a proper face of ψ, write $\omega < \psi$. The standard triangulation of an n-cube ψ consists of all the n-simplices with vertices $(\hat{\omega}_n, \hat{\omega}_{n-1}, \hat{\omega}_{n-2}, ..., \hat{\omega}_1, \hat{\omega}_0)$ where ω_i is a face of dimension i and $\omega_{i-1} < \omega_i$; of course, $\omega_n = \psi$ and $\hat{\omega}_0 = \omega_0$ is a vertex of ψ. In other words, we obtain the simplices of the standard triangulation by all possible choices of sequences of baricentres, where we start with a vertex of ψ, then choose the baricentre of an edge that has that vertex as an end point, then the baricentre of a square that has that edge, then the baricentre of a cube (a real 3-dimensional one) that has that square as a face, and so on until we reach the baricentre of ψ: for instance, in dimension 3 we have the standard decomposition of a cube into 48 simplices. We leave it to the reader to work out the relation between simplicial and cubical complexes, in the opposite direction.

The equivalent of a simplicial map is a *cubical map*. Given two cubes ψ_1 and ψ_2, not necessarily of the same dimension, a cubical map $\varphi : \psi_1 \longrightarrow \psi_2$ will be completely determined by choosing for each vertex v of ψ_1 a vertex $\varphi(v) \in \psi_2$ to be its image. The definition will be given by induction on the dimension of ψ_1. Let $\dot{\psi}$ denote the boundary of the cube ψ, that is the union of all proper faces of ψ: clearly we can represent the cube ψ as the *joint* $\dot{\psi} * \hat{\psi}$, the union of all line segments $\hat{l}x = \overline{x\hat{\psi}}$, $x \in \dot{\psi}$. If the dimension of ψ_1 is 0, that is ψ_1 is just a vertex, then φ is simply a choice of a vertex $\varphi(\psi_1)$ in ψ_2. If the dimension of ψ_1 is n, let $\omega \leq \psi_2$ be the face of *smallest dimension* that contains the set of vertices $\varphi(\psi_1^0)$; by induction we have defined φ on each of the $(n-1)$-cubes that make up the boundary $\dot{\psi}_1$ and of course $\varphi(\dot{\psi}_1) \subset \omega$: we extend φ to the interior of ψ_1 by setting $\varphi(\hat{\psi}_1) = \hat{\omega}$ and sending each segment $\hat{l}x$, $x \in \dot{\psi}_1$, linearly to the segment $\overline{\varphi(x)\hat{\omega}}$ (which may be be reduced to a point). The reader is advised to construct and visualize some cubical maps of the 3-cube $\mathbb{I}^3$ to itself, for instance the one that keeps all the vertices fixed except one which is sent to the opposite vertex. Of course, a map $\varphi : K \longrightarrow L$ between cubical complexes is said to be a *cubical map* if the restriction to each cube of K is a cubical map into a cube of L. By the definition it's clear that a restriction of a cubical map to a subcomplex still is a cubical map.

Since a cubical map $\varphi : \psi_1 \longrightarrow \psi_2$ sends vertices to vertices - in fact, by definition, it is determined by those images $\varphi(\psi_1^0) \subset \psi_2^0$ - and sends the baricentres of faces of ψ_1 to baricentres of faces of ψ_2, and in face of the definition of the standard triangulation of a cube given above, it is easy to see that *a cubical map is a simplicial map for the standard triangulations.* Let $\sigma = (\hat{\omega}_n, \hat{\omega}_{n-1}, \hat{\omega}_{n-2}, ..., \hat{\omega}_1, \hat{\omega}_0)$ be an n-simplex of the standard triangulation of the cube where, as above, $\hat{\omega}_n = \hat{\psi}_1$, $\hat{\omega}_0 = v_0$ is a vertex of ψ_1, and ω_i is a face of dimension i with $\omega_{i-1} < \omega_i$; the simplex can be seen as a sequence of joints: $\sigma_1 = v_0 * \hat{\omega}_1 \subset \omega_1 \subset \dot{\omega}_2$, $\sigma_2 = \sigma_1 * \hat{\omega}_2 \subset \omega_2 \subset \dot{\omega}_3$, ..., $\sigma = \sigma_n = \sigma_{n-1} * \hat{\psi}_1$, corresponding to a sequence of faces of increasing dimension $v_0 < \sigma_1 < \sigma_2 < \cdots < \sigma_{n-1} < \sigma_n = \sigma$. Following the definition of cubical map, let for each $i = 0, 1, ..., n$, $\varkappa_i$ be the face of smallest dimension of ψ_2 that contains the images $\varphi(\omega_i^0)$ of the vertices of ω_i; since $\omega_{i-1} < \omega_i$ and therefore $\omega_{i-1}^0 \subset \omega_i^0$, $\varkappa_i$ contains $\varkappa_{i-1}$, that is $\varkappa_{i-1}$ is a face, but perhaps not a proper face, of $\varkappa_i$; we have thus a sequence of faces of ψ_2, of increasing - perhaps not strictly - dimension, all of which contain the vertex $\varphi(v_0)$: $\varphi(v_0) = \varkappa_0 \leq \varkappa_1 \leq \cdots \leq \varkappa_n \leq \psi_2$; therefore the baricentres $\hat{\varkappa}_0 = \varkappa_0$, $\hat{\varkappa}_1$, ..., $\hat{\varkappa}_n$, not necessarily distinct, are vertices of a top-dimensional simplex ρ of the standard triangulation of ψ_2 that contains the vertex $\varphi(v_0) = \varkappa_0$. Since, by definition of cubical map, $\varphi(\hat{\omega}_i) = \hat{\varkappa}_i$ we have that the vertices of σ are sent to vertices of ρ; furthermore, the segments that make up the above sequence of joints that we considered as forming σ are sent linearly to segments in the simplex ρ: but a simplicial map between the two simplices σ and ρ has exactly this characteristic property.

We now turn to *(cubical) baricentric subdivisions* of cubical complexes. Let $I_0 = [0, 1/2]$ and $I_1 = [1/2, 1]$; the (first) baricentric (cubical) subdivision of the n-cube $\psi = \mathbb{I}^n$ consists of the 2^n embedded n-cubes $I_{i_1} \times I_{i_2} \times \cdots \times I_{i_n}$ where $(i_1, i_2, ..., i_n) \in \{0, 1\}^n$; clearly these cubes have for vertices the original vertices of $\mathbb{I}^n$ plus the baricentres of its faces: more specifically, for each vertex v of $\mathbb{I}^n$ there is exactly one cube of the subdivision, ψ_v, that has for vertices, besides v, the baricentres of all the cubes in $St(v)$, $\psi_v^0 = \{v\} \cup \{\hat{\omega} : \omega \in St(v)\}$; in particular the baricentre $\hat{\psi} = (1/2, ..., 1/2) \in \mathbb{I}^n$ is a vertex of all the cubes of the subdivision. Analogously to the two vertices v and $\hat{\psi}$ of ψ_v, for any face $\pi \leq \psi_v$ of dimension k there are exactly two vertices, say $u_b = \hat{\omega}_b$ and $u_t = \hat{\omega}_t$, $\omega_b, \omega_t \in St(v)$, such that $\dim(\omega_t) - \dim(\omega_b) = k$ and such that for any other vertex $u = \hat{\omega}$ we have $\omega_b < \omega < \omega_t$; we call u_b and u_t, the bottom and top dimensional vertices of π, respectively; note that baricentre of the cube π is the midpoint of the segment $\overline{u_b u_t}$. Given an arbitrary n-cube, $\psi \subset \mathbb{E}^n$, its (first) baricentric (cubical) subdivision consists of the 2^n embedded n-cubes which are the images of the cubes of the baricentric subdivision of $\mathbb{I}^n$ we've just defined, by any cubical homeomorphism $\varphi : \mathbb{I}^n \longrightarrow \psi$. Given a cubical complex K, we denote its iterated n-th baricentric (cubical) subdivision by $K^{[n]}$.

Let $\varphi : \psi \longrightarrow \psi'$ be a cubical map between two cubes and consider the first baricentric subdivisions, $\psi^{[1]}$ and $(\psi')^{[1]}$. Let ψ_v, $v \in \psi^0$, be a cube of $\psi^{[1]}$ and consider the cube $\psi'_{\varphi(v)} \in (\psi')^{[1]}$; for any vertex of ψ_v, which is of the form $\hat{\omega}$, $\omega \in St(v)$, we have, by definition of cubical map, that $\varphi(\hat{\omega}) = \hat{\varkappa}$, where $\varkappa$ is the face of smallest dimension that contains all the vertices $\varphi(\omega^0)$, therefore $\varkappa \in St(\varphi(v))$ and so $\hat{\varkappa}$ is a vertex of the cube $\psi'_{\varphi(v)}$: in conclusion, φ sends the vertices of the cube ψ_v to vertices of the cube $\psi'_{\varphi(v)}$; in fact the restriction of φ to ψ_v is a cubical map between the two subdivision cubes. Following the given definition of cubical map, we proceed by induction on the dimension of the faces of ψ_v. Let $\pi \leq \psi_v$ be a face of dimension k and let $u_b = \hat{\omega}_b$ and $u_t = \hat{\omega}_t$, $\omega_b, \omega_t \in St(v)$ be its bottom and top dimensional vertices; since for any other vertex $u = \hat{\omega} \in \pi^0$ we have $\omega_b < \omega < \omega_t$, the smallest dimension face of ψ' that contains all the vertices $\varphi(\omega_t^0)$, say $\varkappa_t \in St(\varphi(v))$, also contains the vertices $\varphi(\omega^0)$ and $\varphi(\omega_b^0)$. Therefore if $\varphi(\hat{\omega}_b) = \hat{\varkappa}_b$, $\varphi(\hat{\omega}) = \hat{\varkappa}$, $\varphi(\hat{\omega}_t) = \hat{\varkappa}_t$ we have $\varkappa_b \leq \varkappa \leq \varkappa_t$; as a consequence if $\varkappa'$ is the face of $\psi'_{\varphi(v)}$ of smallest dimension that contains all the vertices $\varphi(\pi^0)$,then $\varkappa' \subset \psi'_{\varphi(v)} \cap \varkappa_t$ and are its bottom and top dimensional vertices, respectively. Now, the baricentre of π is the midpoint of the segment $\overline{u_b u_t}$, and likewise the midpoint of the segment $\overline{\hat{\varkappa}_b \hat{\varkappa}_t}$ is the baricentre of π'; since φ sends $\overline{u_b u_t}$ linearly to $\overline{\hat{\varkappa}_b \hat{\varkappa}_t}$, we have that $\varphi(\hat{\pi}) = \hat{\pi}'$; it remains to be seen that for each $x \in \dot{\pi}$, $\overline{x\hat{\pi}}$ is sent linearly to $\overline{\varphi(x)\hat{\pi}'}$: from the construction of the standard triangulation of ψ, we have that each cube ψ_v, as well as each of its proper faces, is a union of simplices of that triangulation - this means that the standard triangulation of ψ is also a triangulation of $\psi^{[1]}$ - furthermore all the segments $\overline{u_b u_t}$ joining bottom and top dimensional vertices are in fact edges of the standard triangulation; this implies that the segment $\overline{x\hat{\pi}}$ is contained in a simplex of the standard triangulation, and since φ is also simplicial, as we saw above, we achieve the conclusion. We have thus proved the following result, showing that - in contrast with what happens with simplicial maps, exemplified in Remark 21 - cubical maps behave well under baricentric subdivisions:

Lemma 22 *Let K and L be cubical complexes and $\varphi : K \longrightarrow L$ a cubical map. Then, for all $n \in \mathbb{N}$, $\varphi : K^{[n]} \longrightarrow L^{[n]}$ is cubical, and simplicial relative to the standard triangulations of $K^{[n]}$ and $L^{[n]}$.*

Note that the standard triangulation of a baricentric (cubical) subdivision, $K^{[1]}$, is a subdivision of the standard triangulation of K, but is not the (simplicial) baricentric one.

The notion of *simplicial approximation* to a continuous map translates from the simplicial to the cubical setting, mutatis mutandi: given cubical complexes K and L and a continuous map $f : K \longrightarrow L$, a cubical map $g : K \longrightarrow L$ is called a *cubical approximation* to f if for each vertex c of K, $f(St_K(c)) \subset St_L(g(c))$.

The proofs of Proposition 15 and Theorems 16,17 (see [3, Chapter 2]) translate in a straightforward manner to give proofs of their cubical analogues, which in turn entail that for any contractible cubical complex K, there is $k \in \mathbb{N}$ and a *cubical contraction* $H : (K \times I)^{[k]} \longrightarrow K$.

Let $H : (K \times I)^{[k]} \longrightarrow K$ be a cubical contraction of the cubical space K, and let $M = (K \times I)^{[k]}$; recall that by the previous lemma, for all $n \in \mathbb{N}$, $H : M^{[n]} \longrightarrow K^{[n]}$ is a cubical contraction. We denote by $\mathcal{Q}$ (or by $\mathcal{Q}_H$ or $\mathcal{Q}_{(k)}$) the space of all cubical contractions $J : M^{[n]} \longrightarrow K^{[n]}$, with the sup-metric: we repeat that our basic idea is to be able to modify a given contraction, say H, by local small moves involving sufficiently fine (cubical) baricentric subdivisions; since each such J is a simplicial map between the *standard triangulations* of $M^{[n]}$ and $K^{[n]}$ we can regard $\mathcal{Q}$ as a subspace of $\mathcal{S}$: we refer to the simplicial maps $J : (K \times I)^{sub} \longrightarrow K^{sub}$ and corresponding simplicial subdivisions, $(K \times I)^{sub}$ and K^{sub}, so obtained as being *cubically framed* (or as having *cubical support*). We thus have the inclusions $\mathcal{Q} \subset \mathcal{S} \subset \mathcal{C}$.

Let $\mathcal{G} = \overline{\mathcal{Q}}$ (or $\mathcal{G}_H$ or $\mathcal{G}_{(k)}$) be the closure of $\mathcal{Q}$ in the space $\mathcal{C}$, the space of all contractions of the cubical space K. We should draw your attention to the following: as in a corollary to the Simplicial Approximation Theorem (see the paragraph after Theorem 16) its cubical analogue, the *Cubical Approximation Theorem* entails that any continuous map $f : K \longrightarrow L$ between cubical complexes, can be arbitrarily ε-approximated, $\varepsilon > 0$, by a cubical map $g : K^{[k]} \longrightarrow L^{[j]}$ for suitable $k, j \in \mathbb{N}$, in particular any contraction in $\mathcal{C}$ can be arbitrarily ε-approximated by a cubical one; but note that the space $\mathcal{Q}$ was defined just with the baricentric subdivisions $M^{[n]}$ and $K^{[n]}$, we are not allowed to vary the upper-indices separately: in fact it will became clear, by the results that follow and taking into account some previous examples, that $\mathcal{G} = \overline{\mathcal{Q}}$ is a much restricted space of $\mathcal{C}$.

$\mathcal{G}$ is a *compact subspace* of $\mathcal{C}$. Since we are assuming all spaces K to be compact, the topology of uniform convergence in $\mathcal{C}$, given by the sup-metric, is the same as the compact-open topology (the c-topology) and by the *Arzela-Ascoli Theorem* (see [8, Chapter XII]) the compactness of $\mathcal{G} = \overline{\mathcal{Q}}$ is equivalent to the *equicontinuity* of $\mathcal{Q}$ - and then $\mathcal{G}$ is also equicontinuous. This is what we will analyse next.

Let $\varphi : \psi_1 \longrightarrow \psi_2$ be a cubical map between cubes; it's clear that φ has a *bounded stretching-factor*, that is, there is some positive constant $c_\varphi > 0$ such that for all line segments $r \subset \psi_1$ the ratio of the lengths $l(\varphi(r))/l(r)$ doesn't exceed c_φ; recall that φ is simplicial relative to the standard triangulations of the two cubes: now, in each n-simplex σ, where $n = \dim(\psi_1)$, of the standard triangulation of ψ_1, φ being linear in σ we can find a direction of maximal stretch with stretching factor say c_σ: obviously $c_\varphi = \max_\sigma \{c_\sigma\}$ is the required constant. Since there are only a finite number of cubical maps between two cubes, we can obviously take the maximum $\max_\varphi \{c_\varphi\}$ as a common stretching-factor bound for all of them which we denote by $c(\psi_1, \psi_2)$. But a much stronger result holds: it is possible to find a *global stretching-factor bound* which is hereditary under iterated baricentric subdivisions, that is, there is a constant $C(\psi_1, \psi_2) > 0$ such that for all $n \in \mathbb{N}$ and for all cubes $\omega \in \psi_1^{[n]}$, $\varkappa \in \psi_2^{[n]}$, $c(\omega, \varkappa) \leq C$. We can assume without loss of generality that $\dim(\psi_1) = \dim(\psi_2) = m$: if one of them has smaller dimension, consider it as a face of a cube ψ of the greater dimension; in the case $\psi_2 < \psi$, it is clear that a global bound for the pair (ψ_1, ψ) is also a bound for the pair (ψ_1, ψ_2) since $\psi_2^{[n]}$ is a subcomplex of $\psi^{[n]}$ and so in the set of constants $c(\omega, \varkappa)$, where $\omega \in \psi_1^{[n]}$ and $\varkappa \in \psi^{[n]}$, $\varkappa$ ranges in particular over $\psi_2^{[n]}$; the other case, $\psi_1 < \psi$, is perfectly analogous. Suppose there was not such a constant $C(\psi_1, \psi_2)$, that is, for each positive constant $k > 0$ we could find $n_k \in \mathbb{N}$, two cubes $\omega_k \in \psi_1^{[n_k]}$, $\varkappa_k \in \psi_2^{[n_k]}$ and a cubical map : $\omega_k \longrightarrow \varkappa_k$ such that $c_{\varphi_k} > k$. Fix two cubical homeomorphisms $h_1 : \mathbb{I}^m \longrightarrow \psi_1$ and $h_2 : \mathbb{I}^m \longrightarrow \psi_2$; to each cube $\omega_k \in \psi_1^{[n_k]}$ there corresponds exactly one cube in $(\mathbb{I}^m)^{[n_k]}$, the cube $\mu_k = h_1^{-1}(\omega_k)$ and likewise for $\varkappa_k \in \psi_2^{[n_k]}$ with $\nu_k = h_2^{-1}(\varkappa_k)$; consider the cubical map : $\mu_k \longrightarrow \nu_k$ defined as the composition $\phi_k = h_2^{-1} \circ \varphi_k \circ h_1$ where we are still using h_1 for the restriction to μ_k; since these are homeomorphisms, it is clear that the stretching factor for ϕ_k is greater or equal to $c_{\varphi_k}/(c_{h_1} \times c_{h_2})$, that is $c_{\phi_k} > k/a$ where $a = c_{h_1} \times c_{h_2}$ is a constant; in conclusion, we found cubes $\mu_k, \nu_k \in (\mathbb{I}^m)^{[n_k]}$ with arbitrarily large stretching factors $c(\mu_k, \nu_k) \geq c_{\phi_k} > k/a$: but this is a contradiction since $\mathbb{I}^m$ is highly symmetric under baricentric subdivisions, as a consequence any cube in a subdivision , $(\mathbb{I}^m)^{[n]}$, is similar by a factor of $2n$ to $\mathbb{I}^m$ or one of its faces and therefore $c(\mu_k, \nu_k) \leq c(\mathbb{I}^m, \mathbb{I}^m)$. Note that the same argument could not be carried on in the simplicial context - this is a main reason to have chosen to go cubical - since this last property doesn't hold for baricentric subdivisions of the standard simplex Δ^n: in successive subdivisions the shapes of the simplices keep changing, with the emergence of new simplices not similar to any in previous subdivisions, as becomes apparent in next figure where the three stages for the third baricentric subdivision of Δ^2 are represented; the two shaded 2-simplices, with vertex at the baricentre of Δ^2, show that in the successive subdivisions we have triangles that have, at that vertex, angles with ever decreasing amplitudes.

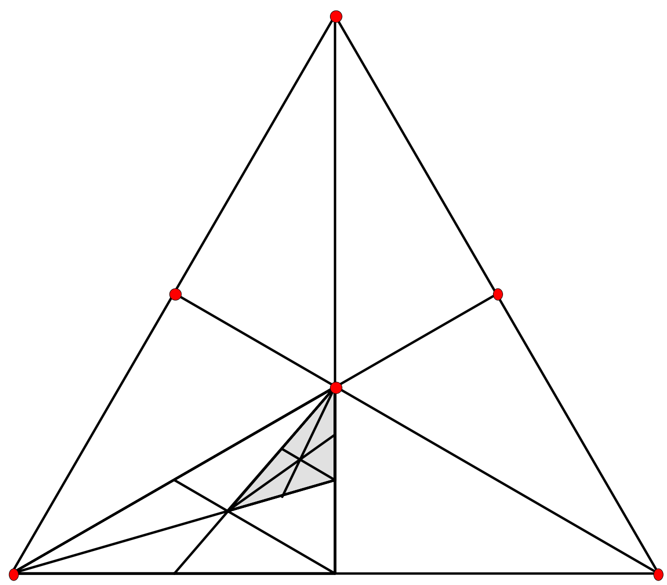

Figure 15

Let's now see the *equicontinuity* of $\mathcal{Q}$; recall that in the cubical complexes K and $M = (K \times I)^{[k]}$ we consider the *length-metric* given locally, in each cube of the decomposition, by the transport metric from a euclidian embedding of that cube. Let p be a *geodesic* in $K \times I$ from the point x to the point y: then the distance between x and y equals the length of p: $d(x, y) = l(p)$. Consider an arbitrary element of $\mathcal{Q}$, $J : M^{[n]} \longrightarrow K^{[n]}$; obviously, p is a polygonal path made up of a succession of line segments each one of which is contained in a cube of $M^{[n]}$ (a top-dimensional one if we assume K to be a full complex): $p = p_1 \cup p_2 \cup \cdots \cup p_j$, with $p_i \in \omega_i$, $\omega_i \in M^{[n]}$ and of course we have $l(p) = \sum_{i=1}^{j} l(p_i)$. $J(p)$ is a polygonal path joining $J(x)$ to $J(y)$, $J(p) = J(p_1) \cup J(p_2) \cup \cdots \cup J(p_j)$ where, for each $i = 1, ..., j$, $J(p_i)$ is a polygonal path in the cube $\varkappa_i = J(\omega_i)$. Of course that $l(\ J(p_i)) \leq l(p_i) \times c(\omega_i, \varkappa_i)$ and if ψ and π are cubes that are ancestors, in M and K respectively, of the cubes ω_i and $\varkappa_i$, we have that $c(\omega_i, \varkappa_i) \leq C(\psi, \pi)$. If we take the maximum of the stretching factor bounds for all pairs of cubes of M and K, $B = \max_{(\psi,\pi)\in M\times K} \{C(\psi, \pi)\}$, we have

$$l(J(p)) = \sum_{i=1}^{j} l(J(p_i)) \leq \sum_{i=1}^{j} l(p_i)c(\omega_i, \varkappa_i) \leq B \sum_{i=1}^{j} l(p_i) = Bl(p) = Bd(x,y)$$

But clearly $d(J(x), J(y)) \leq l(J(p))$ and so we have proved that for each $J \in \mathcal{Q}$ and for any two points $x, y \in M$, we have $d(J(x), J(y)) \leq Bd(x,y)$ thus setting the (uniform) equicontinuity of $\mathcal{Q}$ - in fact we've proved a sharper result: $\mathcal{Q}$ is *equi-Lipschitz.* Let's state the relevant conclusion:

Theorem 23 *Let K be a contractible cubical complex and $H : M = (K \times I)^{[k]} \longrightarrow K$ a cubical contraction. Let $\mathcal{Q}$ be the space of all cubical contractions $J : M^{[n]} \longrightarrow K^{[n]}$, $n \in \mathbb{N}$, and $\mathcal{G} = \overline{\mathcal{Q}}$ the closure in $\mathcal{C}$, the space of all contractions of K.*

Then $\mathcal{G}$ is compact.

Consider again the energy and volume functions, $E, V : \mathcal{S} \longrightarrow [0, +\infty)$ and let Φ stand for the restriction of either of them to $\mathcal{Q} \subset \mathcal{S}$. Of course, Φ is lower-semicontinuous, being the restriction of a lower-semicontinuous function and previous examples (represented in Figure 9) that showed that V, E are not (upper-semi)continuous can be readily readjusted to triangulations (of $I \times I$) with *cubical support* and cubical maps, showing that Φ is not (upper-semi)continuous.

We extend Φ to $\mathcal{G} = \overline{\mathcal{Q}}$, denoting the extension by the same letter: given $J \in \mathcal{G} - \mathcal{Q}$, define $\Phi(J)$ by

$$\begin{aligned} \Phi_\varepsilon(J) &= \inf\{\Phi(H) : H \in \mathcal{Q} \wedge d(J,H) < \varepsilon\} \\ \Phi(J) &= \lim_{\varepsilon \longrightarrow 0} \Phi_\varepsilon(J) \end{aligned}$$

We need to check that the limit in the definition does exist: since, obviously, $\varepsilon < \varepsilon' \Rightarrow \Phi_\varepsilon(J) \geq \Phi_{\varepsilon'}(J)$ what we need to check is that $\{\Phi_\varepsilon(J); \varepsilon > 0\}$ is bounded above, for then $\lim_{\varepsilon \longrightarrow 0} \Phi_\varepsilon(J) = \sup_{\varepsilon > 0} \Phi_\varepsilon(J)$. In fact more is true: both functions, energy and volume, are actually bounded above in $\mathcal{Q}$. Consider the constant B we constructed above in the proof of the equicontinuity of $\mathcal{Q}$. Going back to the definition of the energy function: given any $J : M^{[n]} \longrightarrow K^{[n]}$, consider the length function $L_J : K \longrightarrow [0, +\infty)$, $L_H(x) = l(J(S_x)$, where $S_x = \{x\} \times I$ is the stalk over x; S_x has a certain length in $M^{[n]}$, say $l(S_x)$; from the definition of B as a global stretching factor bound for all $J \in \mathcal{Q}$, it is clear that $L_J(x) \leq Bl(S_x)$ and therefore

$$E(J) = \int_K L_J(x) \leq B \int_K l(S_x)$$

(where the last integral could be taken as the total volume of the complex M). Likewise, in the definition of the volume function, if we look at the area-function $A_J(t) = A(J(M_t))$, where $M_t = K \times \{t\}$ is the slice at level t, we will have $A_J(t) \leq B^m A(M_t)$ where $A(M_t)$ is the area of the slice and m is the dimension of K; therefore

$$V(J) = \int_0^1 A_J(t)dt \leq B^m \int_0^1 A(M_t)dt$$

(where the last integral could also be taken as the total volume of M). Note that, since Φ is lower-semicontinuous, for all $J \in \mathcal{Q}$ we also have that $\Phi(J) = \lim_{\varepsilon \longrightarrow 0} \Phi_\varepsilon(J)$.

It is easy to see that the extension $\Phi : \mathcal{G} \longrightarrow [0, +\infty)$ is lower-semicontinuous. Suppose, on the contrary, that Φ was lower-semidiscontinuous at point J; we can distinguish the two cases, $J \in \mathcal{G} - \mathcal{Q}$ and $J \in \mathcal{Q}$. Assume first that $J \in \mathcal{Q}$: if Φ is lower-semidiscontinuous at J then there is some $\varepsilon_0 > 0$ such that for all $\delta > 0$ there exists $H_\delta \in \mathcal{G}$ such that $d(J; H_\delta) < \delta$ and $\Phi(H_\delta) < \Phi(J) - \varepsilon_0$; since Φ is lower-semicontinuous in $\mathcal{Q}$, we have that $H_\delta \in \mathcal{G} - \mathcal{Q}$; by the definition of $\Phi(H_\delta)$, and considering $\Phi_\delta(H_\delta) \leq \Phi(H_\delta) < \Phi(J) - \varepsilon_0$, there is $G \in \mathcal{Q}$ such that $d(H_\delta, G) < \delta$ and $\Phi(G) < \Phi(J) - \varepsilon_0$; but then $d(J, G) < 2\delta$ and since δ is arbitrary Φ would be lower-semidiscontinuous at J, contradiction. In the second case, for $J \in \mathcal{G} - \mathcal{Q}$ we would have as in the first case, for all $\delta > 0$ a $H_\delta \in \mathcal{G}$ with $d(J; H_\delta) < \delta$ and $\Phi(H_\delta) < \Phi(J) - \varepsilon_0$, ε_0 some fixed positive constant; either $H_\delta \in \mathcal{Q}$ or we could get, as before, $G \in \mathcal{Q}$ with $d(J, G) < 2\delta$ and $\Phi(G) < \Phi(J) - \varepsilon_0$; but then, by the definition of the extension of Φ, we would have $\Phi_\delta(J) \leq \Phi_{2\delta}(J) < \Phi(J) - \varepsilon_0$ which is a contradiction since δ is arbitrary.

Since the energy function $E : \mathcal{G} \longrightarrow [0, +\infty)$ is lower-semicontinuous and $\mathcal{G}$ is compact, E attains a minimum value, say m_E. Let $\mathcal{M}_E = E^{-1}(m_E)$; then $\mathcal{M}_E$ is a closed, therefore compact subspace of $\mathcal{G}$ and so the volume function restricted to $\mathcal{M}_E$ attains a minimum value, say $m_{E,V}$. We will refer to any contraction $H \in \mathcal{G}$ with $E(H) = m_E$ and $V(H) = m_{E;V}$ as having *minimal energy-volume* (in this order).

We could, of course, reverse the roles of energy and volume to get m_V, $\mathcal{M}_V$ and $m_{V;E}$, and so consider contractions with minimal volume-energy. To study the relations between these two notions of minima - namely to search for functions with minimal energy *and* minimal volume, that is to investigate if $\mathcal{M}_E \cap \mathcal{M}_V \neq \emptyset$ - would lead us astray of our main purposes, so we will not pursue the subject here. For technical reasons, in the study of coalescence in section 5.3 it turns out to be more convenient to work with m_E instead of m_V.

Remark 24 *We could have worked, all the way long, directly with cubical complexes and their maps, without ever considering the relations - via standard triangulations - with simplicial complexes, namely the analysis that led to Lemma 22; we could do well without those relations, and that Lemma, by adopting a weaker understanding of the space $\mathcal{Q}$, analogous to our previous considerations about the space $\mathcal{S}$ in Remark 21: we could consider the subdivisions $M^{[n]}$ to be not necessarily cubical - in the sense that the intersection of any two cubes wouldn't need to be a common proper face - and the maps $J : M^{[n]} \longrightarrow K^{[n]}$ not necessarily cubical, but simply "linear" in the sense that in each cube $\psi \in M^{[n]}$ they are the restriction of a cubical map on a cube ancestor to ψ (of course, since a cubical map between cubes is in no sense "linear" - as we pointed out earlier, affine coordinates were not to be used - the use of this word is just an abuse of language meant to draw the parallel with Remark 21). We would still need to prove that a cubical map $\varphi : \psi \longrightarrow \omega$ between cubes has a bounded stretching factor, c_φ: the proof we gave invoked the relation to the standard triangulation and associated simplicial map, but other more direct arguments could be given; from there, the argument to arrive at the global bound $C(\psi, \omega)$, used to settle the equicontinuity of $\mathcal{Q}$, doesn't involve any simplicial stuff.*

Beside the need for the proofs of the Cubical Approximation Theorems (the analogues of Proposition 15 and Theorems 16 and 17) - which, any way, would always, most naturally, beg the "reference" *to the simplicial setting, we would also need to rephrase, in cubical terms, the definitions of the Energy and Volume functions and the proofs of their lower-semicontinuity - which is something that could be easily done, as the reader may care to check.*

5 Shrinking complexity II: coalescence

In this section we pay attention to contractions $H : K \times I \longrightarrow K$ in the case where K is a (finite cubical) *manifold* and our main purpose is to apply the concepts and results of the previous section, the concepts of energy and volume and the existence of minimal energy-volume contractions, to the study of *coalescence*. We start by recalling some simple facts about compact contractible manifolds.

If K is a contractible n-manifold then, by some well known homological arguments (see books on algebraic topology, for instance, [3], [9] or [10]) K is *orientable* and has non-empty boundary, ∂K - which is then an orientable, closed and *connected* $(n-1)$-manifold. If K is contractible, that is if K is *homotopy equivalent* to a point, then all its *homotopy groups* and *homology groups* vanish in dimensions greater than 0: $\pi_n(K) = 0$ and $H_n(K) = 0$, $\forall n \geq 1$ - actually, since K is a cellular space, by the well known *theorems of Whitehead and Hurewicz*, the contractibility of a connected K is equivalent to his having trivial homotopy groups. In particular $\pi_1(K) = 0$, that is K is *simply-connected*, and therefore K is orientable; If K is orientable and *closed*, that is $\partial K = \emptyset$, the sum of all its oriented n-simplices defines a *fundamental homology class* that generates $H_n(K) \cong \mathbb{Z}$. That the boundary of a contractible manifold K is connected is also a consequence of what follows.

Let K be a contractible m-manifold, $B = \partial K$ its boundary and $H : K \times I \longrightarrow K$ an element of $\mathcal{C}$; consider the restriction of H to $\partial K \times I$, which we denote also by H: then $H : \partial K \times I \longrightarrow K$ defines a *contraction of* ∂K *in* K. Note that the existence of a contraction of ∂K in K doesn't imply that ∂K is itself contractible - just think of $S^n = \partial B^n$ - or even simply connected. Although in dimension three the situation is quite simple - in this case it is easy to see that ∂K is a 2-sphere, S^2, and furthermore that the contractibility of K is equivalent to ∂K being contractible in K (you can read proofs of these facts in section 2 of [6]) in dimension four there are examples of contractible manifolds with complicated boundary - see [2] for a description of the *Mazur*-manifold, a 4-dimensional contractible manifold - with an embedded dunce-hat as a spine - whose boundary is the *Poincaré homology 3-sphere* (also known as the *dodecahedron space*; see [14]).

It is easy to see that any contraction of ∂K in K, $H : \partial K \times I \longrightarrow K$ - we are not assuming now that H is the restriction of any contraction of K, and we leave aside the problem of knowing if such an extension always exists - doesn't miss any point of K, that is it must be surjective: for all $x \in K$, there is some $(y, t) \in \partial K \times I$ such that $H(y, t) = x$. Suppose, on the contrary, there is some $x \in K - H(\partial K \times I)$; since $H(\partial K \times I)$ is compact and therefore closed, there is a closed ball $D(x, \varepsilon)$ disjoint from $H(\partial K \times I)$; let L be the complement in K of the interior of $D(x, \varepsilon)$: then L is a manifold with boundary the (disjoint) union of ∂K and an $(m-1)$-sphere, $\partial D(x, \varepsilon) = S^{m-1}$, and ∂K is still contractible in L. Consider the double of L, $DL = L \vee L/\sim$, the identification space obtained from the disjoint union of two copies of L by identifying the two copies of the boundary through the identity map: there are two embedded copies of L in DL that intersect along the common boundary, $\partial L = \partial K \cup$, and clearly ∂K is still contractible in DL; consider any properly embedded arc, l, in L with one end point in ∂K and the other in S^{m-1}: the double of this arc in DL is an embedded closed arc, say Dl, that intersects ∂K, as well as S^{m-1}, transversely once; but this is a contradiction: it would mean that the $(m-1)$ homology cycle correspondent to ∂K and the 1-cycle correspondent to the arc Dl would have intersection number 1; but by well known homological arguments the intersection number depends only on the corresponding homology classes - it is given by the cap product of one with the Poincaré dual of the other - and if ∂K is contractible in DL then the cycle is nulhomologous and so has intersection 0 with any other cycle. Note that this argument generalizes straightforwardly to show that, as stated above, the boundary of a contractible manifold must be connected.

5.1 The geometric condition

Let K be a cubical complex, which is a contractible m-manifold (therefore it is a *full* complex); recall that in K we consider the *length metric* defined via the compatible transport-metrics in the cubes, which are given through embeddings of those cubes in $\mathbb{E}^m$.

We will assume that the metric satisfies the following local condition, which we will refer to, simply, as the *geometric condition* - after the usual behaviour of straight lines in classical geometries: for each point $x \in K$, there is a neighbourhood $N_g(x)$ such that any two *geodesics* in it either do not intersect or intersect only once. In particular there will be *uniqueness of geodesics*: for any two points $y, z \in N_g(x)$ there is a unique geodesic in $N_g(x)$ joining z and y. We call $N_g(x)$ a *geometric neighbourhood*.

Note that if a point $x \in K$ is in the interior of the manifold and is not in the $(m-2)$-skeleton - that is, x is in the interior of an m-cube or in the interior of an $(m-1)$-cube which is not included in ∂K - then x has a neighbourhood with a euclidian structure, that is isometric to an euclidian ball in $\mathbb{E}^m$, and so in particular verifies the geometric condition. But in points of $K^{(m-2)}$ or on the boundary ∂K, the geometric condition may fail; this can be easily seen in examples in dimensions two and three. In dimension 2, when K is a 2-disc, we may have vertices in the interior of the *hyperbolic type*; those are the points $x \in K^0$ where the sum of the angles at x of all the squares (2-cubes!) in $St(x)$ is greater than 2π; for these points the geometric condition fails, as the reader is asked to show in the next exercise:

Exercise 25 *Show that at an hyperbolic point in the interior of the 2-disc, the geometric condition fails, more specifically there are distinct geodesics that meet at the hyperbolic point and coincide, in a common segment, after that (Hint: look at a slice of $St(x)$ with angle 2π)*

But even when all the points in the interior of a 2-disc verify the geometric condition, what happens in particular when the disc is embedded in $\mathbb{E}^2$ and so is euclidian in the interior, there may be points in the boundary that fail to satisfy the condition if the disc is not *convex*, as the next figure shows.

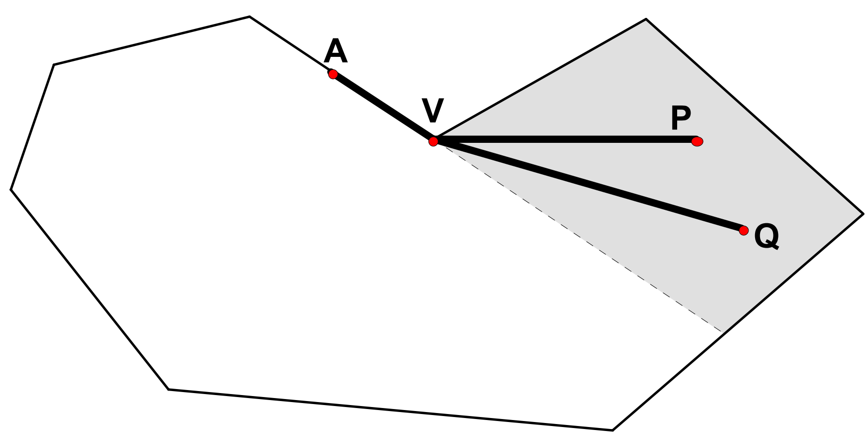

Figure 16

If you look at point A, for instance, it's clear that for all points P, Q in the interior of the shaded region the two geodesics from those points to A, represented thick, meet at vertex V and, from there, have the common segment $\overline{VA}$.

In dimension 3, when K is a *homotopy* 3-*ball* and then $\partial K \cong S^2$, as pointed out before (see section 2 of [6]), we have the analogous failure of the geometric condition at interior vertices $x \in K^0$ where the sum of the solid angles at x of all the cubes in $St(x)$ is greater than 4π - hyperbolic points - and at points in the interior of edges, $e \in K^1$ - hyperbolic edges - where the sum of the angles at e of all the 3-cubes that contain e is greater than 2π; of course we also have the analogous failure of the geometric condition at points on the boundary, due to a lack of convexity (with the straightforward 3-dimensional analogue of the example in Figure 16).

But, keeping our main objective in sight, it is possible to circumvent the failure of the geometric condition at interior points of K, broadening the scope of our previous results: without getting involved in details that would force us to detour from our main path into differential and smoothing technicalities (but see [15, Chapter 3 -and the references there in] for a starter), we will be content with sketching the general procedure; we will do that while focusing only in dimension 3. Given a hyperbolic point, $x \in K$, take a closed ball $B(x, \varepsilon)$ in the interior of $St(x)$: for each 3-cube $\sigma \in St(x)$, $B(x, \varepsilon) \cap \sigma$ is the cone on a spherical triangle with vertex x (recall that, through a specific embedding, we are considering σ as a cube in $\mathbb{E}^3$); $B(x, \varepsilon)$ is the joint $S(x, \varepsilon) * x$ consisting of all linear segments joining x to the points of the sphere $S(x, \varepsilon)$ which is made up of all the spherical triangles $S(x, \varepsilon) \cap \sigma$, $\sigma \in St(x)$. The procedure consists of *pinching* all those segments at x, as suggested - in dimension 2 - in the next figure; in broadly terms, the local metric in each sphere $S(x, \delta), \delta \in (0, \varepsilon)$ is substituted by the local metric of the sphere of radius $r(\delta)$ where r is a suitable function. When the hyperbolic vertex is an end point of an hyperbolic edge, e, this pinching is combined with the pinching that we also have to perform along such edges: these, in turn, are done by considering a banana-shape region around the edge (right side of figure 17) and pinching the circular sections of that region which are perpendicular to e - that is, in each of those circles we pinch all the radius at its centre point, which lies in e.

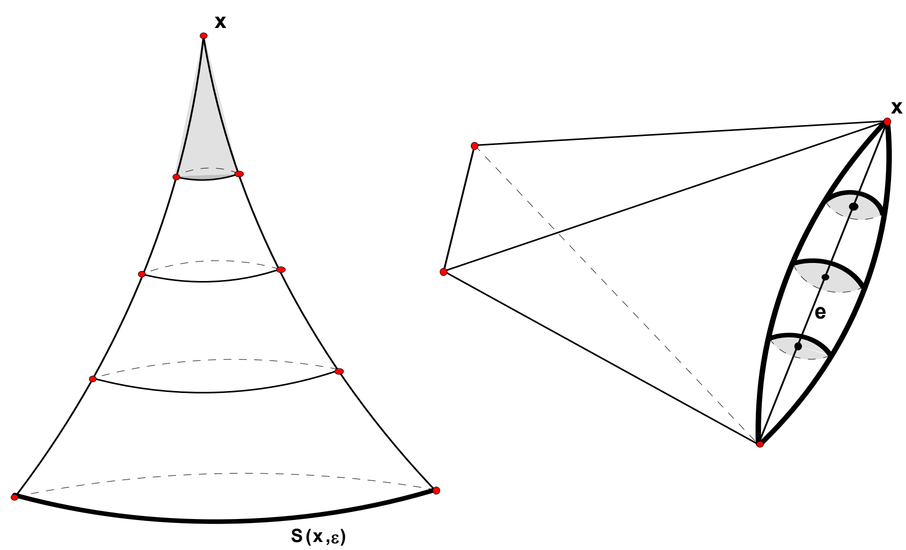

Figure 17

At the end of this process we've traded the euclidian structure in the complement of K^1, we had at the beginning, for another (differentiable) metric structure with varying curvature, and satisfying the geometric condition, and the hyperbolic vertices and edges for vertices and edges, respectively, of a cone type where the geometric condition may still fail, as the reader may care to check (in a neighbourhood of a cone point geodesics are not unique). Finally we *round up* (or smooth) those cone singularities. Of course, we are now referring to the geodesics for the new metrics on the cubes.

Remark 26 *We have to explain how our previous results - that were all formulated in terms of euclidian cubes, their baricentric subdivisions and cubical maps - may adapt to the new metric setting. Given an euclidian cube* $\psi \in K$*, let* ψ_p *denote the correspondent pinched cube; the process of pinching gives us a canonical homeomorphism* $p_\psi : \psi \longrightarrow \psi_p$ *(that is actually a diffeomorphism in the complement of the set of hyperbolic points and edges) sending* ψ^i *to* ψ_p^i*; looking at* $K \times I$*,* p_ψ *extends naturally to a canonical homeomorphism of cubes* $p_{\psi\times I} : \psi \times I \longrightarrow \psi_p \times I$*, with* $p_{\psi\times I} = p_\psi \times I$*. Now, given any two cubes* ψ_p *and* ω_p *(either of* K *or of* $K \times I$*) we define the cubical maps* $\phi : \psi_p \longrightarrow \omega_p$ *to be just all the compositions* $\psi_p \xrightarrow{p_\psi^{-1}} \psi \xrightarrow{\varphi} \omega \xrightarrow{p_\omega} \omega_p$*, where* $\varphi : \psi \longrightarrow \omega$ *is a cubical map, and we define the (baricentric) subdivisions* $\psi_p^{[k]}$ *to be the images by the canonical map* p_ψ *of the corresponding subdivisions of* ψ*.*

With the broadening of the notion of cubical complex - to include length-metrics from pinched cubes also - and of the notions of cubical map and baricentric subdivision as well, the definitions of volume and energy of a contraction, the proof of the lower-semicontinuity of the volume and energy functions, and the proof of the equicontinuity of $\mathcal{Q}$*, all adapt in a straightforward manner: in all calculations of lengths, areas and integrals all we have to do is to go back and forth between the cubes* ψ *and their counterparts* ψ_p*, via the canonical map* p_ψ*; since in some of the proofs we made essential use of the existence of* stretching-factor bounds *for cubical maps we need to know that each canonical map* $p_\psi : \psi \longrightarrow \psi_p$ *also has such a bound: that is indeed the case, since the pinching process can be carried through with bounded derivative (avoiding the emergence of cusps at the singular vertices and edges).*

While we can circumvent the absence of the geometric condition at interior points of K, the same isn't true for points in the boundary. Although the pinching process can still be carried over for hyperbolic points and hyperbolic edges that lie in ∂K, the lack of convexity can not be resolved in that way: on the contrary, the pinching at those hyperbolic points and edges may introduce non-convexity at the boundary, as the left side of figure 17 suggests - seen here as representing the (solid) pinched neighbourhood of such a point. We say that the boundary of K is *convex* - or perhaps more appropriately that K is *convex at the boundary* - if for each $x \in \partial K$ there is a neighbourhood, $N(x, \varepsilon)$, such that for all points $a, b \in N(x, \varepsilon) - \partial K$, that is for all points a, b in $N(x, \varepsilon)$ which are in the interior of K, the geodesic $\overline{ab}$ is contained in $N(x, \varepsilon) - \partial K$; we leave it as an exercise for the reader to relate this definition with the general definition of convexity of K: K is said to be *convex* if its interior $K - \partial K$ is convex, that is, for all interior points a, b, any geodesic $\overline{ab}$ is also contained in the interior of K. In dimension 3, as we've already pointed out, the boundary ∂K must be a 2-sphere: that allows us easily, to get a metric for the homotopy-ball K that makes it convex at the boundary, as we shall see shortly. But we can not hope to be able to do that in general, for $\dim(K) \geq 4$: in this case ∂K is a *homology* 3-*sphere*, a (closed) 3-manifold with the same homology as the 3-sphere, S^3 (this can be easily calculated using the exact sequences of homology and cohomology and the *Poincaré-Lefschetz duality* - see, for instance, [9, Theorem 3.26] [4]). Now, the convexity at

[4]Where it is referred to as *"...[forming] the cornerstone of the subject of geometric topology."*

the boundary points to the existence of a metric in ∂K with non-negative curvature, which in turn points to a spherical geometry and so to a finite fundamental group; but the only known homology 3-sphere which has finite fundamental group is the *dodecahedral space* - that bounds the Mazur-manifold, as already mentioned (in fact, it does have a spherical geometry: see [15], [16]); on the other hand, it is a consequence of Freedman's results that any homology 3-sphere bounds a contractible compact 4-manifold (see [17, section 9.3]).

Let's now see the specially simple situation of dimension 3. Let K be a homotopy 3-ball and $\partial K \cong S^2$ its boundary; in a standard three dimensional procedure, we can cap off K by gluing a 3-ball, B^3, to its boundary, thus getting a homotopy sphere $H^3 \cong K \cup_{\partial K} B^3$ (see [6, section 2] for a detailed description of this relation between homotopy 3-balls and homotopy 3-spheres). Consider any cubical decomposition of H^3 with compatible embeddings $h_\psi : \psi \longrightarrow \mathbb{E}^3$, one for each cube (they could all be the mapped to the standard cube $\mathbb{I}^3$). Choose any cube ψ in the decomposition: by the homogeneity of manifolds there is a self homeomorphism of H^3 that sends the embedded B^3 to ψ and so K is homeomorphic to $H^3 - \mathring{\psi}$: we can thus consider that the boundary of K has the standard cubical decomposition of S^2 as the 2-skeleton of a cube, $\dot{\psi} = \psi^2$. Look at the euclidian cube $\psi' = h_\psi(\psi) \subset \mathbb{E}^3$ and consider any "standard" cube $Q = k(\mathbb{I}^3) + a$ - obtained from $\mathbb{I}^3$ by a dilation followed by a translation - that contains ψ' in its interior; consider the closed region in between the two cubes, $C = Q - \mathring{\psi}'$; by the *PL Schonflies Theorem for* $\mathbb{E}^3$ (see [7, Chapter XIV]) C is homeomorphic to $S^2 \times I$: the two sphere components are $\dot{\psi}'$ and $\dot{Q}$; we can extend the cubical decomposition of $\partial C = \dot{\psi}' \cup \dot{Q}$ - that consists of sixteen quadrilaterals, eight for each component - to a cubical decomposition of C. We now add the *collar* C to K by the component $\dot{\psi}'$, by gluing through the homeomorphism h which is the restriction of h_ψ^{-1} to $\dot{\psi}'$. Let $L = K \cup_h C$: by standard regular neighbourhood theory, L and K are homeomorphic. L has a natural cubical decomposition, consisting of all the previous cubes of H^3 and respective compatible embeddings, except ψ that has been removed, plus all the cubes of C with their natural embeddings in $\mathbb{E}^3$: since the gluing was done with the restriction of h_ψ^{-1} we still have compatibility. C embeds in L as a neighbourhood of $\partial L \equiv \dot{Q}$ and clearly there is convexity at the boundary. Note that, since C is embedded in $\mathbb{E}^3$, the construction also implies that any hyperbolic vertex or any hyperbolic edge in L belong to the subspace K and therefore are in the interior of L: we can then treat these occurrences with the pinching process, as explained above, without disturbing the convexity at the boundary.

Note that in the argument just given, we wouldn't need to go through the capping off process if we had an embedding, $g : (\partial K)^{sud} \longrightarrow \mathbb{E}^3$, of some subdivision of ∂K into three space, that is an isometry for the intrinsic metric in $(\partial K)^{sud}$ that is, such that the restriction to each 2-cube (or simplex if we work with triangulations instead) is an isometry to a quadrilateral (or triangle) in $\mathbb{E}^3$. We would then proceed as before, with any "standard" cube Q containing the 2-sphere $g(\partial K)$ in its interior. We leave it as a problem the existence of such an embedding:

Problem 27 *Let J be a cubical (or simplicial) complex, homeomorphic to S^2, and for each cube (simplex) ψ of the decomposition let $h_\psi : \psi \longrightarrow \psi' \subset \mathbb{E}^2$ be an embedding into a quadrilateral (triangle) of 2-space, with all embeddings compatible along common edges: if $e = \psi \cap \omega$ is a common edge, then $h_\psi(e)$ and $h_\omega(e)$ are congruent segments.*

a) Is there a global embedding $g : J \longrightarrow \mathbb{E}^3$ such that, for each ψ, $g(\psi)$ is congruent to ψ'?

b) If not, is there a global embedding with the property required in a) for some cubical (simplicial) subdivision J^{sub} ?

The subject of *realizability* of (combinatorial, abstract) triangulations is a vast, rich and useful one, I have been introduced to only recently [5]: a classical result is *Steinitz's Theorem* (see [18, Chapter 4]); *a)* asks for a strong version of this theorem and is therefore false - construct a counter-example with rectangular triangles; there remains *b)* to be investigated.

5.2 Volume and energy variations

Due to the close relation between the contractibility of a compact manifold K and the contraction of its boundary it is natural to see how the definitions of volume and energy may adapt for contractions $H : \partial K \times I \longrightarrow K$, including the analysis of the lower-semicontinuity of the respective functions. This is important if we seek to analyse the contractions of ∂K in K independently of any ambient contractions of which they are restrictions. Let's denote by $\mathcal{C}_\partial$ the space of all contractions $H : \partial K \times I \longrightarrow K$, with the sup-metric, and $\mathcal{S}_\partial$, $\mathcal{Q}_\partial$ and $\mathcal{G}_\partial$ the subspaces with definitions analogous to those for $\mathcal{S}$, $\mathcal{Q}$ and $\mathcal{G}$, respectively.

For the energy there is no need for any substantial change: the definition of the length function $L_H : \partial K \longrightarrow [0, +\infty)$, the definition of the energy $E(H) = \int_{\partial K} L_H(x)$ and the proof that the energy function $E : \mathcal{S}_\partial \longrightarrow [0, +\infty)$ is lower-semicontinuous adapt *mutatis mutandi* (just with notational changes; in fact, they would adapt exactly the same way if instead of ∂K we took any other complex L)

In the case of the volume, some relevant changes are called for. First, in the definition of the function $A_H(t)$ which measures the area of the image of the slice at time t: given a contraction $H : [\partial K \times I]^{sub} \longrightarrow K^{sub}$, where $\dim(K) = m$ and writing as before $M = [\partial K \times I]^{sub}$, the slices $M_t = \partial K \times \{t\}$ are now of dimension $(m-1)$ instead of m and so in the definition $A_H(t) = \sum_{\sigma_t : \sigma \in M^m}$ we must take $A(H(\sigma_t))$ to be the $(m-1)$ (instead of the m) volume in the simplex $H(\sigma) \in K$; of course this volume is non zero only if $\dim(H(\sigma)) \geq m-1$. As before, this area function $A_H(t)$ is clearly a continuous function of t, and the definition of the volume is exactly the same: $V(H) = \int_0^1 A_H(t)dt$.

Second, and also because of this dropping of the dimension, we have to adjust one step of the proof of the lower-semicontinuity of the volume function $V : \mathcal{S}_\partial \longrightarrow [0, +\infty)$, in the argument surrounding Figure 10, namely when we claimed there were $\alpha_0 > 0$ and $\varepsilon_0 > 0$ such that for all $t' \in [t, t+\alpha_0] \cup [t-\alpha_0, t]$ and for all $J \in S$ with $d(H, J) < \varepsilon_0$, we had $J(\sigma_{t'}) \supset (\Delta - C_{t_0})$, that is the images of all the discs α_0-close to σ_t by all contractions ε_0-close to H in $\mathcal{S}$, cover the complement in Δ of the collar C_{t_0}; in fact, because we now have $\dim(\sigma) = m$, we don't have the situation where a neighbourhood $N(\Delta, \gamma)$ of the quadrilateral $\Delta = H(\sigma_t)$ is simply the union of Δ with a neighbourhood of the sphere $\Sigma = \dot{\Delta}$, as represented in Figure 10.

We have now the situation represented in Figure 18 (to avoid drawing three dimensional neighbourhoods, we've reduced the dimension of the picture by 1): the image Δ is here represented by segment $\overline{PQ}$ and the complement in Δ of the collar C_{t_0}, which we'll call Q is represented by the thick segment; shaded is a typical neighbourhood $N(\Delta, \gamma)$, for which there is a retraction $r : N(\Delta, \gamma) \longrightarrow \Delta$ such that for each point $d(x, r(x)) \leq \gamma$. Represented in darker shade is $V = r^{-1}(Q)$. We can assume, by taking γ sufficiently small, that V has a product like structure and that the retraction in V is like a projection: specifically, for each simplex $\tau \in K$ which contains Δ, $V \cap \tau$ is isometric either to the euclidian product $Q \times [-\gamma, \gamma]$ or to the euclidian product $Q \times [0, \gamma]$, and the retraction is the vertical projection into the first factor Q; the first case occurs when Δ is properly embedded in τ - the situation represented in the figure - and the second when Δ is contained in a $(m-1)$ face of τ; there are two possibilities: either

[5] I thank my Department colleagues from the Discrete Geometry and Combinatorics group, António, Leonor and Rosário, who gave me, over lunch, some tips and told me about Steinitz's theorem and the book by Ziegler.

that face is included in ∂K, in which case V reduces to $Q \times [0, \gamma]$, or is a common face of two m-simplices (K is a manifold), in which case the two products match to give $V = Q \times [-\gamma, \gamma]$.

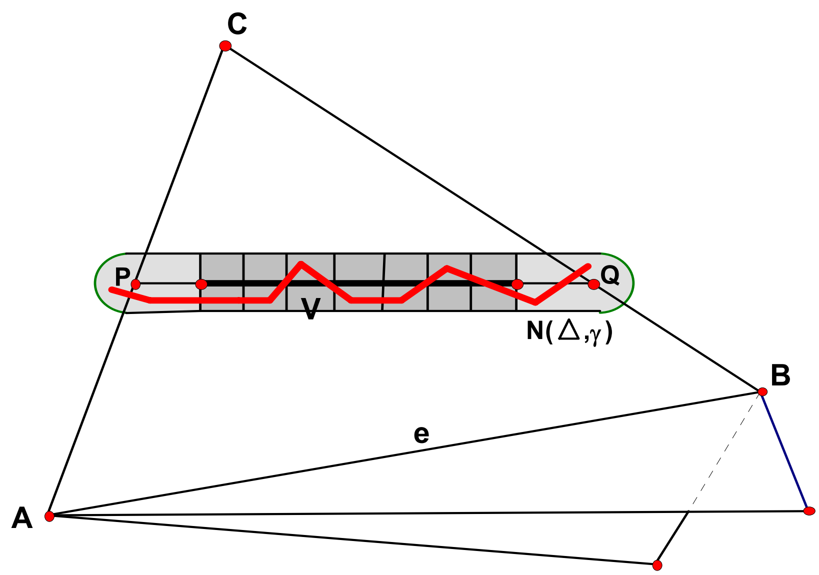

Figure 18

With this add, the proof goes on in a similar fashion: we now claim there are $\alpha_0 > 0$ and $\varepsilon_0 > 0$ such that for all $t' \in [t, t + \alpha_0] \cup [t - \alpha_0, t]$ and for all $J \in S$ with $d(H, J) < \varepsilon_0$, we have $J(\sigma_{t'}) \subset N(\Delta, \gamma)$ and $r(J(\sigma_{t'})) \supset (\Delta - C_{t_0}) = Q$ (the red broken line in the figure represents such a $J(\sigma_{t'})$); the proof of the claim is exactly like before, a clone of the argument used in the case of Bing's house to show it had an opening time greater than 0. Since $r(J(\sigma_{t'})) \supset Q$ and by the definition of V, we have that for all $x \in Q$, $J(\sigma_{t'})$ intersects the stalk of x in V, $x \times [-\gamma, \gamma]$ or $x \times [0, \gamma]$; since the vertical euclidian projection doesn't decrease area, we conclude that $A(J(\sigma_{t'})) \geq A(J(\sigma_{t'}) \cap V) \geq A(Q)$. The rest of the proof of the lower-semicontinuity of the volume function is now identical to the previous one.

Finally, the proof of the equicontinuity of $\mathcal{Q}_\partial$ is totally identical to the previous one; we can therefore consider contractions $H \in \mathcal{G}_\partial$ with minimal energy-volume.

5.2.1 Control energy

For technical reasons it will be useful to consider a variation of the notion of energy which will be called *control-energy* and that, in a certain sense, stands halfway between the two notions of energy and volume: it is still called energy because of its formulation in terms of the lengths of tracks, but it is closer in spirit to the notion of volume, in particular it will be, like the volume, sensitive to reparametrizations. Given a simplicial complex K and a simplicial contraction $H \in \mathcal{S}$, let for each $x \in K$, $l_x^t = L_H^t(x)$ be the length of the terminal part of the stalk at x, in between time t and the end: $L_H^t(x) = l(H(x \times [t, 1]))$. The arguments of section 4.2 that led to the conclusion that $L_H(x)$ is continuous, adapt in a straightforward way to show that $L_H^t(x)$ is a continuous function of both variables, $x$and t. Define now the energy of H *from time* t, $E_t(H)$ by

$$E_t(H) = \int_K L_H^t(x)$$

This is the analogue of the area function $A_H(t)$ in the definition of volume; it is easy to prove, using the continuity of $L^t_H(x)$ that $E_t(H)$ is a continuous function of t. We obtain the control-energy of H, $E_C(H)$, like we did for the volume, by integrating over I:

$$E_C(H) = \int_0^1 E_t(H)dt$$

The control-energy function $E_C : \mathcal{S} \longrightarrow [0, +\infty)$ is also *lower-semicontinuous*. Since the proof is essentially a combination, with minor variations, of the previous proofs for the volume and for the energy functions, it is enough to sketch it. Take an $\varepsilon > 0$; recall that in the case of the length function we claimed that for all x' sufficiently close to x and for all J sufficiently close to H, $L_J(x') > L_H(x) - \varepsilon$; the argument used in the proof of that claim adapts immediately to a proof of the stronger claim: for all x' sufficiently close to x, for all J sufficiently close to H *and for all* t' *sufficiently close to* t, $L^{t'}_J(x') > L^t_H(x) - \varepsilon$. Using this property, the continuity of $L^t_H(x)$ and the compactness of K it is easy to show, in analogy to what we set for the area function $A_H(t)$, that given an $\varepsilon > 0$, for all J sufficiently close to H and for all t' sufficiently close to t, $E_{t'}(J) > E_t(H) - \varepsilon$. Finally we use this last property, the continuity of $E_t(H)$ and the compactness of I to conclude that E_C is lower-semicontinuous, in exactly the same way as we did for the volume function.

Note that the control-energy function $E_C(H)$ is sensitive to reparametrizations: it clear from the definition that if we speed up the contraction H, we get contractions with control-energy as close to 0 as we wish; on the other end if we slow down H we increase the control energy and get contractions with control-energy as close to $E(H)$ as we wish. Restricting to the space $\mathcal{Q}_H \subset \mathcal{S}_H$ is a way of keeping these possible variations of speed within limits. Let m_C be the minimum for the control-energy function in $\mathcal{G}$; as before we can consider functions with minimal energy-(control-energy), or more simply *minimal energy-control*, $m_{E,C}$: these will be specially useful in analysing coalescence.

Recall that the heuristic meaning of the notions of energy and volume is that of measuring the waste due, respectively, to wandering and to folding. The heuristic meaning of the control-energy is the following: a contraction with minimal energy-control $m_{E,C}$ is a function that amongst the functions with minimal energy, and within the bounds in speed imposed by the restriction to the space $\mathcal{G}$, has the points travelling along their tracks as fast as possible.

Note that, as we mentioned already in the case of the energy and volume functions, it is a subject for inquiry the possible relations between the three different minima, m_E, m_C and m_V, namely if, and how, the three sets $E^{-1}(m_E)$, $E_C^{-1}(m_C)$ and $V^{-1}(m_V)$ may intersect.

5.3 Coalescence

In this subsection we look at the subject of coalescence for contractible manifolds. We start by some simple observations relating the existence of a coalescent contraction of K, $H : K \times I \longrightarrow K$, to the existence of a coalescent contraction of ∂K in K, $G : \partial K \times I \longrightarrow K$. Of course the existence of such an H implies the existence of G: just that the restriction of H to $\partial K \times I$. Let's see the implication in the reverse direction. Suppose then that we have a coalescent contraction $G : \partial K \times I \longrightarrow K$; as we recalled in the beginning of the section, G doesn't miss any point of K: for each $x \in K$, there is some point $(y, t) \in \partial K \times I$ such that $G((y, t) = x$; let for each x, let $t_x = \min \left\{ t \in I : \exists y \in \partial K, (y, t) \in G^{-1}(x) \right\}$, that is t_x is the time at which x is reached by the contracting boundary - we call it the *reach-time*. That t_x exists is clear: $G^{-1}(x)$ is closed (and as we just said, non-empty) therefore is compact; its projection into the I factor is a compact subspace of I and therefore has a minimum. We can extend G to a coalescent function $H : K \times I \longrightarrow K$ in a natural way: define $H(x,t)$ by $H(x,t) = x$ for $t \leq t_x$ by $H(x,t) = x$, and for $t \geq t_x$ by $H(x,t) = G(y,t)$ where y is any point of ∂K such that $G(y,t_x) = x$: since G is coalescent, it doesn't matter which y we take, and obviously H is also coalescent. Actually, by definition of coalescence, a coalescent extension of G has to be defined in this way. But it may be discontinuous as an easy example shows.

Example 28 *Consider the next figure: it illustrates a contraction, G, of the boundary of a circle, we next describe. The centre of the circle is labelled O. During the first $1/4$, the arc $\widehat{AB}$ is isotoped through the sector OAB, keeping the end points fixed, until it is sent, at exactly $t = 1/4$, homeomorphically into the union of the two radius $\overline{OA}$ and $\overline{OB}$; therefore for all points in those radius and interior to the circle, the reach-time is $1/4$ (for the end points it is of course 0); next, in the interval $[1/4, 1/2]$, the homotopy takes place inside the image of S^1 at time $1/4$: the radius $\overline{OA}$ and the arc $\widehat{AD}$ are kept fixed; the radius $\overline{OB}$ is shrunk inside itself to the point O, while the arc $\widehat{BD}$ is pulled along and stretched until is it is sent, at time $t = 1/2$, homeomorphically into its union with the radius $\overline{OB}$, with $\widehat{BC}$ sent to $\overline{OB}$ and $\widehat{CD}$ to $\widehat{BD}$.*

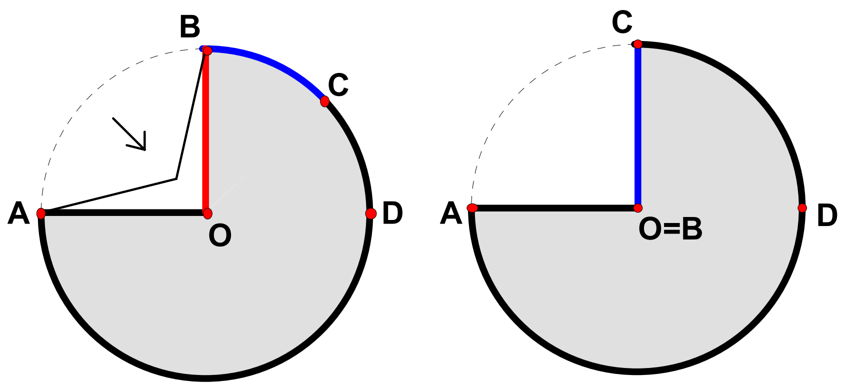

Figure 19

In the third quarter we do nothing, and finally in the final quarter, we homotope to the centre radially. It should be clear that G is discontinuous at the points $(x, 1/2)$ where x is in the interior of radius $\overline{OB}$: they are all sent to the point O while all the points in the interior of the shaded region, in particular those arbitrarily close to $\overline{OB}$, are fixed at least until time $t = 3/4$.

As suggested by the previous example, the reason why the coalescent extension H of G may fail to be continuous is that the function $x \longrightarrow t_x$ that assigns to each x its reach-time, may be discontinuous. Assume that the *reach-time function* is continuous, then H is continuous. Let $(x_0, t_0) \in K \times I$ be an arbitrary point and suppose that H is discontinuous at (x_0, t_0). Let (x_n, t_n) be a sequence converging to (x_0, t_0) such that $H(x_n, t_n)$ doesn't converge to $a_0 = H(x_0, t_0)$. Because K is compact, we can assume, without loss of generality, that $H(x_n, t_n) \longrightarrow b_0 \neq a_0$. Let, for each $n \geq 1$, t_{x_n} be the reach-time of x_n and $y_n \in \partial K$ a *reach-point* for x_n, that is such that.$G(y_n, t_{x_n}) = x_n$. Since we are assuming that the reach-time function is continuous and $x_n \longrightarrow x_0$, we have that $t_{x_n} \longrightarrow t_{x_0}$ where t_{x_0} is the reach-time of x_0. Because $a_0 \neq b_0$, we can assume, passing to a subsequence, if necessary, that for all $n \geq 1$, $t_n \geq t_{x_n}$ and, as a consequence, that $t_0 = \lim_i t_{n(i)} \geq \lim_i t_{x_{n(i)}} = t_{x_0}$ as well: if not, there would be a subsequence $t_{n(i)} \leq t_{x_{n(i)}}, i \geq 1$ and so $t_0 = \lim_i t_{n(i)} \leq \lim_i t_{x_{n(i)}} = t_{x_0}$, therefore $a_0 = H(x_0, t_0) = x_0$; but then, for all $i \geq 1$, $H(x_{n(i)}, t_{n(i)}) = x_{n(i)} \longrightarrow b_0 = x_0$ and so $x_0 = b_0 = a_0$. We can further assume, once again passing to a subsequence, that $(y_n, t_n) \longrightarrow (y_0, t_0)$: because G is continuous, and $(y_n, t_{x_n}) \longrightarrow (y_0, t_{x_0})$ we have that $G(y_0, t_{x_0}) = \lim_n G(y_n, t_{x_n}) = \lim_n x_n = x_0$; therefore y_0 is a reach-point for x_0. Since for all $n \geq 0, t_n \geq t_{x_n}$, we have that $H(x_n, t_n) = G(y_n, t_n)$; but G is continuous and $(y_n, t_n) \longrightarrow (y_0, t_0)$, so we finally have

$$b_0 = \lim_n H(x_n, t_n) = \lim_n G(y_n, t_n) = G(y_0, t_0) = H(x_0, t_0) = a_0$$

which is a contradiction. We have proved the following theorem:

Theorem 29 . *If the* reach-time function $x \longrightarrow t_x$ *is continuous, then G is the restriction of a coalescent contraction of K.*

Note that the proof of the theorem gives us also the following corollary:

Corollary 30 *Let K be a compact manifold with boundary ∂K, and let $G : \partial K \times I \longrightarrow K$ be a contraction for which the reach-time function is discontinuous at $x_0 \in \partial K$. Then there is a sequence $x_n \longrightarrow x_0$ such that for the corresponding sequence of reach-times we have $t_{x_n} \longrightarrow t_0 > t_{x_0}$.*

From now on, we will be considering the spaces of contractions of ∂K in K that we considered above, namely $\mathcal{C}_\partial$, $\mathcal{S}_\partial$, $\mathcal{Q}_\partial$ and $\mathcal{G}_\partial$ and the existence of *minima* for the energy, volume and control-energy functions, m_E, m_C, m_V, as well as for some of their combinations, like $m_{E,C}$; of course, we should note that the discussion of the control-energy above, applies unscathed to those spaces exactly as was the case for the energy function (only in the case of the volume function some adjustments were needed)

We start by considering the heuristic notion of *folding*, that relates especially to the notion of volume. It may be useful, although not strictly necessary, in order to invoke some homological notions and arguments, to recall that to give a contraction $G : \partial K \times I \longrightarrow K$ is equivalent to give a map $\hat{G} : C(\partial K) \longrightarrow K$, from the *cone* $C(\partial K) = \partial K \times I/\partial K \times \{1\}$ to K, which is the inclusion in ∂K - here we are naturally identifying ∂K to $p(\partial K \times \{0\})$, p the identification map - in fact, through *transgression*, G factors as $\hat{G} \circ p$; we leave the details for the reader. This is especially relevant for the visualization in dimension 3 where $\partial K \cong S^2$ and therefore $C(S^2) \cong B^3$.

Let $G \in \mathcal{C}_\partial$, $G : \partial K \times I \longrightarrow K$ and $M = \partial K \times I$; the boundary of M is the union of two disjoint copies of ∂K: $M_0 = \partial K \times \{0\}$ and $M_1 = \partial K \times \{1\}$. Assume that $\dim(K) = \dim(M) = m$. Let $\hat{M}$ denote the cone $C(\partial K) = M/M_1$: as above, we identify $\partial K \equiv M_0$ to its natural embedding in $\hat{M}$. Let C be an $(m-1)$ subcomplex in some subdivision, M^{sud}, and let $\hat{C}$ be its image in $\hat{M}$. Assume that $\hat{C}$ is either a *cycle*, that is, for the homology group $H_{m-1}(\hat{M})$, or a *relative cycle*, for the relative homology group $H_{m-1}(\hat{M}; \partial K)$; using ∂ for the usual *boundary operator* in homology, we have that this is equivalent to C being a relative cycle for the relative homology $H_{m-1}(M; \partial K)$, that is, ∂C is an $(m-2)$ cycle in ∂M (the union of two cycles, one in each of the components M_0, M_1): $\hat{C}$ is a cycle exactly when $\partial C \subset M_1$. Figure 20 illustrates the situation: falsely, because the two components of the boundary, M_0, M_1, which are closed manifolds are represented as intervals; the reader is advised to translate it into dimension 3, where M_0, M_1 are 2-spheres. A good way of doing it is to visualize cycles in a cylinder $B^2 \times I \subset S^2 \times I$, where $B^2 \subset S^2$ is an embedded 2-disc, by spinning the previous figure around the vertical central line: the broken thick lines generate, by revolution, complicated 2-cycles; the analogue of Q' in

3-dimension will be a tube-like object generated by one of the lines from top to bottom. C and C' represent a cycle and, respectively, a relative cycle in M such that $\hat{C}$ and $\hat{C}'$ are both cycles in $\hat{M}$; Q and Q' in turn are sent to relative cycles in $\hat{M}$. Represented shaded are the chains or relative chains that these cycles bound.

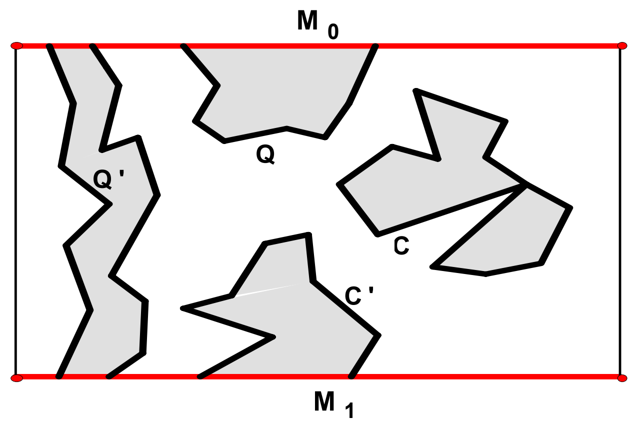

Figure 20

Using the fact that $\hat{M}$ is contractible, being a cone, and that ∂K has the homology of an $(m-1)$-sphere, some standard arguments and routine calculations with homology groups gives us that $\hat{C}$ is nulhomologous (in homology or relative homology) and therefore it separates the space $\hat{M}$, and its boundary cycle, $\partial\hat{C}$, separates ∂K; analogously C separates M. A particular special type of cycles we will consider are the ones corresponding to embeddings of ∂K at different levels, $\partial K \times \{t\}$. Unless we say something to the contrary, we will use simply the word *cycle* to mean a cycle or a relative cycle in codimension 1.

We say that the contraction $G : \partial K \times I \longrightarrow K$ *has a folding*, or simply that *folds*, if there is some cycle C and points x and y in the interior of different (connected) components of $M - C$, such that $G(x) = G(y) \notin G(C)$: we say that the folding is *relative to (or over)* C; we also say that x *and* y *fold over* C. Let's see how this definition relates to the more intuitive notion of folding for simplicial maps. We need some further facts from the simplicial category, about general position. Let $G \in \mathcal{S}_\partial$ be a simplicial contraction of ∂K in K, $\dim(K) = m$; in general we say that a simplicial map $g : L \longrightarrow K$ between two manifolds, such that $\dim(L) \leq \dim(K)$, is in *general position* if g embeds each top-dimensional simplex $\sigma \in L$. Now, given any simplicial map $f : L \longrightarrow K$ and $\varepsilon > 0$, it is possible to change f into a map $g : L^{sub} \longrightarrow K^{sub}$ which is in general position, through a homotopy H that satisfies: for each $x \in L$, and for each $t \in I$, $d(x, H_t(x)) < \varepsilon$ - in particular we will have $d(f, g) < \varepsilon$, for the supreme metric; we can further assume that, for the length metric in K, the track of each point has length less than ε; if f has *maximum stretching factor* C - that is for all paths p, $l(f(p)) \leq Cl(p)$ - by choosing ε sufficiently small relative to $mesh(L)$, we will have that g has *maximum stretching factor* less than $C + \delta$ for any prescribed $\delta > 0$. The proof of this may be modelled in the proofs of Theorems 15 and 16, using the fact that K is a manifold: we leave it as an exercise. In our case, a general position map $g : M^{sub} = [\partial K \times I]^{sub} \longrightarrow K^{sub}$ will not be a contraction, since $g(M_1)$ will be an $(m-1)$ subcomplex and not a point; we consider a *relative* definition instead, by demanding that all top dimensional simplices are embedded by g, except those in $St(M_1)$, which we can see as a collar on M_1: as before, we can arbitrarily approximate any given simplicial contraction by a general position one; since, by taking sufficiently fine subdivisions, we can consider that $g(St(M_1)) \subset St(p)$ is contained in the interior of a neighbourhood, $N(p, \delta)$, homeomorphic to an m-ball, B^m, the restriction to this relative version of general position approximation is not really weakening: once inside $N(p, \delta) \cong B^m$, the mysteries of any contraction are fewer and less. We summarize this in the following:

Lemma 31 *Given any contraction $G \in \mathcal{S}_\partial$ and constant $\varepsilon > 0$, there is a contraction $g \in \mathcal{S}_\partial$ which is in* (relative) general position *and such that G and g are homotopic, through a homotopy with each track having length less than ε, and so $d(G, g) < \varepsilon$ in the supreme metric. Furthermore, we can assume that g has* maximum stretching factor *less than an arbitrary $C + \delta$, $\delta > 0$, where C is the maximum stretching factor for G: we can therefore take $V(g)$, $E(g)$ and $C(g)$ as close to, respectively, $V(G)$, $E(G)$ and $C(G)$ as we want.*

Let C be the maximum stretching factor for the maps in $\mathcal{Q}_\partial$. If we look at those maps as cubically-framed simplicial maps and allowed for an enlargement of the space by including the (relative) general position simplicial approximations given by the Lemma, we would still have a space with maximum stretching factor say $2C$, which therefore is equicontinuous and, consequently, has compact closure. Note that we could also aim at improving the previous lemma, by establishing that for any contraction $G \in \mathcal{Q}_\partial$ a general position approximation could be found within the space $\mathcal{Q}_\partial$, but that wouldn't bring any special benefits, apart from elegance, since we don't really need to enlarge the spaces $\mathcal{Q}_\partial$ and $\mathcal{G}_\partial$. All we need to know is that given any (limit) map $H \in \mathcal{G}_\partial$ we can arbitrarily approximate it, as well as its volume, energy or control-energy, by a (relative) general-position deformation of a cubical map of $\mathcal{Q}_\partial$; we look at these general-position deformations of the maps in $\mathcal{Q}_\partial$ just as auxiliary tools to unveil the folding in a simple way; given a $G \in \mathcal{Q}_\partial$ that approximates $H \in \mathcal{G}_\partial$ within $\varepsilon > 0$ (in either of the four qualities, distance, volume energy or control energy) we call a general-position deformation of G which is an ε-approximation of G - and so is an 2ε-approximation of H - an *ε-shadow* of G and denote it by G^ε_g.

We now turn to the definition of *folding set*. Let $J : [\partial K \times I]^{sub} \longrightarrow K^{sub}$ be a general-position simplicial map, $m = \dim(K)$ - as we've just said, we will be thinking of J as a ε-shadow, G^ε_g, of some $G \in \mathcal{Q}_\partial$, but the definition is valid more generally, for arbitrary general-position maps between manifolds of the same dimension. The folding set is the $(m-1)$-dimensional subcomplex of the domain, Fol_J (or F_J for short) consisting of those $(m-1)$-simplices, σ, such that each point in the interior of σ has a neighbourhood which is embedded by J; equivalently: each $(m-1)$-simplex, σ, is a face of exactly two m-simplices which, by the general position assumption, are sent by J into two m-simplices of K^{sub} with common face $J(\sigma)$; σ is a *folding simplex* exactly when those two m-simplices of K^{sub} are the same.

The folding set Fol_J is a *cycle*: we just have to see we can collect the simplices of Fol_J in pairs $\{\sigma, \tau\}$ such that each pair intersect along a common $(m-2)$ face and that the process exhausts all the $(m-2)$ faces of the simplices in Fol_J; questions of orientability, that is, whether the chosen pairing gives us $\mathbb{Z}$-cycles or $\mathbb{Z}_2$-cycles don't need to concern us here, although an orientation can always be given, due to the trivial homology of $C(\partial K) = M/M_1$. We can easily visualize the argument in dimension 3, but it is valid in general: let α be an *edge* of some 2-simplex $\sigma \in Fol_J$ and consider the set N (see figure 21 bellow) which is the union of all the 3-simplices that have α as edge: those 3-simplices wrap around α in cyclic order when α is not on the boundary; Let s be the circle in the interior of N consisting of the segments that join the baricentre of each

2-simplex that contains α to the baricentre of the next 2-simplex in the cyclic order. When α is on the boundary - note that α must be in M_0 - instead of a circle we have an arc that joins the baricentres of the two 2-simplices of $N \cap M_0$ that contain α.

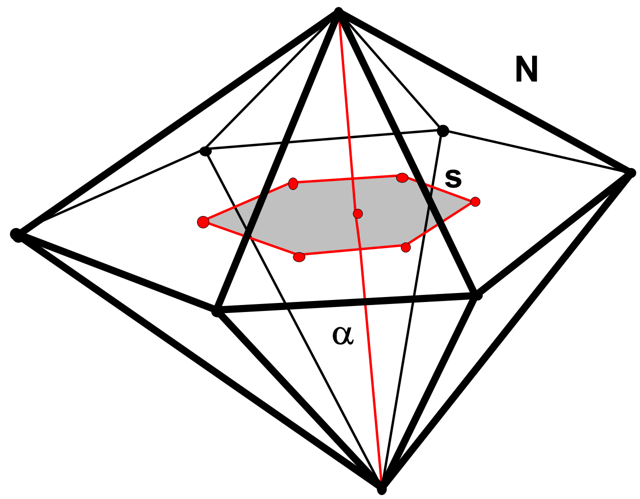

Figure 21

Since J is in general position, N and s are sent, respectively, to a corresponding set N' of 3-simplices that surround the edge $J(\alpha)$ and a corresponding circle $s' \subset N'$; we thus have a map between the two circles s and s': if we consider the circles oriented, to each folding along a 2-simplex of $N \cap Fol_J$ corresponds a change in the orientation of this map; obviously the number of these changes of orientation must be even and we can therefore group the simplices in $N \cap Fol_J$ in pairs as we wanted. The case when α is on the boundary is perfectly analogous: since J is the identity on M_0 we have a map between two oriented arcs which fixes the end points and, just as well, the number of orientation changes must be even. Note that the argument shows that if an edge α is not in Fol_J, then there are no changes of orientation for the map between the two circles s and s' and therefore it must be of the form $z \longrightarrow z^n$, $z \in S^1 \subset \mathbb{C}$, that is, it is an n-covering: this means that on N, J is a branched cover of $N' = J(N)$, of degree n, branched over the edge $J(\alpha)$. It is not difficult to construct examples of contractions in the 3-cube, $K = \mathbb{I}^3$, with edges having this kind of behaviour, with $n > 1$- a good starting point is to do the analogue in dimension 2, where instead of an edge we consider vertices - but, as will shall see bellow, the existence of this kind of edges implies that $Fol_J \neq \emptyset$.

Let's now see the relation between the definition of a *folding contraction* and the definition of *folding-set* for general-position maps. Let $H : \partial K \times I \longrightarrow K$ be a contraction that has a folding relative to the cycle C and let x and y be points in the interior of different (connected) components of $M - C$, such that $H(x) = H(y) \notin H(C)$. Consider the distance, $a > 0$, between $H(x) = H(y)$ and $H(C)$; let $\varepsilon < a/8$: we claim that for any ε-shadow, $J = G_g^{\varepsilon}$, we have $Fol_J \neq \emptyset$. Suppose, on the contrary that $Fol_J = \emptyset$; we argue again in dimension 3 and, for simplicity, we work with the induced maps $\hat{J}, \hat{H} : C(\partial K) \longrightarrow K$ (in dimension three $C(\partial K)$ is a 3-ball) and the corresponding cycle $\hat{C}$: then, since J is in general position, we can consider that $\hat{J}$ is in general position and that $Fol_{\hat{J}} = \emptyset$ as well. Then, $\hat{J}$ sends the set of edges of $C(\partial K)$ to the set of edges of K and so, by the definition of general position, its restriction to the complement of that 1-skeleton is a *covering map* between $C(\partial K) - C(\partial K)^1$ and $K - K^1$; but since the restriction to $\partial K \equiv M_0$ is the identity, the number of sheets of that covering must be one and so it is a homeomorphism: but, as we've seen above, when there are no foldings, we have that for each edge α, J restricted to the set N of all 3-simplices of which α is an edge is a branched cover of N' branched over the edge $J(\alpha)$; the degree n of that cover must therefore be also one and so is a homeomorphism. So we conclude that $\hat{J}$ must indeed be a homeomorphism. But then, $\hat{J}(x)$ and $\hat{J}(y)$ must be in different connected components of $\hat{J}(\hat{C})$ and therefore their distance (which is measured by the length of the shortest path joining them, and which must intersect $\hat{J}(\hat{C})$) is greater than the distance of each one of them to $\hat{J}(\hat{C})$; since J is an ε-shadow, we have that $d(\hat{J}(x), \hat{J}(y)) \leq d(\hat{J}(x), \hat{H}(x)) + d(\hat{J}(y), \hat{H}(y)) < 4\varepsilon < a/2$; on the other hand, we have that $d(\hat{J}(\hat{C}), \hat{H}(\hat{C})) \leq 2\varepsilon$ and therefore $d(\hat{J}(x), \hat{J}(\hat{C})) \geq d(\hat{H}(x), \hat{H}(\hat{C})) - 4\varepsilon = a - 4\varepsilon > a - a/2 = a/2$, which is a contradiction. We state the conclusion as a theorem:

Theorem 32 *Let $H \in \mathcal{C}_\partial$ be a contraction that folds over the cycle C. Then, there is some $\varepsilon > 0$ such that for each $\delta \leq \varepsilon$, each δ-shadow, $J = G_g^\delta$, has non-empty folding set, $Fol_J \neq \emptyset$.*

We need a different proof of this theorem, not as simple as the previous proof but one from which we will be able to draw a fundamental corollary.

We start by noting that if H folds over C, we can assume, without loss of generality, that C is a cycle relative to the boundary M_0. Let x and y be two points in the interior of different connect components of $M - C$ such that $H(x) = H(y) \notin H(C)$. Take one of those components, say the one containing x, O_x: it is easy to see that the closure of O_x has for boundary a cycle, so we can substitute C by that cycle. If this new C is a relative cycle we're done; if not, join a point in the interior of a simplex of C to a point in M_0 by an arc α in $M - \overline{O_x}$ that avoids y and such that $H(\alpha)$ doesn't contain $H(x) = H(y)$; then drill two discs around the end points of α, one in C and the other in M_0, and add a *pipe* around α sufficiently thin (built in a sufficiently fine subdivision of M) so that its image by H still misses $H(x) = H(y)$. We thus get the desired relative cycle C that separates M in two components, one of which contains x and the other contains y - actually, we can further assume that C is a properly embedded manifold, by trading it for any of the boundary components of a sufficiently shallow regular neighbourhood, $N(C)$, but that is not particularly useful. Denote the connected components of $M - C$ by O_x and O_y, and their closures by B_x and B_y, respectively. The homology boundary of C, $C \cap M_0$, separates M_0 and $C \cup (M_0 \cap B_x)$, $C \cup (M_0 \cap B_y)$ are (non-relative) cycles that bound B_x and B_y, respectively. There is in general no reason why the image by H of a separating cycle should still separate: it may be squeezed; but since C is a relative cycle and H is the identity on M_0, $H(C)$ is still a separating cycle - maybe now in several components and not just two.

Let us draw the reader's attention to a detail that, although not necessary for the proof we want to give, helps in developing the feeling for the folding phenomena. Let A be the connected component of $M - H(C)$ that contains $H(x) = H(y) \notin H(C)$; we claim that either $A \subset H(B_x)$ or $A \subset H(B_y)$. Suppose that both $H(B_x)$ and $H(B_y)$ miss points in A: let $p \in A - H(B_x)$ and $q \in A - H(B_y)$. Up to an homotopy of H - which doesn't change the homological status of the various cycles - we can assume that $p = q$, that is $H(B_x)$ and $H(B_y)$ miss the same point of A: consider a closed ball $\mathcal{B}$ contained in A, with p and q in its interior, and an isotopy h of M taking q to p and fixed on $\overline{M - \mathcal{B}}$; consider now the homotopy of the map H which is fixed in B_x and is $h_t \circ H, t \in [0, 1]$ in B_y: then, at the end of this homotopy, both $H(B_x)$ and $H(B_y)$ miss the point p, and since the isotopy was fixed in $\overline{M - \mathcal{B}}$ and so in particular in ∂K, we would get a contraction of ∂K missing a point, which we already know can not happen. Let $I_x = \{x' \in O_x : \exists y' \in O_y, H(x') = H(y')\}$ and I_y similarly defined; note that, since H is the identity on M_0, I_x and I_y are strict subsets of O_x and O_y, respectively; of course they are also non-empty for $x \in I_x$ and $y \in I_y$. By the previous discussion, if $x' \in I_x$, $y' \in I_y$ and

$H(x') = H(y')$, then either x' is in the interior of I_x or y' is in the interior of I_y: if A is the connected component containing $p = H(x') = H(y')$, and $B(p,\varepsilon)$ is an open ball contained in A, $B(p,\varepsilon)$ is contained in at least one of the sets $H(B_x)$ or $H(B_y)$, say $B(p,\varepsilon) \subset H(B_x)$: then, by continuity of H, there is an open neighbourhood of y' in O_y, $B(y',\delta)$, such that $H(B(y',\delta)) \subset B(p,\varepsilon) \subset H(B_x)$; but then, by definition of O_y, $B(y',\delta) \subset I_y$ and so y' is an interior point.

Note that given two points x, y that fold over C, there are many other cycles over which they may fold: let $z = H(x) = H(y)$ and $a = d(z, C)$; by compactness and continuity, there is an $\varepsilon > 0$ such that $H(N(C,\varepsilon)) \subset N(H(C), a/2)$; then each cycle $C' \subset N(C,\delta)$ where $\delta = \min\{\varepsilon, d(x,C), d(y,C)\}$ will be a cycle over which x and y fold. The idea is to maximize the distance $a = d(z, C)$ over those cycles: we will have to take limits of cycles and so consider, more generally, cycles in singular homology. We consider the consecutive baricentric subdivisions, $M^{[n]}$, of M and sequences of cycles which are subcomplexes of these - we can either work with cubical decompositions and their $(m-1)$-faces (if $\dim(K) = m$) or work with triangulations and restrict to those which are cubically-framed. Let C_n be a sequence of (relative) cycles; recall that a subset L is the limit of C_n if for each $\varepsilon > 0$, there is $N(\varepsilon) \in \mathbb{N}$ such that $n \geq N(\varepsilon) \Rightarrow C_n \subset N(C,\varepsilon)$. Let's recall the fact that any given sequence of cycles, $(C_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Let for each $n \geq 1$, $D_n = \bigcup_{i \geq n} C_n$ and $F_n = \overline{D_n}$; clearly, $D_{n+1} \subset D_n$, $F_{n+1} \subset F_n$ and therefore, by Cantor's characterization of compactness, $F = \bigcap F_n \neq \emptyset$, and this is obviously compact. We also have $F = \bigcap_{\varepsilon > 0} N(F,\varepsilon)$, so for any $\varepsilon > 0$, $\exists N(\varepsilon) \in \mathbb{N} : n \geq N(\varepsilon) \Rightarrow D_n \subset F_n \subset N(F,\varepsilon)$, that is C_n converges to F. But if we want a connected limit then we may have to pass to subsequences: consider any connected component L of F; then, for any $\varepsilon > 0$, $N(L,\varepsilon)$ must contain an infinite number of C_n - otherwise it wouldn't contain D_n for n greater than a certain order and so would not contain the intersection of the F_n's, contradiction - and so is the limit of a subsequence. Of course L (or F) still separate any pair of points x, y that are separated by all the C_n: if not we could run a path from x to y in the complement of L; by compactness that path would also miss a neighbourhood $N(L,\varepsilon)$ and so all but a finite number of the C_n's, which is a contradiction.Let $\mathcal{S}$ be the set of cycles - subcomplexes of some $M^{[n]}$ - over which x and y fold and for each $C \in \mathcal{S}$ let $a_C = d(z, H(C))$ where $z = H(x) = H(y)$; we call the number $Fd_{(x,y)} = \sup\{a_C : C \in \mathcal{S}\}$ the *folding distance* of the pair $\{x, y\}$. This number is attained when we consider the extension of $\mathcal{S}$ to the space $\mathcal{F}$ given by all the limits of sequences in $\mathcal{S}$. Let a_{C_n} be an increasing sequence converging to $Fd_{(x,y)}$; by passing to a subsequence if necessary assume that $L = \lim_n C_n$: it is an easy exercise, using the definition of $\lim_n C_n$ and the uniform continuity of H, to see that $d(z, H(L)) = Fd_{(x,y)}$. Let $w \in L$ be any point that minimises the distance, that is, such that $d(z, H(w)) = d(z, H(L)) = Fd_{(x,y)}$. We claim that the point w *behaves like a folding point*, in the following *heuristic* sense: if we consider a small neighbourhood of w, and the two sides in which it is separated by L, they must be send to the same side of $H(L)$ which contains z; otherwise we could distort slightly the cycle L near w so that $H(L)$ would become further apart from z: on the other hand, if the two parts are sent to the same side and we distort L, then $H(L)$ becomes closer to z. Let's make this idea more precise. Consider in $H^{-1}H(w)$ the union, P, of the connected components that intersect L - refer to Figure 22-a where it is represented, very schematic, as the shaded disc. We claim that some point p in the boundary of $P \cup L$ must be a limiting point for the folding sets of the ε-shadows, that is, for every neighbourhood $N(p,\delta)$, $\delta > 0$, there is some $\varepsilon > 0$ such that for every ε-shadow, $J = G_g^{\varepsilon}$, $Fol_J \cap N(p,\delta) \neq \emptyset$. Suppose not: consider a point p in the boundary of P and a open neighbourhood $N(p,\delta) = N_p$ (represented as a box in the previous figure) such that there a sequence of ε-shadows, $J_n = G_g^{\varepsilon_n}$, $\varepsilon_n \longrightarrow 0$, none of which has its folding set, Fol_{J_n}, intersecting N_p.

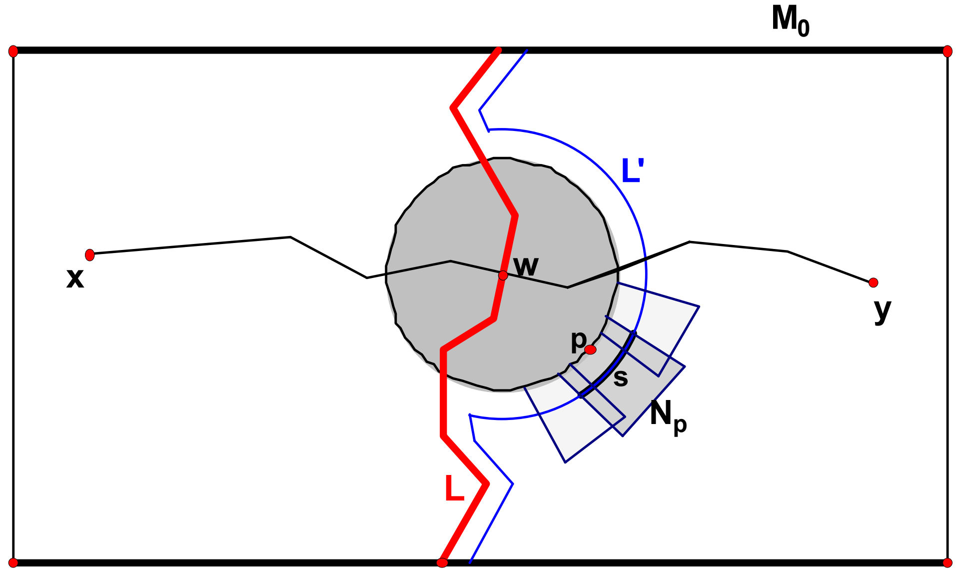

Figure 22-a

By the argument at the end of the previous proof, we know that, for each n, $J_n(N(p))$ must be an open set containing $H(p) = H(w)$: in fact as we saw in that proof, if q is a point which is not in Fol_J and if q not a vertex, then either it is in the complement of the 1-skeleton in which case it has a neighbourhood which is mapped homeomorphically by J, or it is in the interior of an edge in which case it has a neighbourhood N which is sent to to a neighbourhood N' of $J(q)$ through a ramified cover, ramified over a segment. In the case where q is a vertex not in Fol_J, it is easy to see, using the previous condition for the edges that contain q, that $J(St(q))$ also has $J(q)$ has an interior point. Of course that $J_n(N_p)$ is separated by $H(L)$: we will work with the part that doesn't contain z. Cover $P \cup L$ by a finite number of these neighbourhoods, N_p and consider some σ-small regular neighbourhood M of P contained in the union of the N_p's, and the intersection of its manifold boundary, L', with the N_p's - represented by s in the figure which is assumed not to intersect $H^{-1}H(w)$, that is in connected components other than P: we leave it as an exercise to work out the details, why we can consider such an s; note that in each N_p there must not be any other points from different components of $H^{-1}H(w)$: recall that $J_n(N_p)$ ramifies over a segment, say l', image of a segment l: look at the transverse sections to this segment: they are sent to transverse sections of l', preserving the linear order of their intersections with the segments; if one of those sections contains some point of $H^{-1}H(w)$, since $H = \lim_n J_n$ all sections bellow that one will have to contain points of $H^{-1}H(w)$, forming a path to p; we could also argue in a more indirect way: when we have a general position map, say one of the J_n's, and a connected open set, A, extending to the boundary M_0 and disjoint from the folding set, Fol_{J_n}, since it is the identity in M_0, its restriction to A must be an embedding, that is, there is no real ramification in A. We don't necessarily have the same J_n's for all the N_p's, but starting with the ones next to the boundary we have that the J_n are embeddings in those, that is, H is approximated by homeomorphisms there; by a *continuation* argument we can extend to all the other N_p's to conclude that all the J_n must have ramifying number one as well, and therefore are embeddings. Since $J_n(s)$ is contained in the part of $H(L)$ that doesn't contain z and $H = \lim J_n$, the same happens with $H(s)$; furthermore, $d(w, H(s)) > 0$, otherwise

$s \cap P \neq \emptyset$, contradicting our construction. We can now change L, by substituting $L \cap M$ by $L \cup s$, thus getting a new cycle L'that still separates x and y, such that $d(H(L'), H(w)) > 0$ and such that around $H(w)$, $H(L')$ is further apart from z than $H(w)$ - see Figure 22-b.

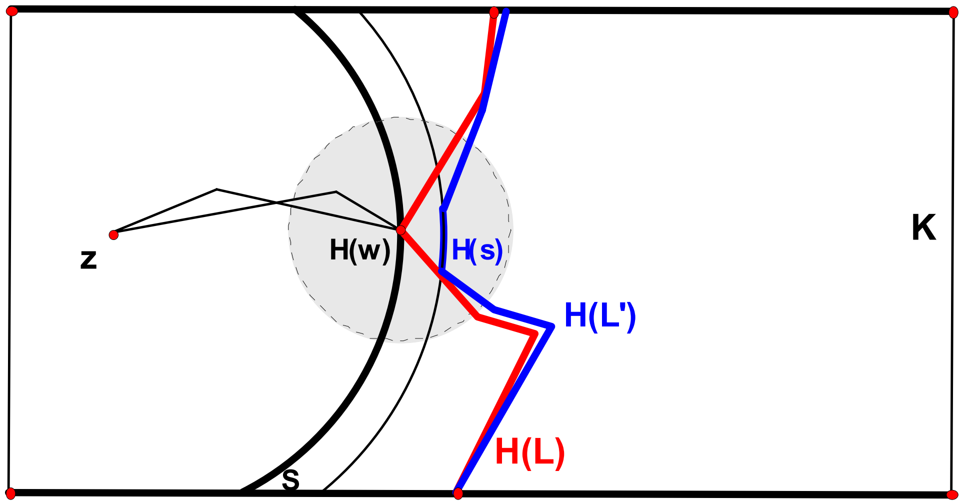

Figure 22-b

If we repeat the process for the other points $w' \in L$ that might minimize the distance, that is such that $d(H(w'), z) = Fd_{(x,y)}$ - and by compactness it is enough to do it a finite number of times - we would get a new cycle L' such that $d(H(L'), z) > Fd_{(x,y)}$ which is a contradiction.

In conclusion, for some point $w \in L$ for which, $d(z, H(w)) = Fd_{(x,y)}$, there is some point p in the boundary of $L \cup P$, where P is the set of the connected components of $H^{-1}H(w)$ that intersect L, which is a limiting point of the folding sets. We define the folding set of H, $Foll_H$, as the set of all these points, that is, the *limit of the folding sets* of all the ε-shadows: denoting by $\mathcal{S}_H^\varepsilon$ the set of all ε-shadows of H,

$$x \in Foll_H \Leftrightarrow \forall \delta > 0, \exists \varepsilon > 0, [J \in \mathcal{S}_H^\varepsilon \Rightarrow Fol_J \cap B(x, \delta) \neq \emptyset]$$

The previous discussion of an alternative proof for the last theorem was aimed at establishing that $Foll_H$ is non-empty - it is possible that more simple proofs exist of that fact. It is clear that $Foll_H$ is closed and therefore compact; furthermore, the previous argument can be easily adapted to show that there must be an *infinite* number of limiting points of the folding sets (not necessarily on $L \cup P$) : if not, we would also have only a finite number of those limiting points on $L \cup P$ and could still construct the cycle L', going around some small neighbourhoods of those points. In fact more is true: a continuation argument, like the one above, shows that no point in $Foll_H$ can be *isolated*: suppose that p is a point which has a closed ball- neighbourhood, $N_p = D(p, \delta)$, such that $D(p, \delta) \cap Foll_H = \{p\}$; given an arbitrary $\varepsilon > 0$, consider another small ball $B(p, \sigma) \subset D(p, \delta)$ such that $H(D(p, \sigma))$ has diameter smaller than $\varepsilon/2$; since p is the only point of $Foll_H$ in $D(p, \delta)$, we can cover the region $D(p, \delta) - B(p, \sigma)$ by a finite number of neighbourhoods N_{q_i}, of points $q_1, ...q_n$, for each one of which there there is an $\varepsilon/2$-shadow J_i with no folding set in it; looking at the successive level spheres $S(p, \rho)$, for $\delta \geq \rho \geq \delta$, we have by a continuation argument that $H(S(p, \rho))$ is ε-approximated by an n ramified cover and so we can $\varepsilon/2$ approximate H on $D(p, \delta) - B(p, \sigma)$ by an n ramified cover J. We can then extend J to a ramified cover defined on the whole of $D(p, \delta)$, and since $\delta(H(D(p, \sigma)) < \varepsilon/2$, J is an ε-approximation of H with no folding set in N_p. A similar argument shows, more generally, that no connected component of $Foll_H$ is 0-dimensional: supposing the contrary, consider $D(p,\delta)$ and $B(p,\sigma)$ as above and such that $D(p,\delta) \cap Foll_H$ and $B(p,\sigma) \cap Foll_H$ are a *clopen* sets in $Foll_H$.

Corollary 33 *Let $H \in \mathcal{C}_\partial$ be a contraction that* folds over *C. Then, the folding set of H, $Foll_H = \lim\{Fol_J, J \in \mathcal{S}_H^\varepsilon\}$, is a non-empty compact subset of M, with no 0-dimensional connected components.*

We should note that although the folding sets of ε-shadows of H are cycles, which are two dimensional complexes and separate M, the folding set of H may not separate M, indeed it can even be one dimensional, as the next example shows.

Example 34 *Refer to Figure 23-a, bellow:this is to help us describe a map in 2-dimensions, from the square to itself, that has easy to visualize analogues in three dimensions. This map, H, will be the limit of a sequence of maps in general position, and will have for folding set, $Foll_H$, the segment s, pictured on the left side; its image, $H(s)$, will be the segment s' on the right side.*

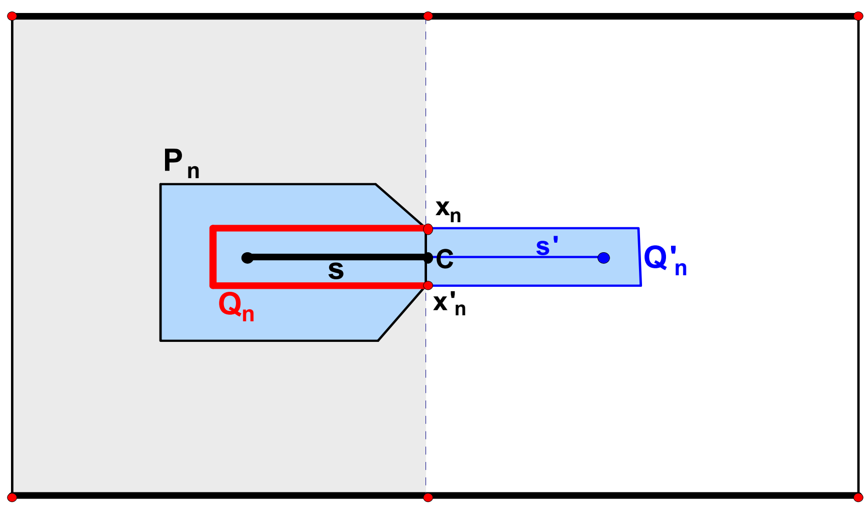

Figure 23-a

The points x_n, x'_n, on the vertical (dashed) segment that divides the square in two halves, are equidistant from the endpoint C of s, and represent a sequence of pairs of points converging to C; as they approach C, the rectangular region Q_n approaches s: so Q_n represents a sequence of chained rectangles whose intersection is precisely the segment s. The region labelled P_n, includes Q_n and represents a sequence of hexagons where all the vertices but x_n and x'_n remain fixed: in the limit intersection it gives a pentagon, say P, with the four fixed vertices and C as the fifth vertex. For each n, let J_n be a general position map defined in the following way: J_n is the identity on the right side of the square; it sends the quadrilateral Q_n to Q'_n, its mirror reflection on the line $x_n x'_n$, thus folding over the segment $\overline{x_n x'_n}$; the region between P_n and Q_n, more precisely the closure $\overline{P_n - Q_n}$ which is homeomorphic to a disc is also sent homeomorphically to Q'_n, and so the black part of its boundary - the arc consisting of the five sides of the hexagon P_n exterior to Q_n - is sent homeomorphically to the segment $\overline{x_n x'_n}$; finally the region on the left side of the square which is exterior to P_n is sent homeomorphically to the whole of the left side, keeping the five sides not in P_n fixed - and sending the others to $\overline{x_n x'_n}$, of course, matching what was already defined for P_n. Clearly, Fol_{J_n} is the boundary of Q_n - the three red sides plus *the segment $\overline{x_n x'_n}$; if $H = \lim_n J_n$, H is the identity on the right side, squeezes the boundary of the pentagon $P = \lim_n P_n$ into the point C, and sends its its interior to the segment s', homeomorphically in s; clearly $Foll_H$ reduces to the segment s.*

There are direct analogues of this in dimension 3 (and above), with parallelepipeds Q_n converging to a segment s, as well as obvious variations: for instance, parallelepipeds Q_n converging to a square S, thus getting a 2-dimensional $Foll_H$; or more complicated situations with parallelepipeds Q_n converging to a set with mixed local dimensions (for instance, a combination of some 3-cubes, squares and segments)

It should be noted that the local dimensions of the image of the folding set, $H(Foll_H)$, may vary and decrease in relation to the corresponding local dimensions of $Foll_H$, by the effect of collapsing parts. We refer to Figure 23-b to describe an example that illustrates this:

Example 35 *The figure is similar to the previous one, but this time the points x, x' and the regions Q and P remain unaltered. Let J_0 be constructed as in the previous example, by sending Q and $\overline{P - Q}$ homeomorphically to Q', the mirror reflection of Q in the segment $\overline{x, x'}$; for each n, let J_n be defined as follows: J_n agrees with J_0 outside the region P, and in P is obtained from J_0 by composing with an embedding of Q' in itself, obtained by lowering the middle thirds of the top and bottom sides - as suggested by the segments $\overline{ba}$ and $\overline{a_n b_n}$ in the picture - and keeping the vertical sides fixed. In the limit, those middle thirds converge to the middle third of the central segment (the bisector of the lateral sides) and for $H = \lim_n J_n$, $H(P) = H(Q)$ is the union of that segment with the two triangles, shown in thick blue lines. As before, Fol_{J_n} consists of the four sides of Q and since this is fixed with n, $Foll_H$ is that same set.*

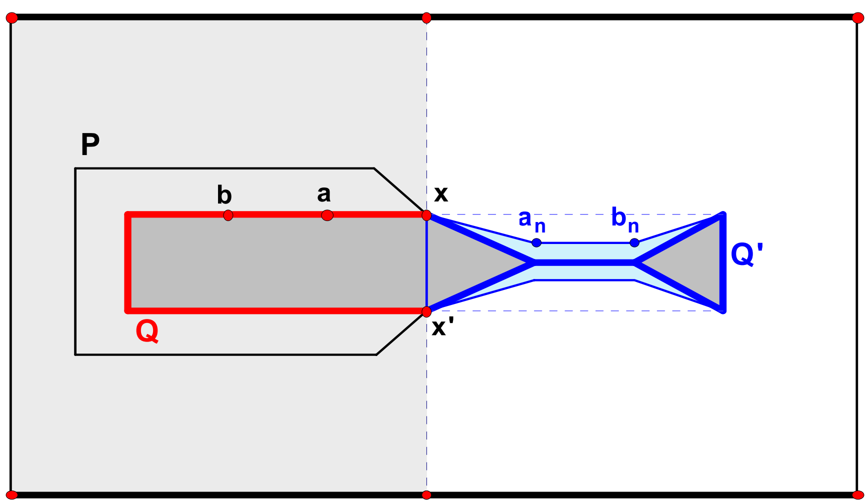

Figure 23-b

As before, this example may be varied in dimension 3 and above.

5.3.1 Under minimal energy-control

We consider now the problem of coalescence for a contraction $H \in \mathcal{G}_\partial$ under the assumption that it has *minimal energy-control*, $m_{E,C}$.

Since energy and control both relate to the lengths of the tracks (for the control function is in fact a mixed relation with length and time) we start by describing a move that, under certain circumstances, allows us to decrease the energy; although, as we shall see, this is a forbidden move in our settings, we think its description will contribute to a better understanding of the spaces of contractions, $\mathcal{G}_\partial$, and of other permitted moves we shall be using. We assume the manifold K verifies the *geometric condition*: recall this implies, in particular, that each point has a neighbourhood - which is said to be *geometric* - where geodesics are unique. Let $(x, t) \in M = \partial K \times I$, $p = H(x, t)$ and $N_g(p)$ a geometric neighbourhood; let $N = B(x, r+\delta) \times [t-\varepsilon, t+\varepsilon]$ be a neighbourhood of (x, t) in M, such that $H(N) \subset N_g(p)$. For each $y \in B(x, 2\varepsilon)$ let $Y_0 = H(y, t-\varepsilon)$ and $Y_1 = H(y, t+\varepsilon)$; so the track of y during time $t \in [t-\varepsilon, t+\varepsilon]$ is a path in $N_g(p)$, $\alpha_y : [t-\varepsilon, t+\varepsilon] \longrightarrow N_g(p)$, from point Y_0 to point Y_1: it may not be the (unique) geodesic in $N_g(p)$ joining the two points; denote that geodesic, parametrized by $[t-\varepsilon, t+\varepsilon]$, by γ_y. The idea for the "forbidden" move is to homotope the path α_y into γ_y, relatively to the end points Y_0 and Y_1, and do it along geodesic arcs: for each $t \in [t-\varepsilon, t+\varepsilon]$ let $g_t(s)$, $s \in [0, 1]$, be the unique geodesic from $\alpha_y(t)$ to $\gamma_y(t)$, and define the homotopy by $G_y(t, s) = g_t(s)$; from uniqueness of geodesics we get the continuity of G_y: denoting by $\gamma_{a,b}(s)$, $s \in [0, 1]$, the unique geodesic in $N_g(p)$ joining point a to point b, and given sequences $a_n \longrightarrow a$, $b_n \longrightarrow b$ and $s_n \longrightarrow s$, we have that $\gamma_{a_n,b_n}(s_n) \longrightarrow \gamma_{a,b}(s)$. We can now change H, just in the interior of N, so as to change all the paths α_y into the geodesic arcs γ_y, for all $y \in B(x, r)$; we use the sphere-band $S(x, r+u)$, $0 \leq u \leq \delta$, to execute the transition continuously: for each $y \in B(x, r+\delta)$ let $k(y) = \min\{0, d(x, y) - r\}$; then $k(y)$ is constant 0 in $B(x, r)$ and increases from 0 to δ in $B(x, 2r) - B(x, r)$; define the new H in N by $H(y, t) = G_y(t, 1 - k/\delta)$; H remains the same outside N. This type of move can be used to decrease the energy: suppose that for x, α_x is different, and therefore longer than γ_x: let $c = l(\alpha_x) - l(\gamma_x) > 0$ be the difference in length between the two paths; by uniform continuity and the uniqueness of geodesics, we can assume that for a sufficiently small $r > 0$, $l(\alpha_y) - l(\gamma_y) > c/2$ for all $y \in B(x, 2r)$. Suppose we modify H as before in $N = B(x, r+\delta) \times [t-\varepsilon, t+\varepsilon]$ for some $\delta < r$: for each point $y \in B(x, r)$ the length of its track decrease at least $c/2$; for the points y in $B(x, r+\delta) - B(x, r)$ we don't know if the lengths of the tracks decreased, since in the previous homotopies $G_y(t, s)$ we haven't asserted that the lengths of the successive paths, $\rho_s(t) = G_y(t, s)$, decrease as s increases; nevertheless it's not difficult to prove that they must have an upper bound, say m - this we leave as an exercise for the reader; finally the points y outside $B(x, r+\delta)$ have their tracks unchanged. Therefore the energy $E(H)$ has a decrease of at least $c/2 \times Vol(B(x, r))$ due to the change in $B(x, r)$ and doesn't increase more then $m \times Vol(B(x, r+\delta) - B(x, r))$ by the change in $B(x, r+\delta) - B(x, r)$; since $Vol(B(x, r+\delta) - B(x, r))$ converges to 0 as $\delta \longrightarrow 0$, we can effectively decrease $E(H)$.

Note that we worked directly with H and the lengths of its tracks, something which, in strict sense, we shouldn't do because the energy, $E(H)$, for $H \in \mathcal{C}_\partial, \mathcal{G}_\partial$ is defined, via the simplicial (cubical) approximations, through a limiting process; but it would be easy to adapt the previous argument to the true formal setting, working with suitable simplicial approximations of H and with paths sufficiently close to the geodesics $\gamma_y(t)$ (there is no reason why such a geodesic should be a simplicial path, so we can only approximate it)

Exercise 36 *Prove the previous assertion about the existence of an upper-bound for the lengths of the s-parameter family of paths $\rho_s(t) = G_y(t, s)$.*

The reason why the move we've just described is a *bootleg* move in our settings is that the deformation of the paths $\alpha_y(t)$ into the geodesics $\gamma_y(t)$ may be done at the expense of stretching in some directions beyond our permitted limit, given by the existence of a *maximal stretching factor* in the space $\mathcal{Q}_\partial$ that implied equicontinuity and the compactness of $\mathcal{G}_\partial$. This kind of bootleg stretching may be easily seen while revisiting *Exercise 20*. We revert to the space S

in that exercise, but for simplicity of drawing we substitute the square by a circle and the two pyramids by two cones - see Figure 24. Exercise 20 consisted in showing that there is no contraction of S with minimal energy; analogously, there is no contraction of the boundary circle $J = \partial S \cong S^1$ in S with minimal energy: the geometric argument is exactly the same.

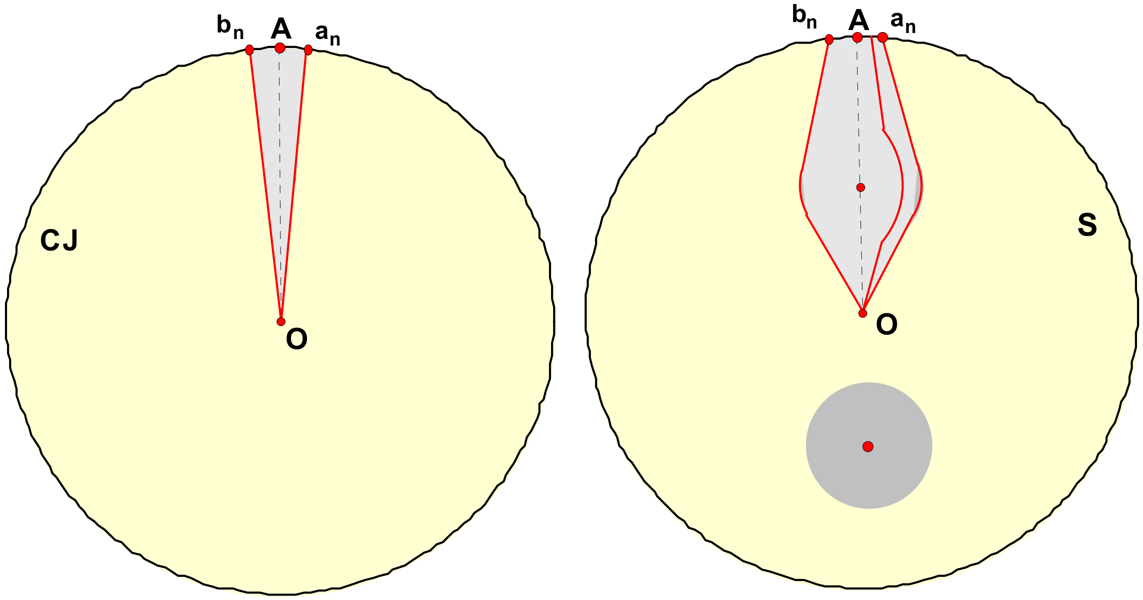

Figure 24

We identify a contraction of J in S with an extension of the inclusion map $J \hookrightarrow S$ to a map on the cone $CJ \cong D^2$ - represented on the left of the picture. The space S is represented on the right, viewed from the top: the two cones are the shaded circles. Consider only contractions of J in S, $H : CJ \longrightarrow S$, that end at the central point $O \in S$ - that is $H(O) = O$ - leaving aside the problem of how the choice of the end point $H(O)$ may vary the lower bounds - and therefore the infimum - for the energy function (we are not solving Exercise 20!). The obvious idea to get a contraction with minimal energy is to construe it in a way that, for each point $x \in J$, the radius $\overline{xO}$ in CJ is sent to a geodesic in S joining x to O; the problem is that, in S, we don't have uniqueness of such geodesics: it is clear, by symmetry, that there are two points of J, the point A and its symmetric relative to O, say A' (which is not represented) for which there are two geodesics joining them to O; those two geodesics from A to O will be symmetric in relation to the dashed line AO. All the other points $x \in J$ have a unique geodesic, say g_x to O: for most of them that geodesic is just like the radius $\overline{xO}$; for the points in the shadows of the two cones, when lit from point O - these shadows are two arcs symmetric relatively to the points A and A' - the geodesic bends around the cone, as illustrate by the three red lines; if we mark two points, a_n, b_n, symmetric relatively to A and moving along the shadow-arc in the direction of A, their geodesics will change continuously and symmetric relatively to the line AO, from the two radius tangent to the base of the cone into the unions of two segments in the base disc with an arc in the cone, until a maximum height is reached for the two geodesics for A. We can consider contractions with ever smaller energy corresponding to sequences of points $a_n, b_n \longrightarrow A$ and $a'_n, b'_n \longrightarrow A'$: for each n, define H_n in the following way: for the points x outside the arcs $\widehat{a_n A b_n}$ and $\widehat{a'_n A' b'_n}$ the radius $\overline{xO}$ is sent to the unique geodesic g_x using parametrizations by length; for the points x in those arcs we send the two discs made up of the unions of the radii $\overline{xO}$ - one of them is represented as the shaded circular sector on the left of the picture - homeomorphically and symmetrically relative to the lines AO, $A'O$ into the discs in S bounded by the arc and the geodesics g_{a_n}, g_{b_n} - represented on the right by the shaded region around one of the cones - in particular, for all n, the radius $\overline{AO}$ is sent into the line AO that goes all the way up through the summit of the cone. It is clear that $E(H_n)$ tends to the infimum $m_E = \int_J l(g_x)dx$, where $l(g_x)$ is the length of the geodesic.

Note that for each n the circular sector is stretched so as to fit the disc bound by the geodesics g_{a_n}, g_{b_n} and cover the top of the cone: therefore, ever smaller regions like the dark shaded one are stretched without bound in the direction transverse to AO; so we don't have a *maximum stretching factor* for this set of contractions, H_n. What happens with the tracks of the contractions H_n near the top of the cones, explains why the previous move is forbidden: assume we smooth the cones at the vertices to have the geometric condition and look at a neighbourhood N_V of the top point, V: in Figure 25 bellow, the cone is represented by the larger shaded disc and N_V by the smaller, darker one.

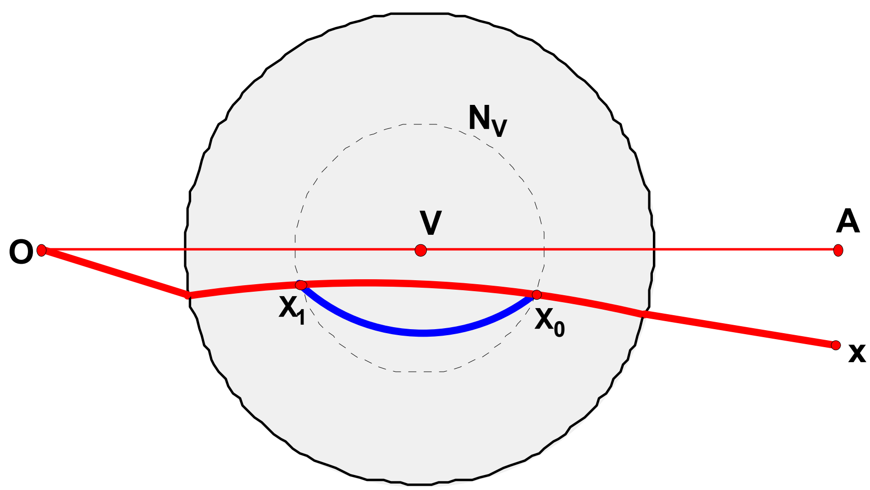

Figure 25

We can assume, by symmetry, that the track of point A, along the line AO is a geodesic in N_V; but for other points x near A, the part of the track in N_V, between points X_0 and X_1, runs higher than the geodesic joining those two points, which is represented by the blue arc: the bootleg move corresponds to the stretching in the direction transverse to AO.

There are special situations where the previous bootleg move becomes legal because it involves only the stretching of things that were previously shrunk: this is a basic heuristic that will be used several times in the sequel. We may see how that works, in a consequence of assuming *minimal energy-control* that we examine next. Let $H \in \mathcal{G}_\partial$ have minimal energy-control, $m_{E,C}$. We claim there are no stoppages all along the contraction, that is, no point ever stops-and-goes while tracing its track: for every $x \in \partial K$, if there is an interval $[a,b] \subset [0,1]$, $a < b$, such that $H_t(x) = H_a(x)$ for all $t \in [a,b]$ then $b = 1$. Note we are not claiming that a point doesn't stop before time $t = 1$: but if it does stop it is for good, at the end point of the contraction. Assume the contrary: for some point $x_0 \in \partial K$, there is an interval $[a,b] \subset [0,1)$ such that $H_t(x_0) = H_a(x_0) = y_0$, $\forall t \in [a,b]$, and such that $H(x_0 \times (b,1]) \neq \{y_0\}$, that is, there is $t > b$ with $H_t(x_0) \neq y_0$. Without loss of generality we can assume that $b = \sup\{t : H(x_0 \times [a,t]) = y_0\}$: so, given an arbitrary $\varepsilon > 0$ we can pick $t(\varepsilon)$, $b < t(\varepsilon) < b + \varepsilon$, such that $H_{t(\varepsilon)}(x_0) \neq y_0$. Consider a closed neighbourhood $N = D(x_0, \delta) \times [a - \delta, t(\varepsilon) + \delta]$ of the stalk segment $x_0 \times [a, t(\varepsilon) + \delta]$, where $D(x_0, \delta)$ is an $(m-1)$-ball and therefore N is homeomorphic to an m-ball, $m = \dim(K)$; construct a self-homeomorphism h of $M = \partial K \times I$ which is the identity on the complement of N and in N is defined as follows: it is stalk preserving; for each $x \in D(x_0, \delta/2)$ the point (x, b) is sent to the point $(x, t(\varepsilon))$ and accordingly the stalk-segments $x \times [a, b]$ and $x \times [b, t(\varepsilon) + \delta]$ are sent linearly to $x \times [a, t(\varepsilon)]$ and $x \times [t(\varepsilon), t(\varepsilon) + \delta]$, respectively, the first one being stretched in the process, and the second one shrunk (in the segments $x \times [a - \delta, a]$ nothing is done); in the points $x \in D(x_0, \delta) - D(x_0, \delta/2)$ we do a similar stretch and shrunk of $x \times [a, b]$ and $x \times [b, t(\varepsilon) + \delta]$, respectively, but varying the point image of (x, b) between $(x, t(\varepsilon))$ and (x, b) linearly as a function of $d(x, x_0) - \delta/2$. See Figure 26:

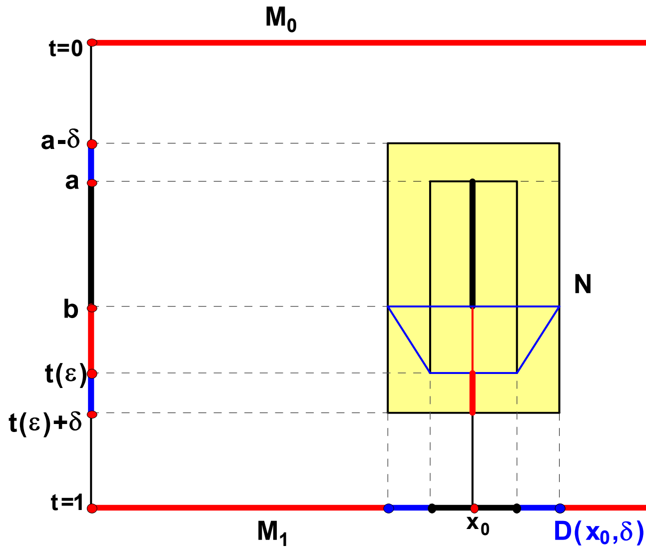

Figure 26

in N, every vertical segment $x \times [a, t(\varepsilon) + \delta]$ is sent homeomorphically into itself, stretching the segment inside the darker shaded region $D(x_0, \delta) \times [a, b]$, into the longer segment inside the lighter shaded region; actually the homeomorphism h so construed is the identity outside $D(x_0, \delta) \times [a, t(\varepsilon) + \delta]$. Let $J = H \circ h$ and let $z_0 = H(x_0, t(\varepsilon))$; by the choice of $t(\varepsilon)$ we have $d(y_0, z_0) > 0$; by uniform continuity of H, we can take δ such that for all $x \in D(x_0, \delta)$ we have $d(H(x, b), H(x_0, b)) = d(H(x, b), y_0) < d(y_0, z_0)/4$ and $d(H(x, t(\varepsilon)), H(x_0, t(\varepsilon))) = d(H(x, t(\varepsilon)), z_0)$, and so $d(H(x, b), H(x, t(\varepsilon)) \geq C = d(y_0, z_0)/2$. Because we are dealing with a length metric, this means that for all $x \in D(x_0, \delta)$ the length of its track in the time interval $[b, t(\varepsilon)]$ is at least C. Recall, from subsection 5.2.1, that for each $x \in \partial K$, $l_x^t = L_H^t(x)$ represents the length of the terminal part of the stalk at x, in between time t and the end, $L_H^t(x) = l(H(x \times [t, 1]))$, and that the energy of H *from time* t, $E_t(H)$ is $E_t(H) = \int_K L_H^t(x)$; clearly, since the homeomorphism h sends each stalk homeomorphically into it self through an increasing function of t (which is actually the identity for all $x \notin D(x_0, \delta)$), we have that $L_J^t(x) \leq L_H^t(x)$ for all t; more specifically, we have that for all $x \in D(x_0, \delta/2)$, $L_J^b(x) < L_H^b(x) - C$ and therefore for the energies from time b we have that $E_b(J) < E_b(H) - C \times Vol(D(x_0, \delta/2))$; since $E_t(J) \leq E_t(H), \forall t$, the two functions are continuous and for $t = b$ $E_b(J)$ is strictly less than $E_b(H)$, integrating over $[0, 1]$ we get $E_C(J) < E_C(H) = m_{E,C}$ (of course, the energy didn't change because the tracks remained the same for H and J: $E(H) = E(J)$). To see that this contradicts the minimallity of $m_{E,C}$, proving our claim, we have to see that $J \in \mathcal{G}_\partial$, which means it can still be approximated by functions in $\mathcal{Q}_\partial$; this is where the contrast with the bootleg move lies: heuristically, since the segment $x \times [a,b]$ is shrunk by H to the point y_0, and h only stretches $x \times [a,b]$ and the segments parallel to it and within δ-distance, by a maximum of $t(\varepsilon) - b$, by choosing $t(\varepsilon)$ sufficiently close to b and δ sufficiently small we have that the new stretching, due to composing with h doesn't surpass the stretching factor bound that exists in $\mathcal{Q}_\partial, \mathcal{G}_\partial$. We can state the following

Theorem 37 *Let $H \in \mathcal{G}_\partial$ have minimal energy-control, $m_{E,C}$. Then for every $x \in \partial K$ and every interval $[a,b] \subset [0,1]$, $a < b$, such that $H_t(x) = H_a(x)$, $\forall t \in [a,b]$ we must have $b = 1$.*

Folding under $m_{E,C}$

The essential step to which we now proceed is the analysis of the folding properties of H, under the assumption of minimal energy, $m_{E,C}$. To simplify the discussion and spare us some technical details, we may assume, without loss of generality, the following condition on the spaces of contractions we've been dealing with:

Condition 38 (Easy Boundary/End) *From now on we will assume that all contractions $H : \partial K \times I \longrightarrow K$ are the identity on some fixed collar $\partial H \times [0,\alpha]$ and are such that $H^{-1}(H(\partial H \times [0,\alpha])) = (\partial H \times [0,\alpha])$. This means that the mess in the contraction starts away from the boundary; this is easily achieved by adding collars to both $\partial K \times I$ and K and reparametrizing all the old contractions.*

Analogously, when $\partial K \cong S^{m-1}$, $m = \dim(K)$, we may also assume that H has an easy end *made up of embeddings, that is, for some $\beta < 1$, H is an embedding of $\partial K \times [\beta, 1)$ into $D(C, 1-\beta) - \{C\}$ where C is the end point of the contraction; equivalently, if we consider H as a map on the cone $C(\partial K)$, that there are closed balls $D(V)$, V the vertex of the cone, and $D(C)$ such that H sends $D(V)$ homeomorphically to $D(C)$ (we are not claiming that $H^{-1}(H(D(C))) = D(C)$).*

Clearly these subspaces of $\mathcal{C}_\partial$, $\mathcal{Q}_\partial$ and $\mathcal{G}_\partial$ are closed and so we have the minima m_V, m_E, m_C and their possible combinations just as before.

With this condition we have that the folding sets, either of ε-shadows or of the limit contraction H, all lie in $\partial K \times [\alpha, \beta]$. Given an ε-shadow of H, $J = G_g^\varepsilon \in S_H^\varepsilon$, we know that its folding set, $Fol_J \subset \partial K \times [\alpha, \beta]$, separates $M = \partial K \times I$; let IF_J be the *internal part of* Fol_J, the union of the closures of all the components of $M - Fol_J$ that do not contain M_0 nor M_1 (these may or may not be separated by Fol_J). Since $IF_J \subset \partial K \times [\alpha, \beta]$, if the stalk of a point x, $S_x = \{x\} \times I$, enters IF_J, then it must leave it again; it is important to understand the basic behaviour of the track of such a point, in relation to the folding of the map J and the image of the internal part, $J(IF_J)$. Let $S_x \cap IF_J \neq \emptyset$ and $t_0 = \min\{t : (x,t) \in IF_J\}$, $t_1 = \max\{t : (x,t) \in IF_J\}$; then $\alpha \leq t_0 \leq t_1 \leq \beta$, where α and β are the constants in the easy boundary/end condition (note that it may happen that $t_0 = t_1$: that's the case when $S_x \cap IF_J$ is a single point). Because J is a branched cover in the complement of Fol_J and the segments $x \times [0, t_0)$ and $x \times (t_1, 1]$ are in $M - IF_J$, the *external part of* Fol_J where J is a local homeomorphism (because J is the identity in M_0 and has an *easy end*) they are embedded by J. This means that while tracing its track, point x enters $J(IF_J)$ along an embedded path $J(x \times [0, t_0))$ until it reaches $J(x, t_0) \in J(Fol_J)$ where it sort of bounces back - although this image of bouncing doesn't apply well to all situations as we'll see soon - and after that there is such a bouncing at each time S_x crosses Fol_J, until the last one at $J(x, t_1)$ after which x follows the embedded arc $J(x \times [t_1, 1))$ to leave $J(IF_J)$ for good. Figure 27 illustrates the situation:

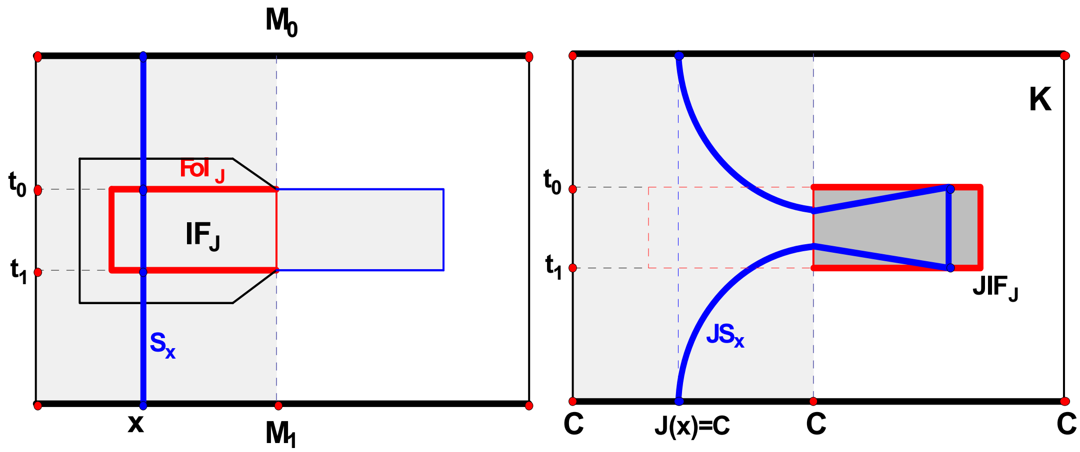

Figure 27
The map J and its folding set - represented as the rectangle with thick red sides in the left frame - are defined as in Examples 34 and 35 above (Figures 23 - a) and b)).

Recall that for $H \in \mathcal{G}_\partial$ its folding set is $Foll_H = \lim \{Fol_J : J \in S_H^\varepsilon\}$, the *limit of the folding sets* of all the ε-shadows. In face of Examples 34 and 35 we can not define an *internal part of* $Foll_H$ in terms of the connected components of $M - Foll_H$, as we did for the ε-shadows, since $Foll_H$ may not separate M; nevertheless its natural to take the limit of the internal parts of the ε-shadows, $IFl_H = \lim \{IF_J : J \in S_H^\varepsilon\}$: this exists since, by definition, for each $J \in S_H^\varepsilon$ we have $Fol_J \subset IF_J$ and so $Foll_H \subset IFl_H$, but the inclusion may or may not be strict. We examine next the image of $H(IFl_H)$; the analysis is by connected component so, to avoid introducing notation and without loss of generality, we may *assume* IFl_H *is connected*. We distinguish cases according to the dimension of $H(IFl_H)$: in 3-dimension we have three cases, due to the following

Exercise 39 *Show that* $\dim H(IFl_H) > 0$ *- since we are assuming* IFl_H *connected this means simply that* $H(IFl_H)$ *is not a point (Hint: look back at Corollary 33 and the argument that* $Fool_H$ *has no* 0*-dimensional components)*

Suppose first that $\dim H(IFl_H) = 3$. For simplicity, we will argue heuristically as if H was one of its ε-shadows: of course, proper argument would have to be done through those ε-approximations in the appropriate $\varepsilon - \delta$ rigour. Consider the set D of points in the boundary of $H(IFl_H)$ which have 3-dimensional neighbourhoods in $H(IFl_H)$: these must be images of points in the folding set $Foll_H$. Let $R = Foll_H \cap H^{-1}(D)$ and $p = (x, t) \in R$; let $B(H(p))$ be a ball in K that satisfies the *geometric condition*, and let $N(H(p)) = B(H(p)) \cap H(IFl_H)$ be a three dimensional neighbourhood as required in the definition of set D; Suppose that the stalk S_x that contains point p is *transverse* to $Foll_H$ at p: this means that sufficiently close stalks also intersect $Foll_H$ transversely near p. Let $M(p) \subset H^{-1}(N(H(p)))$ be an open ball around p. Since H folds this $M(p)$ along $M(p) \cap Foll_H$ into a neighbourhood of $H(p)$ in $N(H(p))$, all the tracks of points sufficiently close to x will bounce back in $D \cap N(H(p))$, inside the *geometric ball* $B(H(p))$ - see Figures 28-a and 28-b; we don't need to describe them in detail: what is intended to be the folding set and its image is similar to Examples 34 and 35. By the geometric condition, these tracks do not follow the shortest paths, and it's clear we can decrease length, by pushing D slightly into $N(H(p))$; since this pushing can obviously be achieved through some cubical subdivisions, that is realized inside $\mathcal{Q}_\partial$, we would get a contraction $H' \in \mathcal{G}_\partial$ with energy $E(H') < m_E$.

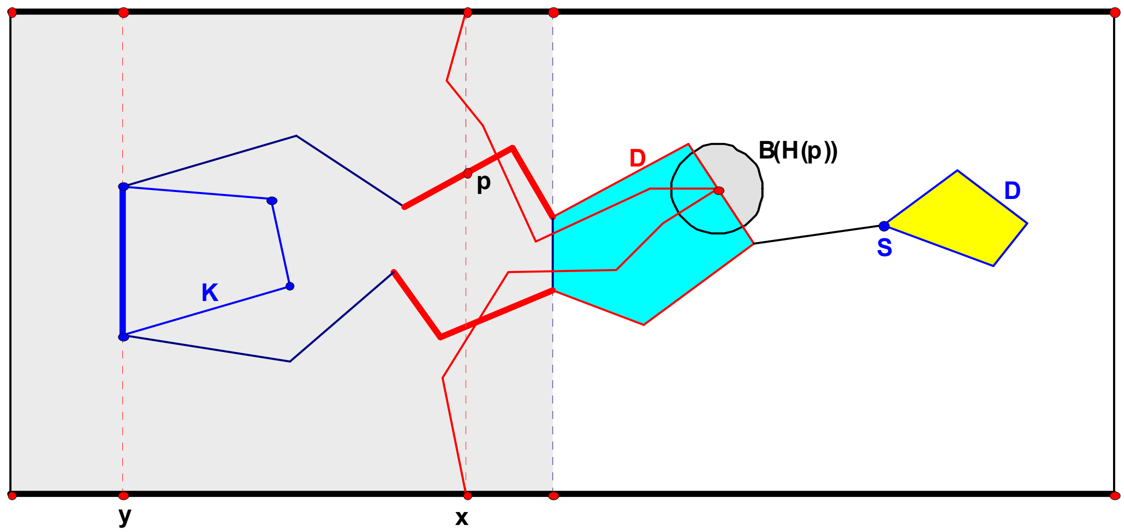

Figure 28-a

It may happen that there are no points in R of the type we've just considered, where the stalks are transverse to $Foll_H$. This situation is suggested in Figure 28 by the yellow coloured regions; it may happen that D is made up of tracks of the contraction: that is the case when the folding of H in points p for which there is a 3-dimensional neighbourhood $N(H(p))$, only happens in vertical parts of $Foll_H$, parts made out of stalk segments - as S_y in the picture. In this case the other points in $Foll_H$ where the stalks are transverse are all sent to parts of $H(IFl_H)$ of smaller dimension, one or two: this is suggested in Figure 28 by the crunching of the middle gray section, on the left, which is sent by H into the middle segment on the right; of course, in this situation the tracks which traverse D much enter and leave it (recall what was said above and illustrated in Figure 27) through a subcomplex S, of dimension 0 or 1; its pre-image $K = H^{-1}(S)$ is a subcomplex of dimension greater or equal to 2 (recall we are arguing in three dimensions). Consider the track segments traversing S near a point q, and a *geometric ball* $B(q)$ - recall that H is supposed to have minimal energy-control $m_{E,C}$ and so, by Theorem 37, we can assume that all these segments are embedded - since they all come from a 1 or 2 dimensional subcomplex in $B(q)$, to fill in a 3-dimensional region, then by the geometric condition, not all them can be geodesic arcs; pick a point $p_0 = (x_0, t_0)$ such that the correspondent segment track near q, say $H(x_0 \times [t_0 - \delta, t_0 + \delta])$, is not a geodesic and let $H(p_0) = q_0$; since $H^{-1}(q_0)$ is a subcomplex of K of dimension at least 1, we can consider an arc γ embedded in K, between points p_1 and p_2 and going through p_0 and such that for all $p' = (x, t) \in \gamma$ we have $H(p') = q_0$ and $H(x \times [t - \delta, t + \delta])$ is not a geodesic (note that by Theorem 37 and since $H(\gamma) = \{q_0\}$ all the segment stalks $x \times [t - \delta, t + \delta]$ are transverse to γ). We can now apply the *bootleg move* to decrease length in a suitable small neighbourhood N of the arc γ and note that, as in the case of the proof of Theorem 37, the move can be made legal since it is done near the arc γ that was shrunk by H into point q_0. Again we would reduce the energy bellow the minimum m_E, and therefore we can conclude that $H(IFl_H)$ can not have dimension 3.

Suppose now that $\dim H(IFl_H) < 3$. The two cases are dealt with simultaneously. The argument starts with the following note, which applies independently of $\dim H(IFl_H)$: consider $P = H^{-1}H(IFl_H)$; then it is the case that $P - IFl_H$ must have non-empty interior: this says simply that the *external part* of $Foll_H$, $EFl_H = M - IFl_H$, must contribute with some nontrivial region surrounding IFl_H for the folding to actually take place - referring back to Example 34, look at the region $P = \lim_n P_n$ surrounding the folding set $Foll_H = s$.

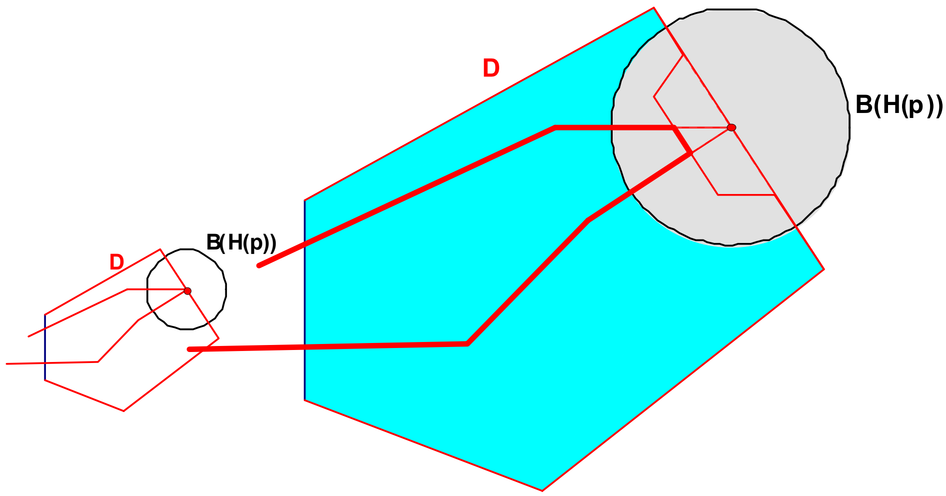

Figure 28-b

We give a general sketch of the argument. Let $J_n, n \geq 1$ be any sequence of ε-shadows approximating H, such that $d(H, J_n) < 1/n$. We look at the internal parts, IF_{J_n}, and their full inverse images, $P_n = J_n^{-1} J_n(IF_{J_n})$. By the definition of folding for simplicial (cubical) maps, we have that $P_n - IF_{J_n} \neq \emptyset$: as in Example 34, P_n consists of IF_{J_n} and an external part made of *phantom regions*; of course the situation in general is much more complicated than that example, with several phantom regions that may overlap; we may also have regions that come from a simpler type of folding that doesn't involve cusp lines or points in Fol_{J_n} - as the cusp points x_n and x'_n in Figure 23-a - as shown in the example the next figure describes:

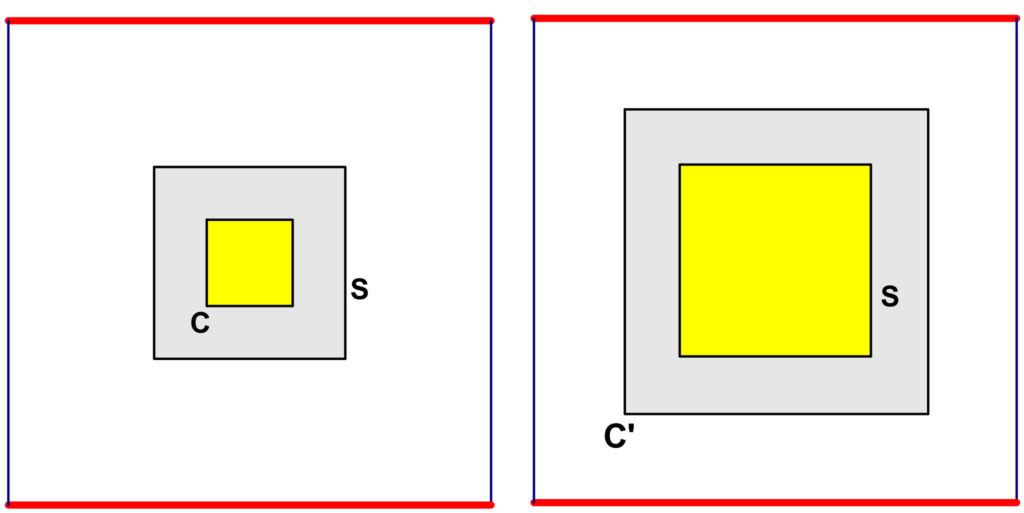

Figure 29

The folding is easy to describe: it is the identity outside the outer square S, and sends the inner (yellow) square C to the larger square C' that surrounds S; the yellowish-gray band between C' and S suggests the double layer of the folding. In any case, each phantom region of J_n has an *outer boundary* (its boundary in EF_{J_n}), which is identified to a part of Fol_{J_n} - through the relation $\sim_H$ induced by the map H ($x \sim_H y \Leftrightarrow H(x) = H(y)$) and an *inner boundary* in Fol_{J_n}, this one is identified by $\sim_H$ either to another part of Fol_{J_n} (we may have overlapping regions) or to another outer boundary of another phantom region: we thus have a sequence of three regions that consecutively fold into the phantom region.

We want to show that the limit of P_n, which is of course contained in P, has non-empty interior. Given a phantom region of P_n, say F_n, let its *depth* be $D(F_n) = \max d(x,y)$ where d is the length metric considered inside F_n and x and y are points in its outer and inner boundaries, respectively: this measures how deep F_n extends, independently of its distance from IF_{J_n}. Take a limit of such F_n, say $F = \lim F_n$; without loss of generality we can assume that all F_n have minimum depth, say D: if the depths of the phantom regions decreased arbitrarily with n, then for every $\delta > 0$ we we could find an order $N(\delta)$ such that for all $n \geq N(\delta)$ the phantoms of J_n would all have depth less than δ; but then it is easy to see, using the uniform continuity of the approximations (the equicontinuity), that we could remove all the folding corresponding to those phantoms by moving points by less than 2δ, thus obtaining arbitrarily close approximations to H without any folding near $Foll_H$, which contradicts the definition. If F has non-empty interior then $P \supset F$ has non-empty interior - we'll see further down the case of $P - IFL_H$. Suppose now that P, and therefore F, have empty interiors; besides having a minimum for the depths of the F_n's, we can also assume, without loss of generality, that under J_n the lengths of the paths that run in F_n between its outer and inner boundaries, don't decrease arbitrarily with n: otherwise, as in the above case of not having minimal depth, we could as well remove all the foldings near $Foll_H$. Therefore the images $J_n(F_n)$ sustain *projected depth* $D_p(F_n) = \max \left\{ l(J_n(\gamma_{x,y}) \right\}$ - where x and y are a pair of points in the outer and inner boundaries of F_n, respectively, and $\gamma_{x,y}$ is an F_n-geodesic joining such a pair - for which there is a minimum D_p; as a consequence we have that, considering the images by J_n of the outer and inner boundaries, $J_n(\partial_{out}(F_n))$ and $J_n(\partial_{in}(F_n))$, at least one of them must have projected depth greater or equal to $D_p/2$: let's denote it indistinctly by $J_n(\partial\_(F_n))$. Recall that either of these images is also the image of a part of Fol_{J_n}, in the triple folding process already mentioned; suppose that $z_n \in Fol_{J_n}$ is a point such that $J_n(z_n) \in J_n(\partial\_(F_n))$ and such that, with some other point, maximizes $D_{J_n}(F_n) \geq D_p$. By equicontinuity, $\exists \delta > 0$ such that $d(x, z_n) < \delta \Rightarrow d(J_n(x), J_n(z_n) < D_p/2$, for all n; but then, because of the projected depth and the definition of folding of J_n we have that $J_n(x) \in H(F_n)$, and so $B(z_n, \delta) \subset P_n$. Letting $z_n \longrightarrow z \in P$ - by passing to a subsequence if necessary - we have that $B(z, \delta/2) \subset \lim P_n \subset P$ and so P has non-empty interior.

To see that $P - IFl_H$ has non-empty interior: since we already know that $\mathring{P} \neq \emptyset$, either $P - IFl_H$ has non-empty interior or IFl_H has non-empty interior. Suppose that $Int(IFl_H) \neq \emptyset$; in the previous approximation by J_n we can assume that for n sufficiently large IF_{J_n} has non-empty interior and so $\lim IF_{J_n} \subset IFl_H$ has non-empty interior and therefore its boundary must be 2-dimensional. Then there must be some limit $F = \lim F_n$ of phantom parts with non-empty interior, otherwise some phantom parts would have to have arbitrarily small depth, something we've already seen can not happen. We can state:

Lemma 40 *Let $H \in \mathcal{G}_\partial$. Let IFl_H be the internal part of its folding and $EFl_H = M - IFl_H$ the external part. Then $EFl_H \cap H^{-1}H(IFl_H)$ has non-empty interior.*

We can now resume the argument for the case $\dim H(IFl_H) < 3$. Let $p = (x_0, t_0)$ be a point in the interior of $P = H^{-1}H(IFl_H)$ and $N(p) = N(x_0, \varepsilon) \times [t_0 - \delta, t_0 + \delta]$ a neighbourhood contained in P. By Theorem 37, and the easy-end condition we are now assuming, we can suppose that all segment tracks $x \times [t_0 - \delta, t_0 + \delta]$ are embedded by H; but then, since $H(N(p)) \subset H(IFl_H)$ and this is one or two dimensional, in each level $N(x_0, \varepsilon) \times t$, $t \in [t_0 - \delta, t_0 + \delta]$, we must have a 1-dimensional stratum where H is constant (at least 1-dimensional: of course H may even crunch the whole level). Look at any path, for instance $x(s) \times \{t_0\}$, $x(s) \in N(x_0, \varepsilon)$, $s \in I$, where H is constant (there is a one parameter family of such paths at each level): this means that the points $x(s)$ are all coalesced at time t_0; by the easy-end condition they must *de-coalesce* at some future time: say $t_1 = \sup\{t : H_t(x(s) = H(x(s'), \forall s, s' \in I\}$ is the last time when all of them are coalesced. We can then apply the *bootleg move* to decrease length in some geometric ball around $H(x(s), t_1)$, using some sufficiently small neighbourhood of the path $l = \{(x(s), t_1) : s \in I\}$: again the bootleg move becomes legal since we will be only expanding what was previously shrunk. Note that in alternative, and due to the easy-boundary condition, we could as well go back in time and reason in the same way. This finishes the analysis of folding under the assumption of minimal energy-control: we came to the following conclusion that in fact it doesn't take place:

Theorem 41 *Let $H \in \mathcal{G}_\partial$ have minimal energy-control, $m_{E,C}$. Then $Foll_H = \emptyset$.*

Coalescence under $m_{E,C}$

The final analysis of coalescence is now much faster. We start by drawing some consequences of last theorem. Let H have minimal control-energy and so, by the last theorem, empty folding set; it is clear that $Foll_H = \emptyset$ implies that for all $x \in M$, $H^{-1}H(x)$ must be connected: if C_1, C_2 were two disjoint connected components of $H^{-1}H(x)$, we could take a regular neighbourhood $N(C_1)$ of C_1 disjoint from C_2; then the boundary of $N(C_1)$ would be a cycle - actually an embedded oriented manifold (a surface in dimension 3) - over which H folded. We state this as a lemma:

Lemma 42 *Let $H \in \mathcal{G}_\partial$ have minimal energy-control, $m_{E,C}$. Then for all $x \in M$, $H^{-1}H(x)$ is connected.*

Remark 43 *The study of the folding set, in particular the definition of $Foll_H$, and last Lemma suggests a relation between empty-foldness and the subject of manifold decompositions and approximation by homeomorphisms - see [19]. A natural conjecture is that for empty $Foll_H$ the sets $K_x = H^{-1}H(x)$ are* cellular*, that is, have arbitrarily close neighbourhoods, $N_\varepsilon \subset N(K_x, \varepsilon)$, which are 3-balls, $N_\varepsilon \cong B^3$.*

Suppose now that under H, which we are assuming has minimal energy-control, the points x and y coalesce at time t_0. Looking in a geometric ball B around $H(x, t_0) = H(y, t_0)$, consider $\delta > 0$ such that $A_x = H(x \times [t - \delta, t + \delta])$ and $A_y = H(y \times [t - \delta, t + \delta])$ are contained in B - recall that by Theorem 37 we can assume these segments of the tracks are embedded arcs; if they are geodesic, then by the geometric condition they intersect transversely: this means, looking at the images of the level-spheres $S^2 \times \{t\}$, that the sphere crosses itself transversely at time t_0. But this transverse crossing of the levels can only happen with some folding, as suggested in the next figure. We can construct a folding cycle C over which the points (x, t_0) and (y, t_0) fold by substituting two small hemispheres, one extending into the past and the other into the future, for neighbourhoods in $S^2 \times \{t_0\}$ of those two points. This means that at a coalescence point the sphere levels can only touch tangent, without crossing.

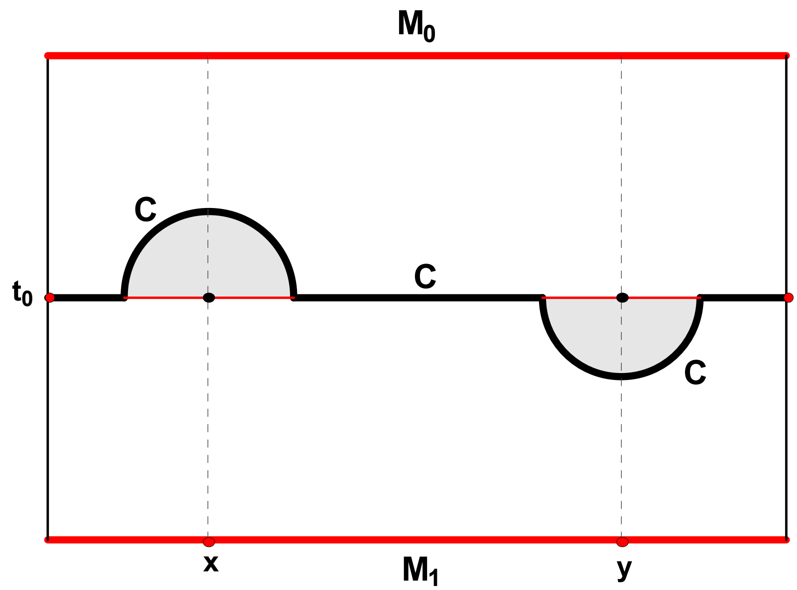

Figure 30

Therefore at least one of the arcs A_x or A_y is not a geodesic, so we can decrease length using the bootleg move: since, by the previous Lemma, $P = H^{-1}H((x,t_0)) = H^{-1}H((y,t_0))$ is connected we can consider an arc l in P joining the two points (x,t_0) and (y,t_0); as before, the bootleg move becomes legal if we perform it in a sufficiently small neighbourhood of l; we can thus state:

Theorem 44 *Let $H \in \mathcal{G}_\partial$ have minimal energy-control, $m_{E,C}$. Then H is coalescent.*

In relation to the heuristic arguments we've advanced this is an understatement, because those arguments apply in greater generality: more than coalescence, we have the effect of tracks passing through the same points at different times. Therefore, we just leave it as a research exercise the problem of relating (in some direct way) the minimal energy-control of H, and consequent coalescence, with the possible continuity of the *reach-time function* (recall Theorem 29)

Problem 45 *Let $H \in \mathcal{G}_\partial$ be a coalescent contraction with minimal energy-control, $m_{E,C}$. Try to prove that the reach-time function, $x \longrightarrow t_x$ (cf. Theorem 29) is continuous. Or find a counter-example...*

5.4 Concluding remarks

We want to conclude with some notes intended to provide some guide lines for future research on the topic of coalescence.

Volume *versus* Energy: Note that we ended up not using the volume function in the analysis of coalescence, although we have used the control-energy function E_C, which is an hybrid of energy and volume. But the idea of volume was the original one and, in any case, the proof of the lower-semicontinuity of E_C is analogous to the one for the volume function - both functions are parameter sensitive - and therefore some equivalent amount of space would have to be provided for that proof, if we didn't have already the previous one to rely on. Besides, the volume function, and other variations may be useful in analysing contractions in more general settings, where we don't have any local geometric condition.

Simplicial *versus* Differential: It is clear that while many of the basic heuristics have their original motivation in a combinatorial and finite setting, thus the simplicial/cubical approximation approach, the need to take limits in the space of functions, as well as the geometric condition, suggest that the differential setting - perhaps including some dynamical systems, with the folding related to singularities - would be better suited to deal with the $(\varepsilon - \delta)$ institutional monster.

Aftermath: Besides the throwing in of formal pageants around some of the heuristics - which is cried for by the aforementioned creature - and the working out of some of the problems left over (e.g. Exercise 14, Problems 27 and 45), we single out the following main themes: the relation of folding with approximation by homeomorphisms as already mentioned in Remark 43; the relation of coalescence with collapsible complexes, with an eye on Zeeman's conjecture (see [2]): as a start, generalize the results of sections 2 and 3 to arbitrary complexes, in particular try to identify some essential features an *intrinsic* contraction of a non-collapsible contractible complex must exhibit and that may be removed when crossing with I.

Non-compactness: The fantastic Whitehead example, [20], of an open subset W of $\mathbb{R}^3$ which is contractible but is not homeomorphic to an open 3-ball, is also an example of a manifold for which no coalescence contraction exists: in fact the usual proof that W is not homeomorphic to $\mathbb{R}^3$ consists of ascertaining that a certain curve is not contained in any closed 3-ball - is not *engulfed* (see for instance [21, Theorem 14.2] or, for a more general treatment, [22]): but surely coalescence implies engulfing; this is another problem for investigation (For more on this fascinating land of exotic contractible 3-manifolds see also [23], [24] and [25])

About 'Shrinking Complexity': The title of the paper has a double meaning, either as a verbal noun (gerund) or a compound noun; the reader is advised to choose the one that better suits what he/she feels has be achieved in the paper.

This brings us to an end.

Acknowledgements

To quote Brower in the beginning is an appropriate tribute to the great topologist, for as we pointed out in the introduction, Sections 2 and 3 are just very simple variations on the subject of map-degree. But the chosen quotation [6] reflects to perfection the spirit and circumstances surrounding the making of this paper. In fact, most of the ideas for the paper - both heuristic and technical - came during our strides along the seafront in Afife with our faithful labradors, Tao, Eta and Bright; and they came as relaxing diversions from our musing over the foundations of mathematics and the philosophy of mind [7]. The writing of this paper was no exception to the usual frustration: the price we always pay for trying to give our most promising and fantastic ideas their right share of objectivity and reality is to end up with the clearest notion of how feeble and poor they reveal themselves in the process. I want to thank my friends for their patience with the many times I've used the writing of this paper as an excuse for excepting

[6] While we can understand Brower's comments about logic from an historical perspective, with today's knowledge we can only condone them under a relativistic interpretation: Brower didn't live to see the dawn of Geometric-Logic/Topos Theory where the deep intuitionists's intuition about Bivalence, Intensionality and Choice are fully vindicated. With the second part of the quotation we fully agree.

[7] At a certain time we were attending a postgraduate seminar on the philosophy of mind, in Faculdade de Letras (FLUP). We would like to thank Sofia Miguens, who run the seminar, and our fellow students there, for the formidable intellectual thrill and impetus we got from those Thursday evenings.

myself from other tasks. Having to be done, for practical reasons, far from the sea breezes, the writing of this paper was an unhealthy business: I take it to that the many mistakes and pieces of nonsense the paper surely contains.

> What we "are tempted to say"... is, of course, not philosophy; but it is its raw material. Thus for example, what a mathematician is inclined to say about the objectivity and reality of mathematical facts, is not a philosophy of mathematics, but something for philosophical treatment.
>
> The philosopher treats a question; like an illness. (Wittgenstein, [26])

> The philosopher is the man who has to cure himself of many sicknesses of the understanding before he can arrive at the notions of the healthy human understanding.
>
> If in life we are surrounded by death, so in healthy understanding we are surrounded by madness. (Wittgenstein, [27]) [8]

References

[1] L. E. J. Brouwer, Consciousness, Philosophy and Mathematics, (1948) in Benacerraf, Paul and Putnam, Hilary (ed.), *Philosophy of mathematics*, second edition (1983), CUP

[2] E. C. Zeeman, On The Dunce Hat, Topology, vol.2 (1964), 341-358.

[3] C. R. F. Maunder, *Algebraic Topology*, Van Nostrand Reinhold Company, London (1972).

[4] C. Rourke and B. Sanderson, *Introduction to piecewise-linear topology*, Springer-Verlag (1972).

[5] T. B. Rushing, *Topological embeddings*, Academic Press (1973).

[6] E. F. Rêgo, On the Mechanics of the Poincaré Conjecture - an Heuristic Tour, CMUP-preprints, 13 (2003) (available in pdf-format from *www.fc.up.pt/cmup*)

[7] R. H. Bing, *The Geometric Topology of 3- Manifolds*, Amer. Math. Soc. Colloquium Publications, vol. 40, Amer. Math. Soc. (1983).

[8] J. Dugundji, *Topology*, Allyn and Bacon, Inc., Boston (1966)

[9] J. F. Davis and P. Kirk, *Lecture Notes in Algebraic Topology*, Graduate Studies in Mathematics, vol. 35, Amer.Math.Soc. (2001).

[10] G. E. Bredon, *Topology and Geometry*, Springer-Verlag (1993).

[11] J. W. Alexander, Combinatorial analysis situs, Trans. Am. Math. Soc. 28 (1926), 301-329

[12] O. Veblen, *Analysis situs*, Am. Math. Soc. Colloquium Publications No. 5, Part II (1922)

[13] E. C. Zeeman, Relative simplicial approximation, Proc. Camb. Phil. Soc. 60 (1964), 39-43

[8] This most wonderful combination of quotations from Wittgenstein was taken from [28, page 29]

[14] R. C. Kirby and M. G. Scharlemann, *Eight Faces of the Poincaré Homology 3-Sphere*, in Geometric Topology (ed. J. C. Cantrell), Academic Press (1979), 113-146.

[15] W. P. Thurston (edited by Silvio Levy), *Three-Dimensional Geometry and Topology*, Princeton University Press (1997).

[16] P. Scott, The Geometries of 3-Manifolds, Bull. London Math. Soc., No. 15 (1983), 401-487.

[17] M. H. Freedman and F. Quinn, *Topology of 4-Manifolds*, Princeton University Press (1990).

[18] G. M. Ziegler, *Lectures on Polytopes*, Springer-Verlag (1995)

[19] R. J. Daverman, *Decompositions of manifolds*, Pure and Appl. Math. vol. 124, Academic Press (1986).

[20] J. H. C. Whitehead, *A certain open manifold whose group is unity*, Quart. J. Math. (2), 6(1935), 268-279.

[21] J. Hempel, *3-manifolds*, Ann. of Math. Studies No. 86, Princeton Univ. Press (1976).

[22] D. R. McMillan, Jr., *Some contractible open 3-manifolds*, Trans. Amer. Math. Soc. 102 (1962), 373-382.

[23] R. Myers, *Contractible open manifolds which are not covering spaces*, Topology 27 (1988), 27-35.

[24] D. G. Wright, *Contractible open manifolds which are not covering spaces*, Topology 31 (1992), 281-301.

[25] D. G. Wright and F. C. Tinsley, *Some contractible open manifolds and coverings of manifolds in dimension three*, Topology and its Applications 77 (1997), 291-301.

[26] L. Wittgenstein, *Philosophical Investigations*, 2nd edn., Edited by G. E. M. Anscombe and R. Rhees, Oxford: Blackwell (1967)

[27] L. Wittgenstein, *Remarks on the Foundations of Mathematics*, 2nd edn., Edited by G. E. M. Anscombe, R. Rhees and G. H. von Wright, Cambridge, MA: MIT Press (1978)

[28] D. G. Stern, *Wittgenstein on Mind and Language*, Oxford University Press (1995)

Centro de Matemática da Universidade do Porto, Departamento de Matemática Pura
Faculdade de Ciências, Universidade do Porto
Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
E-mail address: eerego@fc.up.pt