Geometry & Topology 21 (2017) 441–484

Mean curvature flow of Reifenberg sets

OR HERSHKOVITS

In this paper, we prove short time existence and uniqueness of smooth evolution by mean curvature in $\mathbb{R}^{n+1}$ starting from any n–dimensional (ε, R)–Reifenberg flat set with ε sufficiently small. More precisely, we show that the level set flow in such a situation is non-fattening and smooth. These sets have a weak metric notion of tangent planes at every small scale, but the tangents are allowed to tilt as the scales vary. As for every n this class is wide enough to include some fractal sets, we obtain unique smoothing by mean curvature flow of sets with Hausdorff dimension larger than n, which are additionally not graphical at any scale. Except in dimension one, no such examples were previously known.

53C44

1 Introduction

A family of smooth embeddings $\phi_t\colon M^n \to \mathbb{R}^{n+1}$ for $t \in (a,b)$ is said to evolve by mean curvature if it satisfies the equation

$$\frac{d}{dt}\phi_t(x) = \overrightarrow{H}(\phi_t(x)), \tag{1.1}$$

where $\overrightarrow{H}$ is the mean curvature vector. Equivalently, by the first variation formula, mean curvature flow is the negative gradient flow of the area functional.

If a compact hypersurface $M \subseteq \mathbb{R}^{n+1}$ is of type C^2, it follows from standard parabolic PDE theory that there exists a unique mean curvature flow (abbreviated MCF) starting from M for some finite maximal time T, and that in fact (see for instance [17]),

$$\lim_{t\to T} \max_{x\in M_t} |A(x,t)| = \infty. \tag{1.2}$$

The question of mean curvature flow (and geometric flows in general) with rough initial data, ie when the C^2 assumption is weakened, has been researched extensively; see for instance Ecker and Huisken [5; 6], Wang [21], Simon [19], Koch and Lamm [14], Lauer [15] and Clutterbuck [2]. In the case that M is merely Lipschitz, short time existence was proved by Ecker and Huisken in the celebrated paper [6]. Their proof is based on the fact that in the C^1 case, M can be written locally as a graph of a

Published: 10 February 2017 DOI: 10.2140/gt.2017.21.441

C^1 function, and the ellipticity of the graphical mean curvature equation is controlled by an interior gradient estimate. Note that even in the Lipschitz case, the n–dimensional Hausdorff measure is still finite, although the gradient of the area functional may not be. In a different direction, in [15] Lauer was recently able to show that when $n=1$, for any Jordan curve γ in $\mathbb{R}^2$, if $m(\gamma)=0$ (where m is the two dimensional Lebesgue measure) then the level set flow (see Definition 1.8) is non-fattening and smooth. In a very different direction, Clutterbuck showed in her PhD thesis [2] that on a mean convex domain $\Omega \subseteq \mathbb{R}^n$, for any continuous function $u_0\colon \Omega \to \mathbb{R}$ which vanishes on the boundary, there exists a solution $u\colon \Omega \times (0,T] \to \mathbb{R}$ to the *graphical mean curvature flow equation* (see (5.48)) for some $T>0$, satisfying $u=0$ on $\partial\Omega \times (0,T)$, that converges to u_0 at $t \to 0$ in C^0.

The current paper deals with the existence and uniqueness of smooth flows in $\mathbb{R}^{n+1}$ starting from a class of sets which is general enough to include some sets of Hausdorff dimension larger than n.

Definition 1.3 (Reifenberg flat sets [18]) A compact, connected set $X \subseteq \mathbb{R}^{n+1}$ is called (ε, R)*–Reifenberg flat* if for every $x \in X$ and $0<r<R$ there exists a hyperplane P such that

$$d_H(B(x,r)\cap P, B(x,r)\cap X) \leq \varepsilon r. \tag{1.4}$$

Here d_H is the Hausdorff distance.

The point is that the approximating hyperplanes may tilt as the scales vary. In [18], Reifenberg showed that provided ε is sufficiently small, an (ε, R)–Reifenberg flat set is a topological submanifold. As stated above, the Reifenberg condition is weak enough to allow some fractal sets. For instance, as described in Toro [20], a variant of the Koch snowflake, at which the angles in the construction are β instead of $\frac{\pi}{3}$, is (ε, R)–Reifenberg with $\varepsilon = 4\sin\beta$. Note that the snowflake is not graphical at any scale. An analogue of this can be done in every dimension.

Before diving into more technicalities, we can already state a form of our main theorem.

Theorem 1.5 *There exist some $\varepsilon_0, c_0>0$ such that if X is (ε, R)–Reifenberg flat for $0<\varepsilon<\varepsilon_0$ then there exists a smooth solution to the mean curvature flow $(X_t)_{t\in(0,c_0R^2)}$ attaining the initial value X in the following sense:*

$$\lim_{t\to 0} d_H(X, X_t) = 0. \tag{1.6}$$

Moreover, the flow (X_t) is unique (in a sense that will be explained shortly).

Thus, the mean curvature flow provides a canonical smoothing of Reifenberg flat sets.

Remark 1.7 As described above, there exist previous results implying smoothing by mean curvature flow of sets with Hausdorff dimension larger than n [15; 2]. However, those results assumed either $n = 1$ or a *global* graph structure. Our results allow n to be arbitrary and apply to sets which are *not graphical at any scale*.

To state the uniqueness result more accurately, we need the following definition.

Definition 1.8 (level set flow [13; 12]) A family $(X_t)_{t\in[0,b]}$ of closed subsets of $\mathbb{R}^{n+1}$ is said to be a *weak set flow* starting from X_0 if it satisfies the avoidance principle with respect to any smooth mean curvature flow. More precisely, for any smooth mean curvature flow $(\Delta_t)_{t\in[t_0,t_1]}$ with $0 \le t_0 \le t_1 \le b$ such that

$$\Delta_{t_0} \cap X_{t_0} = \varnothing, \tag{1.9}$$

we have

$$\Delta_t \cap X_t = \varnothing \tag{1.10}$$

for every $t \in [t_0, t_1]$. The *level set flow* is the maximal weak set flow starting from X.

The level set flow was defined in Evans and Spruck [8] and Chen, Giga and Goto [1] using the language of viscosity solutions for PDEs in order to develop a theory for weak solutions of mean curvature flow. The more geometric definition above is from Ilmanen [12], where the equivalence was also shown to hold. If X_0 is a smooth submanifold, the level set flow will coincide with the classical evolution by mean curvature flow for as long as the latter is defined. An advantage of working with the level set flow is that it is defined and unique for all time (so it is indifferent to singularities), and it allows one to flow any closed set. The drawback of it is that the X_t may develop an interior (in $\mathbb{R}^{n+1}$), even if X_0 was the boundary of an open set. The development of an interior is referred to as "fattening" and is the right notion of non-uniqueness in this setting.

We are now ready to state the full version of our main theorem.

Theorem 1.11 (main theorem) *There exist some $\varepsilon_0, c_0 > 0$ such that if X is (ε, R)–Reifenberg flat for $0 < \varepsilon < \varepsilon_0$ then the level set flow $(X_t)_{t\in[0,c_0R^2]}$ starting from X is a (non-vanishing) smooth evolution by mean curvature flow for $t \in (0, c_0R^2)$ that satisfies*

$$\lim_{t\to 0} d_H(X, X_t) = 0. \tag{1.12}$$

In particular, the level set flow does not fatten.

Remark 1.13 It follows immediately from the proof of Theorem 1.11 that the convergence to the initial data is in fact in the parametrized C^0 sense. See Remark 3.79.

We will now give an outline of the argument. Our first goal will be to construct a smooth solution to the mean curvature flow $(X_t)_{t\in(0,c_0R^2)}$ which converges to X in the Hausdorff sense as $t\to 0$. To do that, we first approximate the set X by smooth hypersurfaces at each scale, according to the following theorem, implicit in [11] (see also Section 2).

Theorem 1.14 [11; 18] *There exist some constants $c_1, c_2 > 0$ such that if X is (ε, R)–Reifenberg flat for $0<\varepsilon<\varepsilon_0$, then there exists a family of hypersurfaces $(X^r)_{0<r<R/4}$ such that:*

1. $d_H(X^r, X) \le c_1\varepsilon r$.
2. $|A| \le c_2\varepsilon/r$ *for every $x\in X^r$, where A is the second fundamental form of X^r.*
3. *For every $x\in X$, $r\in(0,R/4)$ and $s\in(r,R/4)$, $B(x,s)\cap X^r$ can be decomposed as*

$$B(x,s)\cap X^r = G\cup B, \tag{1.15}$$

where G is connected and $B\cap B(x,(1-20\varepsilon)s)=\varnothing$.

We want to construct a smooth evolution of X by taking a limit of the flows emanating from the X^r. In order to do that, we derive the following uniform estimates for the evolutions of the hypersurfaces X^r.

Theorem 1.16 (uniform estimates) *For every $\Lambda>0$ there exist some ε and c_0, c_1, c_2, c_3 such that if X is (ε, R)–Reifenberg flat, and considering the approximating surfaces X^r from Theorem 1.14, each X^r flows smoothly by mean curvature for time $t\in[0,c_0R^2]$, and for every $t\in[c_3r^2, c_0R^2]$ we have the following properties:*

1. *Denoting by $A^r(t)$ the second fundamental form of X^r_t,*

$$|A^r(t)| \le \frac{c_1}{\sqrt{t}}. \tag{1.17}$$

2. X^r_t *approximates X in the Hausdorff sense,*

$$d_H(X^r_t, X) \le c_2\sqrt{t}. \tag{1.18}$$

3. *For every $x\in X$ and $s\in(\sqrt{t}/c_1, R/4)$ we have*

$$B(x,s)\cap X^r_t = G\cup B, \tag{1.19}$$

where G is connected and $B\cap B\big(x,\tfrac{9}{10}s\big)=\varnothing$.

Moreover, the constants c_1, c_2 *satisfy*

$$c_1 c_2 < \min\{10^{-6}, \Lambda^{-1}\}, \quad c_1^2 < \tfrac{1}{80}. \tag{1.20}$$

For reasons that will become apparent soon, it is convenient to make the following definition.

Definition 1.21 (approximate solutions, physical solutions) An MCF $(Y_t)_{t\in[0,c_3 R^2)}$ that satisfies estimates (1.17)–(1.19) for every $t \in [c_3 r^2, c_0 R^2]$ will be called a (Λ) *r–approximate evolution of X*. If it satisfies the estimate for every $t \in (0, c_0 R^2]$, we say it is a (Λ) *physical solution*.

Theorem 1.16 is the technical and functional heart of the argument; in addition to being the hardest to prove, this is the last place in the argument at which the Reifenberg assumption is explicitly used. While additional work is required in order to show the uniqueness part, both existence and uniqueness are logical consequences of the uniform estimates of Theorem 1.16. Thus, if one is able to prove a statement analogous to Theorem 1.16 for some class of sets, existence and uniqueness will follow from the work in our paper.

The proof of Theorem 1.16 is by iteration; the idea is the following. Interpolating the Hausdorff bounds and the curvature bounds, we get that X^r is locally a graph of a function u with a small gradient over the hyperplane approximating X at scale r. Letting X^r flow for a short yet substantial time (compared to r^2), we will be able to extend those C^0 and C^2 estimates and interpolate again to provide a gradient bound which will be, say, 1000000 times bigger than the initial gradient bound. Using a new interior estimate for the graphical MCF (Theorem 3.1) we will get an improved bound on the second fundamental form for the evolved hypersurface. By bounding the displacement of X^r_t from X^r and by the Reifenberg property of X, we will see that X^r_t can serve as a good candidate for $X^{\theta r}$ for some fixed $\theta > 1$. This will allow us to iterate. Most of the above strategy is carried out in Lemma 3.25.

Obtaining the desired improved curvature bound, while well expected, is not a trivial task. The existing estimates of Ecker and Huisken ([6], see also Theorem 2.18) are not good enough when the *norm of the gradient* $|\nabla u|$ is small, as they depend on the so-called *gradient function* $\sqrt{1+|\nabla u|^2}$. There are several approaches to deriving sharp estimates in our situation. It turns out that the optimal result using only the norms of the gradient is not sufficiently good to perform iteration even in dimension one. Indeed, even local regularity, which in the graphical case reduces to a weighted L^1 estimate, does not give a sufficient bound in high dimensions. The standard Schauder estimate, using the C^0 norm of u in a cylinder turns out to give an insufficient estimate too, since a priori, u remains comparable to its initial value for too little time. Instead, thinking of the graphical mean curvature flow as a non-homogeneous heat equation on a thick cylinder, we will have to use both the C^0 norm at the initial time slice to control the contribution from the initial data, and the C^1 bound in the cylinder, to control the contribution from the boundary and the effect of the non-linearity. As described above, the gradient bound in the cylinder will be, say, 1000000 times bigger than at the beginning, but as the contribution of the boundary is so weak and as the non-linearity is so marginal (for small gradients, after a short while) this will not matter. The resulting estimate, Theorem 3.1, is somewhat technical and will therefore not be stated precisely in the introduction.

Once Theorem 1.16 is established the existence part of Theorem 1.11 follows immediately from compactness. It is easy to see that the limiting flow will actually be a weak set flow in the sense of Definition 1.8 and that it will satisfy the physicality assumption of Definition 1.21.

The main ingredient in the proof of the uniqueness of the flow is the following separation estimate.

Theorem 1.22 (separation estimate) *There exist $\Lambda > 0$ and $C > 0$ such that if Y_t^r is an r–approximate evolution of X and Z_t^s is an s–approximate evolution of X with $s \le r$, then with $c_0(\Lambda), \dots, c_3(\Lambda)$ from Theorem 1.16,*

$$d_H(Y_t^r, Z_t^s) \le C r^{1/2} t^{1/4} \quad \textit{for every} \quad t \in [c_3 r^2, c_0 R^2]. \tag{1.23}$$

The following partial uniqueness result is an immediate corollary of Theorem 1.22 and Theorem 1.16.

Corollary 1.24 *With the choice of Λ as in Theorem 1.22, and with $\varepsilon(\Lambda)$, $c_0(\Lambda)$, $\dots, c_3(\Lambda)$ as in Theorem 1.16, if X is (ε, R)–Reifenberg flat and X_t^r are the outputs of Theorem 1.16, the full limit*

$$\lim_{r \to 0} X_t^r \tag{1.25}$$

exists and is in fact the unique physical evolution of X.

The idea of the proof of Theorem 1.22 is that the conditions of Definition 1.21 imply that for fixed $t > 0$ and small r and s, Z_t^s is a graph of a function u over Y_t^r in a tubular neighborhood of the latter. In Lemma 4.19 we derive a PDE for u, a derivation which is the parabolic analogue of (and is based on) a similar calculation done for the minimal surface case in Colding and Minicozzi [3]. Once this is done, a bootstrap argument based on the maximum principle, starting from the "crude" bounds coming from (1.18), will give an estimate for $d_H(Y^s_t, Y^r_t)$.

While Theorem 1.11 is much stronger than Corollary 1.24, concluding it using what we described already is very easy. The only thing one needs to note is that in Theorem 1.14, the approximating hypersurfaces can be chosen to be either entirely in the compact domain bounded by X or in its complement (see Corollary 2.11). Those inward and outward r–approximate evolutions of X form barriers to the level set flow from inside and outside. By Theorem 1.22 they converge to the same thing, from which Theorem 1.11 follows.

The paper is organized as follows. In Section 2 we sketch the proof of Theorem 1.14 and collect some more auxiliary results. In Section 3 we prove Theorem 1.16 assuming the interior estimate, Theorem 3.1. In Section 4 we prove Theorem 1.22 and conclude the proofs of Theorem 1.11. In Section 5 we close the argument by proving Theorem 3.1.

Acknowledgements The author wishes to thank his advisor Bruce Kleiner for suggesting the problem, as well as for many useful discussions. He also wishes to thank the anonymous referee for comments and suggestions. He further wishes to thank Jacobus Portegies for carefully reading and commenting on an earlier version of this note, as well as for his many suggestions. The author wishes to thank Robert Haslhofer, for his many suggestions and comments about an earlier version of this notes, and Jeff Cheeger for his generous support during the months in which this project was initiated. The author was partially supported by NSF grants DMS 1406407 and DMS 1105656.

2 Preliminaries

In this section we record some known theorems and derive several simple auxiliary results which will be used later.

We first remark on the proof of Theorem 1.14, as it appears in [11]. The first reason we do not regard this theorem as a black box is that condition (3) of Theorem 1.14 does not appear in [11] explicitly. It is, however, a transparent corollary of their construction. The second reason is that we will need to generalize it a bit to force the approximating surfaces to be entirely inside or outside X; see Corollary 2.11. Let X be an (ε, R)–Reifenberg flat set, and let Ω be the domain bounded by X (recall that by [18] X is a topological manifold, and so by Jordan's separation theorem there exists such a domain).

Sketch of proof of Theorem 1.14 [11] The approximating surfaces X^r are constructed as level sets of mollifications at scale r of the characteristic function of the domain bounded by X. More precisely, choose a smooth radial bump function $\widetilde{\phi} \in C_0^\infty(B(0,1))$ such that

(i) $0 \le \widetilde{\phi} \le 1$

(ii) $\widetilde{\phi}|_{B(0,1/2)} = 1$

and let c_1 be a constant such $\phi = c_1\widetilde{\phi}$ satisfies $\int \phi(x)\,dx = 1$. Define

$$\phi_r(x) = \frac{1}{r^n}\phi\Big(\frac{x}{r}\Big). \tag{2.1}$$

Now, since X is in particular a topological manifold (as was shown in [18]), by the Jordan separation theorem we can set χ to be the characteristic function of the domain Ω bounded by X and let $\chi_r(x) = \phi_r \star \chi(x)$. The approximating sets X^r are defined to be

$$X^r = \chi_r^{-1}\big(\tfrac{1}{2}\big). \tag{2.2}$$

(1) To show that X^r satisfies the first condition of Theorem 1.14 for $r < R/4$, choose $x^r \in X^r$, $x \in X$ which is closest to x^r, T a hyperplane that satisfies the Reifenberg condition at x at scale $4r$ and ν a normal to T such that

$$\Omega \cap B(x,4r) \supseteq \{y : \langle y-x,\nu\rangle \ge 4r\varepsilon\} \cap B(x,4r), \tag{2.3}$$

$$\Omega^c \cap B(x,4r) \supseteq \{y : \langle y-x,\nu\rangle \le -4r\varepsilon\} \cap B(x,4r). \tag{2.4}$$

By the definition of X^r, it is clear that $B(x^r,r) \subseteq B(x,4r)$ and (2.3), (2.4) imply that $|\langle x^r - x,\nu\rangle| \le 4r\varepsilon$, else $\chi_r(x^r)$ would be too low/high. This also implies that $d(x^r,X) < 8r\varepsilon$. Similarly for any $x \in X$ and T, ν as above, $\chi_r(x+4r\varepsilon\nu) > \frac{1}{2}$ while $\chi_r(x-4r\varepsilon\nu) < \frac{1}{2}$, so by the intermediate value theorem we also have $d(x,X^r) \le 4r\varepsilon$. Thus,

$$d_H(X^r,X) \le 8\varepsilon r. \tag{2.5}$$

(2) To show that X^r satisfies the second condition of Theorem 1.14, choose x, T, ν as above and let $\{e_1,\dots,e_n\}$ be an orthonormal basis for T. Setting $S = \{y : |\langle z-x,\nu\rangle| \le 4r\varepsilon\}$ and taking $y \in B(x,2r) \cap S$ one can compute, splitting the convolution integrals according to S and the two half spaces from (2.3) and (2.4) (see [11, Lemma 2.2]), that if ε is sufficiently small,

$$\begin{aligned} &\langle \nabla\chi_r(y),\nu\rangle \ge \frac{c_1}{r}, \qquad |\langle \nabla\chi_r(y),e_i\rangle| \le \frac{c_2\varepsilon}{r},\\ &|\mathrm{Hess}\,\chi_r(\nu,\nu)| \le \frac{c_2}{r^2}, \quad |\mathrm{Hess}\,\chi_r(e_i,\nu)| \le \frac{c_2\varepsilon}{r^2}, \quad |\mathrm{Hess}\,\chi_r(e_i,e_j)| \le \frac{c_2\varepsilon}{r^2}, \end{aligned} \tag{2.6}$$

for some constants c_1, c_2. Defining $C(x,r) = \{y : |T(y-x)| \le r\}$, the first inequality of (2.6) together with what was done in (1) shows that in $B(x,2r) \cap C(x,r)$, the set X^r is a graph of a function u over T (thinking of T as an affine hyperplane passing through x), with

$$|u| \leq 4r\varepsilon. \tag{2.7}$$

The other inequalities of (2.6) further imply that

$$|A| \leq \frac{c_3 \varepsilon}{r}. \tag{2.8}$$

(3) Note that condition (2.8) together with X^r being a graph over T in $B(x,2r) \cap C(x,r)$ imply that for $\varepsilon > 0$ small enough, $\mathrm{inj}(X^r) \geq r/2$, where inj denotes the injectivity radius of the normal exponential map. Now, considering the scale $r/4$, if ε is small enough, the same T as the one for scale r will work. By using parts (1) and (2) for $X^{r/4}$ and in particular, it being a graph over T at scale $r/4$, one sees by interpolating (2.7) and (2.8) that $X^{r/4}$ is a graph of a function u over X^r with $|u(y)| \leq 6r\varepsilon$ for every $y \in X^r$ (see also Lemma 4.16). To put it differently, there is a homeomorphism $\phi_r \colon X^r \to X^{r/4}$ with

$$d(\phi_r(y), y) \leq 6r\varepsilon. \tag{2.9}$$

Composing those maps we see that there is a homeomorphism $\phi_r^k \colon X^r \to X^{r/4^k}$ with

$$d(\phi_r^k(y), y) \leq 12r\varepsilon. \tag{2.10}$$

Now, note that the analysis in (1), (2) shows that for every $x \in X$ one can write $X^r \cap B(x,r) = G \cup B$, where G is connected and $B \cap B(x,(1-8\varepsilon)r) = \varnothing$. Thus, using (2.10) we see that one can write $X^{r/4^k} \cap B(x,r) = G \cup B$, where G is connected and $B \cap B(x,(1-20\varepsilon)r) = \varnothing$.

This conclude the (sketch of the) proof of the theorem. □

The following slight strengthening of Theorem 1.14 will play a role in the final stage of our argument for uniqueness.

Corollary 2.11 *There exist some constants c_1, $c_2 > 0$ such that if X is (ε, R)–Reifenberg flat for $0 < \varepsilon < \varepsilon_0$ and bounds a domain Ω, then there exists a family of surfaces $(X_\pm^r)_{0<r<R/4}$ such that:*

(1) $d_H(X_\pm^r, X) \leq c_1 \varepsilon r$.

(2) $|A(x)| \leq c_2 \varepsilon / r$ *for every* $x \in X_\pm^r$.

(3) *For every* $x \in X$, $r \in (0, R/4)$ *and* $s \in (r, R/4)$, $B(x,s) \cap X^r_\pm$ *can be decomposed as*

$$B(x,s) \cap X^r_\pm = G \cup B, \tag{2.12}$$

where G *is connected and* $B \cap B(x,(1-40\varepsilon)s) = \varnothing$.

(4) $X^r_- \subseteq \Omega$ *and* $X^r_+ \subseteq \bar{\Omega}^c$.

Proof Let N be the exterior unit normal to the X^r from Theorem 1.14. When ε is small enough, conditions (1) and (3) of Theorem 1.14 imply that X^r has a tubular neighborhood of thickness $r/4$ (see Lemma 4.4 for a proof in an analogous situation). Moreover, computing the third partials for χ_r from the above theorem shows that the X^r also satisfy the estimate

$$|\nabla A| \le c_3 \frac{\varepsilon}{r^2}. \tag{2.13}$$

Thus, considering

$$X^r_\pm = \{x \pm 10\varepsilon r N(x) : x \in X^r\} \tag{2.14}$$

we see that $X^r_\pm$ satisfy conditions (1)–(3). Let $x \in X$ and let T, ν be as in (1) in the proof of Theorem 1.14. Then since for every $y \in B(x,r) \cap X^r$ the tangent space to X^r at y is almost parallel to T by (2) in the proof, we see by part (1) of the proof that one of the $X^r_\pm \cap B(x,r)$ will lie in Ω and the other will lie in $\bar{\Omega}^c$. Thus, $X^r_\pm \cap X = \varnothing$, with one of the $X^r_\pm$ in each component. □

We will now recall the gradient estimate of Ecker and Huisken. Let X be a hypersurface in $\mathbb{R}^{n+1}$ such that in $B(x,r)$ it can be parametrized locally as a graph of a function u over the first n coordinates. If $\nu(x)$ is the unit normal to X at x, the gradient function $v\colon X \cap B(x,r) \to \mathbb{R}_+$ is defined to be

$$v(x) = \frac{1}{\nu \cdot e_{n+1}} = \sqrt{1 + |\nabla u|^2}. \tag{2.15}$$

Theorem 2.16 (Ecker–Huisken gradient estimate [6], see also [4]) *Let* $(X_t)_{t \in (t_0,t_1)}$ *be a solution for the mean curvature flow in* $\mathbb{R}^{n+1}$ *and suppose that in the ball* $B(x_0,r)$, X_{t_0} *is locally a graph over the first* n *coordinates. Then* X_t *is locally a graph over the first* n *co-ordinates in* $B(x_0, \sqrt{r^2 - 2n(t-t_0)})$, *and*

$$\left(1 - \frac{|x-x_0|^2 + 2n(t-t_0)}{r^2}\right) v(x,t) \le \max_{x \in B(x_0,r) \cap X_{t_0}} v(x,t_0). \tag{2.17}$$ □

Controlling the gradient in a space-time neighborhood allows one to control the second fundamental form (and its derivatives) if one allows the hypersurface to evolve a little bit.

Theorem 2.18 ([6], see also [4]) *There exists a universal constant $C = C(n)$ such that if $v \leq v_0$ in $B(x,r) \times (t_0, t_0 + r^2)$ then*

$$|A|^2 \leq C\frac{v_0^4}{r^2}, \quad |\nabla A|^2 \leq C\frac{v_0^4}{r^4}, \quad |\partial_t A|^2 \leq C\frac{v_0^4}{r^6}, \tag{2.19}$$

in $B(x, r/2) \times (t_0 + 3r^2/4, t_0 + r^2)$. □

The above classical curvature estimate of Ecker and Huisken will not be good enough for our purposes. We will be interested in considering the case where $|\nabla u|$ is very small, in which the above estimate is not very useful. Indeed, since $v = \sqrt{1+|\nabla u|^2}$ the above estimate will yield that $|A|^2$ is, up to a constant, smaller than $1/r^2$, even if $|\nabla u|$ were very small (in the entire neighborhood). When $|\nabla u| = 0$, however, $|A| = 0$ so there is clearly a big gap to be filled in that regime. The entire Section 5, culminating in the proof of Theorem 3.1, will deal with proving a curvature estimate suitable for small $|\nabla u|$ (and small initial $|u|$).

Controlling the second fundamental form at a certain time allows one to control it for a little bit.

Lemma 2.20 *For every $\delta > 0$ there exists $c = c(\delta) > 0$ such that if X_t flows by mean curvature and*

$$|A(0)| \leq \alpha, \tag{2.21}$$

then

$$|A(t)| \leq (1+\delta)\alpha \tag{2.22}$$

for $0 \leq t \leq c/\alpha^2$.

Proof Since under the mean curvature flow

$$\frac{d}{dt}|A|^2 = \Delta|A|^2 - 2|\nabla A|^2 + 2|A|^4 \leq \Delta|A|^2 + 2|A|^4 \tag{2.23}$$

(see [17], for instance), by the maximum principle we obtain that

$$|A(t)| \leq \frac{\alpha}{\sqrt{1-2\alpha^2 t}} \tag{2.24}$$

for as long as the denominator does not vanish. The result follows. □

We will conclude this section with the following simple geometric lemma for the interpolation of Hausdorff bounds and curvature bounds.

Lemma 2.25 (interpolation) *For every $\delta > 0$ and $\alpha > 0$ there exists $\beta_0(\alpha,\delta) > 0$ such that for every $\beta < \beta_0$ the following holds. Assume $p \in X$ where X is a hypersurface such that in $B(p,r)$ we have*

(1) $|A| \le \alpha/r$,

(2) $d_H(P \cap B^n(p,r), X \cap B^n(p,r)) \le \beta r$ *for* $P = \mathrm{span}\{e_1,\dots,e_n\}$.

Then inside $B(p,r) \cap (B^n(p,(1-\delta)r) \times \mathbb{R})$, the connected component of p is a graph of a function u over P and we have the estimate

$$|\nabla u| \le \frac{\sqrt{2\beta\alpha - \alpha^2\beta^2}}{1-\alpha\beta} \cong \sqrt{2\beta\alpha}, \tag{2.26}$$

and $|u| \le \beta r$.

Proof Assume without loss of generality that $r = 1$ and $p = 0$ and denote $Q = P^{\perp}$ and $C_{\delta,\beta} = B^n(0,(1-\delta)) \times [-\beta,\beta]$. For β sufficiently small, $C_{\delta/4,\beta} \subseteq B(0,1)$ and $\alpha\beta < 1$. Now, let $x \in C_{\delta/2,\beta}$ and let $\gamma(t)$ be a unit speed geodesic in X with $\gamma(0) = 0$. We may assume without loss of generality, by possibly changing the parametrization according to $t \mapsto -t$, that $\langle \gamma'(0), e_{n+1}\rangle = \max_{v\in Q,\,\|v\|=1}\langle \gamma'(0), v\rangle$ and that $x_{n+1}(\gamma(t)) \ge 0$. Letting $f(t) = x_{n+1}(\gamma(t))$ we find $f'(t) = \langle \gamma'(t), e_{n+1}\rangle$ and

$$f''(t) = \langle \gamma''(t), e_{n+1}\rangle = \langle \gamma''(t), e_{n+1} - \langle \gamma'(t), e_{n+1}\rangle \gamma'(t)\rangle \ge -\alpha\sqrt{1 - f'(t)^2}.$$

The equality case of the above ODE for $f'(t)$ corresponds to a circle of radius $1/\alpha$. Letting $\mu(t)\colon \mathbb{R} \to \mathbb{R}^2$ be a clockwise and unit speed parametrized circle of radius $1/\alpha$ with $\mu(0) = (0,0)$ and $\langle \mu'(0), e_2\rangle = f'(0)$, we see that as long as $x_2(\mu(t))$ is increasing, and as long as $\gamma(t) \in C_{\delta/4,\beta}$, one has $x_{n+1}(\gamma(t)) \ge x_2(\mu(t))$. For β sufficiently small (depending on α and δ) $x_2(\mu(t))$ will reach its maximum at time $0 < T < \delta/4$, so the extra condition $\gamma(t) \in C_{\delta/4,\beta}$ is redundant. Thus $x_2(\mu(t)) \le x_{n+1}(\gamma(t)) \le \beta$, and an easy calculation for circles in the plane gives the bound

$$\tan \angle(T_x X, P) \le \frac{\sqrt{2\beta\alpha - \alpha^2\beta^2}}{1-\alpha\beta} \tag{2.27}$$

for β sufficiently small.

What remains to be shown is that the connected component of p is indeed a graph. Assume there exist $x_1, x_2 \in X \cap C_{\delta,\beta}$ with $x_1 \ne x_2$ but $P(x_1) = P(x_2)$, where we use P both for the hyperplane and for the projection operator to it. Observe that by (2.27), $X \cap \overline{C_{\delta,\beta}}$ is a submanifold with boundary. Let $\gamma\colon [0,a] \to X \cap \overline{C_{\delta,\beta}}$ be a minimizing geodesic between x_1 and x_2. Such a geodesic is always C^1 and is smooth for as long as $\gamma(t)$ is away from the boundary. For such t however $\|P(\gamma''(t))\| \le \sqrt{3\alpha\beta}\alpha$ by

(2.27) and so for β sufficiently small, as $\gamma'(0)$ is almost parallel to P, the projection $P(\gamma(t))$ is almost a straight line until it hits the boundary (at some $t < 4$). Since $\gamma(t)$ is C^1, and intersects the boundary with an exterior normal component, this is a contradiction.

To see that for every $y \in B^n(0, 1-\delta)$ there is some $x \in X$ with $P(x) = y$, note that by the Hausdorff condition, we can find $\bar{x} \in X \cap B(0, (1-\delta/2))$ with $d(\bar{x}, y) \leq \beta$ (when β is small). Taking $\bar{y} = P(\bar{x})$ we see, again, by (2.27) for $\bar{x}$, and the fact that the curvature scale $1/\alpha$ is far bigger than β, that there will exist a point over y as well. □

3 Uniform estimates and existence of smooth evolution

The purpose of this section is to prove Theorem 1.16 which will immediately imply the existence part of Theorem 1.11, ie that at least one weak set flow (see Definition 1.8) of an (ε, R)–Reifenberg flat set is smooth whenever $\varepsilon > 0$ is small enough. In Section 3.1 we will state the interior estimate we will employ and remark on why it is plausible. The proof will be deferred to Section 5. In Section 3.2 we will perform the iteration step that was described in the introduction. In Section 3.3 we will prove Theorem 1.16 and derive the existence of a smooth weak set flow.

3.1 Interior estimate for graphical mean curvature flow

In order to implement the iteration, it would be the most comfortable to work with the following interior estimate for mean curvature flow, which will be proved in Section 5.

Theorem 3.1 (main estimate) *There exists some $c \geq 1$ such that for every $\delta > 0$ and $M > 0$, there exist positive $\tau_0 = \tau_0(M, \delta) \ll 1$ and $\lambda_0 = \lambda_0(M, \delta) \ll 1$ such that for every $0 < \lambda < \lambda_0$ there exists some $\varepsilon_0 = \varepsilon_0(\delta, M, \lambda)$ such that for every $\beta > 0$, $0 < \tau < \tau_0$ and $\varepsilon < \varepsilon_0$, the following statement holds.*

If $u\colon B(p, r) \times [0, \tau r^2] \to \mathbb{R}$ is a graph moving by mean curvature such that

(1) *for every $(x, t) \in B(p, r) \times [0, \tau r^2]$,*

$$|\nabla u(x, t)| \leq M\varepsilon, \tag{3.2}$$

(2) *for every $(x, t) \in B(p, r) \times [0, \lambda\tau r^2]$,*

$$|\nabla u(x, t)| \leq \varepsilon, \tag{3.3}$$

(3) *for every* $(x,t) \in B(p,r) \times [0, \tau r^2]$,

$$|u(x,t)| \leq M^2 \beta r, \tag{3.4}$$

(4) *for every* $(x,t) \in B(p,r) \times [0, \lambda \tau r^2]$,

$$|u(x,t)| \leq \beta r, \tag{3.5}$$

then for every $(x,t) \in B(p,(1-\delta)r) \times [0, \tau r^2]$,

$$|A(p,t)| \leq (c+\delta)\frac{\varepsilon}{\sqrt{t}}, \tag{3.6}$$

$$|A(p,t)| \leq c\frac{\beta r}{t} + \delta\frac{\varepsilon}{\sqrt{t}}. \tag{3.7}$$

Remark 3.8 The statement of the above estimate may appear somewhat confusing. We hope that the following trailer to Section 5 will make it appear more plausible. There is nothing wrong with fixing $r = 1$.

(1) When $|\nabla u|$ is very small, we are dealing essentially with the heat equation, as the non-linearity is very small. The estimate

$$|\nabla u(x,t)| \leq \sqrt{\pi}\,\frac{\mathrm{Vol}(S^{n-1})}{\mathrm{Vol}(S^n)}\,\frac{\|u(-,0)\|_\infty}{\sqrt{t}} \tag{3.9}$$

is what one gets when estimating the first derivative to the physical solution of the heat equation in the full space at time t in terms of the sup norm of the initial time slice. Estimate (3.6) reflects that fact, as the derivative of a solution to the heat equation satisfies the heat equation itself. The first term in estimate (3.7) reflects a similar bound on the second derivative.

(2) Since we are dealing with a domain with boundary, one cannot expect to get the same estimate as for the entire space. However, for τ very small, from the perspective of the point $(0,\tau)$, the 0 time slice — a ball of radius 1 — will look like the entire space. Therefore we should get a constant very close to $\sqrt{\pi}\,\mathrm{Vol}(S^{n-1})/\mathrm{Vol}(S^n)$, as expressed in (3.6). The parameter M in the estimate makes sure that the contribution from the boundary doesn't change the result by too much. A similar reasoning leads to the first term of (3.7).

(3) Since the equation is non-linear, there should also be a term coming from the non-linearity. Since the non-linearity is quadratic in the gradient, when the bound $|\nabla u|$ is small enough, it will contribute as little as any small fraction of that gradient bound.

3.2 Iteration

Before we dive into the iteration lemma, we need the following two similar calculations. The first allows one to extend the curvature and Hausdorff bounds for a short time.

Lemma 3.10 *For every $\delta > 0$ there exists some $\widetilde{c} = \widetilde{c}(n, \delta) > 0$ with the following property. Assume α, β are such that $\alpha\beta < 1$ and Y is a smooth hypersurface such that for some $r > 0$:*

(1) *For every $y \in Y$ there exists a hyperplane P_y such that*

$$d_H(B(y, r) \cap P_y, B(y, r) \cap Y) \leq \beta r. \tag{3.11}$$

(2) *For every $y \in Y$,*

$$|A(y)| \leq \frac{\alpha}{r}. \tag{3.12}$$

Then, denoting by Y_t the mean curvature flow emanating from Y and writing $E = 2\alpha\beta$, for every $0 \leq t \leq (\widetilde{c}/\alpha^2)Er^2$ we have:

(1) $|A(t)| \leq (1 + \delta)\dfrac{\alpha}{r}$.

(2) $d_H(B(p, (1 - 10^{-5}\delta\beta)r) \cap P, B(p, (1 - 10^{-5}\delta\beta)r) \cap Y_t) \leq (1 + \delta)\beta r$.

Proof Assume $r = 1$. Using the global curvature bound α, by Lemma 2.20 we can find $c_1 = c_1(\delta)$ such that if

$$0 \leq t \leq c_1 \frac{1}{\alpha^2}, \tag{3.13}$$

we have

$$|A(t)| \leq \left(1 + \frac{\delta}{4}\right)\alpha. \tag{3.14}$$

Using the curvature bound to estimate the motion, we see that there exists $c_2(n, \delta) < c_1$ such that for $t \leq c_2\beta/\alpha$ the surface moves by at most $10^{-6}\delta\beta$. Since $\alpha\beta < 1$, for every

$$0 \leq t \leq c_2 \min\left\{\frac{\beta}{\alpha}, \frac{1}{\alpha^2}\right\} = c_2 \frac{1}{\alpha^2} 2\alpha\beta \tag{3.15}$$

we have:

(1) $|A(t)| \leq (1 + \delta)\alpha$.

(2) $d_H\big(B(p, (1 - 10^{-5}\delta\beta)) \cap P, B(p, (1 - 10^{-5}\delta\beta)) \cap Y_t\big) \leq (1 + \delta)\beta$. □

See also Lemma 3.44 for how motion bounds are used to obtain Reifenberg flatness at a certain scale.

The second calculation allows one to extend gradient estimates for longer times, gaining a definite large multiplicative error.

Lemma 3.16 *For every $\delta > 0$ and $\alpha > 1$ there exist an $M = M(n,\alpha) > 0$ and a $\beta_0 = \beta_0(n,\alpha,\delta) > 0$ such that for every $0 < \beta < \beta_0$, if we set $E = 2\alpha\beta$ and $T = Er^2$, the following holds. Assume that $(Y_t)_{t\in[0,T]}$ is a mean curvature flow with*

$$|A(0)| \leq \frac{\alpha}{r} \tag{3.17}$$

such that for every $y \in Y_0$ there exists a hyperplane P_y such that

$$d_H(B(y,r)\cap P_y, B(y,r)\cap Y_0) \leq \beta r. \tag{3.18}$$

Then setting $(\phi_t)_{t\in[0,T]}$ to be the parametrized flow starting from Y (so $Y_t = \phi_t(Y)$) and assuming without loss of generality that $y = 0$ and $P_y = \operatorname{span}\{e_1,\dots,e_n\}$, the connected components of $\phi_t(y)$ in $Y_t \cap B(y,r)\cap(B^n(y,(1-\delta)r)\times\mathbb{R})$ is a graph of a function u: $B^n(y,(1-\delta)r)\times[0,T]\to\mathbb{R}$ flowing by mean curvature. Moreover, we have the estimates

$$|\nabla u(z,t)| \leq M\sqrt{E}, \quad |u(z,t)| \leq M^2\beta r, \tag{3.19}$$

for every $z \in B^n(y,(1-\delta)r)$ and $0 \leq t \leq T$. Letting $\tilde{c}$ be the constant from Lemma 3.10, for $z \in B^n(y,(1-\delta)r)$ and $0 \leq t \leq \tilde{c}/\alpha^2 T$ we have the better estimates

$$|\nabla u(z,t)| \leq 2\sqrt{E}, \quad |u(z,t)| \leq 2\beta r. \tag{3.20}$$

Proof Assume without loss of generality that $r = 1$. Since $T = 2\alpha\beta$, according to Lemma 2.20 we have for $\beta < \beta_0(\alpha)$ that

$$|A(t)| \leq 2\alpha \tag{3.21}$$

for every $0 \leq t \leq T$. Thus, we can bound the motion to obtain

$$d(\phi_t(z),\phi_0(z)) \leq 4\sqrt{n}\alpha^2\beta \tag{3.22}$$

for every $0 \leq t \leq T$ and $z \in Y$. Now, when $\beta < \beta_0$ this displacement is very small so as before (see also Lemma 3.44) we get $5\sqrt{n}\alpha^2\beta$ closeness to planes on a slightly smaller ball. We can now use Lemma 2.25 to obtain graphicality for every $0 \leq t \leq T$, as well as the estimates

$$|\nabla u(z,t)| \leq 2\sqrt{2\alpha\cdot 5\sqrt{n}\,\alpha^2\beta} = M(\alpha)\sqrt{E}, \tag{3.23}$$

$$|u(z,t)| \leq 5\sqrt{n}\,\alpha^2\beta \leq M(\alpha)^2\beta. \tag{3.24}$$

The second estimate now follows from Lemma 3.10. □

The main technical point of the argument is carried out in the following lemma.

Lemma 3.25 *For every $\Lambda > 0$ there exist some $\alpha, \beta > 0$ and a "scale change" parameter $\theta > 20$ such that setting $E = 2\alpha\beta$ and $T = Er^2$, the following holds.*

Assume Y is a smooth hypersurface such that for some $r > 0$,

(1) *for every $y \in Y$ there exists a hyperplane P_y such that*

$$d_H(B(y,r) \cap P_y, B(y,r) \cap Y) \le \beta r, \tag{3.26}$$

(2) *for every $y \in Y$,*

$$|A(y)| \le \frac{\alpha}{r}. \tag{3.27}$$

Then Y flows smoothly by mean curvature for time T. Moreover, setting $(\phi_t)_{t\in[0,T]}$ to be the parametrized flow starting from Y and letting Y_t be the flow of hypersurfaces (so $Y_t = \phi_t(Y)$), we have:

A. $d(\phi_0(y), \phi_T(y)) \le \frac{1}{20}\beta(\theta r)$ *for every $y \in Y$, and in particular,*

$$d_H(Y, Y_T) \le \tfrac{1}{20}\beta(\theta r).$$

B. *For every $y \in Y_T$,*

$$|A(y)| \le \frac{\alpha}{\theta r}. \tag{3.28}$$

Moreover, we can choose the parameters in such a way that the following relations hold:

I. $2\theta E < \min\{10^{-6}, \Lambda^{-1}\}$.

II. $\alpha > 100$.

III. $\alpha^3\beta < \frac{1}{640}$.

Proof We want to set things up in a way that will allow us to use Theorem 3.1. Assume without loss of generality that $r = 1$, fix $p \in Y$ and assume without loss of generality that $P_p = \operatorname{span}\{e_1, \dots, e_n\}$.

Step 1 (choice of parameters)

(1) (choosing δ) Take $\delta = 1/(1600\sqrt{n}c)$.

(2) (choosing α) Choose $\alpha = 800\sqrt{n}c^2$, where c is from Theorem 3.1.

(3) (choosing M) Choose $M = M(\alpha)$ from Lemma 3.16.

(4) (choosing τ_0 and λ_0) Having fixed M and δ we can choose $\tau_0(M,\delta)$ and $\lambda_0(M,\delta)$ as in Theorem 3.1.

(5) (choosing λ) Let $\tilde{c}(n,\delta)$ be the constant from Lemma 3.10, and choose

$$\lambda = \lambda(\delta,\alpha,\lambda_0) = \min\left\{\frac{\tilde{c}}{\alpha^2},\lambda_0\right\}. \tag{3.29}$$

(6) (choosing $\tau = T$) Assuming E is small enough, T can be made to be smaller than τ_0 and so we let $\tau = T$. Note that E is still free, so we have just expressed our wish to set $\tau = E$ without really fixing τ.

(7) (choosing ε_0) Choose $\varepsilon_0 = \varepsilon_0(M,\delta,\lambda)$ as in Theorem 3.1.

(8) (choosing θ) Set $\theta = 320\,000nc^3 > 20$.

(9) (choosing β) Choose β such that the following conditions hold:

  (a) $\beta < \beta_0(\alpha,\delta/8)$ of Lemma 2.25.
  (b) $\beta < \beta_0(n,\alpha,\delta/4)$ of Lemma 3.16.
  (c) $\sqrt{E} = \sqrt{2\alpha\beta} < \varepsilon_0/2$ (so that Theorem 3.1 will hold, see Step 5).
  (d) $2\theta E = 4\alpha\beta\theta < \min\{10^{-6},\Lambda^{-1}\}$ (to comply with the statement).
  (e) $\alpha^3\beta < \frac{1}{640}$ (to comply with the statement).

As all those conditions want β to be small, they can be satisfied simultaneously.

(10) (choosing ε) We finally set $\varepsilon = \sqrt{E}$.

Step 2 (initial bounds) By the choice of β, we know that $(\alpha,\beta,\delta/8)$ satisfy the conditions of Lemma 2.25. Therefore, we get that the connected component of p in $Y \cap (B^n(p,(1-\delta/4)) \times [-\beta,\beta])$ is a graph of a function u over $B^n(p,(1-\delta/4))$ with

$$|\nabla u| \le \tfrac{3}{2}\sqrt{2\beta\alpha} = \tfrac{3}{2}\sqrt{E}. \tag{3.30}$$

Step 3 (obtaining bounds for long positive times) If $\beta < \beta_0(n,\alpha,\delta/4)$, then by Lemma 3.16, u remains a function on $B^n(p,(1-\delta/4)) \times [0,T]$, where the following estimates hold:

$$|\nabla u(y,t)| \le M\sqrt{E}, \quad |u(y,t)| \le M^2\beta. \tag{3.31}$$

Step 4 (obtaining bounds for small positive times) According to the second part of Lemma 3.16, for $t \in [0,\lambda T]$, Y_t satisfies the conditions of Lemma 2.25 and so we obtain that for those times, for $y \in B^n(p,(1-\delta/2))$,

$$|\nabla u(y,t)| \le 2\sqrt{E}, \quad |u(y,t)| \le 2\beta. \tag{3.32}$$

Step 5 (applying Theorem 3.1, estimating the curvature) By the first estimate of Theorem 3.1 we have that for every $t \in [0, E]$ and $y \in Y_t$,

$$|A(y,t)| \leq \frac{4c\sqrt{E}}{\sqrt{t}}, \tag{3.33}$$

and by the second estimate of Theorem 3.1 and by our choice of α and δ, at the final time T we have

$$|A(y,T)| \leq \frac{2c\beta}{E} + 2\delta\frac{\sqrt{E}}{\sqrt{E}} = \frac{1}{400\sqrt{n}c} := \alpha'. \tag{3.34}$$

Step 6 (estimating the motion) The curvature bound (3.33) implies the mean curvature bound

$$|H(y,t)| \leq \frac{4c\sqrt{n}\sqrt{E}}{\sqrt{t}}, \tag{3.35}$$

which can be integrated to obtain the motion bound

$$d(\phi_0(y), \phi_T(y)) \leq 8c\sqrt{n}E := \beta'. \tag{3.36}$$

Step 7 (conclusion) Note that by our choice of θ and by (3.34) and (3.36), we have

$$|A(y,T)| \leq \alpha' = \frac{\alpha}{\theta}, \tag{3.37}$$

$$\alpha'\beta' = \tfrac{1}{50}E < \tfrac{1}{20}\alpha\beta, \tag{3.38}$$

implying

$$d(\phi_0(y), \phi_T(y)) \leq \beta' \leq \tfrac{1}{20}\beta\theta. \tag{3.39}$$

□

Remark 3.40 Note that the motion estimate in Step 6 illustrates why we needed estimate (3.7) of Theorem 3.1. Using the optimal form of estimate (3.6) of Theorem 3.1 up to time T, we obtain that, in the language of Step 6, for $t \in [0, T]$,

$$|A(y,t)| \leq \sqrt{\pi}\frac{\mathrm{Vol}(S^{n-1})}{\mathrm{Vol}(S^n)}\frac{\sqrt{E}}{\sqrt{t}}, \tag{3.41}$$

$$|H(y,t)| \leq \sqrt{n}\sqrt{\pi}\frac{\mathrm{Vol}(S^{n-1})}{\mathrm{Vol}(S^n)}\frac{\sqrt{E}}{\sqrt{t}}, \tag{3.42}$$

and thus by integration,

$$d_H(Y_0, Y_T) \leq 2\sqrt{n}\sqrt{\pi}\frac{\mathrm{Vol}(S^{n-1})}{\mathrm{Vol}(S^n)}E. \tag{3.43}$$

Thus, as $\mathrm{Vol}(S^{n-1})/\mathrm{Vol}(S^n) \approx \sqrt{n}$, for the corresponding α' and β' we will necessarily have $\alpha'\beta' > \alpha\beta$ which will preclude us from iterating, this time with Y_T at scale θ (see also Section 3.3).

3.3 Uniform estimates

Before proving Theorem 1.16 we need the following elementary lemma about restricting the triangle inequality to balls.

Lemma 3.44 *Suppose X is a $(\beta/10, R)$–Reifenberg set and let Y be a compact set such that for some $r < R$,*

$$d_H(X,Y) \le \tfrac{1}{10}\beta r. \tag{3.45}$$

Then for every $y_0 \in Y$ there exists a hyperplane P passing through y such that

$$d_H(B(y_0,(1-\beta/5)r)\cap P, B(y_0,(1-\beta/5)r)\cap Y) \le \tfrac{3}{5}\beta r. \tag{3.46}$$

Proof Take $x_0 \in X$ with

$$d(x_0,y_0) \le \tfrac{1}{10}\beta r \tag{3.47}$$

and choose a plane $\overline{P}$ passing through x_0 such that

$$d_H(B(x_0,r)\cap \overline{P}, B(x_0,r)\cap X) \le \tfrac{1}{10}\beta r. \tag{3.48}$$

Let P be the plane parallel to $\overline{P}$ passing through y_0. Now, taking any $y \in B(y_0,(1-\beta/5)r)$ and x with

$$d(x,y) \le \tfrac{1}{10}\beta r, \tag{3.49}$$

we see that $x \in B(x_0,r)$ and so there is a point $\overline{p} \in \overline{P} \cap B(x_0,r)$ with

$$d(\overline{p},y) \le \tfrac{1}{5}\beta r. \tag{3.50}$$

Moving a bit inward, this implies that there exists a $\overline{p}_1 \in \overline{P} \cap B(x_0,(1-3\beta/10)r) \subseteq B(y_0,(1-\beta/5)r)$ with

$$d(\overline{p}_1,y) \le \tfrac{5}{10}\beta r, \tag{3.51}$$

and so taking $p \in P$ closest to $\overline{p}_1$ we obtain $p \in P \cap B(y_0,(1-\beta/5)r)$ with

$$d(p,y) \le \tfrac{3}{5}\beta r. \tag{3.52}$$

In the other direction, given $p \in P \cap B(y_0, (1-\beta/5)r)$ and $\overline{p}_1 \in \overline{P} \cap B(x_0, (1-\beta/5)r)$ closest to p, we can move inward to find $\overline{p} \in B(x_0, (1-5\beta/10)r) \cap \overline{P}$ with

$$d(\overline{p}, p) \le \tfrac{2}{5}\beta r. \tag{3.53}$$

Choosing

$$x \in X \quad \text{with} \quad d(x, \overline{p}) \le \tfrac{1}{10}\beta r, \tag{3.54}$$

$$y \in Y \quad \text{with} \quad d(x, y) \le \tfrac{1}{10}\beta r, \tag{3.55}$$

we get

$$d(p, y) \le \tfrac{3}{5}\beta r, \tag{3.56}$$

and $y \in B(x_0, (1-3\beta/10)r) \subseteq B(y_0, (1-\beta/5)r)$. □

We are now in a position to prove Theorem 1.16.

Proof of Theorem 1.16 Fix α, β, θ, E as in Lemma 3.25, let $r < R$ and recall that $\theta E < 10^{-6}$. Choose $\varepsilon > 0$ sufficiently small that:

(1) $c_1\varepsilon < \frac{1}{10}\beta$.

(2) $\varepsilon < \frac{1}{10}\beta$.

(3) $c_2\varepsilon < \alpha$.

Here c_1, c_2 are the constants from Theorem 1.14. Letting $Y = X^r$, with the above choice of ε we have:

(A_r) For every $y \in Y$,

$$|A(y)| \le \alpha/r \le \frac{\alpha}{(1-\beta/5)r}.$$

(B_r) $d_H(Y, X) \le \frac{1}{10}\beta r$.

(C_r) By Lemma 3.44, for every $y \in Y$ there exists a plane P such that

$$d_H(B(y, (1-\beta/5)r) \cap P, B(y, (1-\beta/5)r) \cap Y) \le \tfrac{3}{5}\beta r \le \beta(1-\beta/5)r. \tag{3.57}$$

Conditions (A_r) and (C_r) allow us to apply Lemma 3.25, so we can let Y flow smoothly for time $T = E(1-\beta/5)^2 r^2$. Moreover, we have:

($\mathrm{A}_{\theta r}$) For every $y \in Y_T$,

$$|A(y)| \le \frac{\alpha}{(1-\beta/5)(\theta r)}. \tag{3.58}$$

($\mathrm{B}_{\theta r}$) For every $y \in Y$,

$$d(\phi_0(y), \phi_T(y)) \le \tfrac{1}{20}\beta(\theta(1-\beta/5)r), \tag{3.59}$$

so by condition $B(r)$ and the fact that $\theta > 20$,

$$\begin{aligned} d_H(X, Y_T) \le \tfrac{1}{10}\beta r + \tfrac{1}{20}\beta(\theta(1-\beta/5)r) &\le \tfrac{1}{20}\beta(\theta r) + \tfrac{1}{20}\beta(\theta(1-\beta/5)r) \\ &< \tfrac{1}{10}\beta(\theta r). \end{aligned} \tag{3.60}$$

($\mathrm{C}_{\theta r}$) If $\theta r < R$, we can use Lemma 3.44 to obtain

$$\begin{aligned} d_H(B(y,(1-\beta/5)\theta r) \cap P, B(y,(1-\beta/5)\theta r) \cap Y_T) &\le \tfrac{3}{5}\beta(\theta r) \\ &\le \beta(1-\beta/5)(\theta r). \end{aligned} \tag{3.61}$$

Thus, we conclude that Y_T satisfies the same conditions as Y, this time with respect to the scale θr, so we can restart the process with Y_T instead of Y at scale θr instead of r and iterate. The iterations will be performed at times

$$t_k = \left(1 - \tfrac{1}{5}\beta\right)^2 (1 + \theta^2 + \cdots + \theta^{2(k-1)}) E r^2 \tag{3.62}$$

starting from $k = 1$, and we will be able to proceed with the iteration for as long as $\theta^k r < R/4$. Thus, at the last step we would have $R/(4\theta) \le \theta^k r < R/4$ and as

$$\tfrac{1}{2}\theta^{2(k-1)} E r^2 \le t_k \le \tfrac{1}{16}\theta^{2k} E r^2 \tag{3.63}$$

(since $\theta > 20$), this gives existence for time duration $c_0 R^2$ with

$$c_0 = \frac{E}{32\theta^4}. \tag{3.64}$$

Using the second inequality of (3.63), the estimates above imply that

$$|A(t_k)| \le \frac{2\alpha}{\theta^k r} \le \frac{\alpha\sqrt{E}}{\sqrt{t_k}}. \tag{3.65}$$

Similarly, letting $(\phi_t)_{t \in [0, c_0 R^2]}$ be the flow starting from Y, using the first inequality of (3.63) we have

$$d(\phi_{t_k}(y), \phi_0(y)) \le \tfrac{1}{2}\beta\theta^k r \le \frac{\beta\theta\sqrt{t_k}}{\sqrt{E}}, \tag{3.66}$$

and in particular,

$$d_H(Y_{t_k}, X) \le \frac{\beta\theta\sqrt{t_k}}{\sqrt{E}} + \tfrac{1}{2}\beta r \le \frac{2\beta\theta\sqrt{t_k}}{\sqrt{E}}. \tag{3.67}$$

The above calculations are only valid for the iteration times, but choosing ε sufficiently small, we can make $A(s)$, $B(s)$, $C(s)$ hold for all Y_t with $t \in [0, Er^2)$ and $r \le s \le \theta r$. Thus, we get that for every $t \in [2Er^2, c_0 R^2]$ we have

$$|A(t)| \le \frac{c_1}{\sqrt{t}}, \tag{3.68}$$

$$d_H(Y_t, X) \le c_2 \sqrt{t}, \tag{3.69}$$

with $c_1 = 2\alpha\sqrt{E}$ and $c_2 = 2\beta\theta/\sqrt{E}$. Note that by Lemma 3.25, $c_1^2 < \frac{1}{80}$ and $c_1 c_2 < \min(10^{-6}, \Lambda^{-1})$.

For the last part of Theorem 1.16, fix $c_3 = 4\alpha^2 E$, let $t \in [c_3 r^2, c_0 R^2]$ and let $s \in (\sqrt{t}/(2\alpha\sqrt{E}), R/4)$. Since $s \ge \sqrt{t}/(2\alpha\sqrt{E}) \ge r$, by the third condition of Theorem 1.14, for every $x \in X$ one can write

$$B(x, s) \cap X_0^r = G \cup B, \tag{3.70}$$

where G is connected and $B \cap B(x, (1-20\varepsilon)s) = \varnothing$, and as $\varepsilon < \frac{1}{10}\beta < \frac{1}{10}E < \frac{1}{10}\theta E$, we have $B \cap B(x, (1-2\theta E)s) = \varnothing$. On the other hand, note that by (3.66), we have the motion bound

$$d(\phi_t(y), \phi_0(y)) \le \frac{2\theta\beta\sqrt{t}}{\sqrt{E}} = \frac{\theta E\sqrt{t}}{\alpha\sqrt{E}} \le 2\theta E s, \tag{3.71}$$

so we obtain that

$$B(x, s) \cap X_t^r = G_t \cup B_t, \tag{3.72}$$

where G_t is connected and $B_t \cap B(x, (1-4\theta E)s) = \varnothing$. In particular, we have that $B_t \cap B\big(x, \frac{9}{10}s\big) = \varnothing$. □

Having established Theorem 1.16, the existence parts of Theorem 1.11 follow easily.

Theorem 3.73 *With the constants ε, c_0, c_1, c_2, c_3 as in Theorem 1.16, if X is an (ε, R)–Reifenberg set, there exists a smooth MCF $(X_t)_{t \in (0, c_0 R^2)}$ satisfying*

$$\lim_{t \to 0} d_H(X_t, X) = 0. \tag{3.74}$$

Moreover, setting $X_0 = X$, this solution is also a weak set flow (see Definition 1.8) and a physical solution (see Definition 1.21).

Proof Using the estimates of Theorem 1.16 and the standard estimates for the higher derivatives of A, the existence of a smooth solution $(X_t)_{t \in (0, c_0 R^2]}$ follows immediately by an Arzelà–Ascoli argument. Since the flow $(X_t)_{t \in (0, c_0 R^2)}$ is smooth, we only need to check avoidance with respect to smooth flows starting at time 0. Take a smooth flow $(\Delta_t)_{t\in[0,T]}$ for $T \leq c_0 R^2$ with

$$d_H(\Delta_0, X) = \delta > 0. \tag{3.75}$$

For r sufficiently small we have

$$d_H(X^r, \Delta_0) \geq \tfrac{1}{2}\delta, \tag{3.76}$$

so by avoidance we will also have

$$d_H(X^r_t, \Delta_t) \geq \tfrac{1}{2}\delta. \tag{3.77}$$

As $X_t = \lim_{j\to\infty} X^{r_j}_t$ with $r_j \to 0$ as $j \to \infty$, this implies

$$X_t \cap \Delta_t = \varnothing. \tag{3.78}$$

□

Remark 3.79 Using estimate (3.66), along with an Arzelà–Ascoli argument that utilizes the uniform curvature estimate of Theorem 1.16, it follows that the convergence is in fact in the parametrized C^0 sense.

4 Uniqueness

In this section we will conclude the proof of Theorem 1.11. In Section 4.1 we will see that at any positive time, the approximating flows of small enough scales can be written as graphs over one another. In Section 4.2 we will derive a PDE for such a situation which will be used to obtain the separation estimate, Theorem 1.22 in Section 4.3. Finally, we will conclude the proof of Theorem 1.11 in Section 4.4 (assuming the interior estimate, Theorem 3.1, which will in turn be proved in Section 5). Throughout this section we will assume freely that the parameters α, β, θ, c_0 satisfy the inequalities of Theorem 1.16.

4.1 Graph representation

Recall that $(Y^r_t)_{t\in(0,c_0R^2)}$ is called a (Λ) r–approximate evolution of X if for $t \in [c_3 r^2, c_0 R^2]$,

$$|A^r(t)| \leq \frac{c_1}{\sqrt{t}}, \tag{4.1}$$

$$d_H(Y^r_t, X) \leq c_2\sqrt{t}, \tag{4.2}$$

and for any $t \in [c_3 r^2, c_0 R^2]$ and $s \in (\frac{\sqrt{t}}{c_1}, R/4)$ and $x \in X$ we have

$$B(x,s) \cap Y^r_t = G \cup B, \tag{4.3}$$

where G is connected and $B \cap B\big(x, \frac{9}{10}s\big) = \varnothing$. We will always assume that the parameters satisfy the relations of Theorem 1.16. To be more precise, we assume:

(1) $c_1 c_2 < \min\{10^{-6}, \Lambda^{-1}\}$

(2) $c_1^2 < \frac{1}{80}$

We start by making the following two observations.

Lemma 4.4 *If $(Y_t^r)_{t\in(0,c_0R^2)}$ is an r–approximate evolution of X, then for every $t \in [c_3 r^2, c_0 R^2]$,*

$$\operatorname{inj}(Y_t^r) \geq \min\Big(\frac{\sqrt{t}}{4c_1}, \tfrac{1}{4}R\Big). \tag{4.5}$$

Here inj *denotes the injectivity radius of the normal exponential map. In particular, Y_t^s is contained in a tubular neighborhood of Y_t^r for every $s \leq r$.*

Proof Set ρ to be the distance function from Y_t^r. The curvature bound (4.1) implies that there are no focal points with $\rho \leq \sqrt{t}/c_1$. By the characterization of the injectivity radius we know that if p is a cut point with $\rho(p) = \operatorname{inj}(Y_t^r) \leq \sqrt{t}/4c_1$, then there exist $y_1, y_2 \in Y_t^r$ such that

$$d(y_1, y_2) \leq \frac{\sqrt{t}}{2c_1}, \tag{4.6}$$

$$(y_1 - y_2) \perp T_{y_i} Y_t^r \quad \text{for} \quad i = 1, 2. \tag{4.7}$$

Fix $x \in X$ such that

$$d(x, y_1) \leq c_2\sqrt{t} = \frac{c_1 c_2 \sqrt{t}}{c_1} < 10^{-6}\frac{\sqrt{t}}{c_1}, \tag{4.8}$$

and consider the intersection $B(x, \sqrt{t}/c_1) \cap Y_t^r = G \cup B$, where the splitting is by (4.3). By (4.6) we know that $y_1, y_2 \in G$ and since y_1 is very close to x, the curvature bound (4.1) together with the perpendicularity of the tangent spaces (4.7) will force the connected component of y_1 to leave the ball before being able to return to y_2, which is a contradiction. □

In fact, we even have the following.

Lemma 4.9 *For $t \in [c_3 r^2, c_0 R^2]$ and $s \leq r$, Y_t^s is a graph of a function u over Y_t^r. By that we mean that*

$$Y_t^s = \{y + u(y,t)N(y,t) : y \in Y_t^r\}, \tag{4.10}$$

where $N(y,t)$ is the normal to Y_t^r at y.

Proof By the above lemma, we know that $Y_t^s \subseteq \mathcal{N}(Y_t^r) = \mathcal{N} = B(Y_t^r, \mathrm{inj}(Y_t^r))$. Denote $Y = Y_t^r$ and $Z = Y_t^s$ and denote by $\pi\colon \mathcal{N} \to Y$ the projection to the nearest point.

In one direction, suppose that there are $z_1 \neq z_2 \in Z$ such that $\pi(z_1) = \pi(z_2) = y \in Y$. By applying (4.2) for both Y and Z, we see that

$$d(z_i, y) \leq 2 \cdot c_2\sqrt{t} \leq 2 \cdot 10^{-6}\frac{\sqrt{t}}{c_1}, \tag{4.11}$$

and that there exists an $x \in X$ such that

$$d(x, y) \leq 10^{-6}\frac{\sqrt{t}}{c_1}. \tag{4.12}$$

By (4.3) we can write

$$B\left(x, \frac{\sqrt{t}}{c_1}\right) \cap Z = G \cup B \tag{4.13}$$

and $z_1, z_2 \in G$. The curvature bounds and distance bounds (4.1), (4.2) will not allow that. In fact, by the following lemma, we have that

$$\tan(\angle(T_yY, T_{z_i}Z)) \leq \tfrac{6}{1000}. \tag{4.14}$$

As $(z_i - y) \perp T_yY$ we see that $(z_2 - z_1)/\|z_2 - z_1\|$ is almost perpendicular to $T_{z_i}Z$ and so z_1 and z_2 are two points in G that are very close to the center of $B(x, \sqrt{t}/c_1)$ with almost parallel tangent planes that lie one above the other. The curvature bound (4.1) prevents that since, as before, it will force the connected component of z_1 to leave the ball before returning to z_2. Thus over every point in Y lies at most one point in Z.

On the other hand, given any $y \in Y$ there is $z' \in Z$ with

$$d(z', y) \leq 2 \cdot 10^{-6}\frac{\sqrt{t}}{c_1}. \tag{4.15}$$

Letting $y' = \pi(z')$ we see again that $\tan(\angle(T_{y'}Y, T_{z'}Z)) \leq \frac{6}{1000}$, so by the curvature bounds (4.1), and since y and y' are very close compared to the scale $\sqrt{t}/c_1$, there will be a point z over y. This completes the proof. $\square$

Lemma 4.16 *Let $M^2 \subseteq \mathcal{N}(M^1)$, where M^1 and M^2 are hypersurfaces in $\mathbb{R}^{n+1}$, and let $\pi\colon \mathcal{N}(M^1) \to M^1$ be the nearest point projection. If we have the bounds $\max(|A^1|, |A^2|) \leq a$, $d_H(M^1, M^2) \leq b$ and $|ab| \ll 1$, then for every $m_2 \in M_2$, setting $m_1 = \pi(m_2)$ we have*

$$\tan(\angle(T_{m_1}M^1, T_{m_2}M^2)) \leq \frac{(2 - ab)\sqrt{ab - a^2b^2/4}}{2(1 - ab/2)^2 - 1} \leq 3\sqrt{ab}. \tag{4.17}$$

Remark 4.18 The proof of the above lemma is similar to that of Lemma 2.25. This time we obtain a lower bound for $|u|$. Assume the tangent spaces are not the same (else there is nothing to prove) and that, without loss of generality, $T_{m_1}M^1 = \operatorname{span}\{e_1,\dots,e_n\}$, $m_1 = 0$ and $m_2 = (0,\dots,0,z)$ with $z \geq 0$. By drawing two balls of radius $1/a$, one starting horizontally that bounds M^1 from above and the other starting parallel to $T_{m_2}M^2$ that bounds M^2 from below, those two barriers will force M^2 to drift apart from M^1 until the angle will be halved. This will give a lower bound on the Hausdorff distance, which should be less than b by assumption. The details are left to the reader.

4.2 The graph PDE for two evolving surfaces

Having discovered that Y_t^s is a graph of a function u over Y_t^r when $s \leq r$ and $t \geq c_3 r^2$, let us compute the evolution equation in such a situation. The computation done here is the parabolic analogue for the one done in [3, Lemma 2.26] for the minimal surface case.

Lemma 4.19 *Let $(M_t^1)_{t\in(0,T)}$ flow by mean curvature and let $\mathcal{N}_t$ be a tubular neighborhood of M_t^1. Let $(M_t^2)_{t\in(0,T)}$ be another flow by mean curvature such that for every $t \in (0,T)$, one has $M_t^2 \subseteq \mathcal{N}_t$. Let $\overrightarrow{N}(x,t)$ denote the inner-pointing unit normal to M_t^1 at x and write M_t^2 as a graph of a function $u(x,t)$ over M_t^1, so*

$$M_t^2 = \{x + u(x,t)N(x,t) : x \in M_t^1\}. \tag{4.20}$$

Then u satisfies an equation of the form

$$u_t = (1+\varepsilon)\operatorname{div}((I+L)\nabla u) + |A|^2 u - Q_{ij}A_{ij}, \tag{4.21}$$

where A is the second fundamental form of M^1, A_{ij} are its components in a local orthonormal frame, and for some constant $D > 0$ we have $|\varepsilon|, |L| \leq D(|A||u| + |\nabla u|)$ and $|Q_{ij}| \leq D(|A||u| + |\nabla u|)^2$.

Proof We want to derive an expression for the mean curvature $\overrightarrow{H}_2$ of M^2, for some fixed time t, in terms of u and its derivatives, and in terms of the geometric quantities of M^1. Once we have done that, the derivation of the equation will follow easily. Let A, H and dVol be the second fundamental form, mean curvature and the volume form of M^1 respectively. Choose $m_1 \in M^1$ and let $E_1,\dots,E_n$ be a local orthonormal basis for M^1 around m_1. Extend $E_1,\dots,E_n$, along with N, to normal fields in a neighborhood of m_1 in $\mathcal{N}$ by parallel translating along N. The vectors

$$X_i = E_i + u_i N - uA_{ik}E_k \tag{4.22}$$

form a basis to the tangent space of M^2. Therefore

$$g_{ij} = \delta_{ij} + u_i u_j - 2uA_{ij} + u^2 A_{ik} A_{jk} = \delta_{ij} - 2uA_{ij} + Q_{ij}, \tag{4.23}$$

and so

$$g^{ij} = \delta_{ij} + 2uA_{ij} + Q_{ij}. \tag{4.24}$$

Now, taking a variation $\tilde{u}(x,s) = u(x) + sv(x)$ for some function $v\colon M^1 \to \mathbb{R}$ which is localized in the above neighborhood and computing the corresponding quantities, we get

$$\begin{aligned} \frac{d}{ds}\Big|_{s=0} \sqrt{\det g_{ij}} &= \tfrac{1}{2}\sqrt{\det g_{ij}}\,\operatorname{trace}\Big(g^{ij}\frac{d}{ds}\Big|_{s=0} g_{ij}\Big) \\ &= \tfrac{1}{2}\sqrt{\det g_{ij}}\,(\delta_{ij} + 2uA_{ij} + Q_{ij})(v_i u_j + u_i v_j - 2vA_{ij} + 2uvA_{ik}A_{jk}) \\ &= \sqrt{\det g_{ij}}\,[\langle \nabla u, \nabla v\rangle - vH - uv|A|^2 + \langle L\nabla u, \nabla v\rangle + Q_{ij}A_{ij}v] \\ &= \langle (I+L)\nabla u, \nabla v\rangle + \sqrt{\det g_{ij}}\,[-vH - uv|A|^2 + Q_{ij}A_{ij}v], \end{aligned} \tag{4.25}$$

where in the last equality we have used

$$\sqrt{\det g_{ij}} = 1 - uH + Q \tag{4.26}$$

to absorb $\sqrt{\det g_{ij}}$ into $(I+L)$. Thus, by integration by parts we get

$$\begin{aligned} -\frac{d}{ds}\Big|_{s=0} \int \sqrt{\det g_{ij}}\,\mathrm{dVol} \\ = \int \sqrt{\det g_{ij}}\, v\,[(1+\varepsilon)\operatorname{div}((I+L)\nabla u) + H + u|A|^2 + Q_{ij}A_{ij}]\,\mathrm{dVol}, \end{aligned} \tag{4.27}$$

where we have used (4.26) to write $1/\sqrt{\det g_{ij}} = 1 + \varepsilon$. On the other hand, by the definition of the mean curvature vector, we obtain

$$-\frac{d}{ds}\Big|_{s=0} \int \sqrt{\det g_{ij}}\,\mathrm{dVol} = \int v\langle N, \overrightarrow{H}_2\rangle \sqrt{\det g_{ij}}\,\mathrm{dVol}, \tag{4.28}$$

so

$$\langle N, \overrightarrow{H}_2\rangle = (1+\varepsilon)\operatorname{div}((I+L)\nabla u) + H + u|A|^2 + Q_{ij}A_{ij}. \tag{4.29}$$

Having done that, deriving the PDE is easy. Indeed, writing M^i as $\phi^i(x,t)$ we get

$$\phi^2(x,t) = \phi^1(x,t) + u(x,t)N(x,t), \tag{4.30}$$

so differentiating with respect to time we obtain

$$\overrightarrow{H}_2 = \overrightarrow{H} + u_t \overrightarrow{N} - u\nabla H, \tag{4.31}$$

so

$$u_t = \langle \overrightarrow{H}_2, N\rangle - H = (1+\varepsilon)\,\mathrm{div}((I+L)\nabla u) + u|A|^2 + Q_{ij}A_{ij}. \tag{4.32}$$ □

4.3 The separation estimate

We now come to the proof of the separation estimate, Theorem 1.22. In fact, we are going to prove a slightly stronger theorem, from which Theorem 1.22 will follow easily. At long last, we are going to fix the parameter Λ, on which the output of Theorem 1.16 and thus the definition of an r–approximate evolution depended.

Theorem 4.33 *There exists a constant $C > 0$ with the following significance. Let $s \le r$ and let Y_t^r and Z_t^s be (Λ) r–approximate and s–approximate evolutions of X with $\Lambda = 2D$, where D is the constant from Lemma 4.19. Then writing Y_t^s as a graph of a function u over Y_t^r for $s \le r$ and $t \in [c_3 r^2, c_0 R^2]$, we have the estimate*

$$|u| \le C r^{1/2} t^{1/4}. \tag{4.34}$$

Proof The idea is to use Equation (4.21) and the maximum principle to bootstrap and obtain more and more improved estimates for u. In order to start the bootstrapping, note that by (4.1) and (4.2) we have

$$|A(t)| = |A^r(t)| \le \frac{c_1}{\sqrt{t}}, \tag{4.35}$$

$$|u(t)| \le 2c_2\sqrt{t}. \tag{4.36}$$

Notice that at a maximum point of u, the $|\nabla u|$ dependence of the bounds on the coefficients in Equation (4.21) disappears. Note further that by Cauchy–Schwartz and estimates (4.35) and (4.36),

$$|Q_{ij}A_{ij}| \le D(|A||u|)^2|A| \le 2Dc_1c_2|u||A|^2 \le |u||A|^2, \tag{4.37}$$

where we have used the relation $c_1c_2 < \Lambda^{-1} = (2D)^{-1}$. Therefore, using estimates (4.35) and (4.36) and applying the maximum principle for Equation (4.21) we obtain

$$\frac{d}{dt}u(x_{\max}, t) \le 2u(x_{\max}, t)|A(x_{\max}, t)|^2 \le 2c_2\frac{2c_1^2}{\sqrt{t}} < \frac{2c_2}{40}\frac{1}{\sqrt{t}}, \tag{4.38}$$

where the last inequality holds since $c_2^2 < \frac{1}{80}$. A similar calculation is done for $u(x_{\min})$. Integrating starting from $t_0 = c_3 r^2$, we get the improved estimate

$$|u(t)| \le 2c_2\big(\sqrt{c_3}r + \tfrac{1}{10}\sqrt{t}\big) \le 2c_2\big(c_3^{1/4} r^{1/2} t^{1/4} + \tfrac{1}{10}\sqrt{t}\big). \tag{4.39}$$

Plugging this back into the inequality, we get

$$\frac{d}{dt}u(x_{\max},t) \le 2c_2\left(\frac{1}{40}\frac{c_3^{1/4}r^{1/2}}{t^{3/4}} + \frac{1}{400}\frac{1}{\sqrt{t}}\right), \tag{4.40}$$

which yields

$$|u(t)| \le 2c_2\big(\big(1+\tfrac{1}{10}\big)c_3^{1/4}r^{1/2}t^{1/4} + \tfrac{1}{100}\sqrt{t}\,\big). \tag{4.41}$$

Continuing the bootstrapping, this will yield

$$|u(t)| \le 4c_2c_3^{1/4}r^{1/2}t^{1/4}, \tag{4.42}$$

as required. □

Proof of Theorem 1.22 This follows directly from Theorem 4.33. □

Proof of Corollary 1.24 Fixing $t>0$, for r sufficiently small we have $t \ge c_3r^2$. By Theorem 1.22, for every $s \le r$,

$$d_H(Y_t^r, Y_t^s) \le C(t)\sqrt{r}. \tag{4.43}$$

Thus, for every sequence $\{r_n\}_{n=1}^{\infty}$ with $r_n > 0$ and $\lim_{n\to\infty} r_n = 0$, $\{Y_t^{r_n}\}_{n=1}^{\infty}$ forms a Cauchy sequence in the Hausdorff sense. Therefore $\lim_{r\to 0} Y_t^r$ exists. The second part follows since every physical solution is an r–approximate evolution for every $r>0$, so given two physical flows, we can use Theorem 1.22 with any choice of r. □

4.4 The level set flow

We start this section by adapting what we have done so far a little bit. According to Corollary 2.11 we can choose approximations X_+^r and X_-^r instead of the X^r, such that if Ω is the bounded open domain with $\partial\Omega = X$, then $X_+^r \subseteq \overline{\Omega}^c$ and $X_-^r \subseteq \Omega$. There are two differences between the estimates for X^r that we have been working with and the ones for $X_\pm^r$. The first one is that the constants c_1, c_2 in Theorem 1.14 and Corollary 2.11 are different. The second difference is in property (3), namely, the marginal connected components of $X^r \cap B(x,s)$ for $s \ge r$ are inside the annulus $(B(x,s) - B(x,(1-20\varepsilon)s))$, while for $X_\pm^r$ they will be in the slightly thicker annulus $(B(x,s) - B(x,(1-40\varepsilon)s))$. These two differences clearly do not influence our arguments in any way so in particular, Theorem 1.16 is still valid for the flows emanating from $X_\pm^r$. Thus, choosing Λ as in Theorem 4.33 and with the $\varepsilon, c_0, \dots, c_3$ of Theorem 1.16, we know that $X_{\pm,t}^r$ exist for time duration c_0R^2, and

$$\lim_{r\to 0} X_{-,t}^r = \lim_{r\to 0} X_{+,t}^r = X_t, \tag{4.44}$$

where X_t is the unique physical flow emanating from X.

Proof of Theorem 1.11 Let $\widetilde{X}_t$ be the level set flow of X. By Theorem 3.73 we know that $X_t \subseteq \widetilde{X}_t$. On the other hand, by the avoidance principle, we know that since $X \cap X^r_{\pm} = \varnothing$ we have

$$\widetilde{X}_t \cap X^r_{\pm,t} = \varnothing. \tag{4.45}$$

Denoting by $\Omega^r_{\pm,t}$ the bounded domains with $\partial\Omega^r_{\pm,t} = X^r_{\pm,t}$ we claim that

$$\widetilde{X}_t \subseteq \overline{\Omega}^r_{+,t} - \Omega^r_{-,t} := N^r_t. \tag{4.46}$$

This is true since, considering

$$T^r = \inf\{0 \leq t \leq c_0 R^2 : \widetilde{X}_t \cap (N^r_t)^c \neq \varnothing\}, \tag{4.47}$$

we clearly have that $T^r > 0$ by avoiding balls, since the two sets start a positive distance apart. Given (x_i, t_i) such that $x_i \in \widetilde{X}_{t_i} \cap (N^r_{t_i})^c$ with $t_i \to T^r$ we see (by avoiding balls and since the evolution of $\partial N^r_t = X^r_{+,t} \cup X^r_{-,t}$ is smooth) that $d(x_i, \partial N^r_{t_i}) \to 0$. By compactness we get $\lim_{j\to\infty} x_{i_j} = x \in \partial N^r_{T^r}$. On the other hand, by the level set definition of the level set flow, $\widetilde{X}_t = f_t^{-1}(0)$ for some continuous function $f\colon [0, c_0 R^2] \times \mathbb{R}^{n+1} \to \mathbb{R}$ (see [8; 1]) and so we see that $x \in \widetilde{X}_{T^r}$. Thus, $x \in \widetilde{X}_{T^r} \cap X^r_{\pm,T^r}$ which contradicts (4.45).

Having established (4.46) and noting that by Theorem 1.22, $d_H(N^r_t, X^r_{+,t}) \leq C(t) r^{1/2}$, we finally conclude that

$$\widetilde{X}_t \subseteq \bigcap_{0<r<R/4} N^r_t \subseteq \lim_{r\to 0} N^r_t = X_t. \tag{4.48}$$

Thus $\widetilde{X}_t = X_t$ and we are done. □

5 Proof of Theorem 3.1

The main purpose of this section is to prove Theorem 3.1, which was an essential ingredient in the proof of Theorem 1.16 and is thus a cornerstone for the entire argument. In Section 5.1 we will state Theorem 1.16 in slightly different terms. In Section 5.2 we will derive an a priori estimate for the non-homogeneous heat equation which is suitable for thick space-time cylinders. In Section 5.3 we will derive a Hölder gradient estimate suitable for our situation and in Section 5.4 we will conclude the proof of Theorem 5.1.

5.1 Main estimates

Theorem 1.16 will follow immediately from the following "centered form" of the estimate, the proof of which will be described through the rest of this section.

Theorem 5.1 (main estimate, II) *There exists some $c > 0$ such that for every $\delta > 0$ and $M > 0$, there exist positive $\tau_0 = \tau_0(M, \delta) \ll 1$ and $\lambda_0 = \lambda_0(M, \delta) \ll 1$ such that for every $0 < \lambda < \lambda_0$, there exists some $\varepsilon_0 = \varepsilon_0(\delta, M, \lambda)$ such that for every $\beta > 0$, $0 < \tau < \tau_0$ and $\varepsilon < \varepsilon_0$ the following statement holds.*

If $u\colon B(p,r) \times [0, \tau r^2] \to \mathbb{R}$ is a graph moving by mean curvature such that

(1) *for every $(x,t) \in B(p,r) \times [0, \tau r^2]$,*

$$|\nabla u(x,t)| \le M\varepsilon, \tag{5.2}$$

(2) *for every $x \in B(p,r)$, we have*

$$|\nabla u(x, \lambda\tau r^2)| \le \varepsilon, \tag{5.3}$$

(3) *for every $(x,t) \in B(p,r) \times [0, \tau r^2]$,*

$$|u(x,t)| \le M^2\beta, \tag{5.4}$$

(4) *for every $x \in B(p,r)$, we have*

$$|u(x, \lambda\tau r^2)| \le \beta, \tag{5.5}$$

then

$$|A(p, \tau r^2)| \le (c+\delta)\frac{\varepsilon}{\sqrt{\tau} r}, \tag{5.6}$$

$$|A(p, \tau r^2)| \le c\frac{\beta}{\tau r^2} + \delta\frac{\varepsilon}{\sqrt{\tau} r}. \tag{5.7}$$

5.2 Estimates for the non-homogeneous heat equation on thick cylinders

The first key point in obtaining the main estimate — Theorem 5.1 — is a non-standard Schauder-type estimate for the non-homogeneous heat equation. Before stating it, we record the following definitions.

Definition 5.8 The parabolic ball of radius r with center (p,t) is defined to be

$$P(p,t,r) = B(p,r) \times [t - r^2, t]. \tag{5.9}$$

When it is clear from the context which point is p, we write $P^r = P(p, r^2, r)$. For

$\tau, \lambda < 1$ we define the narrow r, τ cylinder to be

$$P^{r,\tau} = B(p,r) \times [0, \tau r^2], \tag{5.10}$$

and the λ–truncated narrow r, τ cylinder to be

$$P^{r,\tau,\lambda} = B(p,(1-\sqrt{\lambda})r) \times [\lambda\tau r^2, \tau r^2]. \tag{5.11}$$

Theorem 5.12 *There exists a constant $c > 0$ such that for every $\delta > 0$ and $M > 0$, there exist positive $\tau_0 = \tau_0(M,\delta) \ll 1$ and $\lambda_0 = \lambda_0(M,\delta) \ll 1$ such that for every $0 < \lambda < \lambda_0$ and $0 < \alpha < 1$, there is a constant $C = C(\lambda,\alpha) > 0$ such that for every $0 < \tau < \tau_0$, the following statement holds.*

If u is a solution to the non-homogeneous heat equation

$$u_t - \Delta u = f \tag{5.13}$$

on $B(p,r) \times [0, \tau r^2]$ such that

(1) *for every $(x,t) \in B(p,r) \times [0, \tau r^2]$,*

$$|\nabla u(x,t)| \le M\varepsilon, \tag{5.14}$$

(2) *for every $x \in B(p,r)$, we have*

$$|\nabla u(x, \lambda\tau r^2)| \le \varepsilon, \tag{5.15}$$

(3) *for every $(x,t) \in B(p,r) \times [0, \tau r^2]$,*

$$|u(x,t)| \le M^2\beta, \tag{5.16}$$

(4) *for every $x \in B(p,r)$, we have*

$$|u(x, \lambda\tau r^2)| \le \beta, \tag{5.17}$$

then

$$\begin{aligned}\sqrt{\tau} r|\nabla^2 v(0,\tau r^2)| \le{}& \\ &(c+\delta)\varepsilon + \frac{C(\alpha,\lambda)}{\sqrt{\tau} r}\left(\sup_{z_1 \in P^{r,\tau}} d_{z_1}^2 |f(z_1)| + \sup_{z_1, z_2 \in P^{r,\tau}} d_{z_1,z_2}^{2+\alpha} \frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\right),\end{aligned} \tag{5.18}$$

$$\begin{aligned}\sqrt{\tau} r|\nabla^2 v(0,\tau r^2)| \le{}& \\ &\frac{c\beta}{\sqrt{\tau} r} + \frac{C(\alpha,\lambda)}{\sqrt{\tau} r}\left(\sup_{z_1 \in P^{r,\tau}} d_{z_1}^2 |f(z_1)| + \sup_{z_1, z_2 \in P^{r,\tau}} d_{z_1,z_2}^{2+\alpha} \frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\right),\end{aligned} \tag{5.19}$$

where

$$d((x_1,t_1),(x_2,t_2)) = \sqrt{|x_1-x_2|^2+|t_1-t_2|}, \tag{5.20}$$

and $d_{z_1} = d(z_1,\partial D)$, $d_{z_1,z_2} = \min(d_{z_1},d_{z_2})$.

Remark 5.21 There are two main differences between the standard Schauder estimate and the one above. The first and less important one is that in the standard Schauder estimates the term $|u|_0$ appears on the right-hand side instead of $|\nabla u|_0$. In the elliptic case, an estimate similar to the one above follows from the standard Schauder estimate by integration. In the parabolic case this is not as trivial, as we do not assume anything about $|u_t|$. The second and more important difference is that in our estimate there is a distinction between a leading term coming from the initial time slice and a negligible term coming from the rest of the parabolic boundary. In the standard Schauder estimate there is no such distinction.

In order to prove this version of the Schauder estimate, we need the following two lemmas. The first one is a standard result for Gaussian potentials; see for instance the proof of [9, Chapter 4.3, Theorem 2] and [10, Theorems 4.6 and 4.8].

Lemma 5.22 *For every* $0<\alpha<1$ *there exists a constant* C *with the following property. Let* Φ *be the fundamental solution for the heat equation and let* w *be the Gaussian potential corresponding to* f, *ie*

$$w(x,t) = \int_0^t \int_{B(0,r)} \Phi(x-y,t-s) f(y,s)\,dy\,ds. \tag{5.23}$$

Then $w_t - \Delta w = f$, *and for every* $z\in D\subseteq B(0,r)\times[\tau r^2]$ *we have*

$$\begin{aligned} &|w(z)| + d_z|\nabla w(z)| + d_z^2|\nabla^2 w| \\ &\qquad \le C\left(\sup_{z_1\in D} d_{z_1}^2 f(z_1) + \sup_{z_1,z_2\in D} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\right). \end{aligned} \tag{5.24}$$ □

The following lemma gives a good interior derivative estimate for the heat equation at times that are very close to 0 compared to the initial scale.

Lemma 5.25 *There exists some* $c>0$ *with the following significance. For every* $\delta>0$ *and* $M>0$ *there exists a positive* $\tau_0=\tau_0(M,\delta)\ll 1$ *such that for every* $0<\tau<\tau_0$, *the following statement holds.*

If u *is a solution to the homogeneous heat equation*

$$u_t - \Delta u = 0 \tag{5.26}$$

on $B(p,r)\times[0,\tau r^2]$ *such that*

(1) *for every* $(x,t) \in B(p,r) \times [0, \tau r^2]$,

$$|u(x,t)| \le M\varepsilon, \tag{5.27}$$

(2) *for every* $x \in B(p,r)$,

$$|u(x,0)| \le \varepsilon, \tag{5.28}$$

then

$$|\nabla u(p, \tau r^2)| \le (c+\delta)\frac{\varepsilon}{\sqrt{\tau} r}. \tag{5.29}$$

Proof The proof is modeled on that of [7, Theorem 2.3.8].

By scaling, it suffices to prove that if u is such a solution on $B(0,R) \times [0,1]$ for R sufficiently large, we have

$$|\nabla u(0,1)| \le (c+\delta)\varepsilon. \tag{5.30}$$

First, fix a positive cut-off function $\xi \in C_0^\infty(B(0,R))$ with $\xi|_{B(0,R-1)} = 1$, $0 \le \xi \le 1$ and $|\nabla \xi|, |\nabla^2 \xi| \le C(n)$. Considering the function

$$v(x,t) = \xi(x) u(x,t), \tag{5.31}$$

we have that $v\colon \mathbb{R}^n \times [0,1] \to \mathbb{R}$ is defined and satisfies

$$v_t - \Delta v = -u\Delta\xi - 2\nabla u \cdot \nabla \xi. \tag{5.32}$$

As v is bounded we have the representation formula

$$\begin{aligned} v(x,t) &= \int_{\mathbb{R}^n} \Phi(x-y,t) v(y,0)\, dy \\ &\qquad + \int_0^t \int_{\mathbb{R}^n} \Phi(x-y,t-s)\big(-u(y,s)\Delta\xi(y) \\ &\qquad\qquad - 2\nabla u(y,s)\cdot\nabla\xi(y)\big)\, dy\, ds \\ &= \int_{\mathbb{R}^n} \Phi(x-y,t) v(y,0)\, dy \\ &\qquad + \int_0^t \int_{\mathbb{R}^n} \big(\Phi(x-y,t-s)\Delta\xi(y) \\ &\qquad\qquad + 2\nabla_y \Phi(x-y,t-s)\cdot\nabla\xi\big) u(y,s)\, dy\, ds. \end{aligned} \tag{5.33}$$

Differentiating under the integral sign we obtain

$$\begin{aligned}(5.34)\quad \nabla v(x,t) = \int_{\mathbb{R}^n} \nabla_x \Phi(x-y,t) v(y,0)\,dy \\ + \int_0^t \int_{\mathbb{R}^n} \big(\nabla_x \Phi(x-y,t-s)\Delta\xi(y) \\ + 2\nabla_x\nabla_y \Phi(x-y,t-s)\cdot\nabla\xi\big) u(y,s)\,dy\,ds.\end{aligned}$$

We will bound the two integrals appearing in this expression for $\nabla v(0,1) = \nabla u(0,1)$. The first integral is bounded by

$$\|v(-,0)\|_\infty \cdot \|\nabla\Phi(-,1)\|_{L^1} \le c\varepsilon, \quad \text{where } c = \|\nabla\Phi(-,1)\|_{L^1}.$$

As for the second integral, the integrand is zero outside the annulus $R-1<|y|<R$, so as both $\nabla\Phi$ and $\nabla^2\Phi$ are summable on $(\mathbb{R}^n - B(0,1))\times(0,1)$, and by our assumptions on u and ξ, we get that for R large enough,

$$(5.35)\qquad |\nabla u(0,1)| \le (c+\delta)\varepsilon. \qquad \square$$

The following lemma is proved similarly.

Lemma 5.36 *There exists some $C>0$ such that for every $M>0$, there exists a positive $\tau_0 = \tau_0(M) \ll 1$ such that for every $0<\tau<\tau_0$ and $\beta>0$, the following statement holds.*

If u is a solution to the homogeneous heat equation

$$(5.37)\qquad u_t - \Delta u = 0$$

on $B(p,r)\times[0,\tau r^2]$ such that

(1) *for every $(x,t)\in B(p,r)\times[0,\tau r^2]$,*

$$(5.38)\qquad |u(x,t)| \le M^2\beta,$$

(2) *for every $x \in B(p,r)$,*

$$(5.39)\qquad |u(x,0)| \le \beta,$$

then

$$(5.40)\qquad |\nabla^2 u(p,\tau r^2)| \le C\frac{\beta}{\tau r^2}. \qquad \square$$

Proof of Theorem 5.12 By scaling we can assume $r=1$. Let $v=u-w$ where w is the Gaussian potential from Lemma 5.22. Taking $\lambda \ll 1$ we therefore have the estimate

$$(5.41)\quad \sqrt{\lambda}\sqrt{\tau}|\nabla w|_{P^{1,\tau,\lambda}} \le C(\alpha)\Bigg(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)| + \sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\Bigg),$$

so

$$(5.42)\quad |\nabla w|_{P^{1,\tau,\lambda}} \le \frac{C(\alpha,\lambda)}{\sqrt{\tau}}\Bigg(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)| + \sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\Bigg).$$

Similarly,

$$(5.43)\quad \sqrt{\tau}|\nabla^2 w|_{P^{1,\tau,\lambda}} \le \frac{C(\alpha,\lambda)}{\sqrt{\tau}}\Bigg(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)| + \sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\Bigg)$$

and consequently:

(1) For every $x\in B(p,(1-\sqrt{\lambda})r)$,

$$(5.44)\quad |\nabla v(x,\lambda\tau)| \le \varepsilon + \frac{C(\alpha,\lambda)}{\sqrt{\tau}}\Bigg(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)| + \sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\Bigg).$$

(2) For every $(x,t)\in P^{1,\tau,\lambda}$,

$$(5.45)\quad |\nabla v|_{P^{1,\tau,\beta,\lambda}} \le M\varepsilon + M\frac{C(\alpha,\lambda)}{\sqrt{\tau}}\Bigg(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)| + \sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\Bigg).$$

Note that v solves the heat equation and therefore so do $\partial v/\partial x_i$ for every $i=1,\dots,n$. By choosing τ_0 and λ sufficiently small we obtain, by Lemma 5.25,

$$(5.46)\quad \sqrt{\tau}|\nabla^2 v(0,\tau)| \le (c+\delta)\varepsilon + \frac{C(\alpha,\lambda)}{\sqrt{\tau}}\Bigg(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)| + \sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\Bigg).$$

By (5.43) the desired result for u follows. The second estimate is proved similarly, using Lemma 5.36. □

5.3 Hölder gradient estimate

The second key ingredient in proving Theorem 5.1 is the following Hölder gradient estimate for our equation.

Theorem 5.47 *There exist constants $c_1 > 0$ and $0 < \alpha < 1$, depending only on n, such that if $\Omega \subseteq \mathbb{R}^n \times \mathbb{R}_+$ is a bounded domain and $u\colon \Omega \to \mathbb{R}$ solves the graphical mean curvature equation*

$$u_t = \left(\delta^{ij} - \frac{\partial_i u \partial_j u}{1+|\nabla u|^2}\right)\partial_i\partial_j u \tag{5.48}$$

with $\|\nabla u\|_{\Omega,0} = \varepsilon < 1$, then

$$\sup_{z_1,z_2\in\Omega} d^{\alpha}_{z_1,z_2}\frac{|\nabla u(z_1)-\nabla u(z_2)|}{d(z_2,z_1)^{\alpha}} \le c_1\varepsilon. \tag{5.49}$$

Proof This follows from tracing how the constants are formed in the interior Holder gradient estimate for quasilinear parabolic equations of general form, noticing that the derivatives of the coefficients, as well as the ellipticity of the equation obtained during the proof, are controlled; see [16, XII.3] or [10, 13.3] in the elliptic case. As our situation is simpler than the general one, and for the convenience of the reader, we carry it out here, following [16, XII.3] and [10, 13.3]. Before embarking on the proof, let us recall that the oscillation of a real function f on a domain Ω is defined by

$$\mathrm{osc}_{\Omega} f = \sup_{\Omega} f - \inf_{\Omega} f. \tag{5.50}$$

Equation (5.48) is of the form

$$u_t = a^{ij}(\nabla u)\partial_i\partial_j u, \tag{5.51}$$

with

$$a^{ij}(p) = \delta^{ij} - \frac{p_i p_j}{1+|p|^2}. \tag{5.52}$$

Differentiating with respect to k and regrouping, we get

$$\partial_t(\partial_k u) = \partial_i(a^{ij}\partial_k\partial_j u) + a^{ijl}(\partial_k\partial_l u)(\partial_i\partial_j u) \tag{5.53}$$

for $a^{ijl} = \partial_{p_l}a^{ij} - \partial_{p_j}a^{il}$. Considering the function $v = |\nabla u|^2$, we thus obtain

$$\partial_t v = \partial_i(a^{ij}\partial_j v) + a^{ijl}(\partial_l v)(\partial_i\partial_j u) - 2a^{ij}(\partial_r\partial_j u)(\partial_r\partial_i u), \tag{5.54}$$

so fixing a parameter γ and considering

$$w = w^{\pm} = w^{\pm}_k = \pm\gamma\partial_k u + v, \tag{5.55}$$

we get

$$-\partial_t w + \partial_i(a^{ij}\partial_j w) = -a^{ijl}(\partial_l w)(\partial_i\partial_j u) + 2a^{ij}(\partial_r\partial_j u)(\partial_r\partial_i u). \tag{5.56}$$

Since $|\nabla u| \le 1$, the explicit form of the coefficients $a^{ij}(p)$ given in (5.52) implies that $\lambda|\xi|^2 \le a^{ij}\xi_i\xi_j$, $|a^{ij}| \le \Lambda$ and $a^{ijl} \le c_2$ hold with $\lambda = \frac{1}{2}$, $\Lambda = 1$ and $c_2 = 6$. Thus, by ellipticity and by Schwarz's inequality we get, for some $c_3 = c_3(n)$,

$$-(\partial_t w) + \partial_i(a^{ij}\partial_j w) \ge -c_3|Dw|^2. \tag{5.57}$$

Now, assume $(x_0, t_0) \in \Omega$ and $r > 0$ are such that $P(x_0, t_0, 4r) \subseteq \Omega$ and set

$$\overline{w} = \overline{w}_k^{\pm} = \sup_{P(x_0,t_0,4r)} w_k^{\pm}. \tag{5.58}$$

Writing $W = \overline{w} - w$ we get that, for every non-negative $\zeta \in C_0^1(\Omega)$,

$$\int -(\partial_t W)\zeta - (a^{ij}\partial_j W)(\partial_i\zeta) \ge \int -c_3|\nabla W|^2\zeta, \tag{5.59}$$

so replacing ζ with $e^{c_4 W}\zeta$ for some $c_4 = c_4(n)$ we get, using bounded ellipticity,

$$-(\partial_t W) + \partial_i(\overline{a}^{ij}\partial_j W)_i \ge 0 \tag{5.60}$$

for

$$\overline{a}^{ij} = e^{c_4 W}a^{ij}. \tag{5.61}$$

Note that if $\gamma \le c_5(n)$ a bound on the ellipticity of $\overline{a}^{ij}$ will still be determined, regardless of ε (as long as $\varepsilon < 1$). If we are in such a regime, since $W \ge 0$ in $P(x_0, t_0, r)$, by the Moser–Harnack inequality we get that for some $c_6 = c_6(n)$,

$$r^{-n-2}\int_{P(x_0,t_0-4r^2,r)} (\overline{w} - w)\,dx \le c_6 \inf_{P(x_0,t_0,r)} W = c_6\Big(\overline{w} - \sup_{P(x_0,t_0,r)} w\Big). \tag{5.62}$$

Now, choose $\gamma = 10n\varepsilon$ and note that for any subdomain $\Omega_0 \subseteq \Omega$, choosing k such that $\operatorname{osc}_{\Omega_0}(\partial_k u) \ge \operatorname{osc}_{\Omega_0}(\partial_i u)$ for all $i = 1, \ldots, n$ yields

$$8n\varepsilon \operatorname{osc}_{\Omega_0}(\partial_k u) \le \operatorname{osc}_{\Omega_0}(w_k^{\pm}) \le 12n\varepsilon \operatorname{osc}_{\Omega_0}(\partial_k u), \tag{5.63}$$

which will yield (for $w^{\pm} = w_k^{\pm}$)

$$\begin{aligned} \inf_{\Omega_0}(\overline{w}^+ - w^+ + \overline{w}^- - w^-) &\ge 10n\varepsilon\Big(\sup_{\Omega_0} u_k - \inf_{\Omega_0} u_k\Big) + 2\inf_{\Omega_0} v - 2\sup_{\Omega_0} v \\ &\ge 6n\varepsilon \operatorname{osc}_{\Omega_0}(u_k) \\ &\ge \tfrac{1}{2}\operatorname{osc}_{\Omega_0}(w^{\pm}). \end{aligned} \tag{5.64}$$

Thus, there exists some $c_7 = c_7(n)$ such that

$$\operatorname{osc}_{P(x_0,t_0,4r)}(w^{\pm}) \le \frac{c_7}{r^{n+2}} \int_{P(x_0,t_0-4r^2,r)} (\overline{w}^{\pm} - w^{\pm})\,dx \tag{5.65}$$

holds for *at least one* of w^+, w^-. Combining this with (5.62), there exists $c_8 = c_8(n)$ such that, assuming the above holds for $w = w^+$, and denoting oscillations and sup/inf over the parabolic ball $P(x_0, t_0, \rho)$ by osc_ρ, $\sup_\rho$ and $\inf_\rho$, we have

$$\operatorname{osc}_{4r}(w) \le c_8\Big(\sup_{4r} w - \sup_{r} w\Big) \le c_8(\operatorname{osc}_{4r} w - \operatorname{osc}_r(w)). \tag{5.66}$$

Recapitulating, we see that there exists some $0 < \lambda < 1$ depending only on n such that if $P(x_0, t_0, 4r) \subseteq \Omega$ and $\omega_i^{\pm}(r) = \operatorname{osc}_{P(x_0,t_0,r)}(w_i^{\pm})$, then:

(1) For every $i \in \{1, \dots, n\}$, we have

$$\omega_i^{\pm}(r) \le \omega_i^{\pm}(4r). \tag{5.67}$$

(2) If k is chosen such that $\operatorname{osc}_{P(x_0,t_0,4r)}(\partial_k u) \ge \operatorname{osc}_{P(x_0,t_0,4r)}(\partial_i u)$ for $i \in \{1, \dots, n\}$, then for some sign $\sigma \in \{+,-\}$ we have

$$\omega_k^{\sigma}(r) \le \lambda \omega_k^{\sigma}(4r), \tag{5.68}$$

$$8n\varepsilon \operatorname{osc}_{P(x_0,t_0,4r)}(\partial_k u) \le \omega_k^{\sigma}(4r) \le 12n\varepsilon \operatorname{osc}_{P(x_0,t_0,4r)}(\partial_k u). \tag{5.69}$$

After proving the oscillation decay estimate (5.68), we can easily conclude the proof of the theorem. Let $z_1 = (x_1, t_1)$, $z_2 = (x_2, t_2) \in \Omega$ and recall that $d_{z_1} = d(z_1, \partial\Omega)$ and $d_{z_1,z_2} = \min(d_{z_1}, d_{z_2})$. If $d(z_1, z_2) \ge \frac{1}{4} d_{z_1,z_2}$ the conclusion of the theorem holds trivially, so we may assume that $d(z_1, z_2) < \frac{1}{4} d_{z_1,z_2}$. Assume without loss of generality that $t_1 \ge t_2$, set $R = \frac{1}{4} d_{z_1,z_2}$, and let $m \in \mathbb{N}$ be such that

$$4^{-m-1} R \le d(z_1, z_2) < 4^{-m} R. \tag{5.70}$$

Thus we have a sequence of parabolic balls

$$z_1, z_2 \in P(z_1, 4^{-m} R) \subseteq P(z_1, 4^{-m+1} R) \subseteq \cdots \subseteq P(z_1, R) \subseteq \Omega. \tag{5.71}$$

Setting $(x_0, t_0) = z_1$ so that $\omega_i^{\pm}(r) = \operatorname{osc}_{P(z_0,r)}(w_i^{\pm})$, we can compare oscillations in $P(z_1, 4^{-j+1} R)$ and $P(z_1, 4^{-j} R)$ for $j = 1, 2, \dots, m$ according to (5.67), (5.68) and (5.69). Since at any step one of the $w_i^{\pm}$ decreases by a factor of λ, and as there are only $2n$ possibilities for index and sign, we can assume without loss of generality that (5.68), (5.69) happened with $k = 1$ and $\sigma = +$ at least $m/(2n)$ times. This happened for the last time comparing the oscillations for the parabolic balls $P(z_1, 4^{-j+1}R)$ and $P(z_1, 4^{-j}R)$ with $j \geq m/(2n)$. Thus,

$$\begin{aligned}\operatorname{osc}_{P(z_1,4^{-j+1}R)}(\partial_i u) \leq \operatorname{osc}_{P(z_1,4^{-j+1}R)}(\partial_1 u) &\leq \frac{C}{\varepsilon}\omega_1^+(4^{-j+1}R) \\ &\leq \frac{C}{\varepsilon}\lambda^{m/2n}\omega(R) \\ &\leq C\lambda^{m/2n}\varepsilon.\end{aligned} \tag{5.72}$$

On the other hand, $d(z_1, z_2) \geq 4^{-m-1}R$ implies

$$m \geq \frac{1}{\log 4}\log\Big(\frac{R}{4d(z_1, z_2)}\Big), \tag{5.73}$$

so with a suitable choice of $0 < \alpha < 1$ this and (5.72) imply

$$|\nabla u(z_1) - \nabla u(z_2)| \leq C\left(\frac{d(z_1, z_2)}{d_{z_1,z_2}}\right)^{\alpha}\varepsilon \tag{5.74}$$

as required. □

5.4 Proof of the main estimate

Before coming back to proving the main estimate, we first derive a crude estimate for the higher derivatives. This estimate ignores the fact that we have control over a thick cylinder and treats it as a union of small parabolic balls.

Lemma 5.75 *There exists a constant c such that if $\sup_{B(p,r)\times[0,\tau r^2]}|\nabla u| \ll 1$, then*

$$d_z|\nabla^2 u(z)| \leq c, \tag{5.76}$$

$$d_{z_1,z_2}^{1+\alpha}\frac{|\nabla^2 u(z_1) - \nabla^2(u(z_2))|}{d(z_1, z_2)^{\alpha}} \leq c. \tag{5.77}$$

Proof The first estimate is a direct application of the Ecker–Huisken curvature estimate of Theorem 2.18 for balls of radius d_{z_1}. For the second part, if $d(z_1, z_2) \geq \frac{1}{4}d_{z_1,z_2}$, as before there is nothing to prove. Otherwise, using Theorem 2.18 once more, we obtain

$$d_z^2|\nabla^3 u(z)| \leq c, \quad d_z^3|\partial_t\nabla^2 u(z)| \leq c, \tag{5.78}$$

so by integrating, first along space and then along time, we get

$$\begin{aligned}|\nabla^2 u(z_1) - \nabla^2 u(z_2)| &\leq \frac{c}{d_{z_1,z_2}^2}d(z_1, z_2) + \frac{c}{d_{z_1,z_2}^3}d(z_1, z_2)^2 \\ &\leq \frac{c}{d_{z_1,z_2}^2}d(z_1, z_2)^{\alpha}d_{z_1,z_2}^{1-\alpha}.\end{aligned} \tag{5.79}$$

□

Proof of Theorem 5.1 Assume again that $r = 1$. Let $\tau_0(M,\delta)$, $\gamma_0 = \gamma_0(\delta)$ and $\lambda_0 = \lambda_0(M,\delta)$ be the constants from Theorem 5.12. Fix $\lambda < \lambda_0$, $\tau < \tau_0$ and $\gamma < \gamma_0$ such that the assumptions of the theorem are satisfied. The graphical mean curvature equation is of the form

$$u_t - \Delta u = f, \tag{5.80}$$

with

$$f = \frac{(\partial_i u)(\partial_j u)}{1+|\nabla u|^2}\partial_i\partial_j u. \tag{5.81}$$

Setting α as in Theorem 5.47, by Theorem 5.12 we have

$$\begin{aligned}\sqrt{\tau}|\nabla^2 v(0,\tau)| \\ \le (c+\delta)\varepsilon+\frac{C}{\sqrt{\tau}}\biggl(\sup_{z_1\in P^{1,\tau}} d_{z_1}^2|f(z_1)|+\sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha}\biggr).\end{aligned} \tag{5.82}$$

By our assumptions and Lemma 5.75,

$$d_z^2|f(z)| \le C d_z(d_z|\nabla^2 u|)|\nabla u|^2 \le C\sqrt{\tau}(M\varepsilon)^2. \tag{5.83}$$

Similarly, by our assumptions, the Hölder gradient estimate of Theorem 5.47 and Lemma 5.75,

$$\sup_{z_1,z_2\in P^{1,\tau}} d_{z_1,z_2}^{2+\alpha}\frac{|f(z_2)-f(z_1)|}{d(z_2,z_1)^\alpha} \le C\sqrt{\tau}(M\varepsilon)^2. \tag{5.84}$$

Thus, the contribution of the non-linearity is at most quadratic in the gradient (for small gradients) so for $\varepsilon < \varepsilon_0$ it will be smaller than δ. This concludes the proof of the first estimate. The proof of the second estimate is similar, utilizing the second estimate of Theorem 5.12. □

References

[1] **Y G Chen**, **Y Giga**, **S Goto**, *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations*, J. Differential Geom. 33 (1991) 749–786 MR

[2] **J Clutterbuck**, *Parabolic equations with continuous initial data*, PhD thesis, Australian National University, Canberra (2004) arXiv

[3] **T H Colding**, **W P Minicozzi, II**, *A course in minimal surfaces*, Graduate Studies in Mathematics 121, Amer. Math. Soc., Providence, RI (2011) MR

[4] **K Ecker**, *Regularity theory for mean curvature flow*, Progress in Nonlinear Differential Equations and their Applications 57, Birkhäuser, Boston, MA (2004) MR

[5] **K Ecker**, **G Huisken**, *Mean curvature evolution of entire graphs*, Ann. of Math. 130 (1989) 453–471 MR

[6] **K Ecker**, **G Huisken**, *Interior estimates for hypersurfaces moving by mean curvature*, Invent. Math. 105 (1991) 547–569 MR

[7] **L C Evans**, *Partial differential equations*, 2nd edition, Graduate Studies in Mathematics 19, Amer. Math. Soc., Providence, RI (2010) MR

[8] **L C Evans**, **J Spruck**, *Motion of level sets by mean curvature, I*, J. Differential Geom. 33 (1991) 635–681 MR

[9] **A Friedman**, *Partial differential equations of parabolic type*, Prentice-Hall, Englewood Cliffs, NJ (1964) MR

[10] **D Gilbarg**, **N S Trudinger**, *Elliptic partial differential equations of second order*, revised 2nd edition, Grundl. Math. Wissen. 224, Springer, Berlin (1998) MR

[11] **G Hong**, **L Wang**, *A geometric approach to the topological disk theorem of Reifenberg*, Pacific J. Math. 233 (2007) 321–339 MR

[12] **T Ilmanen**, *The level-set flow on a manifold*, from "Differential geometry: partial differential equations on manifolds" (R Greene, editor), Proc. Sympos. Pure Math. 54, Amer. Math. Soc., Providence, RI (1993) 193–204 MR

[13] **T Ilmanen**, *Elliptic regularization and partial regularity for motion by mean curvature*, Mem. Amer. Math. Soc. 520, Amer. Math. Soc., Providence, RI (1994) MR

[14] **H Koch**, **T Lamm**, *Geometric flows with rough initial data*, Asian J. Math. 16 (2012) 209–235 MR

[15] **J Lauer**, *A new length estimate for curve shortening flow and low regularity initial data*, Geom. Funct. Anal. 23 (2013) 1934–1961 MR

[16] **G M Lieberman**, *Second order parabolic differential equations*, World Scientific, River Edge, NJ (1996) MR

[17] **C Mantegazza**, *Lecture notes on mean curvature flow*, Progress in Mathematics 290, Birkhäuser, Basel (2011) MR

[18] **E R Reifenberg**, *Solution of the Plateau Problem for m–dimensional surfaces of varying topological type*, Acta Math. 104 (1960) 1–92 MR

[19] **M Simon**, *Deformation of C^0 Riemannian metrics in the direction of their Ricci curvature*, Comm. Anal. Geom. 10 (2002) 1033–1074 MR

[20] **T Toro**, *Doubling and flatness: geometry of measures*, Notices Amer. Math. Soc. 44 (1997) 1087–1094 MR

[21] **M-T Wang**, *The mean curvature flow smoothes Lipschitz submanifolds*, Comm. Anal. Geom. 12 (2004) 581–599 MR

Department of Mathematics, Stanford University
380 Serra Mall, Stanford, CA 94305, United States

or.hershkovits@stanford.edu

Proposed: Tobias H. Colding
Seconded: Bruce Kleiner, John Lott

Received: 19 May 2015
Accepted: 26 December 2015