A STOCHASTIC TARGET APPROACH TO RICCI FLOW ON SURFACES

ROBERT W. NEEL AND IONEL POPESCU

ABSTRACT. We develop a stochastic target representation for Ricci flow and normalized Ricci flow on smooth, compact surfaces, analogous to Soner and Touzi's representation of mean curvature flow. We prove a verification/uniqueness theorem, and then consider geometric consequences of this stochastic representation.

Based on this stochastic approach we give a proof that, for surfaces of non-positive Euler characteristic, the normalized Ricci flow converges to a constant curvature metric exponentially quickly in every C^k-norm. In the case of C^0 and C^1-convergence, we achieve this by coupling two particles. To get C^2-convergence (in particular, convergence of the curvature), we use a coupling of three particles. This triple coupling is developed here only for the case of constant curvature metrics on surfaces, though we suspect that some variants of this idea are applicable in other cases too. At any rate, this triple coupling provides a purely probabilistic approach to getting second-order derivative estimates for second-order PDEs. Finally, for $k \geq 3$, the C^k-convergence follows relatively easily using induction and coupling of two particles.

1. INTRODUCTION

In [23], Soner and Touzi give a characterization of various extrinsic geometric flows (with ambient space $\mathbb{R}^n$), including mean curvature flow, as stochastic target problems. More specifically, they introduce the relevant target problems and then prove associated verification theorems, namely theorems showing that if the curvature flow has a smooth solution for an interval of time $t \in [0, T)$, then the solution agrees with the solution to the stochastic target problem on this interval. In the first part of this paper, we develop a similar characterization of Ricci flow (and normalized Ricci flow) on compact surfaces, including the relevant verification theorems (see Theorem 3). We then briefly brief discuss time-dependent bounds on the solution to both normalized and un-normalized Ricci flow and estimates on the blow-ups of solutions to Ricci flow in the cases of non-zero Euler characteristic, all obtained from the stochastic formulation of the flow. In the remainder of the paper, we use this stochastic representation to prove that, for a smooth, compact surface of non-positive Euler characteristic, given that a smooth solution to the normalized Ricci flow exists for all time (which is well-known from the literature), it converges to a constant curvature metric exponentially fast in C^∞ (see Theorem 22 for a precise statement).

Ricci flow on smooth, compact surfaces is essentially completely understood, so none of the geometric results in this paper are new. Nonetheless, one feature of our approach is that probability often provides an appealing intuition, as in the case of Brownian motion and heat flow. Thus, if Ricci flow is thought of as a kind of "heat equation for curvature," it is natural to want to extend the analogy to include a diffusion interpretation. For example, it's nice to see the convergence of a

2010 *Mathematics Subject Classification.* Primary 58J65, 53C44; Secondary 60H30.

Key words and phrases. Ricci flow, stochastic target problem, coupling.

The first author gratefully acknowledges support from an NSF Postdoctoral Research Fellowship. The second author was partly supported by the Marie Curie Action Grant PIRG.GA.2009.249200 and also by a grant of the Romanian National Authority for Scientific Research CNCS - UEFISCDI, project number PN-II-RU-TE-2011-3-0259.

manifold under normalized Ricci flow to a constant curvature limit as the equidistribution of the metric, and as a result of the curvature, from a probabilistic perspective.

More generally, one might ask about the potential merits of developing stochastic techniques for Ricci flow (or other curvature flows). One obvious point to be made here is that one gets a representation of the solution and, at least in the theory of linear second order PDEs, this has turned out to be extremely versatile in extracting properties of the solutions. As we will see, the stochastic tools we employ are good enough to give a different proof of a main result in the theory of Ricci flow on surfaces with the bonus that we see the "averaging property of the curvature" as a consequence of coupling, which is a probabilistic manifestation of ergodicity. Another motivation for such an endeavor is that the stochastic target formulation is fairly insensitive to regularity, and thus potentially useful in formulating notions of weak solutions. Indeed, in a second paper, Soner and Touzi [22] show that generalized solutions to various extrinsic curvature flows can also be understood in terms of stochastic target problems. Also stemming from these ideas, we note that stochastic approaches to PDEs can lend themselves to the development of probabilistic numerical schemes (as in [12]), but we do not touch this subject here.

Our framework is not the most general one. We presumably could have worked in a little more generality, but to keep the ideas as appealing and clear as possible, we decided to study surfaces, which are the traditional starting point for studying Ricci flow.

We point out that, as noted in [6], stochastic target problems of certain kind are equivalent to second-order backward stochastic differential equations. As discussed there, second-order backward SDEs are natural stochastic objects to associate with fully non-linear PDEs. Thus, one could presumably recast the results of this paper in those terms. Nonetheless, we have chosen to adopt the stochastic target approach because it seems more geometrically intuitive and visually appealing, and because it puts Ricci flow and mean curvature flow in a similar framework.

There are few papers on stochastic analysis and Ricci flow, for instance [19, 18, 2, 17, 9, 1]. The ones that are somewhat close to our work are [9] and [1]. These papers investigate the Brownian motion (and the associated parallel transport) with respect to a time changing metric on a manifold of any dimension, not only on surfaces. Using stochastic analysis they also develop a Bismut-like formula to represent the gradient of solutions to heat-type flows with respect to the time-dependent metric. In particular, this leads to gradient estimates for the corresponding solutions.

At any rate, though the Bismut formula is very useful for gradient estimates, we do not know how to get a nice and useful version of this formula for second-order derivatives. This is one of the reasons we prefer to deal with an alternative probabilistic tool, namely coupling. This idea for dealing with the second-order derivatives comes from [10], where a coupling of three particles is used to estimate second-order derivatives of harmonic functions on Euclidean domains. This triple coupling indicated by Cranston uses a certain symmetry to get a key cancellation in the estimation of the Hessian. This symmetry is not surprising in the flat case. However, there are obvious technical challenges for a similar construction on manifolds, and the way it works in the flat case does not seem to work on arbitrary manifolds. Nevertheless, it turns out that we can construct such a triple coupling which has enough good properties in the case of surfaces of constant curvature.

We continue with a few more observations about the present work. We do not prove the existence of solutions to the target problem directly; rather, the verification theorems proceed from the assumption that the Ricci flow admits a smooth solution. In the case of normalized Ricci flow, we have long-time existence as proved in [3] and [14]. However, an immediate consequence of such a verification theorem is that the solution (to the flow) is unique.

In contrast to the standard proof of the convergence to constant curvature, we deal directly with the metric itself (and its derivatives), rather than introducing an auxiliary PDE satisfied by the curvature. We use uniformization to work with an underlying metric which has constant curvature and is in the same conformal class as the initial metric. One might hope to extend these arguments to more general situations, but for a first paper on this approach uniformization makes the analysis cleaner and reveals the power of the coupling in a nice way.

The outline of the paper is as follows. We first describe the stochastic target problem in Section 2 giving a fair amount of detail, since it is a somewhat non-standard control problem. Then, in Section 3 we prove the verification/uniqueness theorem, namely that, if there is a smooth solution to the Ricci flow (or normalized Ricci flow) on some interval of time, then it agrees with the solution to the stochastic target problem.

Section 4 is a short section showing how one can use the representation to prove that the unnormalized Ricci flow develops singularities (in certain cases) either in finite time or in infinite time. In Section 5, we develop the a priori bounds for the stochastic target problem. As a consequence, we obtain the exponential convergence in the C^0-norm of the normalized flow in the case of $\chi(M) < 0$ (as usual, $\chi(M)$ denotes the Euler characteristic of M). We also include a short discussion of the blow up of the unnormalized Ricci flow in the cases $\chi(M) > 0$ and $\chi(M) < 0$, which is in tune with the previous section's findings, although this time assuming uniformization.

Section 6 introduces and proves the main result on mirror coupling for the time changed Brownian motions associated to the target problems. This coupling makes sense for short times, but the main challenge is to show that the coupling extends beyond the cut locus. This is done using the geometric structure of the cut locus on surfaces of Euler characteristic less than or equal to 0. We should also point out that there is a coupling of Brownian motions constructed with respect to time-varying metrics (such as Ricci flow) in [17], but it differs from our situation here.

In Section 7 we start the main analysis of the convergence of normalized Ricci flow. We prove the nontrivial fact that in Euler characteristic zero, the normalized flow converges exponentially fast in the C^0-topology. This uses the result from the previous section combined with the comparison of the distance process with a Bessel process in order to estimate the coupling time, which is a fundamentally probabilistic idea. Combining this result with those coming from the a priori estimates proves that, for non-positive Euler characteristic, the flow convergence in the C^0-topology exponentially fast.

The next task is to prove that the convergence takes place also in C^1, or in other words that the gradient of the metric converges exponentially fast. This is done in Section 8, again using coupling. However, the point here is a little different. We use the coupling for particles started close to one another and estimate the coupling time in terms of the gradient of the metric and the initial distance. This in turn yields a functional inequality satisfied by the C^0-norm of the gradient which is contained in Lemma 12. It turns out that this functional inequality is strong enough to produce the exponential convergence.

Going forward, Section 9 is dedicated to the triple coupling used in a crucial way for the Hessian estimates. We exploit in an essential way the constant curvature properties of the underlying metric. We have two mirror coupled particles x and y and another middle particle z which is moving on the geodesic between them which is described by the distance ρ_1 from z to x, or alternatively, the distance ρ_2 from z to y. One of the main interests is the symmetry with respect to swapping ρ_1 and ρ_2. The other thing thrust of the investigation is as follows. Assuming that x and y are time changed Brownian motions, we study the conditions under which z is a time changed Brownian motion with a drift. This is a key point in the Hessian estimates.

Section 10 covers the Hessian estimates. Here we use the results from the previous sections, for instance, the exponential decay of the flow in the C^1-topology and the triple coupling. As in the case of the gradient we end up with a functional inequality for the C^0-norm of the Hessian as in Lemma 21. It turns out that this suffices to conclude the exponential convergence.

The last section proves the C^k-convergence of the flow. This is done essentially using the Ricci flow equation and induction. It is important to mention here that in the flat case, we still have to use the coupling.

A few words about the sphere case, which definitely requires some finer analysis. There are several obstacles we have to overcome. On one hand, the a priori estimates give bounds which blow up in finite or infinite time. However, these estimates are simply bounds of a stochastic differential equation in terms of the ODE in which the martingale is killed off, and eventually can likely be refined. Further, in the case of non-positive Euler characteristic, there is a unique stationary solution to the normalized Ricci flow with a given volume (in a given conformal class), and thus one has to prove that the flow converges to this uniquely determined solution. In the case of the sphere this is not the case and thus convergence is harder to establish, because we do not know beforehand toward which stationary solution the flow wants to converge (this is related to the issue of Ricci solitons). Therefore, the strategy we used in this paper for $\chi(M) \leq 0$ needs some refinements if it's to address the case of positive Euler characteristic.

2. Stochastic target formulation

2.1. **Ricci flow.** Consider a smooth, compact Riemannian surface (M, h), that is, M is a smooth, compact manifold of dimension two and h a smooth Riemannian metric on M. Any other smooth metric in the same conformal class as h can be written as $g = \overline{u}h$ for some smooth, positive function $\overline{u}$. The Ricci curvature of any metric metric g is given by

$$2\mathrm{Ric}_g = R_g = 2K_g g, \tag{1}$$

where R_g is the scalar curvature and K_g is the Gauss curvature. The *Ricci flow* is defined as the evolution of the metric g_t according to

$$\partial_t g_{ij} = -2\mathrm{Ric}_{ij} \tag{2}$$

where Ric is the Ricci tensor. From this, it is easy to see that the Ricci flow preserves the conformal class in two dimensions, and thus it becomes an evolution equation for the conformal factor $\overline{u}_t$. In particular, the Ricci flow corresponds to $\overline{u}$ evolving by

$$\partial_t \overline{u}_t = \Delta_h \log \overline{u}_t - 2K_h \tag{3}$$

where K_h is the Gauss curvature of (M, h). In passing from (2) to (3), we have already used the fact that if $g = uh$, for two metrics, g and h, then (see [8, Exercise 2.8])

$$R_g = \frac{1}{u}(R_h - \Delta_h \log u) \tag{4}$$

where the Δ_h is the Laplacian with respect to the metric h.

This is a non-linear parabolic equation, and thus the usual probabilistic methods of solution (diffusions, Feynman-Kac, etc.) don't apply. Instead, we will adopt a stochastic target approach modeled on the approach of [23] to mean curvature flow, as mentioned above.

To be more concrete, we assume that the initial metric on M can be written as $g_0 = \overline{u}_0 h$ for some smooth, positive $\overline{u}$ and some metric h. There are two natural choices for h. Of course, we can let $h = g_0$ and $\overline{u}_0 \equiv 1$. Alternatively, the uniformization theorem implies that there is a metric in the same conformal class as g_0 which has constant curvature of -1, 0, or 1. Then we can take h to be this metric, in which case $\overline{u}_0$ is determined by the condition that $g_0 = \overline{u}_0 h$. We will find the flexibility of this set-up to be useful.

As usual, we also wish to introduce the *normalized Ricci flow*, which is defined as

$$\partial_t g_{ij} = -2\mathrm{Ric}_{ij} + 2rg_{ij} \tag{5}$$

where r is the average of the Gauss curvature on M with respect to the metric g. Written in terms of the conformal factor, this is

$$\partial_t \bar{u}_t = \Delta_h \log \bar{u} - 2K_h + 2r_t\bar{u}_t.$$

Under this flow, the surface is continually rescaled to preserve the area. Indeed, the Gauss-Bonnet Theorem tells us that the integral of the scalar curvature is

$$\int K_g dA_g = 2\pi\chi(M)$$

where $\chi(M)$ is the Euler characteristic of M and A_g is the area element of the metric g. Consequently, if r_t is the average of the Gauss curvature for g_t, then

$$r_t = \frac{2\pi\chi(M)}{\mathrm{area}(M, g_t)}$$

where $\mathrm{area}(M, g)$ stands for the area of M with the metric g. From here, a straightforward calculation gives that

$$\partial_t \,\mathrm{area}(M, g_t) = \partial_t \int \bar{u}_t dA_h = \int \partial_t \bar{u}_t dA_h = -2\int K_h dA_h + 2r_t \int \bar{u}_t dA_h = 0,$$

which shows that the area is preserved under this evolution and, in particular, r_t does not depend on t. Therefore the flow (5) preserves the area and

$$r = \frac{2\pi\chi(M)}{\mathrm{area}(M, g_0)}. \tag{6}$$

We can now translate (5) into an equation satisfied by the conformal change $\bar{u}_t$ as (recall that $g_t = \bar{u}_t h$)

$$\partial_t \bar{u}_t = \Delta_h \log \bar{u} - 2K_h + 2r\bar{u}_t \tag{7}$$

with r the constant from (6).

As is implicit in the above, we see that the set of all smooth metrics (on M) in a given conformal class corresponds to the set of smooth sections of a one-dimensional bundle over M. More concretely, fixing a "reference metric" h and writing any other (smooth) metric (in the same conformal class) as $\overline{u}h$ induces a global coordinate u on fibers of this bundle making the total space E diffeomorphic to $M \times (0, \infty)$. Further, $\overline{u}$ is given as the composition of the lift from M to E (corresponding to the section) with u. This helps to explain the notation: u is a coordinate on the fibers, and $\overline{u}$ is the expression of a section in this coordinate. Because our bundle admits natural global coordinates, we will almost always work in these coordinates, and thus we won't have much occasion to consider sections in a coordinate-free notation.

Viewed in this light, it is natural to introduce a new coordinate on the fibers. Let $p = (1/2)\log u$. Then any other metric in the same conformal class as h can be written as $g = e^{2\overline{p}}h$ for some smooth function $\overline{p} : M \to \mathbb{R}$, which is given by the composition of the lift $M \to E$ (corresponding to the section) with p. This coordinate makes the bundle into a real line bundle. In particular, the metric h corresponds to the zero section, and fiberwise addition corresponds to composition of conformal changes. However, we won't need the vector space structure on fibers in what follows; we really just view the fibers as having a smooth structure. In terms of the coordinate p, the Ricci flow equation becomes

$$\partial_t \overline{p}_t = e^{-2\overline{p}_t} \left(\Delta_h \overline{p}_t - K_h \right), \tag{8}$$

and the normalized Ricci flow equation becomes (see also [20, Equation 1.3.1])

$$\partial_t \overline{p}_t = e^{-2\overline{p}_t} \left(\Delta_h \overline{p}_t - K_h \right) + r. \tag{9}$$

with r the constant defined in (6) and thus depending only on the are of M with respect to the initial metric g_0.

At this point, we see that there is a one-to-one correspondence between metrics in the same conformal class as h, sections of E over M, and functions $\overline{p}$ (where all of these objects are assumed to be smooth). Further, there is a one-to-one correspondence between smooth sections and smooth hypersurfaces of E that intersect each fiber once and do so transversely; under composition with p this is the same as the correspondence between smooth functions on M and their graphs in $M \times \mathbb{R}$. Viewing metrics as hypersurfaces in the total space E provides a framework for studying Ricci flow which is fairly similar to that of mean curvature flow and well-suited for the stochastic target approach. Our next task is to define the appropriate target problem.

2.2. **The target problem.** Let $\Gamma(0)$ be the hypersurface corresponding to the initial metric g_0. In spite of our previous efforts to distinguish between sections over M from their description in a particular coordinate, in what follows we will fix the global coordinate p on fibers, thus identifying the fibers with $\mathbb{R}$, and formulate everything in those terms. In particular, $\Gamma(0)$ corresponds to the graph of $\overline{p}_0$. The stochastic target problem is, for any time t, the problem of determining the set of points such that the controlled process, starting from such a point, can be made to hit $\Gamma(0)$ (the "target") in time t almost surely. Obviously, this requires specifying the allowed controls and the processes they give rise to. We will generally explain things for the Ricci flow and then indicate the analogous results for the normalized Ricci flow in situations where there are no additional complications.

We start with the infinitesimal picture in normal coordinates. We choose any point $(q, \hat{p}) \in M \times \mathbb{R}$ and let (x_1, x_2) be normal coordinates around q. Thus (x_1, x_2, p) are coordinates on a neighborhood of $\{q\} \times \mathbb{R}$. We assume that the controlled process is currently at $(q, \hat{p})$, say at time τ. The (x_1, x_2)-marginal of the controlled process will be (infinitesimally) Brownian motion on M (with fixed reference metric h), time-changed by $2e^{-2\hat{p}}$. The control consists of choosing a lift of the tangent plane to M at q into the tangent space to E at $(q, \hat{p})$. The controlled process has its martingale part diffusing (infinitesimally) along this lifted plane in the unique way that gives the right (x_1, x_2)-marginal, and has its drift along the fiber at rate $e^{-2\hat{p}} K_h$ (plus an additional $-2\pi\chi(M)/\operatorname{area}(M,h)$ for the normalized Ricci flow). More precisely, the control consists of a choice of $(a_1, a_2) \in \mathbb{R}^2$, for which the processes evolves (infinitesimally, assuming the process is at $(q, \hat{p})$ at time τ) according to

$$\begin{bmatrix} dx_{1,\tau} \\ dx_{2,\tau} \\ dp_\tau \end{bmatrix} = \begin{bmatrix} e^{-\hat{p}} & 0 \\ 0 & e^{-\hat{p}} \\ e^{-\hat{p}} a_1 & e^{-\hat{p}} a_2 \end{bmatrix} \begin{bmatrix} \sqrt{2}\, dW^1_\tau \\ \sqrt{2}\, dW^2_\tau \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ e^{-2\hat{p}} K_h(q) \end{bmatrix}.$$

where W^1 and W^2 are one-dimensional Brownian motions. Here we have written $K_h(q)$ to emphasize that the curvature depends on the point in M. The $\sqrt{2}$ factors (in front of the Brownian differentials) are needed because the Ricci flow is defined using the Laplacian, instead of half the Laplacian, and rather than use a non-standard normalization for the Ricci flow, we choose to speed up our Brownian motions (this is analogous to the usual discrepancy between the analysts' and the probabilists' versions of the heat equation). This is the controlled process, at least infinitesimally, corresponding to the Ricci flow. For the normalized Ricci flow, the set of controls is the same, but the process evolves according to

$$\begin{bmatrix} dx_{1,\tau} \\ dx_{2,\tau} \\ dp_\tau \end{bmatrix} = \begin{bmatrix} e^{-\hat{p}} & 0 \\ 0 & e^{-\hat{p}} \\ e^{-\hat{p}}a_1 & e^{-\hat{p}}a_2 \end{bmatrix} \begin{bmatrix} \sqrt{2}\, dW^1_\tau \\ \sqrt{2}\, dW^2_\tau \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ e^{-2\hat{p}}K_h(q) - r \end{bmatrix}.$$

We point out that, for both the Ricci flow and the normalized Ricci flow, the (infinitesimal) diffusion matrix is

$$\begin{bmatrix} 2e^{-2p_\tau} & 0 & 2e^{-2p_\tau}a_1 \\ 0 & 2e^{-2p_\tau} & 2e^{-2p_\tau}a_2 \\ 2e^{-2p_\tau}a_1 & 2e^{-2p_\tau}a_2 & 2e^{-2p_\tau}(a_1^2+a_2^2) \end{bmatrix}$$

in (x_1, x_2, p) coordinates at $(q, \hat{p})$, of course.

Having given the infinitesimal picture, we now extend this to a global description. While it is tempting to simply assert that this follows immediately from the local description, we prefer to give a more explicit formulation. There is more than one way to do this, but we choose to use the bundle of orthonormal frames on (M, h). The immediate difficulty with extending the above local picture is that, except in special cases (more on which below), we cannot find coordinates which are normal at more than one point at a time, or even a global orthonormal frame. The solution we have in mind is to use the bundle of orthonormal frames to supply each point along the evolving process with an orthonormal frame and its associated normal coordinates. In particular, let $\mathcal{O}(M)$ be the bundle of orthonormal frames over (M, h), consisting of points $(q, \mathfrak{e}(q))$ where $q \in M$ and $\mathfrak{e}(q)$ is an orthonormal basis for T_qM with metric h. We identify $\mathfrak{e}(q)$ with the corresponding linear isometry from $\mathbb{R}^2$ to T_qM. Let e_1 and e_2 be the standard basis for $\mathbb{R}^2$ and let $\mathfrak{E}(e_i)$ be the corresponding canonical vector fields. Further, we let $\pi : \mathcal{O}(M) \to M$ be the usual projection and $\pi_* : T\mathcal{O}(M) \to TM$ be the induced push-forward map on tangent spaces.

The connection with the previous infinitesimal picture comes from the following relationship between the canonical vector fields and normal coordinates. Choose a point $q \in M$ and a frame $\mathfrak{e}(q)$ over q, and let (x_1, x_2) be normal coordinates (for (M, h)) in a neighborhood of q such that $\partial_{x_i} = \mathfrak{e}(q)(e_i)$ at q. Obviously, $\pi_*\left[\mathfrak{E}(e_i)|_{(q,\mathfrak{e}(q))}\right] = \partial_{x_i}|_q$. Moreover, let s be a smooth section of $\mathcal{O}(M)$ in a neighborhood of q which is equal to $\mathfrak{e}(q)$ at q and horizontal at q, meaning that $\partial_{x_i}s$ are horizontal vectors at q. Then $\pi_*\left[\mathfrak{E}(e_i) \circ s\right]$ agrees with ∂_{x_i} to first-order around q. (Indeed, to show that such a section s exists, start with normal coordinates and apply the Gram–Schmidt process to $\{\partial_{x_1}, \partial_{x_2}\}$ at every point in a neighborhood of q.)

We also recall the connection between the bundle of orthonormal frames and Brownian motion on (M, h). We have that $\left(\mathfrak{E}(e_1)^2 + \mathfrak{E}(e_2)^2\right)/2$ is Bochner's Laplacian on $\mathcal{O}(M)$, and the corresponding martingale problem is well-posed (in the sense of Stroock and Varadhan, namely that there is a unique solution for any initial point). We use $\tilde{B}_\tau$ to denote such a process. Projecting $\tilde{B}_\tau$ to M gives Brownian motion on M, which we denote B_τ. This is the well-known Eells-Elworthy-Malliavin construction of Brownian motion on M and we refer the reader to [15] or[24] for a detailed account on the subject. Moreover, the process $\tilde{B}_\tau$ on $\mathcal{O}(M)$ should be thought of as the horizontal lift of B_τ on M, and thus as giving Brownian motion equipped with parallel transport. In particular, this is how we will typically understand $\tilde{B}_\tau$, as Brownian motion on M endowed with parallel transport. Finally, we note that the solution to the martingale problem for Bochner's Laplacian can be realized as the (unique) strong solution to the natural SDE driven by a standard Brownian motion on $\mathbb{R}^2$, or equivalently, two independent, one-dimensional Brownian motions. That is, $\tilde{B}_\tau$

can be realized as the solution to

$$d\tilde{B}_\tau = \mathfrak{E}(e_1) \circ dW_\tau^1 + \mathfrak{E}(e_2) \circ dW_\tau^2$$

where $\circ dW$ indicates that the differential is to be understood in the Stratonovich sense.

We now have the necessary background to give the global formulation of the stochastic target problem for Ricci flow (and the related target problem for normalized Ricci flow). We write points in E as $(x,p) \in M \times \mathbb{R}$ and the controlled process (for the Ricci flow) as $Y_\tau = (x_\tau, p_\tau)$. As suggested above, the M-marginal x_τ will be Brownian motion on M, time-changed by p, and thus we know from the above that we have parallel transport of frames (for T_xM) along the paths x_t (note that the frame is always orthonormal relative to the metric h). In particular, if we choose a frame $\mathfrak{e}(x_0)$ at the starting point, then we let $\mathfrak{e}(x_\tau)$ denote the parallel transport of this frame along x_τ. Abstractly, the control consists in choosing a lift of $T_{x_\tau}M$ to $T_{(x_\tau,p_\tau)}E$. In terms of our evolving frame, such lifts can be identified with points of $\mathbb{R}^2$. This the time to formally introduce the control process. In what follows, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space where the Brownian motion (W^1, W^2) is defined and the reference filtration involved here is $\mathcal{F}_\tau$, the one generated by the Brownian motion.

Definition 1. *For a fixed time $t > 0$, an admissible control process A is a bounded map $A : [0,t] \times M \times \Omega \to \mathbb{R}^2$ which is continuous in the first two coordinates, and such that for each $(x,\tau) \in M \times [0,t]$, $A(\tau,x) : \Omega \to \mathbb{R}^2$ is $\mathcal{F}_\tau$-measurable. We write this in components $A = (a_1, a_2)$.*

We will explain below in the first remark of this section why we require the control to be bounded.

If we start our process from a point $Y_0 = (x_0, p_0)$ equipped with a frame $\mathfrak{e}(x_0)$ of $T_{x_0}M$, then it evolves according to the SDE (note that we're using both Itô and Stratonovich differentials)

$$\begin{aligned} dx_\tau &= e^{-p_\tau} \left[\sum_{i=1}^{2} \mathfrak{e}(x_\tau)(e_i)\sqrt{2} \circ dW_\tau^i \right] \\ dp_\tau &= e^{-p_\tau} \left[\sum_{i=1}^{2} a_i \sqrt{2}\, dW_\tau^i \right] + e^{-2p_\tau} K_h(x_\tau)\, d\tau. \end{aligned} \tag{10}$$

Here we see that $\mathfrak{e}(x_\tau)(e_i)$ is just the projection onto M of $\mathfrak{E}(e_i)$ and to ease the notation we will also use the shortcut $\mathfrak{e}(x_\tau)(e_i) = \mathfrak{e}_i(x_\tau)$, or even more simply $\mathfrak{e}_i$, if there is no confusion generated by dropping x_τ. In particular, the horizontal lift of x_τ, which we write $\tilde{x}_\tau = (x_\tau, \mathfrak{e}(x_\tau))$ evolves according to

$$d\tilde{x}_\tau = e^{-p_\tau} \left[\sum_{i=1}^{2} \mathfrak{E}(e_i)\sqrt{2} \circ dW_\tau^i \right] \quad \text{on } \mathcal{O}(M),$$

and the first line of (10) is just the projection of this onto M. We choose to write (10) in this fashion in order to emphasize that we're ultimately only interested in the evolution of the surface in E and not in the frame; the frame is only used as a convenience in order to express the control and the corresponding SDE. We do this despite that fact that (10) requires evolving the frame $\mathfrak{e}(x_\tau)$ as well.

The mixing of Itô and Stratonovich differentials in (10) is a result of the fact that horizontal Brownian motion (or just Brownian motion on M) is not easily written globally in Itô form. To clarify this, we give the following equivalent characterization, which is just a consequence of Itô's formula. For any smooth function $\varphi : [0,T] \times M \times \mathbb{R} \to \mathbb{R}$ (assuming that the process (x_τ, p_τ) exists for $\tau \in [0,T]$),

(11)

$$d\varphi(\tau, x_\tau, p_\tau) = e^{-p_\tau} \sum_{i=1}^{2} (\mathfrak{e}_i(x_\tau)\varphi + a_i\varphi')\sqrt{2}dW_\tau^i + \left(\partial_\tau\varphi + e^{-2p_\tau}K_h(x_\tau)\varphi' + e^{-2p_\tau}\Delta_h\varphi + e^{-2p_\tau}\sum_{i=1}^{2} a_i^2\varphi'' + 2e^{-2p_\tau}\sum_{i=1}^{2} a_i(\mathfrak{e}_i\varphi)\varphi'\right)d\tau$$

where all the "inside" functions are evaluated at (τ, x_τ, p_τ), $\mathfrak{e}_i(x)\varphi$ signifies the derivative (along $\mathfrak{e}_i(x)$) with respect to the second variable of φ, $\partial_\tau\varphi$ is the derivative with respect to τ variable, and the prime is the partial derivative with respect to p. Note that if we let (x_1, x_2) be appropriate normal coordinates at a point, then applying this to x_1, x_2, and p shows that, at that point, this agrees with the infinitesimal picture described above.

We now take a moment to discuss what we mean by asserting the the controlled process arises from the control via the SDEs just mentioned. We understand these (systems of) SDEs in the weak sense, that is the choice of driving Brownian motions (W_τ^1, W_τ^2) is part of the solution, not prescribed in advance. Of course, for an arbitrary choice of controls, a solution need not exist, and if it does, it may not be unique in law. We will have more to say about this later, after we introduce the target problem.

Now that we've specified the admissible controls A_τ and described the evolution of controlled process $Y_\tau(A)$ that a choice of control gives rise to, it's time to explain how this gives rise to a subset of E.

Definition 2. *We define the reachable set at a given time $t \in [0,\infty)$, denoted $V(t)$, to be the set of points in E for which there exists an admissible control such that the controlled process, started at this point and with this control, is in $\Gamma(0)$ at time t almost surely.*

We follow Soner and Touzi [23] in calling this the reachable set, even though it's the set of points you can reach a fixed target from, not the set of points you can reach from a fixed starting point. In order for this to be well-defined, we need to show that $V(t)$ doesn't depend on the initial choice of frame. Suppose A_τ is a control such that $Y_\tau(A)$, started from $y \in E$ with initial frame $\mathfrak{e}(y)$, hits $\Gamma(0)$ at time t almost surely (so that $y \in V(t)$). If $\tilde{\mathfrak{e}}(y)$ is any other (orthonormal) frame at y, then there is some $r \in O(2)$ such that $\mathfrak{e}(y) = r\tilde{\mathfrak{e}}(y)$. It's clear that $A_\tau r$ is such that $Y_\tau(Ar)$, started from $y \in E$ with initial frame $\tilde{\mathfrak{e}}(y)$, hits $\Gamma(0)$ at time t almost surely. Thus a point of E is in the reachable set or not independent of what frame we use to express the controlled process, and so the $V(t)$ are well-defined.

For a point in the reachable set, we will indicate the control in the definition by $\hat{A}$, if necessary indicating the point in $V(t)$ by writing $\hat{A}(x_0, p_0)$ or $\hat{A}(Y_0)$, and call it a successful control (this seems linguistically more appropriate than optimal control). In light of the fact that this depends on the initial choice of frame, a successful control should really be thought of as a family of controls indexed by $O(2)$. However, since the dependence on the initial frame is so simple and not our primary focus, we will generally gloss over this. We will also write $Y_\tau(\hat{A})$ as $\hat{Y}_\tau$. Thus, the defining property of a point in $V(t)$ and the associated successful control is that if we start the process at this point in $V(t)$, then $Y_t(\hat{A}) \in \Gamma(0)$ almost surely. This necessarily requires that, for a successful control $\hat{A}$, there exists a solution to Equation (10) and thus a corresponding process $Y_\tau(\hat{A})$ for all time $t \in [0,t]$. In particular, one might imagine that some choice of control gives rise to a solution under which p_τ blows up prior to t (x_τ can't blow up since M is compact), but such a control cannot be a successful control by definition. The definition does not require that a successful control gives rise to a solution $Y_\tau(\hat{A})$ which is unique in law, despite the fact that our notation makes it look as though Y_τ is always determined by A. (So it is conceivable that a successful control might give rise to another solution Y'_τ that doesn't almost surely hit the target.) Nonetheless, we will see below that, as long as a smooth solution to the Ricci flow exists, there is essentially only one choice of successful control starting from a given point of $V(t)$, that it is well-behaved, and that this control uniquely determines $\hat{Y}_\tau$.

Finally, we recall that the stochastic target problem is the determination of the reachable sets $V(t)$. We note that $V(0) = \Gamma(0)$; understanding $V(t)$ for positive t and its relationship to Ricci flow is the topic of the next section. Looking ahead, what we will prove is that, assuming the Ricci flow has a smooth solution for some interval of time, that solution agrees with the solution to the stochastic target problem in the sense that $V(t) = \Gamma(t)$ at all times in this interval.

Naturally, we have an analogous set-up which we associate with the normalized Ricci flow. The set of admissible controls remains the same, but now the controlled process, which we denote $Y^n_\tau(A)$ (the "n" in the superscript standing for "normalized") evolves according to

$$\begin{aligned} dx_\tau &= e^{-p_\tau}\left[\sum_{i=1}^{2}\mathfrak{e}(x_\tau)(e_i)\sqrt{2}\circ dW^i_\tau\right] \\ dp_\tau &= e^{-p_\tau}\left[\sum_{i=1}^{2}a_i\sqrt{2}\,dW^i_\tau\right] + \left(e^{-2p_\tau}K_h(x_\tau) - r\right)\,d\tau. \end{aligned} \tag{12}$$

Note that the only difference from Y_τ is that the drift of p_τ has an extra term.

We denote the corresponding reachable sets by $V^n(t)$. We also have the analog of Equation (11) where $e^{-2p_\tau}K_h$ there is replaced by $e^{-2p_\tau}K_h - r$:

$$\begin{aligned} d\varphi(\tau, x_\tau, p_\tau) =& e^{-p_\tau}\sum_{i=1}^{2}(\mathfrak{e}_i(x_\tau)\varphi + a_i\varphi')\sqrt{2}dW^i_\tau \\ &+\left(\partial_\tau\varphi + \left(e^{-2p_\tau}K_h(x_\tau) - r\right)\varphi' + e^{-2p_\tau}\Delta_h\varphi + e^{-2p_\tau}\sum_{i=1}^{2}a_i^2\varphi'' + 2e^{-2p_\tau}\sum_{i=1}^{2}a_i(\mathfrak{e}_i\varphi)\varphi'\right)d\tau. \end{aligned} \tag{13}$$

Remark. We want to discuss why we insist that our control (a_1, a_2) is in L^∞. We begin by describing a simpler situation which illustrates the essential point. Suppose we consider a real-value controlled process given by

$$dx_t = a_t\,dW_t, \quad x_0 = 1,$$

where a_t is an adapted real-valued function which serves as the control. If we consider the goal to be to make the process x_t hit 0 in within time 1 (and we stop the process when it hits 0), then we would like to assert that this is impossible, because, for instance, it would violate the martingale property of x_t. However, without some additional restriction on a_t, this will not be the case. For example, consider the following scheme for controlling the process. For $t \in [0, 1/2)$, we let a be the constant such that the process has probability $1/2$ of hitting 0 by time $t = 1/2$. It is clear that this is possible, since letting a be constant means that x_t is simply a time-changed Brownian motion, and we know that Brownian motion almost surely hits the origin in finite time, no matter where it is started from. Then at $t = 1/2$, the process has hit 0 and been stopped with probability $1/2$. If it hasn't, then $x_{1/2}$ is some positive value. Again, we can find some constant value for a, depending only on $x_{1/2}$, such that if we let a_t equal that constant for $t \in [1/2, 3/4)$, then the process hits 0 in that interval of time with probability $1/2$. Thus, by time $t = 3/4$, the process has hit 0 with probability $3/4$. Now we can iterate this procedure, at each step using up half of the remaining time, in order to get x_t to hit 0 with probability 1 by time $t = 1$. If we do this, the resulting process x_t will no longer be a martingale on the interval $t \in [0, 1]$ but instead merely a local martingale. Part of the point is that this is a simple trick. We can think of a_t as determining a time-change so that x_t is a time-changed Brownian motion, and since we know Brownian motion hits the origin in finite time, if we're allowed to speed up time as much as we'd like we can simply compress the entire lifetime of the Brownian motion prior to the first time it hits the origin into a finite interval.

We now return to the target problem we associate to Ricci flow. In light of the above, if we assumed only that (a_1, a_2) was adapted, we could imagine a similar procedure of choosing the control to be very large so that, from any starting point, we could cause it to hit $\overline{p}_{t-\tau}$ (this is a moving target, but it varies in a smooth fashion and stays bounded) by time t. Once it hits $\overline{p}_{t-\tau}$, we could then "switch" to the successful control described in the next section in order to hit $\overline{p}_0$ as time t. The result would be that every point would be in $V(t)$, which is obviously not what we want. Of course, what we've just described uses a discontinuous control, but one can imaging smoothing it to get a continuous analogue. At any rate, the underlying logic of this "bad" control justifies our wish to avoid unbounded controls.

Requiring that (a_1, a_2) be bounded prevents this kind of easy trick and forces a successful control to respect the geometry of the situation. Of course, one might imagine that there might be other, less restrictive, ways to achieve this, such as requiring the controls to be in some L^p-space for finite p or requiring some natural coordinate to be a martingale, as opposed to merely a local martingale. Indeed, if one were to extend this stochastic target formulation to include, say, non-compact surfaces, it seems like some weaker assumption on the control would be appropriate. However, for the present paper, we have no need to speculate on what other conditions one might want in other circumstances.

Remark. We close this section by noting that, in the case when (M, h) is flat (and thus either a torus or a Klein bottle), the orthonormal frame bundle is unnecessary. In particular, uniformization implies that (M, h) is isometric to $\mathbb{R}^2$ modulo the action of the group of Deck transformations Λ. If we let x_1 and x_2 be the usual Euclidean coordinates on $\mathbb{R}^2$, then $h = dx_1^2 + dx_2^2$ (after identifying M with $\mathbb{R}^2/\Lambda$). Further, (W_τ^1, W_τ^2) is Brownian motion on (M, h), once we take it modulo Λ. In this case, the set of controls are adapted, time-continuous, bounded maps into $\{(a_1, a_2) : a_i \in \mathbb{R}\}$, and the controlled process simplifies, so that it is given, for both Ricci and normalized Ricci flow, by the SDE

$$\begin{bmatrix} dx_{1,\tau} \\ dx_{2,\tau} \\ dp_\tau \end{bmatrix} = \begin{bmatrix} e^{-p_\tau} & 0 \\ 0 & e^{-p_\tau} \\ e^{-p_\tau} a_1 & e^{-p_\tau} a_2 \end{bmatrix} \begin{bmatrix} \sqrt{2}\, dW_\tau^1 \\ \sqrt{2}\, dW_\tau^2 \end{bmatrix}.$$

Convention. *Throughout this paper very often we will have a fixed time $t > 0$ so that the stochastic target problem is defined on $[0, t]$ or the (normalized) Ricci flow is defined up to time t. Since the process time is always going to be in $[0, t]$, all the stopping times involved will always be minimized with t so that the stopped process is well defined.*

Also, the constants involved in the main estimates may change from line to line in such a way that they do not depend on time t.

3. VERIFICATION AND THE CONNECTION WITH RICCI FLOW

At this point, we've described a pair of closely related stochastic target problems, namely the determination of $V(t)$ and $V^n(t)$, which we associate with Ricci flow and normalized Ricci flow, respectively. However, we've given no justification for these associations. In the present section, we prove that, under the assumption that a solution to the Ricci flow exists, the solution is given by the reachable sets. This justifies the introduction of these particular stochastic target problems in the context of Ricci flow.

Continuing with the notation of the previous section, we suppose that there is a smooth solution $\overline{p}_t$ to the Ricci flow, that is, to Equation (8), with initial condition $\overline{p}_0$ on the interval $t \in [0, T)$ (where we allow the possibility that $T = \infty$). At each time t, we can associate the solution with a section of E over M and thus with a submanifold of the total space E, which is smooth and intersects each fiber once, transversely. We call the resulting submanifolds $\Gamma(t)$ and note that this extends our earlier definition of $\Gamma(0)$. Of course, knowing the $\Gamma(t)$ for $t \in [0, T)$ is equivalent to knowing $\overline{p}_t$. Similarly, suppose there is a smooth solution $\overline{p}^n_t$ to the normalized Ricci flow, that is, to Equation (9), with initial condition $\overline{p}^n_0 = \overline{p}_0$ on the interval $t \in [0, T^n)$ (where, for the same manifold (M, h) with the same initial metric g_0, it is not necessarily true that T and T^n are equal). Then we have the associated submanifolds $\Gamma^n(t)$ of E. The connection between the Ricci flow and normalized Ricci flow (viewed in this way) and the stochastic target problems introduced above is given by the following theorem. Note that both this sort of result and the method of proof mirror that of [23]. The main additional complication, besides the geometric formalism needed for the general statement of the target problem, is that the controls are not restricted to a compact set.

Theorem 3. *Let (M, h) be a smooth, compact Riemannian surface with initial metric $g_0 = e^{2p_0}h$, as above. Suppose that the Ricci flow has a smooth solution $\overline{p}_t$ on $t \in [0, T)$. Then $\Gamma(t) = V(t)$ for all $t \in [0, T)$. Similarly, if the normalized Ricci flow has a smooth solution $\overline{p}^n_t$ on $t \in [0, T^n)$, then $\Gamma^n(t) = V^n(t)$ for all $t \in [0, T^n)$.*

Proof. We start with the Ricci flow. We fix some $t \in (0, T)$ and let τ be the time parameter for the controlled process $Y_\tau(A)$, $\tau \in [0, t]$ (as usual in probabilistic approaches to PDEs, process time runs "backward" compared to PDE time). We consider the square of the vertical distance between the controlled process Y_τ and $\Gamma(t-\tau)$. That is, we consider $\eta(x, p, \tau) = \left(p - \overline{p}_{t-\tau}(x)\right)^2$ along the paths of Y_τ, so that $\eta_\tau = \left(p_\tau - \overline{p}_{t-\tau}(x_\tau)\right)^2$.

Actually, we begin by considering a slightly more general quantity. Let $\xi(x, p, \tau) = p - \overline{p}_{t-\tau}(x)$, and for the moment let $\varphi : \mathbb{R} \to [0, \infty)$ be any smooth function. We wish to consider $\varphi(\xi(x, p, \tau))$; clearly η is just the special case $\varphi(z) = z^2$.

We now apply Itô's formula (11) to $(\varphi(\xi))_\tau$. In the following, $\overline{p}$ is always evaluated at time $t - \tau$ and position x_τ, we write $\mathfrak{e}_i$ for $\mathfrak{e}(x_\tau)(e_i)$, and we suppress other arguments (such as for the controls a_i) as desired to make things more readable. Then we have

$$\begin{aligned} d(\varphi(\xi))_\tau =& \sqrt{2}\varphi' e^{-p_\tau}\left[(a_1 - \mathfrak{e}_1\overline{p})\, dW^1_\tau + (a_2 - \mathfrak{e}_2\overline{p})\, dW^2_\tau\right] \\ &+ \sum_{i=1}^{2} e^{-2p_\tau}\left[\varphi''\left(-\mathfrak{e}_i\overline{p}\right)^2 + \varphi'\left(-\mathfrak{e}_i^2\overline{p}\right)\right] d\tau + \varphi'\partial_t\overline{p}\, d\tau \\ &+ e^{-2p_\tau}\left[\varphi' K_h + \varphi''\left(a_1^2 + a_2^2\right)\right] d\tau + 2e^{-2p_\tau}\varphi''\left[-a_1\mathfrak{e}_1\overline{p} - a_2\mathfrak{e}_2\overline{p}\right] d\tau. \end{aligned} \tag{14}$$

Recall that $\mathfrak{e}_1^2 + \mathfrak{e}_2^2$ is just Δ_h. Then a little algebra and the fact that $\overline{p}$ satisfies Equation (8) allows us to simplify this, yielding

$$\begin{aligned} d(\varphi(\xi))_\tau = & \sqrt{2}\varphi' e^{-p_\tau}\left[(a_1 - \mathfrak{e}_1\overline{p})\, dW^1_\tau + (a_2 - \mathfrak{e}_2\overline{p})\, dW^2_\tau\right] \\ &+ \left\{e^{-2p_\tau}\varphi''\left[(a_1 - \mathfrak{e}_1\overline{p})^2 + (a_2 - \mathfrak{e}_2\overline{p})^2\right] + \varphi'\left(e^{-2\overline{p}} - e^{-2p_\tau}\right)\left(\Delta_h\overline{p} - K_h\right)\right\} d\tau. \end{aligned} \tag{15}$$

We now return to considering η. In this case, this equation specializes to

$$\begin{aligned} d\eta_\tau = & 2\sqrt{2}\left(p_\tau - \overline{p}\right) e^{-p_\tau}\left[(a_1 - \mathfrak{e}_1\overline{p})\, dW^1_\tau + (a_2 - \mathfrak{e}_2\overline{p})\, dW^2_\tau\right] \\ &+ 2e^{-2p_\tau}\left[(a_1 - \mathfrak{e}_1\overline{p})^2 + (a_2 - \mathfrak{e}_2\overline{p})^2\right] d\tau + 2\left(p_\tau - \overline{p}\right)\left(e^{-2\overline{p}} - e^{-2p_\tau}\right)\left(\Delta_h\overline{p} - K_h\right) d\tau. \end{aligned} \tag{16}$$

First we show that any point $(x, \overline{p}_t(x))$ in $\Gamma(t)$ is in $V(t)$. Obviously, this is true for $t = 0$. Now choose $t > 0$. We choose our controls a_1 and a_2 as follows: For $\tau \in [0, t]$, we let a_1 be $\mathfrak{e}_1\overline{p}_{t-\tau}(x_\tau)$ and a_2 be $\mathfrak{e}_2\overline{p}_{t-\tau}(x_\tau)$. Thus, our controls are Markov with respect to the process' position and the time (and the "current" frame, although this is largely just a convention, as discussed above). Intuitively, all we are doing is trying to cause the process to be tangent to the evolving solution given by $\overline{p}$. Our controls are not only Markov in space and time, but they are given by evaluating smooth functions of space and time (and the lift of "space" into the orthonormal frame bundle) along the controlled process, and thus we know that the system of SDEs for Y_τ has a unique strong solution. In particular, Y_τ is uniquely determined by these controls. Using these controls, Equation (16) simplifies to

$$d\eta_\tau = 2\left(p_\tau - \overline{p}\right)\left(e^{-2\overline{p}} - e^{-2p_\tau}\right)\left(\Delta_h \overline{p} - K_h\right)\, d\tau.$$

Because $\overline{p}$ is smooth on $M\times[0,T)$ and M is compact, we know that both $\overline{p}_{t-\tau}(x)$ and $\Delta_h\overline{p}_{t-\tau}(x)-K_h$ are bounded on $(x,\tau)\in M\times[0,t]$. Now choose any $\delta>0$ and let $\theta_\delta = \inf\{\tau : \eta_\tau \geq \delta\}$ be the first hitting time of δ. Also observe that both the controlled process $Y_\tau = (x_\tau, p_\tau)$ and η_τ have continuous paths. If we stop our process at θ_δ, then p_τ is also bounded (this follows from the fact that $\overline{p}$ is bounded and the definition of η). Combining the boundedness of both $\overline{p}$ and p_τ with an easy estimate for the exponential function, we see that $e^{-2\overline{p}} - e^{-2p_\tau}$ is bounded above and below by a constant multiple of $\pm\left(p_\tau - \overline{p}\right)$, respectively. It follows that (for $\tau\leq\theta_\delta$), we have $d\eta_\tau \leq C\eta_\tau\, d\tau$, for some positive constant C depending on t, δ, and the bounds mentioned above. Recalling that $\eta_0 = 0$, because we start our controlled process on $\Gamma(t)$, and integrating gives

$$\eta_{\tau\wedge\theta_\delta} \leq C\int_0^{\tau\wedge\theta_\delta} \eta_s\, ds \quad \text{for } \tau\in[0,t].$$

Then Gronwall's lemma implies that $\eta_{\tau\wedge\theta_\delta}=0$ for all $\tau\in[0,t]$. Because η_τ has continuous paths, this means that $\theta_\delta > t$, and thus we have that $\eta_\tau = 0$ for all $\tau\in[0,t]$. In particular, $\eta_t=0$, and so $Y_t\in\Gamma(0)$. Thus we've shown that $\Gamma(t)\subset V(t)$.

Next, we need to show the opposite inclusion, $V(t)\subset\Gamma(t)$. Again, this is clear for $t=0$, so we fix some $t\in(0,T)$. We have some starting point $(\alpha,\beta)\in M\times\mathbb{R}$, and we assume that there exists a control (a_1,a_2) such that $Y_\tau(a_1,a_2)$ almost surely hits $\Gamma(0)$ at time $\tau=t$.

At this point, we produce a mollified version of η by a judicious choice of φ. In particular, we now let $\varphi:\mathbb{R}\to[0,\infty)$ be a smooth, symmetric function satisfying the following additional properties: φ is non-decreasing on $[0,\infty)$, $\varphi(z)=z^2$ in some neighborhood of 0, and φ is constant on $[A,\infty)$ for an appropriately chosen constant A. It follows that the value of φ on $[A,\infty)$ is positive, φ is 0 only at 0, and all derivatives of φ are bounded. If we now let $\hat{\eta}(x,p,\tau)=\varphi(\xi(x,p,\tau))$, then $\hat{\eta}$ is a mollified version of η, in the sense that they agree for small values of η but $\hat{\eta}$ is bounded, along with all of its derivatives.

Let $D(\tau)=\mathbb{E}\left[\hat{\eta}_\tau\right]$. Then Equation (15) shows that

(17)
$$D(\tau) = D(0) + \int_0^\tau \mathbb{E}\Big[e^{-2p_\tau}\varphi''\left[\left(a_1-\mathfrak{e}_1\overline{p}\right)^2 + \left(a_2-\mathfrak{e}_2\overline{p}\right)^2\right] + \varphi'\left(e^{-2\overline{p}}-e^{-2p_\tau}\right)\left(\Delta_h\overline{p}-K_h\right)\Big]\, ds.$$

Here, of course, the derivatives of φ are evaluated at $\xi(x_\tau,p_\tau,\tau)$. Note that $\mathfrak{e}_1\overline{p}$, $\mathfrak{e}_2\overline{p}$ and $\Delta_h\overline{p}-K_h$ are all bounded. Also, for small ξ we have that $\varphi''=2$ and $\varphi' = 2\left(p_\tau-\overline{p}\right)$, and both of these derivatives are bounded for all ξ. Moreover, both $e^{-2p_\tau}\varphi''$ and $\varphi'\left(e^{-2\overline{p}}-e^{-2p_\tau}\right)$ are bounded because the derivatives of φ are identically zero for $\xi > A$. In addition, for any two constants $C_1, C_2\geq 0$, there is another constant $C_3>0$ such that for any $\xi\in\mathbb{R}$,

$$C_1\varphi''(\xi) - C_2\varphi'(\xi)\xi \geq -C_3\varphi(\xi)$$

Combining all of these with the fact that a_1 and a_2 are bounded, we see that there is some positive constant C (different from the one above) depending on the bounds just mentioned, but not on τ, such that

$$D(\tau) \geq D(0) - C\int_0^\tau D(s)\,ds \quad \text{for all } \tau \in [0,t].$$

This is trivially equivalent to

$$-D(\tau) \leq -D(0) + (-C)\int_0^\tau \left(-D(s)\right)\,ds \quad \text{for all } \tau \in [0,t].$$

Applying Gronwall's inequality for the function $-D(\tau)$ on this interval then yields

$$-D(\tau) \leq \left(-D(0)\right) e^{-C\tau} \quad \text{for all } \tau \in [0,t].$$

In particular, letting $\tau = t$ and multiplying by -1 gives $D(0)e^{-Ct} \leq D(t)$. By assumption, the controlled process hits $\Gamma(0)$ at time t a.s., and thus $D(t) = 0$. Since D is always non-negative (because φ was chosen to be non-negative) and e^{-Ct} is positive, we conclude that $D(0) = 0$. This is equivalent to saying that our initial point (α, β) is in $\Gamma(t)$. Thus we have proven that $V(t) \subset \Gamma(t)$.

The proof for the normalized Ricci flow is almost identical. With the appropriate quantities, $\bar{p}, p_\tau, x_\tau$ and so on, equation (14) becomes

$$\begin{aligned} d(\varphi(\xi))_\tau =& \sqrt{2}\varphi' e^{-p_\tau}\left[\left(a_1 - \mathfrak{e}_1\overline{p}\right)\,dW^1_\tau + \left(a_2 - \mathfrak{e}_2\overline{p}\right)\,dW^2_\tau\right] \\ &+ \sum_{i=1}^{2} e^{-2p_\tau}\left[\varphi''\left(-\mathfrak{e}_i\overline{p}\right)^2 + \varphi'\left(-\mathfrak{e}_i^2\overline{p}\right)\right]\,d\tau + \varphi'\partial_t\overline{p}\,d\tau \\ &+ e^{-2p_\tau}\left[\varphi' K_h - re^{2p_\tau} + \varphi''\left(a_1^2 + a_2^2\right)\right]\,d\tau + 2e^{-2p_\tau}\varphi''\left[-a_1\mathfrak{e}_1\overline{p} - a_2\mathfrak{e}_2\overline{p}\right]\,d\tau. \end{aligned} \tag{18}$$

and then from (9), we get exactly the same equation from (15), thus the rest of the proof is identical. □

From the point of view of control theory, the above result is a verification theorem. From the point of view of PDE theory, this can also be thought of as a uniqueness theorem. In particular, it shows that smooth solutions to the Ricci flow are unique and we state this in the following.

Corollary 4. *If there is a (smooth) solution to (normalized) Ricci flow on the time interval $[0, T)$, then it is unique.*

It bears repeating that the above relies on already knowing that the Ricci flow has a smooth solution on some interval; in other words, it sheds no light on the existence of a solution (to either the Ricci flow or the control problem). On the other hand, this existence is well known in the present case. Cao [3] and Hamilton [14] show that, for a smooth, compact initial surface, the Ricci flow always has a smooth solution on some (non-trivial) interval of time, and the normalized Ricci flow has a smooth solution for all time. (Of course, much more can be said, including the relationship between the normalized and un-normalized flows, but again, this is well-known and can be found in any book on the subject.) For an accessible overview we refer to [7, Chapter 5], which treats the (normalized) Ricci flow on surfaces.

One additional feature of the successfully controlled process is that it provides Brownian motion on M under the backward Ricci flow (or backward normalized Ricci flow, of course), as we now explain. If we put a smooth family of metrics g_τ on a smooth manifold M, then a process B_τ is a Brownian motion on (M, g_τ) if it solves the martingale problem for the time-inhomogeneous operator Δ_{g_τ}. Suppose we have a smooth solution to the Ricci flow, as above, for $t \in [0, T)$, and let g_t be the metric on M corresponding to this solution. Then if we choose a time t (in $(0, T)$) and point $x_0 \in M$, there is a unique point (x_0, p_0) over x_0 (where, of course, we use our standard fiber coordinate p) in $\Gamma(t) = V(t)$. If we now run our successfully controlled process $Y_\tau = (x_\tau, p_\tau)$

starting from this point, we know that it is on $\Gamma(t-\tau)$ for all $\tau \in [0,t)$, or equivalently that $p_\tau = \overline{p}_{t-\tau}(x_\tau)$, for all $\tau \in [0,t]$ almost surely. Then looking at x_τ (which is just the M-marginal) and recalling that $g_t = e^{2\overline{p}_t}h$, a little thought shows that x_τ is a Brownian motion on $(M, g_{t-\tau})$ for $\tau \in [0,t]$. That "process time" runs backward compared to "PDE" time, which manifests itself in the $t-\tau$ parameter (with t fixed and τ increasing) for the metric g, explains why we get Brownian motion on M under backward Ricci flow, as opposed to just Ricci flow.

For clarity, let us temporarily denote x_τ under the successful control as $\hat{x}_\tau$. Then recognizing it as Brownian motion under backward Ricci flow gives a way of representing the solution to the Ricci flow (or normalized Ricci flow) that looks more like the usual representations for parabolic (linear) PDEs. In the special case when h is flat, normalized and un-normalized Ricci flow are the same, and we see that p_τ is a martingale. Further, we have that

$$\overline{p}_t(x_0) = \mathbb{E}^{x_0,t}\left[\overline{p}_0\left(\hat{x}_t\right)\right] \tag{19}$$

where the expectation is taken with respect to the successfully controlled process started from $(x_0, \overline{p}_t(x_0))$ and run until $\tau = t$. This is analogous to solving the heat equation with some initial condition by running Brownian motion and then using it to average the initial condition. The difference is that, for the heat equation, we can construct Brownian motion (or more analytically, the heat kernel) without already having a solution to the heat equation with our initial data. This is because Brownian motion (or the heat kernel) doesn't depend on the initial data, and so we can use it to solve the heat equation in the first place. All of this is a manifestation of the linearity of the heat equation. In the case of Ricci flow, we need to know $\hat{p}_\tau$ in order to determine $\hat{x}_\tau$ (or more accurately, these two are intertwined by the system of SDEs they solve), so we can't first determine $\hat{x}_\tau$ and then use it in the above to solve the Ricci flow.

Also, we can now say a bit more about the recent work of [9] and [1]. They give a lift of Brownian motion on a manifold with time-dependent metric to the frame bundle which gives the parallel transport along the Brownian paths. They then introduce a notion of damped parallel transport which, under the Ricci flow (but not the normalized flow), becomes an isometry as well. This damped parallel transport can be used to produce martingales from solutions to heat problems under the Ricci flow. In our notation, x_τ is the Brownian motion with respect to a time-dependent metric (with an additional factor of $\sqrt{2}$ to get the normalization right, of course), and $\{e^{-p_\tau}\mathfrak{e}(x_\tau)(e_1), e^{-p_\tau}\mathfrak{e}(x_\tau)(e_2)\}$ (which is an orthonormal frame for the time-varying metric) gives the parallel transport along the Brownian path x_τ.

4. THE BLOW UPS OF THE RICCI FLOW FOR THE CASE OF POSITIVE OR NEGATIVE EULER CHARACTERISTIC

This section is dedicated to showing that in the case of the (unnormalized) Ricci flow, there are blow ups either in finite or infinite time if the Euler charaterisitc, and hence the reference metric K_h, is either positive or negative.

Assume now that the Ricci flow has a smooth solution defined on the time interval $[0,T)$. Then, from Theorem 3, we learn that for any fixed time $t \in [0,T)$, $p_\tau = \overline{p}_{t-\tau}(x_\tau)$ where (x_τ, p_τ) is the solution to (10) with the initial conditions $(x, p_0(x))$. On the other hand, taking a smooth function $\varphi : [0,t] \times \mathbb{R} \to \mathbb{R}$ in (11), we obtain that

$$d\varphi(\tau, p_\tau) = e^{-p_\tau}\varphi'(p_\tau)\sum_{i=1}^{2} a_i\sqrt{2}dW_\tau^i + \left[\partial_\tau\varphi(\tau,p_\tau) + e^{-2p_\tau}(\varphi'(\tau,p_\tau)K_h(x_\tau) + \varphi''(\tau,p_\tau)\sum_{i=1}^{2} a_i^2)\right]d\tau.$$

Since the successful control is given by $a_i = \mathfrak{e}_i\bar{p}_{t-\tau}$, we get

$$\sum_{i=1}^{2} a_i^2 = |\nabla p_{t-\tau}(x_\tau)|^2,$$

and this means that

$$\varphi(\tau, p_\tau) - \int_0^\tau \left[\partial_\tau\varphi(\sigma, p_\sigma) + e^{-2p_\sigma}\left(\varphi'(\sigma, p_\sigma)K_h(x_\sigma) + \varphi''(\sigma, p_\sigma)|\nabla\bar{p}_{t-\sigma}(x_\sigma)|^2\right)\right] d\sigma$$

is a martingale. In particular, taking expectation at times $\tau = 0$ and $\tau = t$ and using $p_\tau = \bar{p}_{t-\tau}(x_\tau)$, yields

$$\begin{aligned} \varphi(0, \bar{p}_t(x)) =& \mathbb{E}^{(x,t)}[\varphi(t, p_0(x_t))] \\ & - \int_0^t \mathbb{E}^{(x,t)}\left[\partial_t\varphi(\sigma, p_\sigma) + e^{-2p_\sigma}\left(\varphi'(\sigma, p_\sigma)K_h(x_\sigma) + \varphi''(\sigma, p_\sigma)|\nabla\bar{p}_{t-\sigma}(x_\sigma)|^2\right)\right] d\sigma. \end{aligned} \tag{20}$$

There are two obvious obstructions stemming from this formula. The first one is that if $K_h(x) > 0$ for all $x \in M$, then taking $\varphi(\tau, p) = e^{2p}$, the above formula (20) implies

$$e^{2\bar{p}_t(x)} = \mathbb{E}^{(x,t)}[e^{2p_0(x_t)}] - 2\int_0^t \mathbb{E}^{(x,t)}\left[K_h(x_\sigma) + 2|\nabla\bar{p}_{t-\sigma}(x_\sigma)|^2\right] d\sigma \le \mathbb{E}^{(x,t)}[e^{2p_0(x_t)}] - 2\int_0^t \mathbb{E}^{(x,t)}[K_h(x_\sigma)]d\sigma$$

and thus, upon denoting the uniform norm by $|\cdot|_u$ and taking $K_0 = \inf_{x\in M} K_h(x)$,

$$e^{2\bar{p}_t(x)} \le e^{2|p_0|_u} - 2tK_0.$$

As this is true for any $t \in [0, T)$, the extinction time of the Ricci flow is finite and is certainly at most $e^{2|p_0|_u}/(2K_0)$. Therefore, in the case of positive curvature the flow develops singularities in finite time.

On the other hand if the curvature is negative ($K_h < 0$ on M) then there are some constants $C_1, C_2 > 0$ such that

$$\bar{p}_t(x) \ge \log(C_1 t + 1) - C_2 \text{ for all } x \in M \text{ and } t \ge 0.$$

To see this, take $K_0 = \inf_{x\in M} -K_h(x) > 0$, thus $K_h(x) \le -K_0 < 0$ and then consider $\varphi(\tau, p) = p$ in (20) to deduce that

$$\bar{p}_t(x) = \mathbb{E}^{(x,t)}[p_0(x_t)] - \int_0^t \mathbb{E}^{(x,t)}[e^{-2p_\sigma}K_h(x_\sigma)]d\sigma \ge \inf_{x\in M} p_0$$

which means that $\bar{p}_t(x)$ is bounded below uniformly in $t \ge 0$ and $x \in M$. Now consider the test function $\varphi(\tau, p) = \exp\left(\alpha(t - \tau - \frac{1}{2K_0}e^{2p})\right)$. Since $\bar{p}_t(x)$ is bounded below, this implies that for large enough α, $\varphi''(\sigma, p_\sigma) \ge 0$. On the other hand, $\partial_\tau\varphi(\sigma, p) - K_0 e^{-2p}\varphi'(\sigma, p) = 0$, and this combined with the preceding and the fact that φ' is negative leads to

$$\varphi(0, \bar{p}_t(x)) \le \mathbb{E}^{t,x}[\varphi(t, p_0(x_t))] \le 1,$$

which means that $\bar{p}_t(x) \ge \frac{1}{2}\log(2K_0 t)$ for any $t > 0$ for which $\bar{p}_t$ exists. In particular this shows that either the flow ceases to exist after a finite time, or, if it does exist for all times, $\bar{p}_t(x)$ goes to infinity uniformly over $x \in M$. The moral is that we can not expect the Ricci flow to converge as the time approaches either the extinction time or infinity.

For the flat case, since the curvature is 0, the normalized and the unnormalized Ricci flows are the same and thus we will treat this case as the normalized Ricci flow.

Remark. The blow up in the negative case does not take place in finite but this requires more arguments. However we see this

5. Time-dependent a priori bounds for Ricci flow

We now turn our attention to using the stochastic target representation for the normalized Ricci flow to derive (more accurately, of course, to re-derive) geometric facts about the flow. We will always work with the case where the reference metric h has constant curvature. By uniformization, this is no loss of generality, and it simplifies the analysis considerably. After a preliminary rescaling, we can assume that this constant curvature is either 1, 0, or -1. Further, we can rescale the initial metric g_0 so that it has the same area as h. Thus, without loss of generality, we are in one of three cases (by the Gauss-Bonnet theorem). First, if the Euler characteristic of M is positive, we have that $K_h \equiv r \equiv 1$. If the Euler characteristic of M is zero, we have that $K_h \equiv r \equiv 0$. Finally, when the Euler characteristic of M is negative we have that $K_h \equiv r \equiv -1$. The bounds we have in mind are similar in all three cases, although the differences in sign of K_h result in important differences.

We call these bounds "a priori" because they don't depend on the structure of the reachable set. We elaborate on this after Theorem 6.

We have one more comment about notation before we begin. Because we will be concerned with the normalized Ricci flow for the rest of the paper, we drop the "n" superscripts. Thus, for instance, we let $\bar{p}_t$ denote a solution to the normalized Ricci flow, unless otherwise indicated.

The interesting feature of choosing h to be a metric of constant curvature is that the drift of the SDE satisfied by p_τ doesn't depend on x_τ (although the target always does, except in trivial cases). In particular, we have the following three cases:

$$
\begin{aligned}
r = 1: &\quad dp_\tau = e^{-p_\tau}\left[\sum_{i=1}^{2} a_i \sqrt{2}\, dW_\tau^i\right] + \left(e^{-2p_\tau} - 1\right)\, d\tau \\
r = 0: &\quad dp_\tau = e^{-p_\tau}\left[\sum_{i=1}^{2} a_i \sqrt{2}\, dW_\tau^i\right] \\
r = -1: &\quad dp_\tau = e^{-p_\tau}\left[\sum_{i=1}^{2} a_i \sqrt{2}\, dW_\tau^i\right] + \left(1 - e^{-2p_\tau}\right)\, d\tau
\end{aligned}
\tag{21}
$$

In general, the stochastic target problem for the normalized Ricci flow (and also the Ricci flow itself) gives an equation of the form

$$
dp_\tau = e^{-p_\tau}\left[\sum_{i=1}^{2} a_i dW_\tau^i\right] + U_\tau(p_\tau) d\tau, \tag{22}
$$

where the controls a_i, $i = 1, 2$ are bounded and chosen such that p_t is almost surely on M_0, the section corresponding to $\bar{p}_0$ in the bundle $M \times \mathbb{R}$. In the case at hand we assume that $U_\tau(p)$ is a function $U : [0,t] \times \mathbb{R} \to \mathbb{R}$ which is uniformly locally Lipschitz in the second variable, i.e. for any $L > 0$ there is a constant C_L with $|U_\tau(p) - U_\tau(q)| \leq C_L |p - q|$ for all $\tau \in [0,t]$ and $p, q \in [-L, L]$.

The basic point is that there are natural barriers for p_τ given in terms of equation (22) where the martingale part is set to be equal to 0. To be precise, we define a barrier as a solution q_τ to the ODE

$$
dq_\tau = U_\tau(q_\tau) d\tau. \tag{23}
$$

In this framework we have a general result as follows.

Lemma 5. *Assume that p_τ and q_τ are solutions to (22) and (23) respectively for $\tau \in [0,t]$ with U a uniformly locally Lipschitz function in the second variable on $[0,t] \times \mathbb{R}$.*

If at any time $\tau_1 \in [0,t)$, $p_{\tau_1} < q_{\tau_1}$ with positive probability, then at any later time $\tau_2 \in (\tau_1, t]$, $p_{\tau_2} < q_{\tau_2}$ with positive probability.

Similarly, if at any time $\tau_1 \in [0,t)$, $p_{\tau_1} > q_{\tau_1}$ with positive probability, then at any later time $\tau_2 \in (\tau_1, t]$, $p_{\tau_2} > q_{\tau_2}$ with positive probability.

Proof. The proof is a basic application of stopping time and Gronwall's lemma. We will prove only the first part, the second one being similar.

So, assume that $q_{\tau_1} > p_{\tau_1}$ with positive probability, and therefore that we can choose a constant $L > 0$ such that $L \geq q_{\tau_1} - p_{\tau_1} > 1/L$ with positive probability. We further take L large enough so that $|q_\tau| \leq L$ for all $\tau \in [0,t]$.

Now, for any smooth function $\eta : \mathbb{R} \to \mathbb{R}$, we have

(24)
$$\eta(q_\tau - p_\tau) = \eta(q_{\tau_1} - p_{\tau_1}) + M_\tau + \int_{\tau_1}^{\tau} \left(e^{-2p_s} \eta''(q_s - p_s)(a_1^2(s) + a_2^2(s)) + \eta'(q_s - p_s)(U_s(q_s) - U_s(p_s)) \right) ds$$

where M_τ is a martingale with $M(\tau_1) = 0$. Further, we choose the function $\eta(\xi)$ such that it is non-decreasing, equal to 0 for $\xi \leq 0$, equal to 1 for $\xi \geq 2L$ and $\eta(\xi) = \xi^2$ for small $\xi \geq 0$.

Next, we define the stopping time $\sigma = \inf\{u \geq \tau_1 : p_u \geq q_u\} \wedge t$. With this setup, we will denote for simplicity $\eta_\tau = \eta(q_\tau - p_\tau)$, $\eta'_\tau = \eta'(q_\tau - p_\tau)$ and $\eta''_\tau = \eta''(q_\tau - p_\tau)$. Furthermore, from (24),

$$\mathbb{E}[\eta_{\tau\wedge\sigma}] = \mathbb{E}[\eta_{\tau_1}] + \int_0^\tau \mathbb{E}[\mathbb{1}_{[\tau_1,\sigma]}(s) \left(e^{-2p_s} \eta''_s (a_1^2(s) + a_2^2(s)) + \eta'_s (U_s(q_s) - U_s(p_s)) \right)] ds.$$

Since q_s remains bounded on $[\tau_1, \tau_2]$ and η' has compact support, combined with the property that U_τ is uniformly Lipschitz in the second variable on compact intervals, we can find a constant $C > 0$, such that

$$\eta'_s (U_s(q_s) - U_s(p_s)) \geq -C \eta'_s (q_s - p_s).$$

This, the choice of our function η, the fact that the controls a_i, $i = 1, 2$ are bounded, and that q_s is bounded, yield, in the first place, that $e^{-2p_s}\eta''_s$ is bounded, and also that for some constant $C > 0$,

$$\left(e^{-2p_s} \eta''_s (a_1^2(s) + a_2^2(s)) + \eta'_s (U_s(q_s) - U_s(p_s)) \right) \geq -C \eta_s.$$

To check this, one can reason as follows. For $q_s \leq p_s$, both sides are 0. For $\epsilon > q_s - p_s > 0$ with small ϵ, the first term is non-negative and the second one is bounded below by $-C(q_s - p_s)^2$ which is again a constant times η_s. For $q_s - p_s > \epsilon$, the inequality follows easily as the left hand side is bounded below by some negative constant and η_s is certainly bounded below by ϵ^2.

Summing up the findings we get that

$$\mathbb{E}[\eta_{\tau\wedge\sigma}] \geq \mathbb{E}[\eta_{\tau_1}] - C \int_{\tau_1}^{\tau} \mathbb{E}[\mathbb{1}_{[\tau_1,\sigma]}(s) \eta_s] ds$$

Since σ is the first time $p_s = q_s$, it follows that, $\mathbb{1}_{[\tau_1,\sigma]}(s)\eta_s = \eta_{s\wedge\sigma}$, consequently, for some $C > 0$,

$$\mathbb{E}[\eta_{\tau\wedge\sigma}] \geq \mathbb{E}[\eta_{\tau_1}] - C \int_{\tau_1}^{\tau} \mathbb{E}[\eta_{s\wedge\sigma}] ds \text{ for } \tau \in [\tau_1, \tau_2].$$

At this point an application of Gronwall's inequality results in

$$\mathbb{E}[\eta_{\tau\wedge\sigma}] e^{C(\tau-\tau_1)} \geq \mathbb{E}[\eta_{\tau_1}] > 0,$$

where the hypothesis $q_{\tau_1} > p_{\tau_1}$ with positive probability is translated into the last inequality. For $\tau = \tau_2$ we obtain $\mathbb{E}[\eta_{\tau_2\wedge\sigma}] = \mathbb{E}[\eta_{\tau_2}, \sigma > \tau_2] > 0$ and therefore we conclude that $\{\sigma > \tau_2\}$ has positive probability; stated otherwise, the probability that $q_{\tau_2} > p_{\tau_2}$ is positive.

One technical word is in place here. Namely, the definition from (22) is in the sense of local martingales, but during the proof we look at $\eta(q_\tau - p_\tau)$ and this is actually a true semi-martingale not merely a local one. This is indeed due to the boundedness and continuity of the quantities involved, namely $e^{-p_s}\eta'_s$, $e^{-2p_s}\eta''_s$ and the controls a_i, $i = 1, 2$. $\square$

Next, we solve equation (23) for each of the three cases described in equation (21) (this is straight-forward, as the resulting ODEs are separable). For ease of reference, we will label the resulting equations as $B^K_c(\tau)$ with super- and sub-scripts indicating relevant parameters. In the case $r = 1$, we have that

$$B^1_c(\tau) = \frac{1}{2}\log\left(1 - ce^{-2\tau}\right) \quad \text{for some constant } c \in (-\infty, 1).$$

The choice of c allows any initial condition. Note that $c = 0$ gives the constant solution $B^1_0(\tau) \equiv 0$. For any c, as $\tau \to \infty$, we see that $B^1_c(\tau) \to 0$. The case $r = 0$ gives

$$B^0_c(\tau) = c \quad \text{for some constant } c \in \mathbb{R}.$$

Obviously, the choice of c allows any initial condition. (This is perhaps a bit pedantic, but we include it for the sake of completeness.) Finally, $r = -1$ gives

$$B^{-1}_c(\tau) = \frac{1}{2}\log\left(1 - ce^{2\tau}\right) \quad \text{for some constant } c \in (-\infty, 1).$$

Again, the choice of c allows any initial condition, and $c = 0$ gives the constant solution $B^{-1}_c(\tau) \equiv 0$. This time, though, if $c \neq 0$, then the solution heads to $\pm\infty$ as τ increases (in finite time for negative initial condition, and as $\tau \to \infty$ for positive initial condition).

Continuing, we want to use the previous lemma and a judicious choice of the parameter c to bound the reachable set at time t. Recall that $\overline{p}_0$ gives the initial metric g_0 and serves as the target in the target problem (and which as a section we write as $\Gamma(0)$). The the assumption that g_0 and h have the same area implies that $\max_{x\in M}\overline{p}_0(x) = \alpha \geq 0$ and that $\min_{x\in M}\overline{p}_0(x) = \beta \leq 0$. Further, if either α or β is zero then both are, meaning that $\overline{p}_0 \equiv 0$ and g_0 is just h.

The logic of the proof of the following theorem explains why solutions q_τ of equation (23) are called barriers, in this context.

Theorem 6. *Consider the target problem (for the normalized Ricci flow) where h corresponds to one of the three constant curvature cases as discussed above (and with α and β as just described). For any $t \geq 0$, we have that*

$$\sup_{(x,p)\in V^n(t)} p \leq \begin{cases} \frac{1}{2}\log\left(1 - e^{-2t}\left(1 - e^{2\alpha}\right)\right) & \text{if } r = -1, \\ \alpha & \text{if } r = 0, \\ \frac{1}{2}\log\left(1 - e^{2t}\left(1 - e^{2\alpha}\right)\right) & \text{if } r = 1, \end{cases}$$

and

$$\inf_{(x,p)\in V^n(t)} p \geq \begin{cases} \frac{1}{2}\log\left(1 - e^{-2t}\left(1 - e^{2\beta}\right)\right) & \text{if } r = -1, \\ \beta & \text{if } r = 0, \\ \frac{1}{2}\log\left(1 - e^{2t}\left(1 - e^{2\beta}\right)\right) & \text{if } r = 1 \text{ and } t < -\frac{1}{2}\log\left(1 - e^{2\beta}\right). \end{cases}$$

(If $\beta = 0$, we set $-\frac{1}{2}\log\left(1 - e^{2\beta}\right) = \infty$.)

Proof. We start with the upper bound in the $r = -1$ case. We consider some fixed but arbitrary $t \geq 0$. Let $c' = e^{-2t}\left(1 - e^{2\alpha}\right)$. Then

$$B^{-1}_{c'}(t) = \alpha \quad \text{and} \quad B^{-1}_{c'}(0) = \frac{1}{2}\log\left(1 - e^{-2t}\left(1 - e^{2\alpha}\right)\right).$$

Thus, by the previous lemma, if we start from a point (x_0, p_0) with $p_0 > B_{c'}^{-1}(0)$, we have that $p_t > B_{c'}^{-1}(t) = \alpha$ with positive probability (for any controls). By the definition of α, this means that p_t is not in the target with positive probability. Since this holds for any controls, it follows that (x_0, p_0) is not in the reachable set at time t, which we recall we denote $V^n(t)$. This implies the upper bound on $\sup_{(x,p)\in V^n(t)} p$ given in the theorem.

For the lower bound in the $r = -1$ case, consider $c' = e^{-2t}\left(1 - e^{2\beta}\right)$. Then

$$B_{c'}^{-1}(t) = \beta \quad \text{and} \quad B_{c'}^{-1}(0) = \frac{1}{2}\log\left(1 - e^{-2t}\left(1 - e^{2\beta}\right)\right).$$

Analogously to the argument for the upper bound, the previous lemma implies that no point (x_0, p_0) with $p_0 < B_{c'}^{-1}(0)$ can be in $V^n(t)$. This implies the desired lower bound.

For the $r = 0$ case, analogous arguments apply, using $c' = \alpha$ for the upper bound and $c' = \beta$ for the lower bound.

Finally, we consider the $r = 1$ case. The upper bound is proven just as in the $K = -1$ case, using $c' = e^{2t}\left(1 - e^{2\alpha}\right)$. The proof of the lower bound is similar, except that if $t \geq -\frac{1}{2}\log\left(1 - e^{2\beta}\right)$, we have that $B_c^1(t) > \beta$ for any choice of $c \in (-\infty, 1)$. Thus, these arguments do not produce any lower bound for $\inf_{(x,p)\in V^n(t)} p$ in this case. On the other hand, if $t < -\frac{1}{2}\log\left(1 - e^{2\beta}\right)$, we can let $c' = e^{2t}\left(1 - e^{2\beta}\right)$ and argue just as before. □

In light of the verification theorem, these conclusions can be restated in terms of $\overline{p}_t$. Namely, we can replace $\sup_{(x,p)\in V^n(t)} p$ in the above theorem with $\max_{x\in M} \overline{p}_t(x)$ and $\inf_{(x,p)\in V^n(t)} p$ with $\min_{x\in M} \overline{p}_t(x)$. Nonetheless, there is a reason to state the theorem as above. Suppose we consider the same target problem (or problems, since there are three cases), except that now we allow the target to be any (non-empty) closed set Γ such that $\max_\Gamma p = \alpha \geq 0$ and $\max_\Gamma p = \beta \leq 0$, rather than just a smooth section corresponding a metric g_0 on M. Then we can still ask about the reachable set at time $t \geq 0$. Assuming that it is non-empty, the bounds in the above theorem still hold (with the same proofs). This shows that these bounds don't depend on the verification theorem and the resulting connection with PDEs, or on the structure of the reachable set, such as its smoothness or whether it's a section. (Moreover, similar methods could be employed even if α and β weren't assumed to be non-negative and non-positive, respectively.) It is this sense in which we refer to them as "a priori bounds." Of course, it is likely that these bounds are only *interesting* in light of their connection to the Ricci flow, as given by the verification theorem.

We close this section with some easy observations about this theorem. First of all, if $\alpha = 0$, then $\sup_{(x,p)\in V^n(t)} p = 0$ for all $t \geq 0$, and this holds in all three cases. Similarly, if $\beta = 0$, then $\inf_{(x,p)\in V^n(t)} p = 0$ for all $t \geq 0$, in all three cases. Since one of α or β being zero implies that both are, we conclude that if either α or β is zero, the reachable set only contains points with $p = 0$. On the other hand, every point with $p = 0$ will clearly be in the reachable set (just let the controls be identically zero). Thus we will have $V^n(t) = \{p \equiv 0\}$ for all $t \geq 0$. This corresponds to the basic fact that if g_0 is already a metric of constant curvature, then it is stationary under the normalized Ricci flow.

In the case when α and β are not zero, we see much different behavior for the cases of the three different curvatures. For $r = -1$, the bounds improve as t increases, which we will see makes this the easiest case to deal with. For $r = 0$, the bounds are constant. Finally, for $r = 1$, the bounds get worse as t increases, and the lower bound even ceases to exist in finite time. This corresponds to the well-known observation that the case of the sphere (or projective space) is the hardest case to handle for Ricci flow on compact surfaces.

Remark. It's worth point out that the above argument from Theorem 6 is overkill in the $r = 0$ case, since then the lemma follows directly from the fact that p_τ is a martingale and martingales have constant expectation.

We finish this discussion with the following useful Corollary which plays an important role later on.

Corollary 7. *For the case of $r = -1$, or equivalently, the case $\chi(M) < 0$, the solution $\bar{p}_t$ of the normalized Ricci flow converges to 0 uniformly in the C^0-norm exponentially fast as $t \to \infty$.*

The same arguments work in the case of unnormalized Ricci flow. We record this here as follows.

Theorem 8. *For the unnormalized Ricci flow, as long as the stochastic target is well defined up to time t,*

$$\sup_{(x,p)\in V^n(t)} p \leq \begin{cases} \frac{1}{2}\log\left(e^{2\alpha}+t\right) & \text{if } r=-1, \\ \frac{1}{2}\log\left(e^{2\alpha}-t\right) & \text{if } r=1, \text{ and } t<e^{2\alpha}. \end{cases}$$

and

$$\inf_{(x,p)\in V^n(t)} p \geq \begin{cases} \frac{1}{2}\log\left(e^{2\beta}+t\right) & \text{if } r=-1, \\ \frac{1}{2}\log\left(e^{2\beta}-t\right) & \text{if } r=1, \text{ and } t<e^{2\beta}. \end{cases}$$

The only thing we should point out here is that there is a blow-up in finite time for the case of $r = 1$ and there is also a blow up in finite or infinite time for the case of $r = -1$. This recovers the blow-up results in the previous section, only this time we used uniformization.

Remark. This theorem shows that for the unnormalized Ricci flow, in the negative curvature case, the flow does not blow up in finite time, at least in the C^0 topology. This is already a good indication that the solution is defined for all times and corroborated with the above Theorem shows that the flow blows up at infinite. Thus, this result is probably a better result (in the case of negative constant curvature case) as the one obtained in Section 4.

6. MIRROR COUPLING

For the remainder of the paper, we assume that we have a smooth initial metric and a smooth solution to the normalized Ricci flow for all time (which we do since the initial conditions are smooth on a compact surface). We are interested in studying the convergence to the constant curvature limit according to the stochastic framework we have been developing.

We consider the cases of zero Euler characteristic and of negative Euler characteristic, and we work relative to the underlying metric of constant curvature, as in the previous section. The positive Euler characteristic case (the sphere or projective plane) is well-known to be more difficult. This is largely due to the fact that there are many constant curvature metrics in any given conformal class, so that it's not clear in advance which one will be the limiting metric under normalized Ricci flow (this is related to the issue of solitons). As a result, we don't pursue this case.

We are assuming that we have a smooth solution to the normalized Ricci flow for all time. This means that the reachable set is always a smooth hypersurface transverse to the vertical fibers. From now on, we're only interested in the successfully controlled process, so for notational simplicity we will let (x_τ, p_τ) always denote that process (that is, what we previously denoted $\hat{Y}_\tau = Y_\tau(\hat{A})$). Moreover, if $\bar{p}$ is the smooth solution, we see that $p_\tau = \bar{p}_{t-\tau}(x_\tau)$. One consequence of this is that we can generally restrict our attention to the x_τ process. In particular, if we wish to couple two copies of the successfully controlled process (so that they meet as quickly as possible), it is enough to couple the x_τ marginals, since if the processes meet on the manifold, then they also meet on the fiber. In this sense, what we are doing is equivalent to just considering

Brownian motion on the underlying time-varying manifold, and so we see again that running a Brownian motion along the solution flow (and employing the stochastic techniques that apply in that situation) is subsumed by the more general construction of the stochastic target problem.

A significant part of our results on the long-time convergence of the normalized Ricci flow is based on coupling two copies of the marginal process on M, which we denote by x_τ and y_τ. Recall that x_τ will be time-changed Brownian motion on (M,h), with the time change given by integrating $a = 2e^{-2\bar{p}}$ along the paths, and analogously for y_τ, where we let b denote the instantaneous time-dilation (this is one significant advantage to working relative to this fixed metric). Note that we've incorporated the $\sqrt{2}$ normalization factor into the time-change, so that we really do have Brownian motion with respect to h as the underlying object. This makes the stochastic analysis look a bit more standard.

We wish to implement the mirror coupling for x_τ and y_τ, where the mirror map is with respect to the fixed h metric. Viewed in this way, this is a fairly straight-forward variant of the mirror coupling for two Brownian motions on a smooth (non-varying) Riemannian manifold. We simply generalize to allow our processes to be Brownian motions up to a random but smooth (in terms of the particle's position in space-time) time-change. References for the standard (non-time changed) construction are [15] and [11], and we proceed by modifying this as necessary and by not belaboring the aspects which carry over without modification.

Note that, since we're working only in the cases of non-positive Euler characteristic, a (and thus also b) is bounded above and below by positive constants (depending only on the initial metric) for all time, by the results of the previous section.

First, let C_M be the subset of $M \times M$ consisting of points (x,y) such that $y \in \mathrm{Cut}(x)$ (which is equivalent to $x \in \mathrm{Cut}(y)$), and let D_M be the diagonal subset of $M \times M$. Then let E_M be $M \times M$ minus C_M and D_M. Note that the distance function $\mathrm{dist}(x,y)$ is smooth on E_M, and that the direction of the (unique) minimal geodesic from x to y is smooth on E_M. Let $(x,y) \in E_M$; then the mirror map is the isometry from T_xM to T_yM given by reflection along the minimal geodesic connecting x and y. We see that the mirror map is smooth (on E_M, which is where it is defined). As result, there is no problem in running the mirror coupling as long as the joint process is in E_M. That is, for one-dimensional independent Brownian motions W^1_τ and W^2_τ, consider the system of SDEs

$$\begin{aligned} dx_\tau &= a_\tau \left[\sum_{i=1}^{2} \mathfrak{e}_i(x_\tau) \circ dW^i_\tau \right] \\ dy_\tau &= b_\tau \left[\sum_{i=1}^{2} \Psi_\tau \left[\mathfrak{e}_i(y_\tau) \right] \circ dW^i_\tau \right], \end{aligned}$$

where $\Psi_\tau = \Psi(x_\tau, y_\tau) = m_{x_\tau,y_\tau} \mathfrak{e}(x_\tau) \mathfrak{e}(y_\tau)^{-1}$ with $m_{x,y}$ being the mirror map, namely parallel transport followed by reflection with respect to the perpendicular to the geodesic from x to y. Then the coefficients are smooth in both space and time, so the system admits a unique strong solution, up until the first time the process leaves E_M.

The point of the coupling is to get the particles to meet, so we turn our attention to this issue next. First note that the marginals x_τ and y_τ are time-changed Brownian motions as desired, so we're coupling the right processes. The natural object of study is the distance between the particles, with respect to the fixed metric h. We denote this distance by ρ_τ. It is a (continuous, non-negative) semi-martingale, so we derive the SDE that it satisfies by Ito's formula. This is the standard computation with the factors of a and b included, so we'll be brief. For more on this, see [15, Section 6.5].

The martingale part is easily seen to be $(a+b)\,d\hat{W}_\tau$ for some Brownian motion $\hat{W}_\tau$, whether we're in the $r=0$ or $r=-1$ case. (In what follows, we use $\hat{W}_\tau$ to denote some Brownian motion, which may change from appearance to appearance, in order to more conveniently describe the SDE satisfied by a given process.) As for the drift, the only contribution comes from the second derivative of the distance with respect to the diffusions perpendicular to the geodesic from x to y, which is computed in terms of the index of the appropriate Jacobi field along the geodesic from x to y. We now summarize the computation.

Let γ be the unique minimal geodesic from x to y (parametrized by arclength), and let v be a unit vector field along γ, perpendicular to γ (this determines v uniquely up to sign, and either of choice of sign is fine). Then we want the Jacobi field $J(s)v(\gamma(s))$ where $J:[0,\rho]\to\mathbb{R}$ satisfies

$$\ddot{J}+rJ=0,\quad J(0)=a,\quad J(\rho)=b.$$

When $r\equiv 0$, the solution space to this differential equation is spanned by 1 and s. Taking the boundary conditions into account, we see that the solution is

$$J(s)=a+\frac{b-a}{\rho}s.$$

Similarly, when $r\equiv -1$, the solution space is spanned by $\cosh s$ and $\sinh s$, and the boundary conditions give

$$J(s)=a\cosh s+\frac{b-a\cosh\rho}{\sinh\rho}\sinh s.$$

The index of each of these Jacobi fields is given by

$$\int_\gamma\left(\left(\dot{J}\right)^2-rJ^2\right)ds=J(\rho)\dot{J}(\rho)-J(0)\dot{J}(0),$$

where the right-hand side is obtained from the left via integration by parts and the differential equation satisfied by J. Thus for $r\equiv 0$, the index is

$$b\left(\frac{b-a}{\rho}\right)-a\left(\frac{b-a}{\rho}\right)=\frac{(a-b)^2}{\rho},$$

and for $r\equiv -1$, the index is

$$\begin{aligned} b\left[a\sinh\rho+(b-a\cosh\rho)\frac{\cosh\rho}{\sinh\rho}\right]-a\left[\frac{b-a\cosh\rho}{\sinh\rho}\right]&=\left(a^2+b^2\right)\coth\rho-2ab\frac{1}{\sinh\rho}\\ &=(a-b)^2\coth\rho+2ab\tanh\frac{\rho}{2}.\end{aligned}$$

Putting this together, we see that

$$d\rho_\tau=\begin{cases}(a+b)\,d\hat{W}_\tau+\frac{1}{2}\left[\frac{(a-b)^2}{\rho}\right]d\tau & \text{for } r=0,\\ (a+b)\,d\hat{W}_\tau+\frac{1}{2}\left[(a-b)^2\coth\rho+2ab\tanh\frac{\rho}{2}\right]d\tau & \text{for } r=-1.\end{cases}$$

As mentioned, this holds until the first exit time from E_M. Following the reasoning in [15, Section 6.6], one can show that $\hat{W}_\tau=-\sum_{i=1}^2\langle\mathfrak{e}_i(x_\tau),\dot{\gamma}_\tau(0)\rangle dW_\tau^i$ where γ_τ is the minimal geodesic joining x_τ and y_τ starting at x_τ and running at unit speed. We will come back to these expression in Section 9.

When the particles meet, we've achieved our goal, and we can either stop the process, or allow it to continue to run as $x_\tau=y_\tau$. Either way, there's no problem caused by the process hitting the diagonal. On the other hand, we do need to find a way to continue the process past the first hitting time of the cut locus. Showing that this is possible constitutes the proof of the following theorem.

Theorem 9. *Let $M = (M, h)$ be a compact surface of constant curvature 0 or -1, and let $a = a(x, \tau)$ and $b = b(y, \tau)$ be as above. Then there exists a process (x_τ, y_τ) on $M \times M$, started from any $(x_0, y_0) \notin D$ and run until the first time hitting time of D, such that*

1. *The marginals x_τ and y_τ are time-changed Brownian motions, with times changes given by a and b, respectively.*
2. *The distance (relative to h) between x_τ and y_τ, denoted ρ_τ, satisfies the SDE*

$$d\rho_\tau = \begin{cases} (a+b)\,d\hat{W}_\tau + \frac{1}{2}\left[\frac{(a-b)^2}{\rho}\right] d\tau - L_\tau & \text{for } r \equiv 0, \\ (a+b)\,d\hat{W}_\tau + \frac{1}{2}\left[(a-b)^2 \coth\rho + 2ab\tanh\frac{\rho}{2}\right] d\tau - L_\tau & \text{for } r \equiv -1, \end{cases} \tag{25}$$

where L_τ is a non-decreasing process which increases only when $(x_\tau, y_\tau) \in C_M$ (and the set of τ for which $(x_\tau, y_\tau) \in C_M$ has measure zero almost surely).

Proof. As mentioned, the only issue is extending the construction mentioned above past the first hitting time of C_M. As usual, we proceed by approximation.

Choose small, positive δ. Until the process comes within distance δ of C_M (in the product metric on $M \times M$), we run the mirror coupling as above. When the process hits distance δ from C_M, at time τ_1, we start to run x_τ and y_τ as independent (time-changed) Brownian motions. This continues until the process is distance 2δ from C_M, at time τ_2, when we again run them under the mirror coupling. We continue this procedure, so that we have a joint process $(x^\delta_\tau, y^\delta_\tau)$ which evolves under the mirror coupling on intervals of time $[\tau^\delta_{2n}, \tau^\delta_{2n+1})$ and as independent processes on intervals of time $[\tau^\delta_{2n-1}, \tau^\delta_{2n})$, for non-negative integers n, where the τ_m are the alternating hitting times of the δ and 2δ level sets of the distance from C_M.

It's clear that x^δ_τ and y^δ_τ are time-changed Brownian motions as desired, and that the ρ^δ_τ satisfies the desired SDE when $(y^\delta - x^\delta)_\tau$ is distance more than 2δ from C_M. It's also clear that when x^δ_τ and y^δ_τ are being run independently, ρ^δ satisfies an SDE of the form

$$d\rho^\delta_\tau = u\,d\hat{W}_\tau + v\,d\tau - \hat{L}_\tau,$$

where u and v are bounded (with bound depending only on M and the bounds on a and b) and $\hat{L}_\tau$ is a non-decreasing process which increases only when $(x^\delta_\tau, y^\delta_\tau) \in C_M$ (again, see the references mentioned above).

Suppose we show that, for any $t > 0$, the expected amount of time on the interval $[0, t]$ that the process spends within distance 2δ of C_M goes to zero with δ. (Thus the amount of time the particles spend being run independently goes to zero almost surely.) Then letting δ go to zero, we know there is at least one subsequence along which the process $(x^\delta_\tau, y^\delta_\tau)$ converges to a limiting process (x_τ, y_τ) (by compactness). That this limiting process satisfies the first property in the theorem is immediate, since x^δ_τ and y^δ_τ do for all $\delta > 0$. For the second property, note that the contributions from the $u\,d\tilde{W}_\tau$ term and the $v\,d\tau$ term go to zero by the boundedness of u and v and the fact that the expected length of time over which these terms are integrated goes to zero. It follows that the martingale part and the "regular" part of the drift come entirely from the SDE for ρ induced by the (mirror) coupling, and that the time spent at C_M has measure zero. Finally, the $\hat{L}_\tau$ contribution converges to a term L_τ as indicated.

Thus, to complete the proof, we need only show that the expected amount of time on the interval $[0, t]$ that the process $(y^\delta, x^\delta)_\tau$ spends within distance 2δ of the C_M goes to zero with δ. This is the most tedious part, so we will try not to belabor it. Also, to simplify matters, we will take advantage of the structure of M. In the $r = 0$ case, we can think of M as $\mathbb{R}^2$ modulo some lattice, or equivalently, as a solid parallelogram with opposite sides identified. Then for fixed x, the cut locus is just the identified sides of the parallelogram (centered at x), which means that it is just two intersecting closed geodesics. This implies that C_M consists of two intersecting smooth (compact) hypersurfaces in $M \times M$. The advantage here is that we can understand the time spent near C_M in terms of the time spent near each of these two hypersurfaces surfaces individually. This is nice because, in some (possibly local) tubular neighborhood of such a hypersurface, the (signed) distance to the hypersurface is smooth.

The $r = -1$ case is analogous. Here M is the hyperbolic plane modulo an appropriate group action, or equivalently, a polygon in the hyperbolic plane with sides identified. So for fixed x, the cut locus is the union of a finite number of intersecting closed geodesics, and C_M consists of a finite number of intersecting smooth (compact) hypersurfaces in $M \times M$.

In either case, it's clearly enough to consider one smooth hypersurface component of C_M, which we denote Σ, and prove that the time spent within distance 2δ goes to zero. We fix some $\delta_0 > 0$ such that the signed distance from Σ, which we denote ξ, is smooth on a neighborhood of the closure of the δ_0-neighborhood of Σ. (We will later also assume that δ_0 is small enough to satisfy one other condition.) Now the SDE satisfied by $\xi_\tau = \xi^\delta_\tau$, for $|\xi| < \delta_0$, depends on whether the joint process is being run under independence or under the mirror coupling (of course, ξ_τ depends on δ, but we'll suppress this for ease of notation).

Under independence, the bounded geometry and non-positive curvature of M implies that ξ_τ satisfies an SDE of the form

$$d\xi_\tau = u\, d\hat{W}_\tau + v\, d\tau \quad \text{for } |\xi| < \delta_0,$$

where u and $|v|$ are bounded and u is bounded from below by a positive constant (with all of these bounds depending only on M, δ_0, and the bounds on a and b). Under the mirror coupling, ξ_τ satisfies an SDE of the form

$$d\xi_\tau = \tilde{u}\, d\hat{W}_\tau + \tilde{v}\, d\tau \quad \text{for } 0 < |\xi| < \delta_0.$$

(Note that the process never runs under the mirror coupling when $\xi = 0$.) Here, the bounded geometry and non-positive curvature of M again implies that $\tilde{u}$ and $|\tilde{v}|$ are bounded with bounds depending only on M, δ_0, and the bounds on a and b. We further claim that $\tilde{u}$ is also bounded from below by a positive constant depending only on M, δ_0, and the bounds on a and b. This is the key fact, and it uses the particular structure of the coupling. We now establish this claim.

It's easiest to visualize what's happening by looking at a horizontal slice of $M \times M$. The instantaneous picture is given by Figure 1, which illustrates the mirror map for Brownian differentials obtained from (dW^1_τ, dW^2_τ) by reflection and the minimal geodesics from y to x and the slice of Σ (which we denote Σ_0 and which is a piece of Cut_x), from which the gradient of ξ and its interaction with the generator of the process are determined. Let λ be the distance from y to Σ_0. By symmetry under interchanging x and y, the symmetry of being in the cut locus of a point, and the product structure on $M \times M$, we see that $\xi = \pm\lambda/\sqrt{2}$, with the sign coming form the fact that ξ is a signed distance.

Since the martingale part of $d\xi_\tau$ depends only on the first order structure at a point, we see that we can consider the horizontal and vertical components separately (and that the martingale part is the same, infinitesimally, for both the $K \equiv 0$ and the $K \equiv -1$ cases). If we let φ be the angle between the minimal geodesic from x to y and the minimal geodesic from y to Σ_0, then the martingale part of λ relative to the evolution of y_τ (with x temporarily fixed) is $-b\cos\varphi\, dW^1_\tau - b\sin\varphi\, dW^2_\tau$. As we see from symmetry, the martingale part of λ relative to the evolution of x_τ is $-a\cos\varphi\, dW^1_\tau + a\sin\varphi\, dW^2_\tau$. Combing these, we see that the martingale part of $d\xi_t$ is

$$\pm\frac{1}{\sqrt{2}}\left[-(a+b)\cos\varphi\, dW^1_\tau + (a-b)\sin\varphi\, dW^2_\tau\right],$$

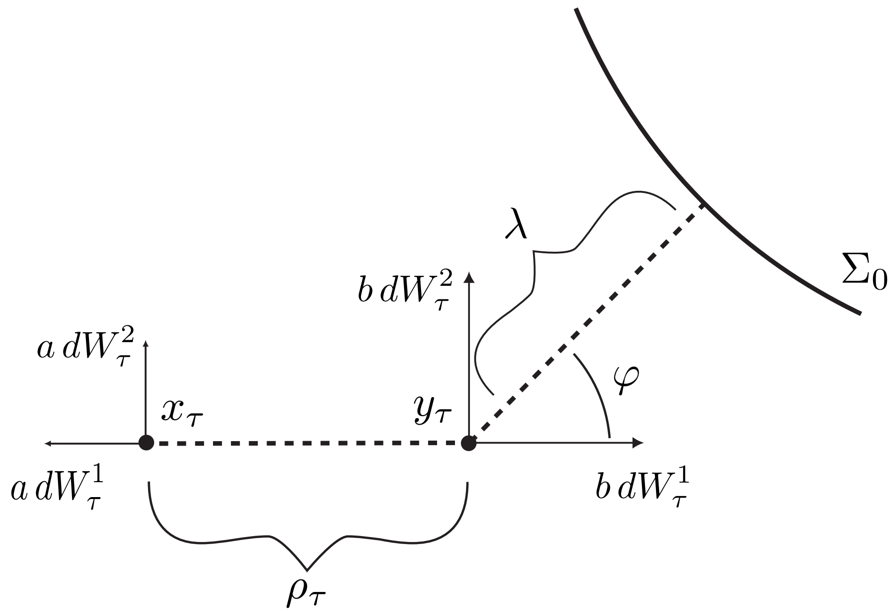

FIGURE 1. The configuration of the joint process relative to the cut locus.

and thus, using that W_τ^1 and W_τ^2 are independent, we have

$$\tilde{u} = \sqrt{(a+b)^2 \cos^2 \varphi + (a-b)^2 \sin^2 \varphi}$$

for an appropriate choice of $\hat{W}$ in the SDE satisfied by ξ_τ.

Next, we need to show that φ is bounded away from $\pi/2$ (on the set $\{0 < |\xi| < \delta_0\}$), with the bound only depending on M and δ_0. To do this, consider the behavior of φ as we let y_τ approach a point in Σ_0. In the limit, the geodesic from x_τ to y_τ cannot be tangent to Σ_0, since for surfaces of non-positive curvature the geodesics from x are transverse to Cut_x. Thus, the corresponding limit of φ is strictly less than $\pi/2$. By compactness (and continuity of φ), all such limiting values of φ are bounded away from $\pi/2$, with bound depending only on the geometry of M. Now φ varies continuously as y_τ approaches any point in Σ_0, so again by compactness, φ will be bounded away from $\pi/2$ on any sufficiently small neighborhood of Σ_0. Thus, by assuming that δ_0 is small enough (this is the other condition on δ_0 mentioned above), we see that φ is bounded away from $\pi/2$ (on the set $\{0 < |\xi| < \delta_0\}$), with the bound only depending on M and δ_0. Continuing, since a and b are bounded below by a positive constant, the claim that $\tilde{u}$ is bounded from below follows.

Thus the SDE satisfied by ξ_τ switches between these two possibilities at the stopping times τ_i^δ. Using the bounds on the coefficients of the SDEs just given, it is a standard exercise in stochastic analysis to show that the time, over $\tau \in [0,t]$, that ξ_τ spends in the interval $[-2\delta, 2\delta]$ goes to zero with δ almost surely, and we omit the details. As noted, this completes the proof. □

7. CONVERGENCE OF FIRST ORDER TO CONSTANT CURVATURE IN THE CASE $\chi(M) = 0$

Now that we have our uniqueness/verification theorem and the general coupling procedure, we begin exploring some of the consequences. As usual, for simplicity, we assume that we have a smooth solution $\overline{p}_t$ for all time $t \geq 0$ on the manifold M. We take here a flat metric h, which is possible under the assumption that $\chi(M) = 0$.

The main result of this section is the following.

Theorem 10. *For M, h, and p_0 as above, suppose that we have a smooth solution $\overline{p}_t$ to Equation (9) for all $t \in [0,\infty)$. Then there exist constants, $c, C > 0$ which depend only on the metrics g_0 and h such that*

$$\sup_{x\in M} |\bar{p}_t(x)| \leq ce^{-Ct}. \tag{26}$$

Proof. Fix a time $t > 0$, a time $s \in [0,t)$ and a point $x \in M$ so that the Ricci flow has a solution on $[0,t]$. The first thing to notice is that $p_\tau = \bar{p}_{t-\tau}(x_\tau)$ is a martingale. Thus we have the following stochastic representation

$$\overline{p}_t(x) = \mathbb{E}\left[\overline{p}_{t-\sigma}(x_\sigma)\right] \tag{27}$$

valid for any stopping time σ with $0 \leq \sigma \leq t$. In particular, setting $\sigma = t$ shows that $\overline{p}_t(x)$ is a weighted average of the values of p_0. Thus

$$\min_M p_0 \leq \min_M \overline{p}_t \leq \max_M \overline{p}_t \leq \max_M p_0 \tag{28}$$

for any t. The main idea for getting (26) is to prove that for some $c, C > 0$,

$$\operatorname{osc} \bar{p}_t \leq ce^{-Ct}. \tag{29}$$

Indeed, if this is true, then combining this with the fact that the integral of $e^{2\bar{p}_t}$ with respect to the volume induced by h is 1, we deduce that there is at least one point $\tilde{x}$ for which $\bar{p}_t(\tilde{x}) = 0$ and from here it is clear that we get (26).

We now choose any two starting points x and y for the processes x_τ and y_τ. Over each of these points, there's exactly one point ($\overline{p}_1(x)$ and $\overline{p}_1(y)$) in the fiber which is in the reachable set Γ_t. We wish to run the controlled process starting from both $(x, \overline{p}_t(x))$ and $(y, \overline{p}_t(y))$, and couple them so that they meet as quickly as possible. Our reachable sets have the semigroup property, i.e. the process (x_τ, p_τ) at time $\tau \in [0,t]$ is on $\Gamma_{t-\tau}$, and since we know that we have a solution until time t, we know that after running the controlled processes for time $\tau \leq t$ they will be on the solution section corresponding to the Ricci flow at time $t - \tau$. This means that if the particles couple on M, they couple in the total space as well, that is, $x_\tau = y_\tau$ implies that $\overline{p}_{t-\tau}(x_\tau) = \overline{p}_{t-\tau}(y_\tau)$ as well.

In light of this, if σ is the coupling time of x_σ and y_σ, the martingale property gives that

$$\begin{aligned} \bar{p}_t(x) - \bar{p}_t(y) &= \mathbb{E}[\bar{p}_{t-\sigma\wedge s}(x_{\sigma\wedge s})] - \mathbb{E}[\bar{p}_{t-\sigma\wedge s}(y_{\sigma\wedge s})] \\ &= \mathbb{E}[\bar{p}_{t-\sigma}(x_\sigma) - \bar{p}_{t-\sigma}(y_\sigma), \sigma \leq s] + \mathbb{E}[\bar{p}_{t-s}(x_s) - \bar{p}_{t-s}(y_s), s < \sigma] \\ &= \mathbb{E}[\bar{p}_{t-s}(x_s) - \bar{p}_{t-s}(y_s), s < \sigma]. \end{aligned} \tag{30}$$

The outcome of this is that

$$\operatorname{osc} \bar{p}_t \leq \mathbb{P}(s < \sigma) \operatorname{osc} \bar{p}_{t-s}. \tag{31}$$

What remains to be controlled here is $\mathbb{P}(s < \sigma)$. While the above is true for any coupling of x_τ and y_τ, we wish to use the mirror coupling, as was introduced in the previous section. The main property of this coupling, for us, is contained in (25) which gives the equation satisfied by the distance function $\rho_\tau = d(x_\tau, y_\tau)$, namely

$$d\rho_\tau = (a+b)\, d\hat{W}_\tau + \frac{1}{2\rho_\tau}(a-b)^2\, d\tau - L_\tau \tag{32}$$

with $a_\tau = e^{-\bar{p}_{t-\tau}(x_\tau)}$, $b_\tau = e^{-\bar{p}_{t-\tau}(y_\tau)}$ and $\hat{W}$ being a one dimensional Brownian motion on the time interval $[0,t]$. Obviously the time τ runs up to σ (the hitting time of 0) or t, whichever comes first and the term L_τ is non-negative. We are interested in estimating the probability this hitting time σ occurs after time s. To this end, the first thing which will be used here is the fact that from (28) we know that a and b are all bounded from above as well from below. So we have two constants $A, B > 0$ which are depending only on p_0, or otherwise the starting metric g_0, with the property that

$$A \le a, b \le B. \tag{33}$$

To move on, we let

$$\lambda(u) = \int_0^u \frac{1}{(a_v + b_v)^2} dv$$

be the time-change making the martingale part of ρ_τ from (32) into a Brownian motion. Then with the notation $\tilde{\rho}_u = \rho_{\lambda(u)}$,

$$d\tilde{\rho}_u = d\tilde{W}_u + \frac{1}{2\tilde{\rho}_u} \frac{(a-b)^2}{(a+b)^2} du - d\tilde{L}_u \tag{34}$$

where a and b are evaluated at time $\lambda(u)$ and the above equation is valid for $u \in [0, t \wedge \lambda^{-1}(t))$, where $\lambda^{-1}(t)$ is the first value of u corresponding to $\lambda(u) = t$. Obviously $cu \le \lambda(u) \le Cu$ for some constants $c, C > 0$ and also because of (33),

$$\left| \frac{a-b}{a+b} \right| \le \frac{B-A}{B+A} = 1 - \delta < 1.$$

Ignoring the L term in (34) and then using standard comparison for ordinary stochastic differential equations, we learn that the process $\tilde{\rho}$ is bounded above by a Bessel process of index $\delta < 2$ and starting at some value ρ_0 bounded by the diameter (with respect to the metric h) of the manifold M. Thus, invoking [13, Equation (15)] which gives the distribution of the hitting time $\tilde{\sigma}$ of 0 for a Bessel process of index $\delta < 2$ starting at $\tilde{\rho}$, we obtain

$$\mathbb{P}(s < \tilde{\sigma}) = \frac{1}{\Gamma(1-\delta/2)} \int_0^{\tilde{\rho}_0/(2s)} y^{-\delta/2} e^{-y} dy.$$

Finally, since $cu \le \lambda(u) \le Cu$ and the diameter of the manifold M is finite, we arrive at

$$\mathbb{P}(s < \sigma) \le \frac{1}{\Gamma(1-\delta/2)} \int_0^{D/s} y^{-\delta/2} e^{-y} dy =: \Lambda(s)$$

where D is a constant which depends only on the initial metric g_0 and some geometry of the underlying metric h (more precisely the diameter of M with respect to h). Hence it turns out that the function Λ is determined by the metrics h and g_0.

To summarize, from (31) and the preceding we now have that

$$\operatorname{osc} \bar{p}_t \le \Lambda(s) \operatorname{osc} \bar{p}_{t-s}.$$

Using this, it is easy to get (29) as follows. For $t \in [0,1]$, we know from (28), that $\operatorname{osc} \bar{p}_t \le \operatorname{osc} p_0$. Now for each $t \in [n, n+1]$, $n \ge 1$, using repeatedly the above inequality, we arrive at

$$\operatorname{osc} \bar{p}_t \le \Lambda(1)^n \operatorname{osc} \bar{p}_{t-n} \le \Lambda(1)^t \operatorname{osc} p_0 / \Lambda(1)$$

which is exactly the exponential decay of (29) since $0 < \Lambda(1) < 1$. □

Remark. It is interesting to point out that we can prove the same exponential decay as in Theorem 10 for the case of $\chi(M) < 0$ using the coupling argument. This decay is, however, already taken care of by the a priori estimates of Corollary 7. Nonetheless, this coupling argument is the one we will employ for the gradient estimates in the following section.

8. Estimates on the gradient decay of the normalized Ricci flow in the case $\chi(M) \leq 0$

We continue under the same assumptions that M is a compact surface with reference metric h of constant curvature 0 or -1 (so M has non-positive Euler characteristic by the Gauss-Bonnet theorem) and g_0 is a smooth initial metric in the same conformal class and with the same area as h, so that the normalized Ricci flow has a smooth solution for all time which is given by $\bar{p}_t$. Now, $\bar{p}_t$ converges in the C^0-norm exponentially fast to 0 as shown in Corollary 7 for the case $\chi(M) < 0$ and Theorem 10 for the case $\chi(M) = 0$. So we have that for some constants $c, C > 0$,

$$\sup_{x \in M} |\bar{p}_t(x)| \leq ce^{-Ct}. \tag{35}$$

Let

$$G_t = \sup_{x \in M} |\nabla \bar{p}_t(x)|$$

The idea is to start with

$$\langle \nabla \bar{p}_t(x), \xi \rangle = \lim_{h \to 0} \frac{\bar{p}_t(\gamma_h(x)) - \bar{p}_t(x)}{h}$$

where ξ is a unit vector in the tangent space at x and $\gamma_t(x)$ is any curve started at x with initial speed ξ. Then we use the coupling to estimate $\bar{p}_t(x) - \bar{p}_t(y)$ for x and y close to one another. Due to the non-linearity of the flow, the estimates coming from the above will still contain the gradient bounds, but in the end, letting x and y come close to one another leads to a functional inequality on G_t, from which we are able to derive the desired estimate.

Theorem 11. *If $\chi(M) \leq 0$ then G_t goes to 0 exponentially fast. As a consequence, $\bar{p}_t$ converges to 0 exponentially fast in C^1.*

Proof. Pick two sufficiently close points $x, y \in M$ and some $t > 0$, and let $\rho_\tau = d(x_\tau, y_\tau)$ for $0 \leq \tau \leq t$ be the distance (measured with respect to the time independent metric h) between the processes x_τ and y_τ started at x, and y respectively. We are going to use mirror coupling for the processes $x.$ and $y.$. Recall that the equations coupling equations satisfied by (x_τ, p_τ) and (y_τ, q_τ) are given by

$$\begin{aligned}
dx_\tau &= e^{-p_\tau} \left[\sum_{i=1}^{2} \mathfrak{e}_i(x_\tau) \sqrt{2} \circ dW_\tau^i \right] \\
dy_\tau &= e^{-q_\tau} \left[\sum_{i=1}^{2} \mathfrak{e}_i(y_\tau) \sqrt{2} \circ d\tilde{W}_\tau^i \right] \\
dp_\tau &= e^{-p_\tau} \left[\sum_{i=1}^{2} a_i \sqrt{2}\, dW_\tau^i \right] + r(e^{-2p_\tau} - 1)\, d\tau \\
dq_\tau &= e^{-q_\tau} \left[\sum_{i=1}^{2} a_i' \sqrt{2}\, d\tilde{W}_\tau^i \right] + r(e^{-2q_\tau} - 1)\, d\tau,
\end{aligned} \tag{36}$$

where $r = 0$ or -1 and $\tilde{W}$ is the Brownian motion given by the mirror coupling.

We consider σ, the coupling time of $x.$ and $y.$. From the fact that $p_\tau + r\int_0^\tau (1 - e^{-2p_u}) du$ is a martingale and $p_\tau = \bar{p}_{t-\tau}(x_\tau)$, we write,

$$\bar{p}_t(x) - \bar{p}_t(y) = \mathbb{E}[p_{t\wedge\tau} - q_{t\wedge\tau}] - r\mathbb{E}\left[\int_0^{t\wedge\tau} (e^{-2p_u} - e^{-2q_u}) du \right] \tag{37}$$

for any stopping time τ. The useful estimates we are interested in are estimates from above of $\bar{p}_t(x) - \bar{p}_t(y)$, and this is good if we assume that $\bar{p}_t(x) - \bar{p}_t(y) > 0$. This is always possible unless $\bar{p}_t$ is constant in which case the gradient is 0, so there is nothing to prove then. Thus, assume that $\bar{p}_t(x) - \bar{p}_t(y) > 0$ for some points x and y (which is the same as $p_0 > q_0$) and take α to be the first time u for which $p_u = q_u$. With this choice of the stopping time, for any $u \in [0, \alpha]$ we know that $p_u \geq q_u$, which thus means $e^{-2p_u} - e^{-2q_u} \leq 0$. This combined with the crucial fact that $r \leq 0$ and the exponential decay of $\bar{p}_t$, implies that for any $s \in [0, t]$,

$$\bar{p}_t(x) - \bar{p}_t(y) \leq \mathbb{E}[p_\alpha - q_\alpha, \alpha \leq s] + \mathbb{E}[p_s - q_s, s < \alpha] \leq ce^{-C(t-s)}\mathbb{P}(s < \alpha).$$

The point is that if σ is the first coupling time, of the processes $x.$ and $y.$, it is obvious that $\alpha \leq \sigma$ and thus

$$\mathbb{P}(s < \alpha) \leq \mathbb{P}(s < \sigma) \text{ for any } s \in [0, t].$$

which in turn yields

$$\bar{p}_t(x) - \bar{p}_t(y) \leq ce^{-C(t-s)}\mathbb{P}(s < \sigma) \quad \text{for any } s \in [0, t]. \tag{38}$$

With this equation our next task becomes the estimate of $\mathbb{P}(s < \sigma)$.

From Theorem 9 we learn that the distance process ρ_τ satisfies

$$d\rho_\tau \leq (e^{-p_\tau} + e^{-q_\tau})dB_t + \frac{(e^{-p_\tau} - e^{-q_\tau})^2}{2\rho_\tau}d\tau \tag{39}$$

in the case $r = 0$ and

$$d\rho_\tau \leq (e^{-p_\tau} + e^{-q_\tau})dB_\tau + \frac{1}{2}\left[\left(e^{-p_\tau} - e^{-q_\tau}\right)^2 \coth\rho_\tau + 2e^{-p_\tau - q_\tau}\tanh\frac{\rho_\tau}{2}\right]d\tau \tag{40}$$

in the case $r = -1$. Here B_t is a one dimensional Brownian motion run in the time interval $[0, t]$.

So far, we have used this strategy of coupling in the proof of Theorem 10, in which, due to the singularity in the drift of the equations (39) and (40), we compared the distance function ρ_τ with a Bessel process. For the gradient estimates, we are going to remove the singularity based on the observation that

$$p_\tau = \bar{p}_{t-\tau}(x_\tau) \text{ and similarly } q_\tau = \bar{p}_{t-\tau}(y_\tau).$$

The upshot of this is that the term $e^{-p_\tau} - e^{-q_\tau}$ is in fact of order ρ_τ. More precisely, due to the boundedness of $\bar{p}$,

$$|e^{-p_\tau} - e^{-q_\tau}| = |e^{-\bar{p}_{t-\tau}(x_\tau)} - e^{-\bar{p}_{t-\tau}(y_\tau)}| \leq Cd(x_\tau, y_\tau)\sup_{x\in M}|\nabla\bar{p}_{t-\tau}(x)| = CG_{t-\tau}\rho_\tau.$$

Since $\rho_\tau \leq D$, where D is the diameter of M, it is straightforward to show that either of (39) and (40) implies

$$d\rho_\tau \leq (e^{-p_\tau} + e^{-q_\tau})dB_\tau + C(1 + G_{t-\tau}^2)\rho_\tau d\tau.$$

To go further from here, consider $\tilde{\rho}_\tau$ the solution to

$$d\tilde{\rho}_\tau = (e^{-p_\tau} + e^{-q_\tau})dB_\tau + C(1 + G_{t-\tau}^2)\tilde{\rho}_\tau d\tau.$$

with the same initial condition $\rho_0 = d(x, y)$ as ρ_τ. Standard arguments (in fact a simple application of Gronwall's Lemma) give that

$$\rho_\tau \leq \tilde{\rho}_\tau$$

which results in the fact that the first hitting time of 0 for ρ is less then or equal to the first hitting time of 0 for $\tilde{\rho}$. Now if $\tilde{\sigma}$ denotes the hitting time of 0 for the process $\tilde{\rho}_t$

$$\mathbb{P}(s < \sigma) \leq \mathbb{P}(s < \tilde{\sigma}) \text{ for all } s \in [0, t]. \tag{41}$$

Therefore the task now is to estimate the latter, and to do this we solve for $\tilde{\rho}$ as

$$\tilde{\rho}_\tau = \left(\rho_0 + \int_0^\tau (e^{-p_v} + e^{-q_v}) e^{-\int_0^v f(z)dz} dB_v\right) e^{\int_0^\tau f(z)dz}$$

with the notation $f(\tau) = C(1+G_{t-\tau}^2)$, for $0 \le \tau \le t$. Consequently the first hitting time of 0 for $\tilde{\rho}$ is the first hitting time of $-\rho_0$ for the time-changed Brownian motion $\int_0^\tau (e^{-p_v} + e^{-q_v}) e^{-\int_0^v f(z)dz} dB_v$. In law, this is the same as the first hitting time of $-\rho_0$ of $B_{c(\tau)}$, with the time change

$$c(\tau) = \int_0^\tau (e^{-p_v} + e^{-q_v})^2 e^{-2\int_0^v f(z)dz} dv.$$

Once again using the boundedness of $\bar{p}$, we can find a constant $C > 0$ such that

$$c(\tau) \ge c'(\tau) := C \int_0^\tau e^{-2\int_0^v f(z)dz} dv \text{ for } \tau \in [0, t]$$

Now, if $\sigma_{-\rho_0}$ is the first hitting time of $-\rho_0$ for the Brownian motion, then the hitting time of $-\rho_0$ for $B_{c(\tau)}$ is given by $c^{-1}(\sigma_{-\rho_0} \wedge c(t))$. This combined with (41) yields that

$$\mathbb{P}(s < t \wedge \tilde{\sigma}) = \mathbb{P}(s < c^{-1}(\sigma_{-\rho_0} \wedge c(t))) \le \mathbb{P}(c(s) \le \sigma_{-\rho_0}) \le \mathbb{P}(c'(s) < \sigma_{-\rho_0}). \tag{42}$$

The distribution of $\sigma_{-\rho_0}$ is actually well understood (see for instance the Remark after Proposition 3.7 in [21]), and its density is given by $\frac{\rho_0}{\sqrt{2\pi x^3}} e^{-\rho_0^2/(2x)}$ on the positive axis, which results in

$$\mathbb{P}(c'(s) < \sigma_{-\rho_0}) = \int_{c'(s)}^\infty \frac{\rho_0}{\sqrt{2\pi x^3}} e^{-\rho_0^2/(2x)} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\frac{\rho_0}{\sqrt{c'(s)}}} e^{-\tau^2/2} d\tau.$$

Going back to (38) and using the preceding, we conclude that for $s \in [0, t]$,

$$\bar{p}_t(x) - \bar{p}_t(y) \le c e^{-C(t-s)} \int_0^{\frac{\rho_0}{\sqrt{c'(s)}}} e^{-\tau^2/2} d\tau,$$

from which, using the fact that $d(x, y) = \rho_0$ and letting ρ_0 go to 0, we fairly easily deduce that

$$G_t \le c \frac{e^{-C(t-s)}}{\sqrt{c'(s)}},$$

which we rearrange as

$$A_t \int_0^s e^{-\int_0^\tau A_{t-u} du} d\tau \le c e^{-Ct} \text{ for all } s \in [0, t] \text{ with } A_\tau = CtG_\tau^2.$$

From here the exponential decay of A_t is taken care of by the following Lemma. $\square$

Lemma 12. *Suppose A_t is non-negative, depends continuously on t for $t \ge 0$ and has the property that for some constants $c, C > 0$,*

$$A_t \int_0^s e^{-\int_0^\tau A_{t-u} du} d\tau \le c e^{-C(t-s)} \textit{ for all } s \in [0, t]. \tag{43}$$

Then there are constants $k, K > 0$ such that

$$A_t \le K e^{-kt} \textit{ for all } t > 0.$$

Proof. For each $n \geq 1$, let $m_n = \sup_{t\in[n,n+1]} A_t$ and $M_n = \sup_{t\in[n-1,n+1]} A_t$. Notice that the exponential decay we are looking for is actually equivalent to $m_n \leq Ke^{-kn}$ for large enough n.

Now, for $t \in [n, n+1]$ and $s \in [0,1]$, we have $t - s \in [n-1, n+1]$ and therefore $-A_{t-u} \geq -M_n$, which combined with (43) yields, for t near the supremum of A_t on $[n, n+1]$, and eventually another constant $c > 0$

$$m_n \int_0^1 e^{-\tau M_n} d\tau = m_n \frac{1 - e^{-M_n}}{M_n} \leq e^{-cn} \text{ for all large } n.$$

which in turn gives

$$m_n \leq \frac{M_n}{1 - e^{-M_n}} e^{-cn}. \tag{**}$$

Now, for each particular n, we have one of the following two alternatives:

(1) $M_n \leq e^{-cn/2}$, in which case it is clear that

$$m_n \leq e^{-cn/2}. \tag{\#}$$

(2) $M_n > e^{-cn/2}$, and in this case $1 - e^{-M_n} > 1 - e^{-e^{-cn/2}} > \frac{1}{2}e^{-nc/2}$ for large enough n, say $n \geq n_0$. From (**), it follows that $m_n \leq 2M_n e^{-cn/2} \leq M_n e^{-c/2}$ for all n large enough, say $n \geq n_1$. This inequality implies that

$$m_n \leq m_{n-1} e^{-c/2} \text{ for all } n \geq n_1. \tag{\#\#}$$

Indeed if the supremum of A_t on the interval $[n-1, n+1]$ is the same as the supremum on $[n, n+1]$, then $M_n = m_n$ and this in turn implies $M_n = 0$, in particular we trivially have (##). If the supremum of A_t on $[n-1, n+1]$ is the same as the supremum on $[n-1, n]$, this gives $M_n = m_{n-1}$ and then (**) gives (##).

Using these two alternatives we argue as follows. Assume that there is a large enough n_2 such that $m_{n_2} \leq e^{-cn_2/2}$. Then an easy induction using the two alternatives above give that $m_n \leq e^{-cn/2}$ for all $n \geq n_2$. If there is no such n_2, this mean that for all $n \geq n_1$ we clearly have the second alternative and in this case $m_n \leq m_{n_1} e^{-nc/2}$. In both cases we obtain the exponential decay we were looking for.

An alternative proof can be given as follows. Take a sufficiently large constant $K > 0$, which will be chosen later. Now we look at $B_t = A_t e^{kt}$. Assume there is a time $t \geq K$ such that B_t is the maximum over the time interval $[0, t]$. We then have $A_\tau \leq A_t e^{-k\tau}$ for $\tau \in [0, t]$ and from (43) with $s = 1$,

$$A_t \int_0^1 e^{-\tau A_t e^k du} d\tau \leq ce^{-Ct},$$

and from this

$$1 - e^{-A_t e^k} \leq ce^k e^{-Ct},$$

which gives that

$$A_t \leq -e^{-k} \log(1 - ce^k e^{-Ct}).$$

If we choose the constant K large enough and k small enough, so that $1/2 < 1 - ce^k e^{-CK}$, then we arrive at

$$A_t \leq ce^{-Ct} \leq Ce^{-kt}.$$

In particular this means that $A_t e^{kt} \leq C$. As this B_t is the maximum of B_τ over $\tau \in [0, t]$, we get that $A_\tau \leq Ce^{-kt}$. The other alternative which remains is that there is no $t \geq K$ for which B_t attains a maximum on $[0, t]$ for $t \geq K$. In this case we deduce that B_t has a maximum for $t \in [0, K]$ and the exponential decay follows again. $\square$.

Remark. As one can see we do not fully use the condition in (43) for all $s \in [0,t]$. It suffices to have this hold for $s \in [0,1]$ and large enough t.

Before we close this section let us point out that the exponential decay of the gradient has the following consequence that we will use later on for the estimates of the higher order derivatives.

Corollary 13. *Under the same assumptions as in Theorem 11,*

$$\mathbb{P}(s < \sigma) \le C\frac{\rho_0}{\sqrt{s}} \text{ for } s \in (0,t]. \tag{44}$$

Proof. This follows by combining (41), (42), and the fact that $c'(s)$ is bounded (due to the gradient estimate). □

9. Triple Coupling

9.1. **Basic idea.** We have just used coupling to prove the exponential convergence of $\bar{p}$ to 0 in the C^1-topology. The next step in our analysis is the estimate of the decay of the Hessian of $\bar{p}$, which, from the Ricci flow equation, implies the convergence of the curvature to a constant. The basic idea starts with writing

$$\langle \operatorname{Hess} \bar{p}_t(z)\xi, \xi\rangle = \lim_{\rho_0 \to 0} \frac{\bar{p}_t(\gamma(-\rho_0)) - 2\bar{p}_t(z) + \bar{p}_t(\gamma(\rho_0))}{\rho_0^2}$$

where ξ is a unit vector at z, and γ is a geodesic running at unit speed started (at $t=0$) at z with velocity ξ. Now we are concerned with three points, $x = \gamma(-\rho_0)$, $y = \gamma(\rho_0)$, and the middle point z. As in the gradient estimate case, we want to write $\bar{p}(x)$, $\bar{p}(y)$, and $\bar{p}(z)$ as integrals of some functions of the associated Brownian motions and then use probabilistic estimates to find bounds for $\bar{p}_t(\gamma(-\rho_0)) - 2\bar{p}_t(z) + \bar{p}_t(\gamma(\rho_0))$ in terms of ρ_0.

There is very little literature on this idea, though it certainly seems that this probabilistic tool is quite useful for estimating second-order derivatives for evolution equations. The only reference to this approach we're aware of is in [10], where it is essentially used to estimate the Hessian of harmonic functions on Euclidean domains.

To make this idea more precise, we will develop a mechanism of triple coupling (that is, a coupling of three particles, as opposed to just two). We will use mirror coupling for the processes corresponding to the particles x and y, taking them as time changed Brownian motions, as in the previous section. Now we wish to include a third particle, namely z, which we want to couple together with x and y. It is natural to want to have this "middle particle" remain on the geodesic joining the other two. We will see that this is possible (at least in the cases we're considering) if we allow it to evolve as time-changed Brownian motion, possibly with drift along the direction of the geodesic.

Instead of starting with a time-changed Brownian motion with drift, z_τ and then trying to figure out the time change and drift necessary so that it stays on the geodesic, we do it the other way around. Namely, since we want the particle z_τ to move on the geodesic, we determine the conditions on the distance to one of the other points so that the corresponding point on the geodesic is a time-changed Brownian motion with a drift along the geodesic. For the purpose of the Hessian estimates, and in light of the gradient decay, this will be sufficient.

9.2. **Rigorous Approach.** Assume we start with an arbitrary Riemannian surface M and that x_τ, y_τ run as time-changed Brownian motions with the time changes a and b, as above. The idea is that the middle point z_τ on the geodesic joining x_τ and y_τ is completely described by specifying the distance $\rho_{1,\tau}$ from z_τ to one of the ends, say x_τ. We use a mirror coupling of the particles x_τ and y_τ and $\rho_{1,\tau}$ will be described in terms of a real-valued SDE. In addition to ρ_1, we will also consider ρ_2, which in intuitive terms is just the distance from the middle particle z_τ to y_τ. We are seeking several key symmetry properties which will play an important role in the economy of the Hessian estimates to follow.

In what follows, as always, fix a time horizon $t > 0$, and assume that $a = a(\tau, x, y, \rho_1, \rho_2)$ and $b = b(\tau, x, y, \rho_1, \rho_2)$ are two positive functions defined on $[0,t] \times M \times M \times [0,\infty) \times [0,\infty)$, which will be time changes for the processes x_τ and y_τ. To describe this, again denote by $m_{x,y} : T_xM \to T_yM$ the mirror map, that is, the parallel transport along the minimal unit speed geodesic $\gamma_{x,y}$ joining x and y (assuming that x, y are not at each other's cut locus) followed by the reflection about the orthogonal direction to the geodesic at y.

The system we start with is the following

$$
\begin{cases}
dx_\tau = a(\tau)\left[\displaystyle\sum_{i=1}^{2} \mathfrak{e}_i(x_\tau)\circ dW_\tau^i\right] \\
dy_\tau = b(\tau)\left[\displaystyle\sum_{i=1}^{2} \Psi_\tau[\mathfrak{e}_i(y_\tau)]\circ dW_\tau^i\right] \\
d\rho_{1,\tau} = -a(\tau)\displaystyle\sum_{i=1}^{2}\langle \mathfrak{e}_i(x_\tau), \dot{\gamma}_\tau(0)\rangle dW_\tau^i + \alpha(\tau)dW_\tau^3 + \beta(\tau)d\tau \\
d\rho_{2,\tau} = b(\tau)\displaystyle\sum_{i=1}^{2}\langle \Psi_\tau[\mathfrak{e}_i(y_\tau)], \dot{\gamma}_\tau(l(\tau))\rangle dW_\tau^i + \tilde{\alpha}(\tau)dW_\tau^3 + \tilde{\beta}(\tau)d\tau,
\end{cases}
\tag{45}
$$

where $\Psi_\tau = m_{x_\tau,y_\tau}\mathfrak{e}(x_\tau)\mathfrak{e}(y_\tau)^{-1}$ is the reflection map acting on T_{y_τ}, γ_τ is the minimal geodesic running at unit speed from x_τ to y_τ, and W^3 is a one-dimensional Brownian motion independent of (W^1, W^2). As a notation, let $l(\tau)$ be the length of the geodesic γ_τ. Here we do not specify what the functions α, $\tilde{\alpha}$, β, $\tilde{\beta}$ are as we will do this along the way, depending on the properties we want to reveal. They are defined, like a and b, on $[0,t] \times M \times M \times [0,\infty) \times [0,\infty)$. The equations for ρ_1 and ρ_2 can be thought of as the equations of the distances from the middle point z_τ to x_τ and y_τ, as indicated in the previous section, and also as discussed for the coupling in [15, Section 6.6]. Notice here an important point– since

$$\langle \Psi_\tau \mathfrak{e}_i(y_\tau), \dot{\gamma}_\tau(l(\tau))\rangle = -\langle \mathfrak{e}_i(x_\tau), \dot{\gamma}_\tau(0)\rangle,$$

the last equation of (45) can be rewritten as

$$d\rho_{2,\tau} = -b(\tau)\sum_{i=1}^{2}\langle \mathfrak{e}_i(x_\tau), \dot{\gamma}_\tau(0)\rangle dW_\tau^i + \tilde{\alpha}(\tau)dW_\tau^3 + \tilde{\beta}(\tau)d\tau. \tag{46}$$

There is no problem with the existence of a solution for the system (45) (as long as the entries $a, b, \alpha, \beta, \tilde{\alpha}$, and $\tilde{\beta}$ are smooth) up to the stopping time $\mathcal{T}$, which is the first time τ when $\rho_{1,\tau}\rho_{2,\tau}$ hits 0 or when $d(x_\tau, y_\tau)$ hits a (small) r_0 smaller than the injectivity radius. This way we have a well-defined system and do not have to worry about the extension beyond the cut locus, as we did in the previous (two particle) coupling case. From now on, during this section we will assume that the time in the system (45) is run until $\mathcal{T}$.

The object of interest to us is the process (x, y, ρ_1, ρ_2). It is clear that this is a diffusion, and it is a relatively straightforward task to determine that the generator of (x, y, ρ_1, ρ_2) is

$$\begin{aligned}
&\frac{a^2}{2}\Delta_x + \frac{b^2}{2}\Delta_y + \frac{a^2+\alpha^2}{2}\partial^2_{\rho_1} + \frac{b^2+\tilde{\alpha}^2}{2}\partial^2_{\rho_2} + ab\langle m_{xy}X_{1,i}, Y_{2,j}\rangle X_{1,i}Y_{2,j}\\
&\quad - a^2\langle X_{1,i}, \dot{\gamma}_{x,y}(0)\rangle X_{1,i}\partial_{\rho_1} - ab\langle X_{1,i}, \dot{\gamma}_{x,y}(0)\rangle X_{1,i}\partial_{\rho_2}\\
&\quad - ab\langle m_{x,y}X_{1,i}, \dot{\gamma}_{x,y}(0)\rangle X_{1,i}\partial_{\rho_1} - b^2\langle m_{x,y}X_{1,i}, \dot{\gamma}_{x,y}(0)\rangle X_{1,i}\partial_{\rho_2}\\
&\quad + (\alpha\tilde{\alpha} - ab\sum_{i=1}^{2}\langle X_{1,i}, \dot{\gamma}_{x,y}(0)\rangle^2)\partial_{\rho_1}\partial_{\rho_2} + \beta\partial_{\rho_1} + \tilde{\beta}\partial_{\rho_2},
\end{aligned}$$

with $X_{1,i}$, $i = 1, 2$ being an orthonormal basis of T_xM and $Y_{2,j}$, $j = 1, 2$ an orthonormal basis of T_yM. In fact, we can choose $X_{1,1} = \dot{\gamma}_{x,y}(0)$ and $X_{1,2} = \xi_1 \in T_xM$, which is perpendicular to $\dot{\gamma}_{x,y}(0)$. Similarly, choose $Y_{2,1} = \dot{\gamma}_{y,x}(0)$ and $Y_{2,2} = \xi_2 = m_{x,y}\xi_1$, or said simply, the parallel transport of ξ_1 along the geodesic $\gamma_{x,y}$. With these choices, the generator simplifies to

$$\begin{aligned}
\mathcal{L} = &\frac{a^2}{2}\Delta_x + \frac{b^2}{2}\Delta_y + \frac{a^2+\alpha^2}{2}\partial^2_{\rho_1} + \frac{b^2+\tilde{\alpha}^2}{2}\partial^2_{\rho_2} + ab(\dot{\gamma}_{x,y}(0)\dot{\gamma}_{y,x}(0) + \xi_1\xi_2)\\
&- a^2\dot{\gamma}_{x,y}(0)\partial_{\rho_1} - ab\dot{\gamma}_{y,x}(0)\partial_{\rho_1}\\
&- ab\dot{\gamma}_{x,y}(0)\partial_{\rho_2} - b^2\dot{\gamma}_{y,x}(0)\partial_{\rho_2}\\
&+ (\alpha\tilde{\alpha} - ab)\partial_{\rho_1}\partial_{\rho_2} + \beta\partial_{\rho_1} + \tilde{\beta}\partial_{\rho_2}.
\end{aligned} \tag{47}$$

The first property we want to see is that $\rho_1 + \rho_2 = \rho$. This property is nothing but the geometric picture that ρ_1 is the distance from z_τ to x_τ while ρ_2 is the distance between z_τ to y_τ.

To do this we recall that the distance ρ_τ between the mirror-coupled processes x_τ and y_τ is given by

$$d\rho_\tau = -(a(\tau) + b(\tau))\sum_{i=1}^{2}\langle \mathfrak{e}_i(x_\tau), \dot{\gamma}_\tau(0)\rangle dW^i_\tau + \frac{1}{2}\mathcal{I}(\tau)d\tau \tag{48}$$

where $\mathcal{I}$ is the index form of the Jacobi field $J(\tau)$ along the geodesic γ_τ which, at the endpoints, has values aE and bE. We use the notation E for the parallel translation of $\xi_1 \in T_xM$ along the geodesic joining x and y. The index form is computed as

$$\mathcal{I}(J, J) = \int_0^{l(\gamma)} |\dot{J}(u)|^2 + \langle R(\dot{\gamma}(u), J(u))\dot{\gamma}(u), J(u)\rangle du,$$

with $l(\gamma)$ being the length of the geodesic γ. Here the curvature tensor is the standard tensor curvature given as in [5]

$$R(X, Y) = \nabla_X\nabla_Y - \nabla_Y\nabla_X - \nabla_{[X,Y]}.$$

On the other hand, from (45),

$$d(\rho_{1,\tau} + \rho_{2,\tau}) = -(a(\tau) + b(\tau))\sum_{i=1}^{2}\langle \mathfrak{e}_i(x_\tau), \dot{\gamma}_\tau(0)\rangle dW^i_\tau + (\alpha(\tau) + \tilde{\alpha}(\tau))dW^3_\tau + (\beta(\tau) + \tilde{\beta}(\tau))d\tau.$$

We clearly see here that ρ_τ and $\rho_{1,\tau} + \rho_{2,\tau}$ have the same martingale part if $\tilde{\alpha} = -\alpha$. The choice for β and $\tilde{\beta}$ is provided by the following result.

Theorem 14. *Assume that*

$$\begin{cases} \tilde{\alpha} = -\alpha \\ \beta(\tau, x, y, \rho_1, \rho_2) = \frac{1}{2} \int_0^{\rho_1} |\dot{J}(u)|^2 + \langle R(\dot{\gamma}(u), J(u))\dot{\gamma}(u), J(u)\rangle du \\ \tilde{\beta}(\tau, x, y, \rho_1, \rho_2) = \frac{1}{2} \int_{l(\gamma)-\rho_2}^{l(\gamma)} |\dot{J}(u)|^2 + \langle R(\dot{\gamma}(u), J(u))\dot{\gamma}(u), J(u)\rangle du \end{cases} \tag{49}$$

where J is the Jacobi field along the geodesic γ from x to y and having values aE at 0 and bE at $l(\gamma)$.

If in addition, $\rho_{1,0} = \rho_{2,0} = \rho_0/2$, then almost surely $\rho_\tau = \rho_{1,\tau} + \rho_{2,\tau}$.

Proof. Take $\tilde{\rho}_{1,\tau} = \rho_\tau - \rho_{2,\tau}$. It is clear now that we have

$$d(\tilde{\rho}_{1,\tau} - \rho_{1,\tau}) = \int_0^{\tilde{\rho}_{1,\tau}} A(u)du - \int_0^{\rho_{1,\tau}} A(u)du$$

with

$$A(u) = \frac{1}{2}\left[|\dot{J}(u)|^2 + \langle R(J(u), \dot{\gamma}(u))\dot{\gamma}(u), J(u)\rangle du\right].$$

From here, the fact that $\tilde{\rho}_{1,0} = \rho_{1,0}$ (or $\tilde{\rho}_{1,0} - \rho_{1,0} = 0$) and standard application of Gronwall's inequality leads to $\tilde{\rho}_{1,\tau} = \rho_{1,\tau}$, which is what we want. □

We return now to the case where the curvature is constant and start with [4, Lemma 3.4] which says that

$$R(X, Y)Z = -r(\langle X, Z\rangle Y - \langle Y, Z\rangle X). \tag{50}$$

We should point out that the Do Carmo [4], takes the curvature to be given by the negative of the curvature we consider here, or for that matter other people. Then the Jacobi field equation becomes

$$\ddot{J} - R(\dot{\gamma}, J)\dot{\gamma} = 0$$

or equivalently,

$$\ddot{J} + rJ - r\langle\dot{\gamma}, J\rangle\dot{\gamma} = 0. \tag{51}$$

Since this Jacobi field is perpendicular to the geodesic, it follows that

$$\begin{cases} \ddot{J} + rJ = 0 \\ J(0) = aE \\ J(l(\gamma)) = bE. \end{cases}$$

The solution is

$$J(s) = (aw_1(s) + bw_2(s))E(s) \text{ for } s \in [0, l(\gamma)] \tag{52}$$

where w_1, w_2 are defined on the interval $[0, l(\gamma)]$ by the following odes

$$\begin{cases} \ddot{w}_1 + rw_1 = 0 \\ w_1(0) = 1 \\ w_1(l(\gamma)) = 0. \end{cases} \text{ and } \begin{cases} \ddot{w}_2 + rw_2 = 0 \\ w_2(0) = 0 \\ w_2(l(\gamma)) = 1. \end{cases} \tag{53}$$

An integration by parts argument together with the equation of the Jacobi field and the constant curvature assumption reveals that

$$\int_0^s |\dot{J}(u)|^2 - \langle R(J(u), \dot{\gamma}(u))\dot{\gamma}(u), J(u)\rangle du = \int_0^s |\dot{J}(u)|^2 - r|J(u)|^2 du$$
$$= \langle \dot{J}(s), J(s)\rangle - \langle \dot{J}(0), J(0)\rangle = (aw_1(s) + bw_2(s))(a\dot{w}_1(s) + b\dot{w}_2(s)) - a(a\dot{w}_1(0) + b\dot{w}_2(0))$$

and

$$\begin{aligned}\int_s^{l(\gamma)}|\dot J(u)|^2-\langle R(J(u),\dot\gamma(u))\dot\gamma(u),J(u)\rangle du&=\langle \dot J(l(\gamma)),J(l(\gamma))\rangle-\langle \dot J(s),J(s)\rangle\\&=a(a\dot w_1(l(\gamma))+b\dot w_2(l(\gamma)))-(aw_1(s)+bw_2(s))(a\dot w_1(s)+b\dot w_2(s)).\end{aligned}$$

A direct consequence of these formulae and the fact that $w_2(s)=w_1(l(\gamma)-s)$, plus a few elementary manipulations, results in

$$\int_{l(\gamma)-s}^{l(\gamma)}|\dot J(u)|^2-\langle R(J(u),\dot\gamma(u))\dot\gamma(u),J(u)\rangle du=(bw_1(s)+aw_2(s))(b\dot w_1(s)+a\dot w_2(s))-a(b\dot w_1(0)+a\dot w_2(0))).$$

Summarizing, the choices of β and $\tilde\beta$ from (49) in the case of constant curvature become more explicitly:

$$\begin{cases}\beta=\frac12\left((aw_1(\rho_1)+bw_2(\rho_1))(a\dot w_1(\rho_1)+b\dot w_2(\rho_1))-a(a\dot w_1(0)+b\dot w_2(0))\right)\\ \tilde\beta=\frac12\left((bw_1(\rho_2)+aw_2(\rho_2))(b\dot w_1(\rho_2)+a\dot w_2(\rho_2))-a(b\dot w_1(0)+a\dot w_2(0)))\right).\end{cases}\tag{54}$$

It goes without saying that here a and b are evaluated at (τ,x,y,ρ_1,ρ_2).

We say that a function $f(\tau,x,y,\rho_1,\rho_2)$ is symmetric in ρ_1 and ρ_2 if $f(\tau,x,y,\rho_1,\rho_2)=f(\tau,x,y,\rho_2,\rho_1)$.

Before we move on to another property of the diffusion (x,y,ρ_1,ρ_2), we close the discussion so far with the following property of the choices of β and $\tilde\beta$ from (49) :

(55) If a and b are equal and symmetric in ρ_1 and ρ_2, then $\beta(\tau,x,y,\rho_1,\rho_2)=\tilde\beta(\tau,x,y,\rho_2,\rho_1)$.

A symmetry which plays a crucial role in the Hessian estimates is the following.

Theorem 15. *If, in equation (45), we take*

$$\begin{cases}a \text{ and } \alpha \text{ symmetric in } \rho_1 \text{ and } \rho_2\\ b=a\\ \tilde\alpha=-\alpha\\ \tilde\beta(\tau,x,y,\rho_1,\rho_2)=\beta(\tau,x,y,\rho_2,\rho_1)\\ \rho_{1,0}=\rho_{2,0},\end{cases}$$

then the processes (x,y,ρ_1,ρ_2) and (x,y,ρ_2,ρ_1) have the same law. In particular, the processes (x,y,ρ_1) and (x,y,ρ_2) have the same law.

Proof. Although this is almost trivial, we say a word about it. If $\mathcal{L}$ is the generator of a diffusion ω_τ on a manifold $\mathcal{M}$ and $\pi:\mathcal{M}\to\mathcal{M}$ is such that for any smooth function $\varphi:\mathcal{M}\to\mathbb{R}$,

$$\mathcal{L}(\varphi\circ\pi)=(\mathcal{L}\varphi)\circ\pi,$$

then uniqueness of the diffusion implies that ω and $\pi(\omega)$ have the same law. This can be easily seen from the martingale characterization of the law of the diffusion. We apply this to the operator $\mathcal{L}$ from (47) and the map $\pi(x,y,\rho_1,\rho_2)=(x,y,\rho_2,\rho_1)$. The rest follows. □

Notice that (cf. (54)), the choices of β and $\tilde\beta$ from Theorem 14 are actually consistent with the conditions of Theorem 15 under the assumptions that a and b are equal and symmetric.

The "middle particle" process we are interested is

$$z_\tau=\gamma_{x_\tau,y_\tau}(\rho_{1,\tau}).\tag{56}$$

The symmetry between ρ_1 and ρ_2 should be interpreted as saying that the reflection of the process z_τ with respect to the middle point of the geodesic γ_{x_τ,y_τ} has the same law as z_τ itself.

Our next objective is the law of z_τ. Before we jump into the heart of the matter, we take up a discussion on the following class of vector fields that play that are the main actors in our computation.

Assume we have a geodesic γ from x to y with length l and consider a smooth, two-parameter geodesic perturbation $f : (-\epsilon,\epsilon)\times(-\epsilon,\epsilon)\times[0,l]\to M$ of γ, i.e. $f(0,0,s)=\gamma(s)$ and for each fixed choice of u and v, the curve $s\to f(u,v,s)$ is a geodesic. One of the things we want to understand is the field

$$\mathcal{K}(s)=\frac{D}{du}\frac{D}{dv}f(u,v,s)\big|_{u=0,v=0}.$$

Let $J_v(s)=\frac{D}{dv}f(u,v,s)\big|_{v=0}$ be the Jacobi field obtained by differentiating f with respect to v. Similarly let $J_u(s)=\frac{D}{du}f(u,v,s)\big|_{u=0}$ be the Jacobi field obtained by differentiating f with respect to u. In order to determine the equation satisfied by $\mathcal{K}$, we recall here [4, Lemma 4.1], which asserts that for any two-parameter family $g(a,b)$ and vector field V along g,

$$\frac{D}{da}\frac{D}{db}V-\frac{D}{db}\frac{D}{da}V=-R\left(\frac{Dg}{db},\frac{Dg}{da}\right)V. \tag{57}$$

Recall here that Do Carmo [4] uses the curvature as the negative of what other people use, as for instance [5].

Now, what we want to do is find a differential equation satisfied by $\mathcal{K}$. As pointed out already, $\mathcal{K}(s)=\frac{D}{du}J_v(s)$ and starting with (51) for J_v, namely,

$$\frac{D^2}{ds^2}J_v+rJ_v-r\langle\dot\gamma,J_v\rangle\dot\gamma=0$$

we take the derivative with respect to u to arrive at

$$\frac{D}{du}\frac{D^2}{ds^2}J_v+r\mathcal{K}-r\left\langle\frac{D}{du}\dot\gamma,J_v\right\rangle\dot\gamma-k\langle\dot\gamma,\mathcal{K}\rangle\dot\gamma-r\langle\dot\gamma,J_v\rangle\frac{D}{du}\dot\gamma=0.$$

To move forward, use that $\frac{D}{du}\dot\gamma=\frac{D}{ds}\frac{D}{du}\gamma=\dot J_u$ to re-write the previous equation as

$$\frac{D}{du}\frac{D^2}{ds^2}J_v+r\mathcal{K}-r\langle\mathcal{K},\dot\gamma\rangle\dot\gamma-r\langle\dot J_u,J_v\rangle\dot\gamma-r\langle\dot\gamma,J_v\rangle\dot J_u=0.$$

Our task now is to commute the derivatives with respect to u and s. For this, use (57) and (50) to justify

$$\begin{aligned}\frac{D}{du}\frac{D^2}{ds^2}J_v&=\frac{D}{ds}\frac{D}{du}\frac{D}{ds}J_v-R\left(\frac{Df}{ds},\frac{Df}{du}\right)\dot J_v=\frac{D}{ds}\frac{D}{du}\frac{D}{ds}J_v-R\left(\dot\gamma,J_u\right)\dot J_v\\&=\frac{D}{ds}\frac{D}{du}\frac{D}{ds}J_v+r\left(\langle\dot\gamma,\dot J_v\rangle J_u-\langle J_u,\dot J_v\rangle\dot\gamma\right).\end{aligned}\tag{*}$$

Now we once again employ (57),

$$\begin{aligned}\frac{D}{ds}\frac{D}{du}\frac{D}{ds}J_v&=\frac{D^2}{ds^2}\frac{D}{du}J_v-\frac{D}{ds}\left(R\left(\frac{Df}{ds},\frac{Df}{du}\right)J_v\right)=\ddot{\mathcal{K}}-\frac{D}{ds}\left(R\left(\dot\gamma,J_u\right)J_v\right)\\&=\ddot{\mathcal{K}}+r\frac{D}{ds}\left(\langle\dot\gamma,J_v\rangle J_u-\langle J_u,J_v\rangle\dot\gamma\right)\\&=\ddot{\mathcal{K}}+r\left(\langle\dot\gamma,\dot J_v\rangle J_u+\langle\dot\gamma,J_v\rangle\dot J_u-\langle\dot J_u,J_v\rangle\dot\gamma-\langle J_u,\dot J_v\rangle\dot\gamma\right).\end{aligned}\tag{**}$$

Putting together (*) and (**) at $u=0$, we obtain

$$\frac{D}{du}\frac{D^2}{ds^2}J_v = \ddot{\mathcal{K}} + r\left(2\langle\dot\gamma, \dot J_v\rangle J_u + \langle\dot\gamma, J_v\rangle \dot J_u - \langle \dot J_u, J_v\rangle\dot\gamma - 2\langle J_u, \dot J_v\rangle\dot\gamma\right),$$

and finally,

$$\ddot{\mathcal{K}} + r\mathcal{K} - k\langle\mathcal{K},\dot\gamma\rangle\dot\gamma + 2r\left(\langle\dot\gamma,\dot J_v\rangle J_u - \langle\dot J_u, J_v\rangle\dot\gamma - \langle J_u,\dot J_v\rangle\dot\gamma\right) = 0. \tag{58}$$

The boundary conditions imposed here are the natural ones

$$\begin{cases}\mathcal{K}(0) = \frac{D}{du}\frac{D}{dv}f(u,v,0)\\ \mathcal{K}(l) = \frac{D}{du}\frac{D}{dv}f(u,v,l).\end{cases}$$

The typical example of the type of perturbation $f(u,v,s)$ that will appear below is given by the following. Take a geodesic γ defined on $[0,l]$ and consider two geodesic curves, $\eta_1(u)$ starting at $\gamma(0)$ and another, $\eta_2(v)$, started at $\gamma(l)$. Then we take $f(u,v,\cdot)$ to be the geodesic run at unit speed from $\eta_1(u)$ to $\eta_2(v)$.

We discussed the case of a two-parameter perturbation of the geodesic γ in the form $f(u,v,s)$ but exactly the same argument works also for the case where $f(u,s)$ is a perturbation with geodesics of γ, and we consider the field

$$\mathcal{K}(s) = \frac{D^2}{du^2}f(u,s)\big|_{u=0}.$$

The main result from the argument above then gives that

$$\ddot{\mathcal{K}} + r\mathcal{K} - r\langle\mathcal{K},\dot\gamma\rangle\dot\gamma + 2r\left(\langle\dot\gamma,\dot J_u\rangle J_u - 2\langle J_u,\dot J_u\rangle\dot\gamma\right) = 0, \tag{59}$$

with $J_u(s) = \frac{D}{du}f(u,s)|_{u=0}$.

We are finally ready for the next result.

Theorem 16. *Assume that*

(60)

$$\begin{aligned}
\alpha(\tau,x,y,\rho_1,\rho_2) &= a(\tau,x,y,\rho_1,\rho_2)w_1(\rho_1) + b(\tau,x,y,\rho_1,\rho_2)w_2(\rho_1)\\
\theta(\tau,x,y,\rho_1,\rho_2) &= \beta(\tau,x,y,\rho_1,\rho_2) + r\bigg(\int_0^{\rho_1}(a(\tau,x,y,\rho_1,\rho_2)w_1(\sigma) + b(\tau,x,y,\rho_1,\rho_2)w_2(\sigma))^2 d\sigma\\
&\qquad - \frac{\rho_1}{l}\int_0^l (a(\tau,x,y,\rho_1,\rho_2)w_1(\sigma) + b(\tau,x,y,\rho_1,\rho_2)w_2(\sigma))^2 d\sigma\bigg)
\end{aligned}$$

with w_1 and w_2 defined by (53). With these choices, the process $z_\tau = \gamma_{x_\tau,y_\tau}(\rho_{1,\tau})$ has the property that, for any smooth function φ on M,

$$\varphi(z_\tau) - \int_0^\tau\left(\frac{\alpha^2(u)}{2}[\Delta\varphi](z_u) + \theta(u)\langle\nabla\varphi(z_u),\dot\gamma_{x_u,y_u}(\rho_{1,u}))\rangle\right)du \tag{61}$$

is a martingale with respect to the filtration generated by W_1, W_2, and W_3. Inside the integral, $\alpha(u)$ and $\theta(u)$ are shorthand for α and θ evaluated at $(u,x_u,y_u,\rho_{1,u},\rho_{2,u})$ In other words, z_τ is a time-changed Brownian motion (with the time change given by α) with a drift in the geodesic direction from x_τ to y_τ.

Proof. The idea of the proof is to start with the generator of the diffusion (x,y,ρ_1) and a function φ and look at the process $\varphi(z_\tau)$. More precisely, we find the bounded variation part of this. It is clear that, in terms of the generator (47), we need to compute the action of each term of this expression on $\varphi(\gamma_{x,y}(s))$. Notice that the part which involves derivatives of ρ_2 simply drops out in this calculation.

For simplicity we will drop the dependence on τ, x, and y in the notation and let $l = d(x,y)$. Thus the geodesic $\gamma_{x,y}$ will appear as γ. Let E denote the parallel vector field along γ which is obtained by parallel translation of ξ_1.

Now we take the terms one by one. Again for simplicity in writing, we will use s instead of ρ_1 as the parameter in the geodesic direction.

(1) We write the Laplacian term as

$$\Delta_x[\varphi(\gamma_{x,y}(s))] = \frac{d^2}{du^2}\varphi(\gamma_{\eta_{u,1},y}(s)) + \frac{d^2}{du^2}\varphi(\gamma_{\eta_{u,2},y}(s))$$

where $\eta_{u,1}$ and $\eta_{u,2}$ are geodesics starting at x and having derivatives given by $\dot\eta_{0,1} = \dot\gamma_{x,y}(0)$ and $\dot\eta_{0,2} = \xi_1$. Then we continue with

$$\begin{aligned}\Delta_x[\varphi(\gamma_{x,y}(s))] &= \langle \mathrm{Hess}\varphi(\gamma(s))\dot\gamma(s), \dot\gamma(s)\rangle + \langle \mathrm{Hess}\varphi(\gamma(s))J_1(s), J_1(s)\rangle \\ &\quad + \left\langle \nabla\varphi(\gamma(s)), \frac{D^2}{\partial u^2}\gamma_{\eta_{u,1},y}(s)\Big|_{u=0}\right\rangle + \left\langle \nabla\varphi(\gamma(s)), \frac{D^2}{\partial u^2}\gamma_{\eta_{u,2},y}(s)\Big|_{u=0}\right\rangle\end{aligned} \tag{62}$$

where J_1 is the Jacobi field along γ given by $J_1(s) = \frac{D}{du}\gamma_{\eta_{u,2},y}(s)|_{u=0}$, which can also be characterized as the Jacobi field with the boundary conditions $J_1(0) = \xi_1$ and $J_1(l) = 0$. On the other hand, if we look at $\mathcal{K}(s) = \frac{D^2}{\partial u^2}\gamma_{\eta_{u,2},y}(s)\Big|_{u=0}$, using (59) we obtain the equation satisfied by $\mathcal{K}$ as

$$\begin{cases}\ddot{\mathcal{K}} + r\mathcal{K} - r\langle \mathcal{K}, \dot\gamma\rangle\dot\gamma + 2r\left(\langle\dot\gamma, \dot J_1\rangle J_1 - 2\langle J_1, \dot J_1\rangle\dot\gamma\right) = 0\\ \mathcal{K}(0) = 0\\ \mathcal{K}(l) = 0.\end{cases}$$

Notice here that the boundary conditions follow from the fact that $\eta_{u,2}$ is a geodesic and that $\gamma_{\eta_{u,2},y}(l) = y$.

Now, the Jacobi field J_1 is given by

$$J_1(s) = w_1(s)E(s)$$

and this in turn gives the equation of $\mathcal{K}$ as

$$\begin{cases}\ddot{\mathcal{K}} + r\mathcal{K} - r\langle \mathcal{K}, \dot\gamma\rangle\dot\gamma = 4rw_1\dot w_1\dot\gamma\\ \mathcal{K}(0) = 0\\ \mathcal{K}(l) = 0.\end{cases}$$

We solve this as

$$\mathcal{K} = w_{1,0}\dot\gamma \text{ with } w_{1,0}(s) = 2r\int_0^s w_1^2(\sigma)d\sigma - \frac{2sr}{l}\int_0^l w_1^2(\sigma)d\sigma. \tag{63}$$

The conclusion is that

$$\Delta_x[\varphi(\gamma_{x,y}(s))] = \langle \mathrm{Hess}\varphi(\gamma(s))\dot\gamma(s), \dot\gamma(s)\rangle + w_1^2(s)\langle \mathrm{Hess}\varphi(\gamma(s))E(s), E(s)\rangle + w_{1,0}(s)\langle\nabla\varphi(\gamma(s)), \dot\gamma(s)\rangle. \tag{64}$$

(2) In the same vein, with very few changes, we can treat the next term, which is the Laplacian Δ_y applied to $\varphi(\gamma(s))$. To this end, start with the following equation,

$$\begin{aligned}\Delta_y[\varphi(\gamma(s))] &= \langle \mathrm{Hess}\varphi(\gamma(s))J_2(s), J_2(s)\rangle \\ &\quad + \left\langle \nabla\varphi(\gamma(s)), \frac{D^2}{\partial u^2}\gamma_{x,\eta_{u,1}}(s)\Big|_{u=0}\right\rangle + \left\langle \nabla\varphi(\gamma(s)), \frac{D^2}{\partial u^2}\gamma_{x,\eta_{u,2}}(s)\Big|_{u=0}\right\rangle\end{aligned} \tag{65}$$

which is an analogue of (62) with the role of x being played by y, $\eta_{u,1}$ a geodesic starting at y with initial speed given by $\gamma(l)$, and $\eta_{u,2}$ a geodesic starting at y with initial speed $\xi_2 = E(l)$. Here J_2 is the Jacobi field which is 0 at 0 and ξ_2 at l. As in the case of J_1, we can show that

$$J_2(s) = w_2(s)E(s).$$

The first term in the second line of (65) is 0 because $\gamma_{x,\eta_{u,1}}(s)$ does not depend on u. The second term can be dealt with in a similar way to that outlined above in dealing with the derivative of the Jacobi field. We skip the details and give the main result. Let

$$\mathcal{K}(s) = \frac{D^2}{\partial u^2}\gamma_{x,\eta_{u,2}}(s)\Big|_{u=0}.$$

From (59), the equation satisfied by $\mathcal{K}$ (with w_2 given by (53)) is

$$\mathcal{K} = w_{0,1}\dot{\gamma} \text{ with } w_{0,1}(s) = 2r\int_0^s w_2^2(\sigma)d\sigma - \frac{2sr}{l}\int_0^l w_2^2(\sigma)d\sigma. \tag{66}$$

Then we have

(67)

$$\Delta_y[\varphi(\gamma_{x,y}(s))] = \langle \mathrm{Hess}\varphi(\gamma(s))\dot{\gamma}(s), \dot{\gamma}(s)\rangle + w_2^2(s)\langle \mathrm{Hess}\varphi(\gamma(s))E(s), E(s)\rangle + w_{0,1}(s)\langle \nabla\varphi(\gamma(s)), \dot{\gamma}(s)\rangle.$$

(3) For the next term, matters are fairly simple. Namely, because we are differentiating with respect to the geodesic parameter s,

$$\partial_s^2[\varphi(\gamma_{x,y}(s))] = \langle \mathrm{Hess}\varphi(\gamma_{x,y}(s))\dot{\gamma}_{x,y}(s), \dot{\gamma}_{x,y}(s)\rangle. \tag{68}$$

(4) Next in line is

$$\dot{\gamma}_{x,y}(0)\dot{\gamma}_{y,x}(0)[\varphi(\gamma_{x,y}(s))] = 0. \tag{69}$$

This produces 0 because

$$\dot{\gamma}_{y,x}(0)[\varphi(\gamma_{x,y}(s))] = 0,$$

which follows from the fact that perturbing y along a curve $\eta_{u,2}$ in the geodesic direction of $\gamma_{x,y}$ yields that $\gamma_{x,\eta_{u,2}}(s) = \gamma_{x,y}(s)$, and thus is independent of u.

(5) Now we deal with

$$\xi_1\xi_2[\varphi(\gamma_{x,y}(s))].$$

To this end, consider the geodesics $\eta_{u,1}$ and $\eta_{v,2}$ which start at x (respectively y) and have the tangent vectors ξ_1 (respectively ξ_2). What we need to compute is

$$\frac{D}{du}\frac{D}{dv}[\varphi(\gamma_{x,y}(s))]\Big|_{u=v=0} = \langle \mathrm{Hess}\varphi(\gamma_{x,y}(s))J_1(s), J_2(s)\rangle + \left\langle \nabla\varphi(\gamma_{x,y}(s)), \frac{D}{du}\frac{D}{dv}\gamma_{\eta_{u,1},\eta_{v,2}}(s)\Big|_{u=v=0}\right\rangle.$$

If we let

$$\mathcal{K}(s) = \frac{D}{du}\frac{D}{dv}\gamma_{\eta_{u,1},\eta_{v,2}}(s)\Big|_{u=v=0},$$

from (58), we obtain

$$\begin{cases} \ddot{\mathcal{K}} + r\mathcal{K} - r\langle \mathcal{K}, \dot{\gamma}\rangle\dot{\gamma} = 2r(w_1\dot{w}_2 + w_2\dot{w}_1)\dot{\gamma} \\ \mathcal{K}(0) = 0 \\ \mathcal{K}(l) = 0, \end{cases}$$

which we solve as

$$\mathcal{K} = w_{1,1}\dot{\gamma} \text{ with } w_{1,1}(s) = 2r\int_0^s w_1(\sigma)w_2(\sigma)\sigma - \frac{2sr}{l}\int_0^l w_1(\sigma)w_2(\sigma)d\sigma. \tag{70}$$

We conclude that

$$\xi_1\xi_2[\varphi(\gamma_{x,y}(s))] = w_1(s)w_2(s)\langle \mathrm{Hess}\varphi(\gamma(s))E(s), E(s)\rangle + w_{1,1}(s)\langle \nabla\varphi(\gamma(s)), \dot\gamma(s)\rangle. \tag{71}$$

(6) Next is the following

$$\begin{aligned}\dot\gamma_{x,y}(0)\partial_s[\varphi(\gamma_{x,y}(s))] &= \dot\gamma(0)\langle \nabla\varphi(\gamma(s)), \dot\gamma(s)\rangle \\ &= \langle \mathrm{Hess}\varphi(\gamma(s))\dot\gamma(s), \dot\gamma(s)\rangle + \langle \nabla\varphi(\gamma(s)), \ddot\gamma(s)\rangle.\end{aligned}$$

Therefore, since γ is geodesic,

$$\dot\gamma(0)\partial_s[\varphi(\gamma_{x,y}(s))] = \langle \mathrm{Hess}\varphi(\gamma(s))\dot\gamma(s), \dot\gamma(s)\rangle. \tag{72}$$

(7) Now,

$$\dot\gamma_{y,x}(0)\partial_s[\varphi(\gamma_{x,y}(s))] = 0, \tag{73}$$

as can be easily seen from the fact that perturbing y in the geodesic direction (say, along η_v) reveals that $\gamma_{x,\eta_v}(s) = \gamma_{x,y}(s)$, so that the derivative with respect to v vanishes.

(8) The last term is easy to deal with and gives

$$\partial_s[\varphi(\gamma_{x,y}(s))] = \langle \nabla\varphi(\gamma(s)), \dot\gamma(s)\rangle. \tag{74}$$

Putting together all the results from (64)-(74) we arrive at (we drop the subscripts x and y)

$$\begin{aligned}\mathcal{L}[\varphi(\gamma(s))] = {}& \frac{\alpha^2}{2}\langle \mathrm{Hess}\varphi(\gamma(s))\dot\gamma(s), \dot\gamma(s)\rangle + \frac{(aw_1(s)+bw_2(s))^2}{2}\langle \mathrm{Hess}\varphi(\gamma(s))E(s), E(s)\rangle \\ &+ \left(\beta + \frac{a^2w_{1,0}+b^2w_{0,1}+2abw_{1,1}}{2}\right)\langle \nabla\varphi(\gamma(s)), \dot\gamma(s)\rangle.\end{aligned} \tag{75}$$

A little simplification follows from

$$a^2w_{1,0}+b^2w_{0,1}+2abw_{1,1} = 2r\left(\int_0^s (a(s)w_1(\tau)+b(s)w_2(\tau))^2 d\tau - \frac{s}{l}\int_0^l (a(s)w_1(\tau)+b(s)w_2(\tau))^2 d\tau\right)$$

which then gives the result of the theorem, for the choice of α as in (60). □

We close this section with the following result summarizing all of the important findings for the next section.

Corollary 17. *Assume that the entries of* (45) *satisfy*

$$\begin{cases} a \text{ is symmetric in } \rho_1 \text{ and } \rho_2 \\ b = a \\ \tilde\alpha = -\alpha \\ \alpha(\tau,x,y,\rho_1,\rho_2) = a(\tau,x,y,\rho_1,\rho_2)w(\rho_1) \\ \beta(\tau,x,y,\rho_1,\rho_2) = \frac{1}{2}a^2(\tau,x,y,\rho_1,\rho_2)(w(\rho_1)\dot w(\rho_1) - \dot w(0)) \\ \tilde\beta(\tau,x,y,\rho_1,\rho_2) = \frac{1}{2}a^2(\tau,x,y,\rho_1,\rho_2)(w(\rho_2)\dot w(\rho_2) - \dot w(0)) \\ \rho_{1,0} = \rho_{2,0} = \rho_0/2 \end{cases} \quad \text{with} \quad \begin{cases} \ddot w + rw = 0 \\ w(0) = 1 \\ w(d(x,y)) = 1. \end{cases} \tag{76}$$

Then

(1) $\rho_{1,\tau} + \rho_{2,\tau} = \rho_\tau$ *almost surely.*

(2) *The diffusions* $(x_\tau, y_\tau, \rho_{1,\tau}, \rho_{2,\tau})$ *and* $(x_\tau, y_\tau, \rho_{2,\tau}, \rho_{1,\tau})$ *have the same law. In particular,* $(x_\tau, y_\tau, \rho_{1,\tau})$ *and* $(x_\tau, y_\tau, \rho_{2,\tau})$ *have the same law.*

(3) *If $z_\tau = \gamma_{x_\tau,y_\tau}(\rho_{1,\tau})$, then for any smooth function φ on M,*

$$\varphi(z_\tau) - \int_0^\tau \left(\frac{\alpha^2(u)}{2}[\Delta\varphi](z_u) + \theta(u)\langle\nabla\varphi(z_u), \dot\gamma_{x_u,y_u}(\rho_{1,u}))\rangle\right) du \tag{77}$$

is a martingale with respect to the filtration generated by W_1, W_2, and W_3, where

$$\theta(\tau,x,y,\rho_1,\rho_2) = \beta(\tau,x,y,\rho_1,\rho_2) + ra^2(\tau,x,y,\rho_1,\rho_2)\left(\int_0^{\rho_1} w^2(\sigma)d\sigma - \frac{\rho_1}{d(x,y)}\int_0^{d(x,y)} w^2(\sigma)d\sigma\right).$$

A word is in place here. The statement of Theorem 15 requires the symmetry of α with respect to ρ_1 and ρ_2. This is not satisfied by the choice in (76) for arbitrary ρ_1 and ρ_2. However, because of the choice of β and $\tilde\beta$ and Theorem 14, we know that (almost surely) $\rho_{1,\tau} + \rho_{2,\tau} = \rho_\tau$. So it suffices to ensure the symmetry of α and $\tilde\alpha$ with respect to ρ_1 and ρ_2 only in the case that $\rho_1+\rho_2 = \rho = d(x,y)$, which follows from the fact that $w(s) = w(d(x,y)-s)$ for $s \in [0, d(x,y)]$.

For a given l, the solution w to (76) is

$$w(s) = \frac{\cos\left((l-2s)\sqrt{r}/2\right)}{\cos(l\sqrt{r}/2)}. \tag{78}$$

In particular, for $r = 0$ or -1 and l bounded, $w(s)$ is bounded and so are all its derivatives.

10. Estimates on the Hessian decay for $\chi(M) \le 0$

For Euler characteristic less than or equal to 0, we know that $\bar p_t$ and $\nabla\bar p_t$ decay exponentially fast. Our goal is now to extend this to the Hessian of $\bar p_t$, resulting in the converge of the metric to the constant curvature metric in C^2. In particular, the curvature converges to a constant.

To estimate the Hessian decay, we proceed in a similar way to the estimation of the gradient, only that now we need to use a coupling procedure for three points rather than two.

Let us denote, for $t > 0$,

$$H_t = \sup_{x\in M} |\mathrm{Hess}\,\bar p_t(x)|.$$

What we want to show is that H_t decays to 0 exponentially fast.

Theorem 18. *For the case $\chi(M) \le 0$, H_t converges to 0 exponentially fast as $t \to \infty$.*

Proof. To begin with, notice that

$$\langle \mathrm{Hess}\,\bar p_t(z)\xi, \xi\rangle = \lim_{\rho_0\to 0} \frac{\bar p_t(\gamma(-\rho_0)) - 2\bar p_t(z) + \bar p_t(\gamma(\rho_0))}{\rho_0^2} \tag{79}$$

where γ is the unique geodesic passing through z and having the initial velocity given by ξ. Thus, similarly to the case of the gradient estimate, we will use the three particle coupling to get a handle on the right-hand side of the above quantity, for sufficiently small ρ_0.

For convenience, fix a time $t > 0$ and let $s \in [0, 1\wedge t]$. Pick two points $x, y \in M$, with $d(x,y) = \rho_0$ small enough, and let z be the middle point on the geodesic between x and y such that $d(x,z) = d(z,y) = \rho_0/2$. Consider the triple coupling described by (45) with the choices from Corollary 17. All the data there is completely described by the choice of the time change a of the processes x_τ and y_τ. In this section we choose

$$a(\tau,x,y,\rho_1,\rho_2) = e^{-\bar p_{t-\tau}(\lambda_{x,y})} \tag{80}$$

where $\lambda_{x,y}$ is the middle point on the geodesic between x and y. This choice does not depend on ρ_1 or ρ_2, and consequently it is symmetric in ρ_1 and ρ_2, as required by Corollary 17. Other choices are possible for the argument here, but we stick with this because it is symmetric with respect to x and y and makes some of the estimates look more natural.

Now, we consider $\bar{p}_{t-\sigma}(z_\sigma)$, where z_τ is defined in the previous section. Again invoking Corollary 17, we learn that

$$d\bar{p}_{t-\tau}(z_\tau) = M_{1,\tau} + \left(-\partial_t \bar{p}_{t-\tau}(z_\tau) + \alpha^2(\tau)\Delta\bar{p}_{t-\tau}(z_\tau) + \theta(\tau)\langle\nabla\bar{p}_{t-\tau}(\tau), \dot{\gamma}_\tau\rangle\right) d\tau.$$

where $M_{1,\tau}$ is a martingale. From the Ricci flow equation, $\partial_t\bar{p}_{t-\tau}(z_\tau) = e^{-2\bar{p}_{t-\tau}(z_\tau)}\Delta\bar{p}_{t-\tau}(z_\tau) + r(1-e^{-2\bar{p}_{t-\tau}(z_\tau)})$ so we continue with

(81)
$$d\bar{p}_{t-\tau}(z_\tau) = M_{1,\tau} + \left(\left(\alpha^2(\tau) - e^{-2\bar{p}_{t-\tau}(z_\tau)}\right)\Delta\bar{p}_{t-\tau}(z_\tau) + \theta(\tau)\langle\nabla\bar{p}_{t-\tau}(\tau), \dot{\gamma}_\tau\rangle - r(1-e^{-2\bar{p}_{t-\tau}(z_\tau)})\right) d\tau.$$

For the semimartingale $\bar{p}_{t-\tau}(x_\tau)$ we have from (45) and the Ricci flow equation that

$$d\bar{p}_{t-\tau}(x_\tau) = M_{2,\tau} + \left((e^{-2\bar{p}_{t-\tau}(\lambda_\tau)} - e^{-2\bar{p}_{t-\tau}(x_\tau)})\Delta\bar{p}_{t-\tau}(x_\tau) - r(1-e^{-2\bar{p}_{t-\tau}(x_\tau)}\right) d\tau \tag{82}$$

where λ_τ is the middle point of the geodesic joining x_τ and y_τ. Similarly for $\bar{p}_{t-\tau}(y_\tau)$,

$$d\bar{p}_{t-\tau}(y_\tau) = M_{3,\tau} + \left((e^{-2\bar{p}_{t-\tau}(\lambda_\tau)} - e^{-2\bar{p}_{t-\tau}(y_\tau)})\Delta\bar{p}_{t-\tau}(y_\tau) - r(1-e^{-2\bar{p}_{t-\tau}(y_\tau)}\right) d\tau. \tag{83}$$

Now, putting these together,

(84)
$$\begin{aligned}
\bar{p}_{t-\tau}(x_\tau) - 2\bar{p}_{t-\tau}(z_\tau) + \bar{p}_{t-\tau}(y_\tau) &= \bar{p}_t(x) - 2\bar{p}_t(z) + \bar{p}_t(y) + M_\tau \\
&+ \int_0^\tau \left(\left(\alpha^2(u) - e^{-2\bar{p}_{t-u}(z_u)}\right)\Delta\bar{p}_{t-u}(z_u)\right) du \\
&+ \int_0^\tau \theta(u)\langle\nabla\bar{p}_{t-u}(u), \dot{\gamma}_u\rangle du \\
&+ \int_0^\tau \left((e^{-2\bar{p}_{t-u}(\lambda_u)} - e^{-2\bar{p}_{t-u}(x_u)})\Delta\bar{p}_{t-u}(x_u) + (e^{-2\bar{p}_{t-u}(\lambda_u)} - e^{-2\bar{p}_{t-u}(y_u)})\Delta\bar{p}_{t-u}(y_u)\right) du \\
&+ r\int_0^\tau (e^{-2\bar{p}_{t-u}(x_u)} - 2e^{-2\bar{p}_{t-u}(z_u)} + e^{-2\bar{p}_{t-u}(y_u)}) du
\end{aligned}$$

where M_τ is a martingale.

From the definition of α in Corollary 17 and the fact that we stop the processes before the distance between x and y hits some small number r_0, it is not hard to prove that there is a constant $C > 0$ such that

$$|\alpha(u) - a(u)| \le C\rho_1 d(x,y),$$

which in turn, using the gradient decay estimates and the fact that $d(z_u, \lambda_u) \le d(x_u, y_u)/2 = \rho_u/2$, leads to

$$\left|\alpha^2(u) - e^{-2\bar{p}_{t-u}(z_u)}\right| \le C\rho_{1,u} + |e^{-2\bar{p}_{t-\tau}(z_u)} - e^{-2\bar{p}_{t-u}(\lambda_u)}| \le C\rho_u^2 + Ce^{-Ct}\rho_u \le C\rho_u.$$

Observe here that we do not need the full power of the exponential decay of the gradient. Just the boundedness suffices for this particular estimate, but used in conjunction with the definition of θ from Corollary 17 it justifies

$$|\theta(u)\langle\nabla\bar{p}_{t-u}(u), \dot{\gamma}_u\rangle| \le c\rho_{1,u}e^{-Ct}.$$

Finally, from the exponential decay of the gradient and elementary arguments,

$$\left|e^{-2\bar{p}_{t-u}(x_u)} - e^{-2\bar{p}_{t-u}(z_u)}\right| + \left|e^{-2\bar{p}_{t-u}(y_u)} - e^{-2\bar{p}_{t-u}(z_u)}\right| \le c\rho_u e^{-Ct}$$

and also

$$\left|e^{-2\bar p_{t-u}(x_u)}-e^{-2\bar p_{t-u}(\lambda_u)}\right|+\left|e^{-2\bar p_{t-u}(y_u)}-e^{-2\bar p_{t-u}(\lambda_u)}\right|\le c\rho_u e^{-Ct}.$$

Now, let σ be the first time u when $\rho_{1,u}$ or $\rho_{2,u}$ becomes 0, and let ζ be the first time u when ρ_u hits r_0, a small number (less than half of the injectivity radius). Replacing τ by $\tau\wedge\sigma\wedge\zeta$ in (84) and then taking the expectation at $\tau=0$ and $\tau=s$, combined with the above estimates, leads to

$$\begin{aligned}|\bar p_t(x)-2\bar p_t(z)+\bar p_t(y)|\le&|\mathbb{E}[\bar p_{t-s\wedge\sigma\wedge\zeta}(x_{s\wedge\sigma\wedge\zeta})-2\bar p_{t-s\wedge\sigma\wedge\zeta}(z_{s\wedge\sigma\wedge\zeta})+\bar p_{t-s\wedge\sigma\wedge\zeta}(y_{s\wedge\sigma\wedge\zeta})]|\\&+ce^{-Ct}\mathbb{E}\left[\int_0^{s\wedge\sigma\wedge\zeta}\rho_u du\right]+c\mathbb{E}\left[\int_0^{s\wedge\sigma\wedge\zeta}\rho_u H_{t-u}du\right]\end{aligned}\tag{85}$$

Next, the stopping time σ is $T_1\wedge T_2$, where T_1 and T_2 are, respectively, the first time ρ_1 hits 0 and the first time ρ_2 hits 0. Now we can write

(86)

$$\begin{aligned}&\mathbb{E}[\bar p_{t-s\wedge\sigma\wedge\zeta}(x_{s\wedge\sigma\wedge\zeta})-2\bar p_{t-s\wedge\sigma\wedge\zeta}(z_{s\wedge\sigma\wedge\zeta})+\bar p_{t-s\wedge\sigma\wedge\zeta}(y_{s\wedge\sigma\wedge\zeta})]\\=&\mathbb{E}[\bar p_{t-s\wedge\zeta}(x_{s\wedge\zeta})-2\bar p_{t-s\wedge\zeta}(z_{s\wedge\zeta})+\bar p_{t-s\wedge\zeta}(y_{s\wedge\zeta}),\zeta<\sigma]\\&+\mathbb{E}[\bar p_{t-s\wedge\sigma}(x_{s\wedge\sigma})-2\bar p_{t-s\wedge\sigma}(z_{s\wedge\sigma})+\bar p_{t-s\wedge\sigma}(y_{s\wedge\sigma}),\sigma\le\zeta]\\=&\mathbb{E}[\bar p_{t-s\wedge\zeta}(x_{s\wedge\zeta})-2\bar p_{t-s\wedge\zeta}(z_{s\wedge\zeta})+\bar p_{t-s\wedge\zeta}(y_{s\wedge\zeta}),\zeta<\sigma]\\&+\mathbb{E}[\bar p_{t-T_1}(y_{T_1})-\bar p_{t-T_1}(x_{T_1}),T_1<T_2\le s\wedge\zeta]+\mathbb{E}[\bar p_{t-T_2}(x_{T_2})-\bar p_{t-T_2}(y_{T_2}),T_2<T_1\le s\wedge\zeta]\\&+\mathbb{E}[\bar p_{t-T_1}(y_{T_1})-\bar p_{t-T_1}(x_{T_1}),T_1<s\le T_2\wedge\zeta]+\mathbb{E}[\bar p_{t-T_2}(x_{T_2})-\bar p_{t-T_2}(y_{T_2}),T_2<s\le T_1\wedge\zeta]\\&+\mathbb{E}[\bar p_{t-s}(x_s)-2\bar p_{t-s}(z_s)+\bar p_{t-s}(y_s),s\le\sigma\le\zeta].\end{aligned}$$

The main point of this expresson is that, due to the symmetry with respect to ρ_1 and ρ_2 from Corollary 17, we have the crucial cancellations

$$\mathbb{E}[\bar p_{t-T_1}(y_{T_1})-\bar p_{t-T_1}(x_{T_1}),T_1<T_2\le s\wedge\zeta]+\mathbb{E}[\bar p_{t-T_2}(x_{T_2})-\bar p_{t-T_2}(y_{T_2}),T_2<T_1\le s\wedge\zeta]=0\tag{87}$$

and also

$$\mathbb{E}[\bar p_{t-T_1}(y_{T_1})-\bar p_{t-T_1}(x_{T_1}),T_1<s\le T_2\wedge\zeta]+\mathbb{E}[\bar p_{t-T_2}(x_{T_2})-\bar p_{t-T_2}(y_{T_2}),T_2<s\le T_1\wedge\zeta]=0.\tag{88}$$

Furthermore, from the exponential decay of $\bar p$ and $\nabla\bar p$, for any $s\in[0,t]$ we have

$$\begin{aligned}&|\mathbb{E}[\bar p_{t-s\wedge\sigma\wedge\zeta}(x_{s\wedge\sigma\wedge\zeta})-2\bar p_{t-s\wedge\sigma\wedge\zeta}(z_{s\wedge\sigma\wedge\zeta})+\bar p_{t-s\wedge\sigma\wedge\zeta}(y_{s\wedge\sigma\wedge\zeta})]|\\\le&|\mathbb{E}[\bar p_{t-s\wedge\zeta}(x_{s\wedge\zeta})-2\bar p_{t-s\wedge\zeta}(z_{s\wedge\zeta})+\bar p_{t-s\wedge\zeta}(y_{s\wedge\zeta}),\zeta<\sigma]|+|\mathbb{E}[\bar p_{t-s}(x_s)-2\bar p_{t-s}(z_s)+\bar p_{t-s}(y_s),s\le\sigma\le\zeta]|\\\le&ce^{-Ct}\mathbb{P}(\zeta\le s\wedge\sigma)+ce^{-Ct}\mathbb{E}[\rho_s,s\le\sigma\wedge\zeta].\end{aligned}$$

where we used the following inequalities

$$\begin{aligned}&|\mathbb{E}[\bar p_{t-s\wedge\zeta}(x_{s\wedge\zeta})-2\bar p_{t-s\wedge\zeta}(z_{s\wedge\zeta})+\bar p_{t-s\wedge\zeta}(y_{s\wedge\zeta}),\zeta<\sigma]|\\&\qquad\le|\mathbb{E}[\bar p_{t-s}(x_s)-2\bar p_{t-s}(z_s)+\bar p_{t-s}(y_s),s<\zeta<\sigma]|+|\mathbb{E}[\bar p_{t-\zeta}(x_\zeta)-2\bar p_{t-\zeta}(z_\zeta)+\bar p_{t-\zeta}(y_\zeta),\zeta\le s\wedge\sigma]|\\&\qquad\le ce^{-Ct}\mathbb{E}[\rho_s,s<\sigma\wedge\zeta]+ce^{-Ct}\mathbb{P}(\zeta\le s\wedge\sigma).\end{aligned}$$

Putting these together into (85), plus a little simplification, gives that for any $s\in[0,t]$

$$\begin{aligned}|\bar p_t(x)-2\bar p_t(z)+\bar p_t(y)|\le&ce^{-Ct}\mathbb{P}(\zeta\le s\wedge\sigma)+ce^{-Ct}\mathbb{E}[\rho_s,s\le\sigma\wedge\zeta]\\&+ce^{-Ct}\int_0^s\mathbb{E}[\rho_u,u\le\sigma\wedge\zeta]du+c\int_0^s H_{t-u}\mathbb{E}[\rho_u,u\le\sigma\wedge\zeta]du.\end{aligned}$$

A further simplification is due to the symmetry with respect to ρ_1 and ρ_2 from Corollary 17, which has the effect that

$$\mathbb{E}[\rho_u,u<\sigma\wedge\zeta]=2\mathbb{E}[\rho_{1,u},u<\sigma\wedge\zeta],$$

and thus for $s \in [0, t]$,

$$
\begin{aligned}
|\bar{p}_t(x) - 2\bar{p}_t(z) + \bar{p}_t(y)| \leq & Ce^{-Ct}\mathbb{P}(\zeta < s \wedge \sigma) + Ce^{-Ct}\mathbb{E}[\rho_{1,s}, s < \sigma \wedge \zeta] \\
& + e^{-Ct}\int_0^s \mathbb{E}[\rho_{1,u}, u < \sigma \wedge \zeta]du + C\int_0^s H_{t-u}\mathbb{E}[\rho_{1,u}, u < \sigma \wedge \zeta]du.
\end{aligned} \tag{89}
$$

The key step forward is the following result.

Theorem 19. *Let W^1, W^2, and W^3 be three independent, one-dimensional Brownian motions, and let $\tilde{\rho}_1$ and $\tilde{\rho}_2$ be two processes such that $\tilde{\rho}_{1,0} = \tilde{\rho}_{2,0} = \tilde{\rho}_0 > 0$ and*

$$
\begin{cases}
d\tilde{\rho}_{1,\tau} = (1 + O(\tilde{\rho}_{1,\tau}))(A_\tau dW^1_\tau + B_\tau dW^2_\tau) + (1 + O(\tilde{\rho}_{1,\tau}))dW^3_\tau + O(1)d\tau \\
d\tilde{\rho}_{2,\tau} = (1 + O(\tilde{\rho}_{2,\tau}))(A_\tau dW^1_\tau + B_\tau dW^2_\tau) - (1 + O(\tilde{\rho}_{2,\tau}))dW^3_\tau + O(1)d\tau
\end{cases} \tag{90}
$$

with $A_\tau^2 + B_\tau^2 = 1$.

Let $\tilde{\sigma}$ be the first hitting time of 0 for the process $\tilde{\rho}_1\tilde{\rho}_2$ and $\tilde{\zeta}$ the first time either $\tilde{\rho}_1$ or $\tilde{\rho}_2$ hits some value $\tilde{r}_0$. Assume that (90) is valid for $\tau \in [0, \tilde{\sigma} \wedge \tilde{\zeta}]$, and in addition that for some constant $C > 0$

$$
\mathbb{E}[\tilde{\rho}_{2,s}, s < \tilde{\sigma} \wedge \tilde{\zeta}] \leq C\mathbb{E}[\tilde{\rho}_{1,s}, s < \tilde{\sigma} \wedge \tilde{\zeta}] \text{ for all } s \in [0, 1 \wedge t]. \tag{91}
$$

Then, there is a constant $C > 0$ such that, for all $s \in [0, 1 \wedge t]$ and sufficiently small $\tilde{\rho}_0 > 0$,

$$
\mathbb{E}[\tilde{\rho}_{1,s}, s < \tilde{\sigma} \wedge \tilde{\zeta}] \leq C\tilde{\rho}_0^2/\sqrt{s} \tag{92}
$$

and

$$
\mathbb{P}(\tilde{\zeta} < s \wedge \tilde{\sigma}) \leq C\tilde{\rho}_0^2. \tag{93}
$$

Proof. If we regard the process $(\tilde{\rho}_{1,\tau}, \tilde{\rho}_{2,\tau})$ as a process in the first quadrant, the equations in (90) give the property that near the axes the process behaves as a Brownian motion.

To give a bit more insight, what we want to do is to compare $\mathbb{E}[\tilde{\rho}_{1,s}, s < \tilde{\sigma}]$ with the analogous quantity in which $\tilde{\rho}_1$ and $\tilde{\rho}_2$ run as independent Brownian motions.

In the simplest case, $(\tilde{\rho}_1, \tilde{\rho}_2)$ is $\sqrt{2}$ times a planar Brownian motion started at $(\tilde{\rho}_0, \tilde{\rho}_0)$ and $\mathbb{E}[f(\tilde{\rho}_{1,s}, \tilde{\rho}_{2,s}), s < \tilde{\sigma}]$ is simply $\varphi(s, \tilde{\rho}_0, \tilde{\rho}_0)$, with φ being the solution to the following PDE on the upper-right quadrant $\Omega = \{(x, y), x, y > 0\}$,

$$
\begin{cases}
\partial_t \varphi = \Delta\varphi \\
\varphi(t, (x, y)) = 0, \quad (x, y) \in \partial\Omega \\
\varphi(0, (x, y)) = f(x, y), \quad (x, y) \in \Omega.
\end{cases} \tag{94}
$$

This solution can be written in terms of the heat kernel, which we discuss now. On the half line, the heat kernel for the Laplacian with the Dirichlet boundary condition is given by

$$
h_t(x, y) = \frac{1}{\sqrt{\pi t}}\left(e^{-\frac{(x-y)^2}{t}} - e^{-\frac{(x+y)^2}{t}}\right)
$$

for all $x, y, t > 0$. On Ω, the heat kernel with the Dirichlet boundary condition is simply

$$
\mathbf{h}_t((x_1, x_2), (y_1, y_2)) = h_t(x_1, y_1)h_t(y_1, y_2).
$$

Turning back to the PDE (94), the solution is given by

$$
\varphi(t, x, y) = \int_0^\infty \int_0^\infty \mathbf{h}_t((x, y), (x_1, y_1))f(x_1, y_1)dx_1\, dy_1.
$$

For the case we are most interested in, namely $f(x, y) = x$, the solution above can be computed as

$$\varphi(s, x, y) = x\,\Phi\left(\frac{y}{\sqrt{s}}\right) \text{ with } \Phi(x) = \frac{2}{\sqrt{\pi}}\int_0^y e^{-u^2}du.$$

Now we go back to the system (90) and take $\varphi(s-\tau, \tilde\rho_{1,\tau}, \tilde\rho_{2,\tau})$ as a semimartingale which, from Itô's formula and $\partial_t\varphi = \frac{1}{2}\Delta\varphi$, becomes

$$\begin{aligned} d\varphi(s-\tau, \tilde\rho_{1,\tau}, \tilde\rho_{2,\tau}) =& \partial_x\varphi\, d\tilde\rho_{1,\tau} + \partial_y\varphi\, d\tilde\rho_{2,\tau} - \partial_t\varphi d\tau + \frac{1}{2}\partial^2_{xx}\varphi\, d\langle\tilde\rho_1\rangle_\tau + \partial^2_{xy}\varphi\, d\langle\tilde\rho_1, \tilde\rho_2\rangle_\tau + \frac{1}{2}\partial^2_{yy}\varphi\, d\langle\tilde\rho_2\rangle_\tau \\ =& M_\tau + O(1)\left(\Phi\left(\frac{\tilde\rho_{2,\tau}}{\sqrt{s-\tau}}\right) + \frac{\tilde\rho_{1,\tau}}{\sqrt{s-\tau}}\Phi'\left(\frac{\tilde\rho_{2,\tau}}{\sqrt{s-\tau}}\right)\right)d\tau \\ &+ \frac{\tilde\rho_{1,\tau}O(\tilde\rho_{2,\tau})\Phi''(\frac{\tilde\rho_{2,\tau}}{\sqrt{s-\tau}})}{s-\tau}d\tau + \frac{O(\tilde\rho_{1,\tau}+\tilde\rho_{2,\tau})}{\sqrt{s-\tau}}\Phi'\left(\frac{\tilde\rho_{2,\tau}}{\sqrt{s-\tau}}\right)d\tau, \end{aligned}$$

where M_τ is a martingale. Since Φ' and $y\Phi''(y)$ are bounded, we deduce that the drift in the above is bounded in absolute value by $\frac{C(\tilde\rho_{1,\tau}+\tilde\rho_{2,\tau})}{\sqrt{s-\tau}}$. Now replacing τ by $\tau\wedge\tilde\sigma\wedge\tilde\zeta$ and evaluating at $\tau = 0$ and $\tau = s$, we are led to

$$\begin{aligned} \mathbb{E}[\tilde\rho_{1,s}, s < \tilde\sigma\wedge\tilde\zeta] \le \mathbb{E}[\varphi(s - s\wedge\tilde\sigma\wedge\tilde\zeta, \tilde\rho_{1,s\wedge\tilde\sigma\wedge\tilde\zeta}, \tilde\rho_{2,s\wedge\tilde\sigma\wedge\tilde\zeta})] &\le \varphi(s, \tilde\rho_0, \tilde\rho_0) + C\mathbb{E}\left[\int_0^{s\wedge\tilde\sigma\wedge\tilde\zeta}\frac{\tilde\rho_{1,\tau}+\tilde\rho_{2,\tau}}{\sqrt{s-\tau}}d\tau\right] \\ &\le C\tilde\rho_0\Phi\left(\frac{\tilde\rho_0}{\sqrt{s}}\right) + C\int_0^s\frac{\mathbb{E}[\tilde\rho_{1,\tau}+\tilde\rho_{2,\tau}, \tau < \tilde\sigma\wedge\tilde\zeta]}{\sqrt{s-\tau}}d\tau. \end{aligned}$$

Denote for simplicity $f(s) = \mathbb{E}[\tilde\rho_{1,s}, s < \tilde\sigma\wedge\tilde\zeta]$ and $g(s) = C\tilde\rho_0\Phi\left(\frac{\tilde\rho_0}{\sqrt{s}}\right)$. Now condition (91) implies

$$f(s) \le g(s) + C\int_0^s\frac{f(\tau)}{\sqrt{s-\tau}}d\tau. \tag{95}$$

This functional inequality is interesting enough to be treated separately, and so we do this formally in the following.

Lemma 20. *Assume $f, g : [0, t] \to [0, \infty)$ are bounded, continuous functions such that for all $s \in [0, 1\wedge t]$*

$$f(s) \le g(s) + C\int_0^s\frac{f(\tau)}{\sqrt{s-\tau}}d\tau. \tag{96}$$

If $g(s) \le C\rho^2/\sqrt{s}$ for all $s \in [0, 1\wedge t]$, then

$$f(s) \le C\rho^2/\sqrt{s} \text{ for all } s \in (0, 1\wedge t].$$

Proof. Rewrite (96) in the form

$$f(s) \le g(s) + C\sqrt{s}\int_0^s\frac{f(\tau)}{\sqrt{s-\tau}}d\tau = g(s) + C\sqrt{s}\int_0^1\frac{f(sw)}{\sqrt{1-w}}dw.$$

Now introduce the random variable W with density $\frac{1}{2\sqrt{1-w}}$ and observe that the right hand side of the above equation becomes $g(s) + C\sqrt{s}\mathbb{E}[f(sW)]$. Hence, the inequality at hand can be re-written as

$$f(s) \le g(s) + C\sqrt{s}\mathbb{E}[f(sW)].$$

Iterating this inequality, one can prove that if we pick an iid sequence $W_1, W_2, \dots$ with the same distribution as W, then for any $n \geq 1$,

$$f(s) \leq \sum_{k=0}^{n} (C\sqrt{s})^k \mathbb{E}[\sqrt{W_1}\sqrt{W_1 W_2} \dots \sqrt{W_1 W_2 \dots W_{k-1}} g(s W_1 W_2 \dots W_k)] + (C\sqrt{s})^n \mathbb{E}[\sqrt{W_1}\sqrt{W_1 W_2} \dots \sqrt{W_1 W_2 \dots W_n} f(s W_1 W_2 \dots W_{n+1})].$$

The random variable W has moments

$$\mathbb{E}[W^k] = \frac{\sqrt{\pi}\Gamma(k+1)}{\Gamma(k+3/2)} \text{ for all } k > -1.$$

Particularly important is the case of $k = -1/2$, so that $\frac{1}{\sqrt{W}}$ is integrable, and in fact $\mathbb{E}[1/\sqrt{W}] = \pi/2$. It is an elementary task to obtain from this that, for some constant $C > 0$,

$$\mathbb{E}[W^k] \leq C/\sqrt{k} \text{ for all } k > 0.$$

Since g is bounded, the series

$$\sum_{k=0}^{\infty} (C\sqrt{s})^k \mathbb{E}[\sqrt{W_1}\sqrt{W_1 W_2} \dots \sqrt{W_1 W_2 \dots W_{k-1}} g(s W_1 W_2 \dots W_k)]$$

is absolutely convergent and $(C\sqrt{s})^n \mathbb{E}[\sqrt{W_1}\sqrt{W_1 W_2} \dots \sqrt{W_1 W_2 \dots W_n} f(s W_1 W_2 \dots W_{n+1})]$ goes to 0 as $n \to \infty$. Consequently,

$$f(s) \leq \sum_{k=0}^{\infty} (C\sqrt{s})^k \mathbb{E}[\sqrt{W_1}\sqrt{W_1 W_2} \dots \sqrt{W_1 W_2 \dots W_{k-1}} g(s W_1 W_2 \dots W_k)].$$

If $g(s) \leq C\rho^2/\sqrt{s}$, the above yields

$$f(s) \leq C \frac{\rho^2}{\sqrt{s}} \sum_{k=0}^{\infty} (C\sqrt{s})^k \mathbb{E}\left[\frac{\sqrt{W_1}\sqrt{W_1 W_2} \dots \sqrt{W_1 W_2 \dots W_{k-1}}}{\sqrt{W_1 W_2 \dots W_k}}\right] = \frac{C\rho^2}{\sqrt{s}},$$

where we used the decay of the moments of W together with the fact that $1/\sqrt{W}$ is integrable to justify that the series is convergent. □

The rest of the proof of (92) follows now from Lemma 20.

We now turn our attention to (93) and observe that, from (90), we easily deduce that

$$\begin{aligned} d\tilde{\rho}_1\tilde{\rho}_2 &= \tilde{\rho}_1 d\tilde{\rho}_2 + \tilde{\rho}_2 d\tilde{\rho}_1 + d\langle\tilde{\rho}_1, \tilde{\rho}_2\rangle_\tau \\ &= dM_\tau + O(\tilde{\rho}_1 + \tilde{\rho}_2)d\tau \end{aligned}$$

with M_τ a martingale. Using this at the times $\tau = 0$ and $\tau = s \wedge \tilde{\sigma} \wedge \tilde{\zeta}$ with $0 \leq s \leq 1 \wedge t$ and integrating, we get

$$\begin{aligned} \tilde{r}_0^2 \mathbb{P}(\tilde{\zeta} < s \wedge \tilde{\sigma}) \leq \mathbb{E}[\tilde{\rho}_{1,s\wedge\tilde{\sigma}\wedge\tilde{\zeta}}\tilde{\rho}_{2,s\wedge\tilde{\sigma}\wedge\tilde{\zeta}}] &\leq \tilde{\rho}_0^2 + C\mathbb{E}\left[\int_0^{s\wedge\tilde{\sigma}\wedge\tilde{\zeta}} (\tilde{\rho}_{1,\tau} + \tilde{\rho}_{2,\tau})d\tau\right] \\ &\leq \tilde{\rho}_0^2 + C\int_0^s \mathbb{E}[(\tilde{\rho}_{1,\tau} + \tilde{\rho}_{2,\tau}), \tau < \tilde{\sigma} \wedge \tilde{\zeta}]d\tau \\ &\overset{(91)\&(92)}{\leq} \tilde{\rho}_0^2 + C\int_0^s \frac{\tilde{\rho}_0^2}{\sqrt{\tau}} d\tau = C\tilde{\rho}_0^2. \end{aligned}$$

which is what we needed. □

Now we go back to (89). We cannot use Theorem 19 to conclude that $\mathbb{E}[\rho_{1,s}, s < \sigma\wedge\zeta] \le C\rho_0^2/\sqrt{s}$ because the equations satisfied by ρ_1 and ρ_2 are not of the form (90). However, if we take $\tilde\rho_{1,s} = \rho_{1,s}e^{\bar p_{t-s}(\lambda_s)}$, $\tilde\rho_{1,s} = \rho_{1,s}e^{\bar p_{t-s}(\lambda_s)}$, then (45) and an application of Itô's formula (followed by several rearrangements) show that $\tilde\rho_1$ and $\tilde\rho_2$ do satisfy (90). In addition, Corollary 17 combined with the fact that $e^{\bar p_{t-s}(\lambda_s)}$ is bounded shows that (91) is also satisfied. Therefore, according to Theorem 19, $\mathbb{E}[\tilde\rho_{1,s}, s < \sigma\wedge\zeta] \le C\rho_0^2/\sqrt{s}$ and this in turn implies

$$\mathbb{E}[\rho_{1,s}, s < \sigma\wedge\zeta] \le C\rho_0^2/\sqrt{s} \text{ and } \int_0^s \mathbb{E}[\rho_{1,u}, u < \sigma\wedge\zeta]du \le C\rho_0^2\sqrt{s}.$$

Using the preceding in (89), we write the resulting equation as

$$|\bar p_t(x) - 2\bar p_t(z) + \bar p_t(y)| \le c\rho_0^2\frac{e^{-Ct}}{\sqrt{s}} + c\rho_0^2\int_0^s \frac{H_{t-u}}{\sqrt{u}}du \quad \text{for any } s\in[0,1\wedge t].$$

Now dividing both sides by ρ_0^2 and then letting ρ_0 tend to 0, we arrive at

$$H_t \le c\frac{e^{-Ct}}{\sqrt{s}} + ce^{-Ct}\int_0^s \frac{H_{t-u}}{\sqrt{u}}du \quad \text{for any } s\in[0,1\wedge t].$$

From here, the rest is taken care of by the following lemma.

Lemma 21. *If $H:[0,\infty)\to[0,\infty)$ is a continuous function such that, for some constant $C>0$,*

$$H_t \le c\left(\frac{e^{-Ct}}{\sqrt{s}} + \int_0^s \frac{H_{t-u}}{\sqrt{u}}du\right), \quad 0 < s \le 1\wedge t, \tag{97}$$

then there are constants $k, K > 0$ such that

$$H_t \le Ke^{-kt} \quad \textit{for all } t > 0.$$

Proof. It suffices to concentrate on the case $t\ge 1$.

Let $m_n = \sup_{t\in[n,n+1]} H_t$ and $M_n = \sup_{t\in[n-1,n+1]} H_t$. Clearly, $m_n \le M_n$ and M_n is either m_n or m_{n+1}.

Now, if we take the t which maximizes H_t on $[n, n+1]$ and use (97), we get that for some constant $C>0$ and any $s\in[0,1]$,

$$m_n \le c\left(\frac{e^{-Cn}}{\sqrt{s}} + \sqrt{s}M_n\right).$$

We want to minimize the right hand side of the above expression over $s\in[0,1]$. For any $a,b>0$, the minimum of $a/\sqrt{s}+b\sqrt{s}$ with $s\in[0,1]$ is attained at $\frac{a}{b}\wedge 1$. Hence

$$m_n \le c\left(\frac{e^{-Cn}}{\sqrt{\frac{e^{-Cn}}{M_n}\wedge 1}} + M_n\left(\sqrt{\frac{e^{-Cn}}{M_n}\wedge 1}\right)\right).$$

We split the analysis according to the following cases:

(1) Case: $e^{-Cn/2} \le M_n$. This leads first to $e^{-Cn}/M_n < e^{-Cn/2} < 1$, and then to

$$m_n \le 2ce^{-Cn/2}\sqrt{M_n} \le 2ce^{-Cn/4}M_n.$$

This is enough to conclude that we can find a large n_1 such that for all $n\ge n_1$ one gets $m_n \le M_n/2$, which means that we cannot have $M_n = m_n$ unless $m_n = m_{n-1} = 0$. Hence $M_n = m_{n-1}$, which in turn implies that for some $k>0$

$$m_n \le e^{-k}m_{n-1} \text{ if } n\ge n_1. \tag{*}$$

(2) Case: $M_n \leq e^{-Cn/2}$. This yields

$$m_n \leq e^{-kn}. \tag{**}$$

Notice that we can arrange the constant $k > 0$ to be the same in (*) and (**) simply by taking the smaller.

By combining (*) and (**), we can show that m_n decays exponentially fast. Indeed, if there is $n_2 \geq n_1$ for which the second alternative holds, then $m_{n_2} \leq e^{-kn_2}$. Then an easy induction and use of both alternatives yields that $m_n \leq e^{-kn}$ for all $n \geq n_2$. On the other hand, if there is no such n_2, that means the second alternative holds, and this means that $m_n \leq m_{n-1}e^{-k}$ for all $n \geq n_1$. This then results in $m_n \leq m_{n_1} e^{-k(n-n_1)}$ and thus in the exponential decay. $\square$

This completes the proof of Theorem 18. $\square$

11. C^k CONVERGENCE OF $\bar{p}$ ON SURFACES WITH $\chi(M) \leq 0$

In the previous two sections, using the same notation and assumptions, we proved there exists a constant $C > 0$ such that

$$\sup_{x\in M} |\bar{p}_t(x)| + \sup_{x\in M} |\nabla \bar{p}_t(x)| + \sup_{x\in M} |\mathrm{Hess}\bar{p}_t(x)| \leq ce^{-Ct} \text{ for all } t > 0. \tag{98}$$

Alternatively stated, $\bar{p}$ converges to 0 exponentially fast in the C^2-norm. In particular, this proves that the metric g_t converges to the constant curvature metric h in the C^2-topology, and thus the curvature of g_t converges uniformly to a constant.

We now complete our discussion of the convergence to the constant curvature metric by extending this to C^∞-convergence. The culmination of the last several sections is the following theorem.

Theorem 22. *Let M be a smooth, compact surface with $\chi(M) \leq 0$, with a reference metric h of constant curvature 0 or -1, and let g_0 be a smooth initial metric in the same conformal class as h and with the same area. Then if we let $\bar{p}_t$ for $t \in [0, \infty)$ be the associated solution to the normalized Ricci flow (as given in equation (9)), we have that*

$$\bar{p}_t \to 0 \quad \text{in } C^\infty\text{, exponentially fast,}$$

in the sense that this convergence takes place exponentially fast in the C^k-norm for all positive integers k. Stated differently, if g_t for $t \in [0, \infty)$ is the family of solution metrics to the normalized Ricci flow (and so the metrics corresponding to $\bar{p}_t$), then $g_t \to h$ in C^∞, exponentially fast.

Proof. We start with the equation

$$\partial_t \bar{p} = e^{-2\bar{p}_t} \Delta \bar{p}_t + r(1 - e^{-2\bar{p}_t}).$$

Now we can assume, by induction, that all derivatives of $\bar{p}_t$ of order l with $0 \leq l \leq k-1$ decay to 0 exponentially fast as t goes to infinity. In light of the C^2-convergence, we may assume that $k \geq 3$.

Taking the kth derivative $\bar{p}_t^{(k)} = \nabla^{(k)} \bar{p}_t$, after commuting the Laplacian with the covariant derivative we obtain

$$\partial_t \bar{p}_t^{(k)} = e^{-2\bar{p}_t} \Delta \bar{p}_t^{(k)} + 2re^{-2\bar{p}_t} \bar{p}_t^{(k)} + Q_t^{(k)} \tag{99}$$

where Q^k depends on the lower order derivatives of $\bar{p}_t$, and thus we may assume by induction that for $k \geq 2$,

$$|Q_t^{(k)}| \leq ce^{-Ct}. \tag{100}$$

The idea now is to write a Feynman-Kac formula for the solution to (99) and get the estimates from this. Indeed, notice that if x_σ is the time changed Brownian motion starting at x which is defined by (12), then

$$\exp\left(2r\int_0^\sigma e^{-\bar{p}_{t-u}(x_u)}du\right)\mathcal{T}_\sigma\bar{p}^{(k)}_{t-\sigma}(x_\sigma)-\int_0^\sigma \exp\left(2r\int_0^u e^{-\bar{p}_{t-v}(x_v)}dv\right)\mathcal{T}_u Q^{(k)}_{t-u}(x_u)du \tag{101}$$

is a martingale, where $\mathcal{T}_u$ is the extension to tensors of the parallel transport along the path $x|_{[u,0]}$ from x_u to $x_0 = x$. From the technical side this expression can be seen in a clear way by lifting the equation (99) to the orthonormal frame bundle, where the lift of $\bar{p}^{(k)}_t$ takes values in a tensor product space of a fixed 2-dimensional Euclidean space. This is standard in stochastic analysis and we do not belabor it.

One result of equation (101) is that evaluation at $\sigma = 0$ and $\sigma = t$ gives

(102)

$$\bar{p}^{(k)}_t(x) = \mathbb{E}\left[\exp\left(2r\int_0^t e^{-\bar{p}_{t-u}(x_u)}du\right)\mathcal{T}_\sigma\bar{p}^{(k)}_0(x_\sigma)\right]-\mathbb{E}\left[\int_0^t \exp\left(2r\int_0^u e^{-\bar{p}_{t-v}(x_v)}dv\right)\mathcal{T}_u Q^{(k)}_{t-u}(x_u)du\right].$$

Notice the first consequence of this, namely that $|\bar{p}^{(k)}_t|$ is bounded for $r \le 0$ (which is the case under consideration). We consider separately the cases $r = -1$ and $r = 0$.

Case: $r = -1$. From the exponential decay of $\bar{p}_t$ and the induction hypothesis (the decay of $Q^{(k)}_t$) it is easy to see that

$$|\bar{p}^{(k)}_t(x)| \le ce^{-Ct} \text{ for all } t \ge 0,$$

and thus the induction is done.

Case: $r = 0$. For the flat case, we still learn from (102) that $\bar{p}^{(k)}_t(x)$ is uniformly bounded in t and x. Since the curvature of the underlying metric h is 0 we know (cf. [16, Theorem 8.1]) that the holonomy groups are trivial (perhaps after lifting to the orientation cover). Stated differently, the parallel transport along loops is the identity.

To finish the argument, we are going to use the coupling technique we already used for the gradient estimates. Start with a fixed point $x \in M$ and a unit vector ξ, and write

$$\bar{p}^k_t(x)\xi = \nabla_\xi \bar{p}^{(k-1)}_t = \lim_{h\to 0}\frac{\mathcal{T}_h\bar{p}^{(k-1)}_t(\gamma(h)) - \bar{p}^{(k-1)}_t(x)}{h}, \tag{103}$$

where here $\mathcal{T}_h$ is the parallel transport from $T_{\gamma(h)}$ to T_x along the geodesic γ started at x with initial velocity ξ.

Now we use the martingale representation (101) with k replaced by $(k-1)$ to see that, for x and y close enough and $\mathcal{T}$ the parallel transport from T_y to T_x along the minimizing geodesic,

$$\mathcal{T}\bar{p}^{(k-1)}_t(y)-\bar{p}^{(k-1)}_t(x) = \mathbb{E}\left[\mathcal{T}\,\mathcal{T}_\sigma\bar{p}^{(k-1)}_{t-\sigma}(y_\sigma) - \mathcal{T}_\sigma\bar{p}^{(k-1)}_{t-\sigma}(x_\sigma)\right]-\mathbb{E}\left[\int_0^\sigma\left(\mathcal{T}\,\mathcal{T}_u Q^{(k-1)}_{t-u}(y_u) - \mathcal{T}_u Q^{(k-1)}_{t-u}(x_u)\right)du\right].$$

Take $t \ge 1$ and let σ be $1\wedge\tau$ with τ the coupling time of x_u and y_u. Now, because the holonomy group is trivial, it follows that

$$\mathbb{E}\left[\mathcal{T}\,\mathcal{T}_{1\wedge\tau}\bar{p}^{(k-1)}_{t-1\wedge\tau}(y_{1\wedge\tau}) - \mathcal{T}_{1\wedge\tau}\bar{p}^{(k-1)}_{t-1\wedge\tau}(x_{1\wedge\tau})\right] = \mathbb{E}\left[\mathcal{T}\,\mathcal{T}_1\bar{p}^{(k-1)}_{t-1}(y_1) - \mathcal{T}_1\bar{p}^{(k-1)}_{t-1}(x_1), 1<\tau\right].$$

From this and the exponential decay of $\bar{p}^{(k-1)}_t$ and $Q^{(k-1)}_t$, we have

$$|\mathcal{T}\bar{p}^{(k-1)}_t(y) - \bar{p}^{(k-1)}_t(x)| \le e^{-Ct}\mathbb{P}(1<\tau) + e^{-Ct}\int_0^1 \mathbb{P}(u<\tau)du.$$

Finally, using the estimate (44), we get

$$|\mathcal{T}\bar{p}_t^{(k-1)}(y) - \bar{p}_t^{(k-1)}(x)| \leq e^{-Ct}d(x,y) + e^{-Ct}\int_0^1 \frac{d(x,y)}{\sqrt{u}}du = Ce^{-Ct}d(x,y).$$

Now taking $y = \gamma(h)$ and considering the limit as h goes to 0 leads to

$$|\bar{p}_t^{(k)}(x)\xi| \leq ce^{-Ct}$$

for any unit vector ξ, which implies the exponential convergence of $\bar{p}_t^{(k)}$. □

REFERENCES

[1] Marc Arnaudon, Kolehe Abdoulaye Coulibaly, and Anton Thalmaier, *Brownian motion with respect to a metric depending on time: definition, existence and applications to Ricci flow*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 13-14, 773–778.

[2] Marc Arnaudon, Koléhè Abdoulaye Coulibaly, and Anton Thalmaier, *Horizontal diffusion in C^1 path space*, Séminaire de Probabilités XLIII, Lecture Notes in Math., vol. 2006, Springer, Berlin, 2011, pp. 73–94.

[3] Huai Dong Cao, *Deformation of Kähler metrics to Kähler-Einstein metrics on compact Kähler manifolds*, Invent. Math. **81** (1985), no. 2, 359–372.

[4] Manfredo Perdigao Do Carmo, *Riemannian geometry*, Birkhäuser, Boston, 1992.

[5] Jeff Cheeger and David G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Publishing Co., Amsterdam, 1975, North-Holland Mathematical Library, Vol. 9.

[6] Patrick Cheridito, H. Mete Soner, Nizar Touzi, and Nicolas Victoir, *Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs*, Comm. Pure Appl. Math. **60** (2007), no. 7, 1081–1110.

[7] Bennett Chow and Dan Knopf, *The Ricci flow: an introduction*, Mathematical Surveys and Monographs, vol. 110, American Mathematical Society, Providence, RI, 2004.

[8] Bennett Chow, Peng Lu, and Lei Ni, *Hamilton's Ricci flow*, Graduate Studies in Mathematics, vol. 77, American Mathematical Society, Providence, RI, 2006.

[9] Koléhè A. Coulibaly-Pasquier, *Brownian motion with respect to time-changing Riemannian metrics, applications to Ricci flow*, Ann. Inst. Henri Poincaré Probab. Stat. **47** (2011), no. 2, 515–538.

[10] M. Cranston, *A probabilistic approach to gradient estimates*, Canad. Math. Bull. **35** (1992), no. 1, 46–55.

[11] Michael Cranston, *Gradient estimates on manifolds using coupling*, J. Funct. Anal. **99** (1991), no. 1, 110–124.

[12] Arash Fahim, Nizar Touzi, and Xavier Warin, *A probabilistic numerical method for fully nonlinear parabolic PDEs*, Ann. Appl. Probab. **21** (2011), no. 4, 1322–1364. MR 2857450 (2012j:65008)

[13] Anja Göing-Jaeschke and Marc Yor, *A survey and some generalizations of Bessel processes*, Bernoulli **9** (2003), no. 2, 313–349.

[14] Richard S. Hamilton, *The Ricci flow on surfaces*, Mathematics and general relativity (Santa Cruz, CA, 1986), Contemp. Math., vol. 71, Amer. Math. Soc., Providence, RI, 1988, pp. 237–262.

[15] Elton P. Hsu, *Stochastic analysis on manifolds*, Graduate Studies in Mathematics, vol. 38, American Mathematical Society, Providence, RI, 2002.

[16] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry. Vol. I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1996, Reprint of the 1963 original, A Wiley-Interscience Publication.

[17] Kazumasa Kuwada, *Convergence of time-inhomogeneous geodesic random walks and its application to coupling methods*, to appear in Ann. of Prob.

[18] Kazumasa Kuwada and Robert Philipowski, *Coupling of Brownian motions and Perelman's $\mathcal{L}$-functional*, J. Funct. Anal. **260** (2011), no. 9, 2742–2766.

[19] ———, *Non-explosion of diffusion processes on manifolds with time-dependent metric*, Math. Z. **268** (2011), no. 3-4, 979–991.

[20] Ma Li, *Ricci-Hamilton flow on surfaces*, Singapore : Global Scientific Pub., 2004.

[21] Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999.

[22] H. Mete Soner and Nizar Touzi, *A stochastic representation for the level set equations*, Comm. Partial Differential Equations **27** (2002), no. 9-10, 2031–2053.

[23] ———, *A stochastic representation for mean curvature type geometric flows*, Ann. Probab. **31** (2003), no. 3, 1145–1165.

[24] Daniel W. Stroock, *An introduction to the analysis of paths on a Riemannian manifold*, Mathematical Surveys and Monographs, vol. 74, American Mathematical Society, Providence, RI, 2000.

DEPARTMENT OF MATHEMATICS, LEHIGH UNIVERSITY, BETHLEHEM, PA, USA
E-mail address: robert.neel@lehigh.edu

SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, 686 CHERRY STREET, ATLANTA, GA 30332, AND "SIMION STOILOW" INSTITUTE OF MATHEMATICS OF ROMANIAN ACADEMY, 21 CALEA GRIVIŢEI, BUCHAREST, ROMANIA
E-mail address: ipopescu@math.gatech.edu, ionel.popescu@imar.ro