

# THE BROWN MEASURE OF THE FREE MULTIPLICATIVE BROWNIAN MOTION

BRUCE K. DRIVER, BRIAN C. HALL, AND TODD KEMP

ABSTRACT. The free multiplicative Brownian motion  $b_t$  is the large- $N$  limit of the Brownian motion on  $\mathrm{GL}(N; \mathbb{C})$ , in the sense of  $*$ -distributions. The natural candidate for the large- $N$  limit of the empirical distribution of eigenvalues is thus the Brown measure of  $b_t$ . In previous work, the second and third authors showed that this Brown measure is supported in the closure of a region  $\Sigma_t$  that appeared work of Biane. In the present paper, we compute the Brown measure completely. It has a continuous density  $W_t$  on  $\bar{\Sigma}_t$ , which is strictly positive and real analytic on  $\Sigma_t$ . This density has a simple form in polar coordinates:

$$W_t(r, \theta) = \frac{1}{r^2} w_t(\theta),$$

where  $w_t$  is an analytic function determined by the geometry of the region  $\Sigma_t$ .

We show also that the spectral measure of free unitary Brownian motion  $u_t$  is a “shadow” of the Brown measure of  $b_t$ , precisely mirroring the relationship between Wigner’s semicircle law and Ginibre’s circular law. We develop several new methods, based on stochastic differential equations and PDE, to prove these results.

## CONTENTS

1. Introduction	2
1.1. $*$ -Distribution and Brown measure	2
1.2. Brownian motions on $U(N)$ , $\mathrm{GL}(N; \mathbb{C})$ , and their large- $N$ limits	4
1.3. The Brown measure of $b_t$	5
1.4. Connection to Physics	6
2. Statement of main results	8
2.1. A formula for the Brown measure	8
2.2. A connection to free unitary Brownian motion	11
2.3. The structure of the formula	14
2.4. Deriving the formula	15
3. Comparison with the eigenvalue distribution of $B_t^N$	16
4. Properties of $\Sigma_t$	19
5. The PDE for $S$	25
6. The Hamilton–Jacobi method	29
6.1. Setting up the method	29
6.2. Constants of motion	33
6.3. Solving the equations	35
6.4. More about the lifetime of the solution	39
6.5. Surjectivity	42

---

Kemp’s research supported in part by NSF CAREER Award DMS-1254807 and NSF Award DMS-1800733.

7. Letting $x$ tend to zero	44
7.1. Outline	44
7.2. Outside $\overline{\Sigma}_t$	46
7.3. Inside $\Sigma_t$	48
7.4. Near the boundary of $\Sigma_t$	52
7.5. Proof of the main result	55
8. Further properties of the Brown measure	56
8.1. The formula for $\omega$	56
8.2. The connection to free unitary Brownian motion	60
References	62

## 1. INTRODUCTION

**1.1. \*-Distribution and Brown measure.** Two of the most fundamental models in random matrix theory are the *Gaussian unitary ensemble* and the *Ginibre ensemble*. Each is a Gaussian random matrix; the Ginibre ensemble has i.i.d. complex Gaussian entries, with variance  $1/N$ , while the Gaussian unitary ensemble is the Hermitian part of the Ginibre ensemble. For our purposes, it is natural to think of these two ensembles as endpoints of *Brownian motion on Lie algebras*. Indeed, the Lie algebra  $\mathfrak{gl}(N; \mathbb{C})$  of the general linear group consists of all  $N \times N$  complex matrices. For each fixed  $t$ , the Brownian motion  $Z_t$  on this space (with an appropriately scaled time parameter) is distributed as the Ginibre ensemble, scaled by  $\sqrt{t}$ . Similarly, the space of Hermitian matrices is equal to  $i\mathfrak{u}(N)$ , where  $\mathfrak{u}(N)$  is the Lie algebra of the unitary group, namely the space of skew-Hermitian matrices. For each fixed  $t$ , the Brownian motion  $X_t$  on this space is distributed as the Gaussian unitary ensemble, scaled by  $\sqrt{t}$ .

Among the earliest results in random matrix theory are the discovery of the large- $N$  limits of the empirical eigenvalue distributions of these ensembles. That is to say, the (random) counting measure  $\frac{1}{N} \sum_{j=1}^N \delta_{\lambda_j}$  of the eigenvalues  $\{\lambda_j\}$  of each ensemble has an almost-sure limit which is a deterministic measure. The eigenvalues of  $X_t$  are real, and their limit empirical eigenvalue distribution is the semicircle law on the interval  $[-2\sqrt{t}, 2\sqrt{t}]$  (cf. [46]); the eigenvalues of  $Z_t$  are complex, and their limit empirical eigenvalue distribution is the uniform probability measure on the disk of radius  $\sqrt{t}$  (cf. [19]). We note for later reference a simple but intriguing link between these two limiting distributions: the push-forward under the “real part” map of the uniform measure on the disk is the semicircular distribution on an interval.

It is convenient to recast the empirical eigenvalue distribution in an analytic form. For an  $N \times N$  matrix  $A = A_N$ , the function

$$\begin{aligned} L_A(\lambda) &= \log(|\det(A - \lambda)|^{1/N}) \\ &= \frac{1}{N} \sum_{j=1}^N \log(|\lambda - \lambda_j|) \end{aligned} \tag{1.1}$$

is subharmonic on  $\mathbb{C}$ . Actually,  $L_A$  is harmonic on the complement of the spectrum, and  $-\infty$  at the eigenvalues. The (distributional) Laplacian of  $L$  is therefore a positive measure; in fact it is (up to a factor of  $2\pi$ ) equal to the empirical eigenvalue

distribution of the matrix  $A$ . Hence, if one can compute an appropriate “large- $N$  limit” of the functions  $L_{A_N}$ , the Laplacian of the limiting function will provide a natural candidate for the limiting empirical eigenvalue distribution.

Free probability theory affords a medium in which to identify abstract limits of random matrix ensembles themselves. The limits are constructed as operators in a tracial von Neumann algebra  $(\mathcal{A}, \tau)$ , and the limit is with respect to  $*$ -distribution. If  $A_N$  is a sequence of  $N \times N$  random matrices, an operator  $a \in \mathcal{A}$  is said to be a **limit in  $*$ -distribution** of  $A_N$  if, for each polynomial  $p$  in two noncommuting variables, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{trace}[p(A_N, A_N^*)] = \tau[p(a, a^*)]$$

almost surely. For  $X_t$  and  $Z_t$ , Voiculescu [45] showed that the large- $N$  limits can be identified as certain *free stochastic processes*, namely the **free additive Brownian motion**  $x_t$  and the **free circular Brownian motion**  $c_t$ . These limiting processes are no longer random: they are one-parameter families of operators with freely independent increments, and  $*$ -distributions that can be described elegantly in the combinatorial framework of free probability.

In particular,  $x_t$  is a self-adjoint operator, and it therefore has a spectral resolution  $E^{x_t}$ : a projection-valued measure so that  $x_t = \int_{\mathbb{R}} \lambda E^{x_t}(d\lambda)$ . The composition  $\mu_{x_t} := \tau \circ E^{x_t}$  is a probability measure on  $\mathbb{R}$  called the *spectral measure* of  $x_t$  (in the state  $\tau$ ). The statement that  $X_t$  has the semicircle law as its limit empirical eigenvalue distribution is equivalent to the statement that  $\mu_{x_t}$  is semicircular.

On the other hand, since the operator  $c_t$  is not normal, there is no spectral theorem to yield a spectral measure. There is, however, a substitute: the Brown measure, introduced in [9]. For any operator  $a \in \mathcal{A}$ , define a function  $L_a$  on  $\mathbb{C}$  by

$$L_a(\lambda) = \tau[\log(|a - \lambda|)], \tag{1.2}$$

where  $|a - \lambda|$  is the self-adjoint operator  $((a - \lambda)^*(a - \lambda))^{1/2}$ . The quantity  $L_a(\lambda)$  is the logarithm of the Fuglede–Kadison determinant [17, 18] of  $a - \lambda$ . It is finite outside the spectrum of  $a$  but may become  $-\infty$  at points in the spectrum. If  $\mathcal{A}$  is the space of all  $N \times N$  matrices and  $\tau = \frac{1}{N} \text{trace}$ , then  $L_a$  agrees with the function in (1.1). In a general tracial von Neumann algebra  $(\mathcal{A}, \tau)$ , the function  $L_a$  is subharmonic.

The **Brown measure** of  $a$  is then defined in terms of the distributional Laplacian of  $L_a$ :

$$\mu_a := \frac{1}{2\pi} \Delta L_a.$$

If  $a$  is self-adjoint, the Brown measure of  $a$  coincides with the spectral measure. If  $a = c_t$  is the free circular Brownian motion at time  $t$ , its Brown measure  $\mu_{c_t}$  is equal to the uniform probability measure on the disk of radius  $\sqrt{t}$ . Indeed, the proof of this claim is essentially contained in Ginibre’s original paper [19].

By regularizing the right-hand side of (1.2), one can construct the Brown measure  $\mu_a$  as a weak limit,

$$d\mu_a(\lambda) = \lim_{x \rightarrow 0^+} \frac{1}{4\pi} \Delta_\lambda \tau[\log[(a - \lambda)^*(a - \lambda) + x]] d\lambda, \tag{1.3}$$

where  $\Delta_\lambda$  is the Laplacian with respect to  $\lambda$  and  $d\lambda$  is the Lebesgue measure on the plane. (See [40, Section 11.5] and [33, Eq. (2.11)].) It is not hard to see that the Brown measure of  $a$  is determined by the  $*$ -moments of  $a$ , but the dependence is

singular: if a sequence of operators  $a_n$  converges in  $*$ -distribution to  $a$ , the Brown measures  $\mu_{a_n}$  need not converge to  $\mu_a$ .

**1.2. Brownian motions on  $U(N)$ ,  $GL(N; \mathbb{C})$ , and their large- $N$  limits.** As we have noted, the Gaussian unitary ensemble and the Ginibre ensemble can be described in terms of Brownian motions on the Lie algebras  $\mathfrak{u}(N)$  and  $\mathfrak{gl}(N; \mathbb{C})$ . Specifically, the Brownian motions are induced by a choice of inner product on these finite-dimensional vector spaces; in both cases, we use the inner product

$$\langle X, Y \rangle_N = N \operatorname{Re} \operatorname{trace}(X^* Y).$$

(The factor of  $N$  in the definition produces the scaling of  $1/N$  in the variances of the two ensembles.) It is natural to consider the counterpart *Brownian motions on the Lie groups*  $U(N)$  and  $GL(N; \mathbb{C})$ . In general, if  $G \subset M_N(\mathbb{C})$  is a matrix Lie group with Lie algebra  $\mathfrak{g} \subset M_N(\mathbb{C})$ , there is a simple relationship between the Brownian motion  $B_t$  on  $G$  and the Brownian motion  $A_t$  on  $\mathfrak{g}$  (the latter being the standard Brownian motion determined by an inner product on  $\mathfrak{g}$ ). It is known as the *rolling map*, and it can be written as a Stratonovitch stochastic differential equation (SDE):

$$dB_t = B_t \circ dA_t, \quad B_0 = I.$$

The solution of this SDE is a diffusion process on  $G$  whose increments (computed in the left multiplicative sense) are independent and whose generator is half the Laplacian on  $G$  determined by the left-invariant Riemannian metric induced by the given inner product on  $\mathfrak{g}$ . Thus, its distribution at each time is the heat kernel on the group. For computational purposes, it is useful to write the SDE in Itô form; the result depends on the structure of the group. Letting  $U_t = U_t^N$  denote the Brownian motion on  $U(N)$  and  $B_t = B_t^N$  the Brownian motion on  $GL(N; \mathbb{C})$ , the corresponding Itô SDEs are

$$\begin{aligned} dU_t &= iU_t dX_t - \frac{1}{2}U_t dt \\ dB_t &= B_t dZ_t. \end{aligned}$$

It is then natural to investigate the large- $N$  limits of these random matrix processes. The candidate large- $N$  limits are the *free* stochastic processes generated by the analogous *free* SDEs:

$$du_t = iu_t ds_t - \frac{1}{2}u_t dt \tag{1.4}$$

$$db_t = b_t dc_t. \tag{1.5}$$

(For the theory of free stochastic calculus, see [7, 8, 37].) In 1997, Biane [4, 5] introduced these processes, the **free unitary Brownian motion**  $u_t$ , and the **free multiplicative Brownian motion**  $b_t$ . (He called  $b_t = \Lambda_t$ , and wrote a slightly different but equivalent free SDE for it.) The main result of [4] was the theorem that the process  $u_t$  is indeed the large- $N$  limit in  $*$ -distribution of the unitary Brownian motion  $U_t = U_t^N$ . Since  $U_t$  and  $u_t$  are unitary (hence normal) operators, this also means that the empirical eigenvalue distribution of  $U_t^N$  converges to the spectral measure of  $u_t$ .

Biane also computed the spectral measure  $\nu_t$  of  $u_t$ . We now record this result, since it relates closely to the results of the present paper. Let  $f_t$  denote the holomorphic function on  $\mathbb{C} \setminus \{1\}$  defined by

$$f_t(\lambda) = \lambda e^{\frac{t}{2} \frac{1+\lambda}{1-\lambda}}. \tag{1.6}$$

Then  $f_t$  has a holomorphic inverse  $\chi_t$  in open unit disk, and  $\chi_t$  extends continuously to the closed unit disk. Biane showed that

$$\chi_t = \frac{\psi_{u_t}}{1 + \psi_{u_t}}$$

where  $\psi_{u_t}(z) = \tau[(1 - zu_t)^{-1}] - 1$  is the (recentered) moment-generating function of  $u_t$ . From this (and other SDE computations) he determined the following result.

**Theorem 1.1** (Biane, 1997 [4, 5]). *The spectral measure  $\nu_t$  of the free unitary Brownian motion  $u_t$  is supported in the arc*

$$\left\{ e^{i\phi} \mid |\phi| < \phi_{\max}(t) := \frac{1}{2}\sqrt{(4-t)t} + \cos^{-1}(1-t/2) \right\}$$

for  $t < 4$ , and is fully supported on the circle for  $t \geq 4$ . The measure  $\nu_t$  has a continuous density  $\kappa_t$ , which is real analytic on the interior of its support arc, given by

$$\kappa_t(e^{i\phi}) = \frac{1}{2\pi} \frac{1 - |\chi_t(e^{i\phi})|^2}{|1 - \chi_t(e^{i\phi})|^2}.$$

See, for example, p. 275 in [5]. In the same papers [4, 5] in which he introduced  $u_t$ , Biane considered the free multiplicative Brownian motion  $b_t$  as well (for example, computing its norm). He conjectured that it should be the large- $N$  limit of the Brownian motion  $B_t = B_t^N$  on  $\mathrm{GL}(N; \mathbb{C})$ , in  $*$ -distribution. This was proved by the third author in [35], with complementary estimates of moments given in [36]. The goal of the present paper is to fully determine the Brown measure of the free multiplicative Brownian motion  $b_t$ , giving the full complex analog of Theorem 1.1.

**1.3. The Brown measure of  $b_t$ .** The main result of this paper is a formula for the Brown measure  $\mu_{b_t}$  of the free multiplicative Brownian motion  $b_t$ . We expect that  $\mu_{b_t}$  coincides with the large- $N$  limit of the empirical eigenvalue distribution of the Brownian motion  $B_t^N$  on  $\mathrm{GL}(N; \mathbb{C})$ . A similar strategy has been used in the analysis of the general circular law (in which the entries are independent and identically distributed but not necessarily Gaussian), beginning from the work of Girko [20] and continuing with results of Bai [2], Götze and Tikhomirov [21], and ending with the definitive version of the circular law established by Tau and Vu [44]. In these works, the authors first identify a limiting object and compute its Brown measure (in this case, uniform on a disk) and then prove that the eigenvalue distribution converges almost surely to the Brown measure.

A previous result [33] of the second and third authors showed that  $\mu_{b_t}$  is supported on a the closure of a certain region  $\Sigma_t$  introduced by Biane in [5]; see Figures 1 and 2. (We reprove that result in the present paper by a different method; see Theorem 7.2 in Section 7.2.) The proof in [33] is based on Biane’s “free Hall transform”  $\mathcal{G}_t$ , introduced in [5]. This transform was conjectured by Biane to be the large- $N$  limit of the generalized Segal–Bargmann transform of Hall [28, 29], a conjecture that was verified independently by Cébron [10] and the authors of the present paper [13]. A key idea in Biane’s work is Gross and Malliavin’s probabilistic interpretation [25] of the transform in [28].

For each  $\lambda$  outside  $\Sigma_t$ , [33] uses  $\mathcal{G}_t$  to construct an “inverse” of  $b_t - \lambda$  and  $(b_t - \lambda)^2$ . These inverses are not necessarily bounded operators, but live in the noncommutative  $L^2$  space, that is, the completion of  $\mathcal{A}$  with respect to the inner

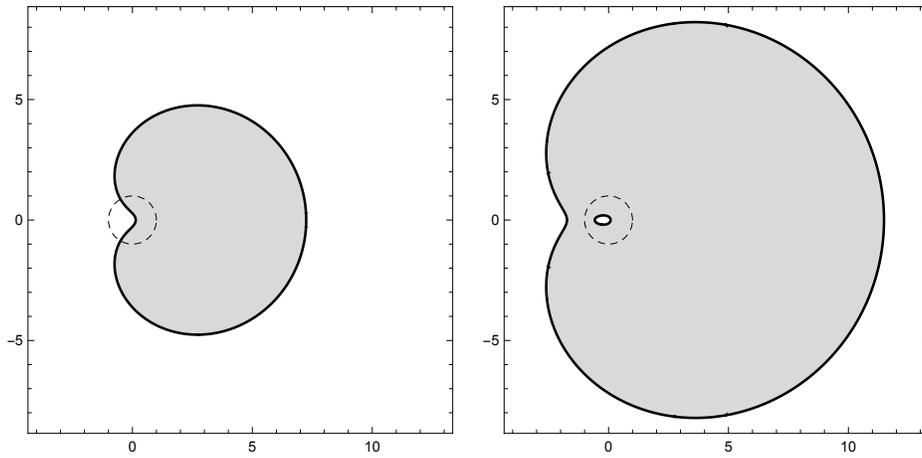


FIGURE 1. The regions  $\Sigma_t$  with  $t = 3$  and  $t = 4.1$  with the unit circle (dashed) indicated for comparison

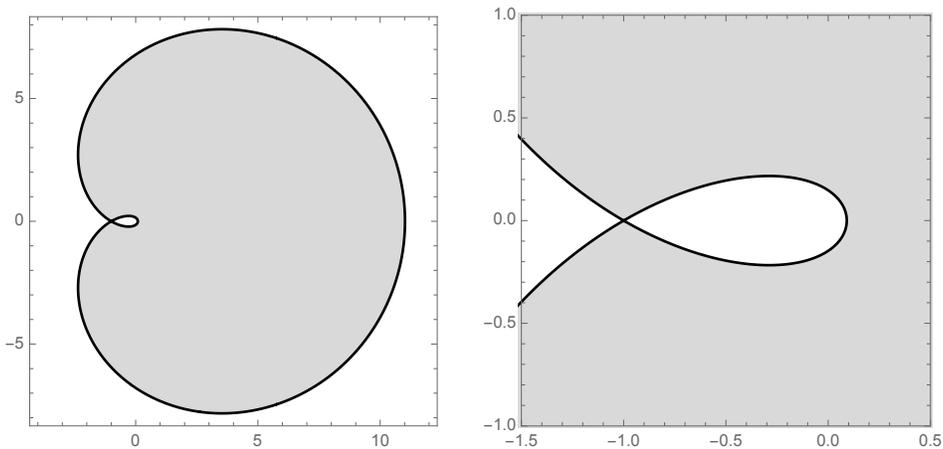


FIGURE 2. The region  $\Sigma_t$  with  $t = 4$  (left) and a detail thereof (right)

product  $\langle a, b \rangle := \tau(a^*b)$ . We then strengthen the standard result that the Brown measure is supported on the spectrum of the operator to show that existence of an  $L^2$  inverse of  $(b_t - \lambda)^2$  guarantees that  $\lambda$  is outside the support of  $\mu_{b_t}$ . We note, however, that the methods of [33] do not give any information about the distribution of the Brown measure  $\mu_{b_t}$  inside the region  $\Sigma_t$ .

**1.4. Connection to Physics.** The eigenvalue distribution of Brownian motion in  $GL(N; \mathbb{C})$ , in the large- $N$  limit, has been studied in the physics literature, first by Gudowska-Nowak, Janik, Jurkiewicz, and Nowak [26] and then by Lohmayer,

Neuberger, and Wettig [39]. At least in the case of [39], the motivation for considering this model is a connection to two-dimensional Yang–Mills theory. Yang–Mills quantum field theory is a key part of the standard model in particle physics and the two-dimensional case can be treated in a mathematically rigorous fashion. Two-dimensional Yang–Mills theory is a much-studied model, in part as a toy model of the four-dimensional theory and in part because of its connections to string theory [22, 23].

Yang–Mills theory with structure group  $G$  describes a random connection on a principal  $G$ -bundle. One typically studies the theory through the associated *Wilson loop functionals*, given by the expectation value of the trace of the holonomy around a loop. Assume at first that  $G$  is compact, that the spacetime manifold is the plane, and that the loop is a simple closed curve in the plane. Then the distribution of the holonomy is described by Brownian motion in  $G$ , with time-parameter proportional to the area enclosed by the curve. (See works of Driver [12] and Gross–King–Sengupta [24] and the references therein.)

Of particular importance is the case  $G = \mathbf{U}(N)$ , with  $N$  tending to infinity; the resulting theory is called the “master field.” The master field in the plane is therefore built around the large- $N$  limit of Brownian motion in  $\mathbf{U}(N)$ , i.e., the free unitary Brownian motion  $u_t$ . (See works of Singer [42], Anshelevich–Sengupta [1], and Lévy [38].) In this context, the change in behavior of  $u_t$  at  $t = 4$ —in which the support of the spectral measure wraps all the way around the circle—is called a “topological phase transition.” (See also [14], [15], [11], and [32] for recent progress constructing a rigorous large- $N$  Yang–Mills theory on surfaces other than the plane.)

Although Yang–Mills theory is typically constructed when the structure group  $G$  is compact, it requires only a small step of imagination to consider also the case  $G = \mathbf{GL}(N; \mathbb{C})$ . Thus, in [39], Brownian motion in  $\mathbf{GL}(N; \mathbb{C})$  is considered as a sort of “complex Wilson loop” computation. It is of interest to determine whether the large- $N$  limit—namely the free multiplicative Brownian motion  $b_t$ —still has a topological phase transition at  $t = 4$ .

The papers [26] and [39] both derive, using nonrigorous methods, the region into which the eigenvalues of  $B_t^N$  cluster in the large- $N$  limit. Both papers find this domain to be precisely the region  $\Sigma_t$  considered in the present work. Since  $\Sigma_t$  wraps around the origin precisely at time  $t = 4$ , the authors conclude that indeed the topological phase transition persists after the change from  $\mathbf{U}(N)$  to  $\mathbf{GL}(N; \mathbb{C})$ . The paper [39] also considers a two-parameter extension of the Brownian motion of the sort considered in [13, 34, 35], and find that the eigenvalues cluster into the domain denoted  $\Sigma_{s,t}$  in [34]. A rigorous version of these results—specifically, that the Brown measure of the relevant free Brownian motion is supported in  $\overline{\Sigma}_t$  or  $\overline{\Sigma}_{s,t}$ —was then obtained by the second and third authors in [33].

We emphasize that the papers [26], [39], and [33] are concerned only with the *region* into which the eigenvalues cluster. Nothing is said there about how the eigenvalues are *distributed* in the region. By contrast, in the present work, we not only prove (again) that the Brown measure of  $b_t$  is supported in  $\overline{\Sigma}_t$ , we actually *compute* the Brown measure (Theorem 2.2). Furthermore, we not only see the same transition at  $t = 4$  for the  $\mathbf{GL}(N; \mathbb{C})$  case as for the  $\mathbf{U}(N)$  case, we actually find a *direct connection* (Proposition 2.6) between the Brown measure of  $b_t$  and the spectral measure of  $u_t$ .

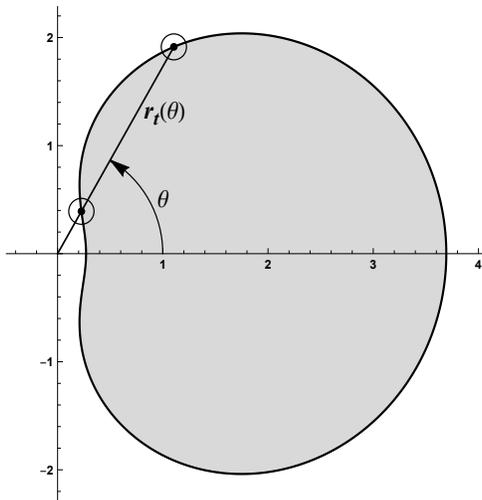


FIGURE 3. We let  $r_t(\theta)$  denote the larger of the two radii where the ray with angle  $\theta$  intersects  $\partial\Sigma_t$ . Shown for  $t = 1.5$

## 2. STATEMENT OF MAIN RESULTS

**2.1. A formula for the Brown measure.** In this paper, we compute the Brown measure  $\mu_{b_t}$  of the free multiplicative Brownian motion  $b_t$ , using completely different methods from those in [33]. To state our main result, we need to briefly describe the regions  $\Sigma_t$ . For each  $t > 0$ , consider the holomorphic function  $f_t$  on  $\mathbb{C} \setminus \{1\}$  defined by (1.6). It is easily verified that if  $|\lambda| = 1$  then  $|f_t(\lambda)| = 1$ . There are, however, *other* points where  $|f_t(\lambda)| = 1$ . We then define

$$F_t = \{\lambda \in \mathbb{C} \mid |\lambda| \neq 1, |f_t(\lambda)| = 1\} \quad (2.1)$$

and

$$E_t = \overline{F}_t. \quad (2.2)$$

**Definition 2.1.** For each  $t > 0$ , we define  $\Sigma_t$  to be the connected component of the complement of  $E_t$  containing 1.

We will show (Theorem 4.1) that  $\Sigma_t$  may also be characterized as

$$\Sigma_t = \{\lambda \in \mathbb{C} \mid T(\lambda) < t\}, \quad (2.3)$$

where the function  $T$  is defined in (4.1). Each region  $\Sigma_t$  is invariant under the maps  $\lambda \mapsto 1/\lambda$  and  $\lambda \mapsto \bar{\lambda}$ . If we consider a ray from the origin with angle  $\theta$ , if this ray intersects  $\Sigma_t$  at all, it does so in an interval of the form  $1/r_t(\theta) < r < r_t(\theta)$  for some  $r_t(\theta) > 1$ . (See Figures 3 and 4.) See Section 4 for more information.

We are now ready to state our main result.

**Theorem 2.2.** For all  $t > 0$ , the Brown measure  $\mu_{b_t}$  of  $b_t$  is absolutely continuous with respect to the Lebesgue measure on the plane and supported in the domain  $\overline{\Sigma}_t$ . In  $\Sigma_t$ , the density  $W_t$  of  $\mu_{b_t}$  with respect to the Lebesgue measure is strictly positive and real analytic, with the following form in polar coordinates:

$$W_t(r, \theta) = \frac{1}{r^2} w_t(\theta) \quad (2.4)$$

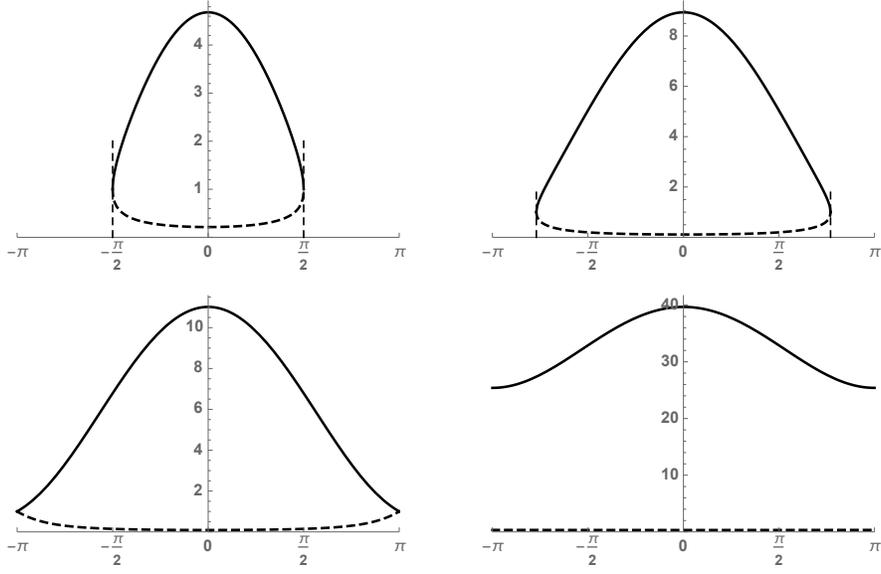


FIGURE 4. Graphs of  $r_t(\theta)$  (black) and  $1/r_t(\theta)$  (dashed) for  $t = 2, 3.5, 4,$  and  $7$

for a certain even function  $w_t$ . This function may be computed as

$$w_t(\theta) = \frac{1}{4\pi} \left( \frac{2}{t} + \frac{\partial}{\partial \theta} \frac{2r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta} \right) \quad (2.5)$$

where  $r_t(\theta)$  is the larger of the two radii where the ray with angle  $\theta$  intersects the boundary of  $\Sigma_t$ .

Since  $\Sigma_t$  is invariant under  $\lambda \mapsto \bar{\lambda}$ , the function  $r_t(\theta)$  is an even function of  $\theta$ , from which it is easy to check that the second term on the right-hand side of (2.5) is also an even function of  $\theta$ . Although we will customarily let  $r_t(\theta)$  denote the larger of the two radii, we note that

$$r \mapsto \frac{2r \sin \theta}{r^2 + 1 - 2r \cos \theta} \quad (2.6)$$

is invariant under  $r \mapsto 1/r$ . Thus, the value of  $w_t$  does not actually depend on which radius is used. It is noteworthy that the one nonexplicit part of the formula for  $w_t$ , namely the second term on the right-hand side of (2.5), is computable entirely in terms of the geometry of the region  $\Sigma_t$ . According to Proposition 8.5,  $w_t$  can also be computed as a logarithmic derivative along the boundary of  $\Sigma_t$  of the function  $f_t$  in (1.6).

It follows from (2.3) that the function  $T$  equals  $t$  on the boundary of  $\Sigma_t$ . It is then possible to use implicit differentiation in the equation  $T(\lambda) = t$  to compute  $dr_t(\theta)/d\theta$  as a function of  $r_t(\theta)$  and  $\theta$ . We may then use this computation to rewrite (2.5) in a form that no longer involves a derivative with respect to  $\theta$ , as follows.

**Proposition 2.3.** *The function  $w_t$  in Theorem 2.2 may also be computed in the form*

$$w_t(\theta) = \frac{1}{2\pi t} \omega(r_t(\theta), \theta).$$

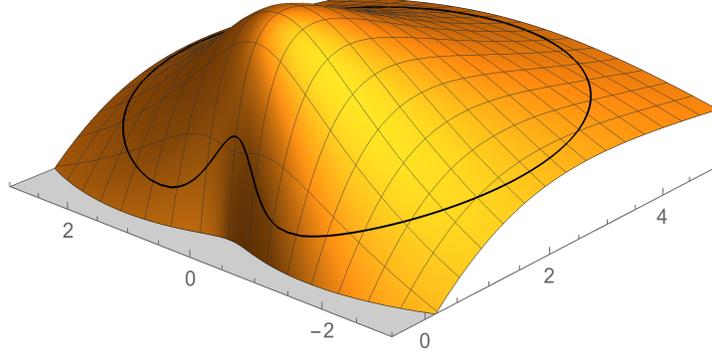


FIGURE 5. The function  $w_t(\theta)$  is computed by evaluating  $\omega$  on the boundary of  $\Sigma_t$  and parametrizing the boundary by the angle  $\theta$ . Shown for  $t = 2$

Here

$$\omega(r, \theta) = 1 + h(r) \frac{\alpha(r) \cos \theta + \beta(r)}{\beta(r) \cos \theta + \alpha(r)}, \quad (2.7)$$

where

$$h(r) = r \frac{\log(r^2)}{r^2 - 1}; \quad \alpha(r) = r^2 + 1 - 2rh(r); \quad \beta(r) = (r^2 + 1)h(r) - 2r.$$

Thus, to compute  $w_t(\theta)$ , we evaluate  $\omega/(2\pi t)$  on the boundary of  $\Sigma_t$  and then parametrize the boundary by the angle  $\theta$ ; see Figure 5. Using Proposition 2.3, we can derive small- and large- $t$  asymptotics of  $w_t(\theta)$  as follows:

$$\begin{aligned} w_t(\theta) &\sim \frac{1}{\pi t}, & t \text{ small;} \\ w_t(\theta) &\sim \frac{1}{2\pi t}, & t \text{ large.} \end{aligned}$$

See Section 8 for details.

The following simple consequences of Theorem 2.2 helps explain the significance of the factor of  $1/r^2$  in the formula (2.4) for  $W_t$ .

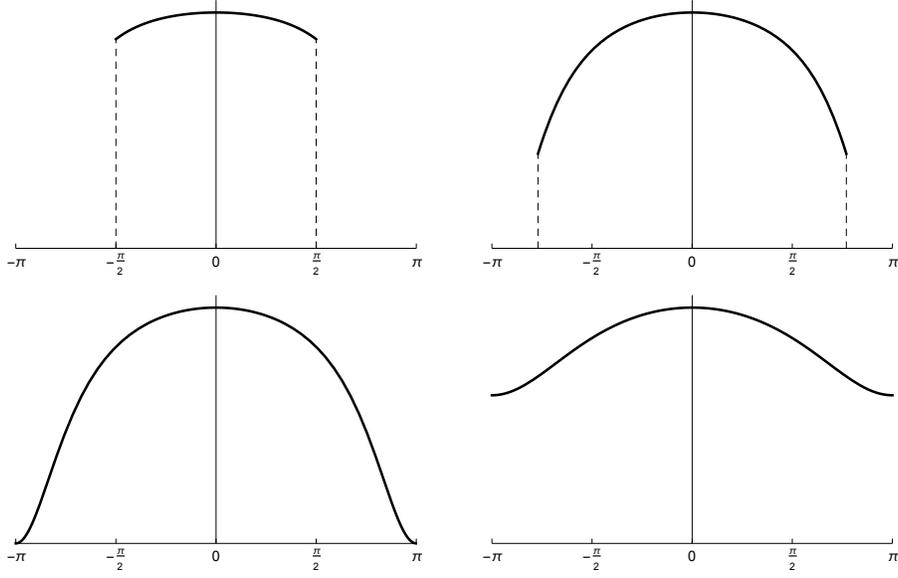
**Corollary 2.4.** *The Brown measure  $\mu_{b_t}$  of  $b_t$  has the following properties.*

- (1)  $\mu_{b_t}$  is invariant under the maps  $\lambda \mapsto 1/\lambda$  and  $\lambda \mapsto \bar{\lambda}$ .
- (2) Let  $\Xi_t$  denote the image of  $\Sigma_t \setminus (-\infty, 0)$  under the complex logarithm map, using the standard branch cut along the negative real axis. We write points  $z \in \Xi_t$  as  $(\rho, \theta)$ . Then for points in  $\Xi_t$ , the pushforward of  $\mu_{b_t}$  by the logarithm map has density  $\omega_t(\rho, \theta)$  given by

$$\omega_t(\rho, \theta) = w_t(\theta),$$

independent of  $\rho$ .

*Proof.* As we have stated above, the region  $\Sigma_t$  is invariant under the maps  $\lambda \mapsto 1/\lambda$  and  $\lambda \mapsto \bar{\lambda}$ . The invariance of  $\mu_{b_t}$  under  $\lambda \mapsto \bar{\lambda}$  follows from the fact that  $w_t$  is even. Then in polar coordinates, the map  $\lambda \mapsto 1/\lambda$  is  $(r, \theta) \mapsto (1/r, -\theta)$ . Now, we


 FIGURE 6. Plots of  $w_t(\theta)$  for  $t = 2, 3.5, 4,$  and  $7$ 

may compute  $\mu_{b_t}$  in polar coordinates as

$$d\mu_{b_t} = \left( \frac{1}{r^2} w_t(\theta) \right) r \, dr \, d\theta = w_t(\theta) \frac{1}{r} dr \, d\theta.$$

If  $s = 1/r$  and  $\phi = -\theta$ , then  $r = 1/s$  and

$$\frac{1}{r} dr = s \left( -\frac{1}{s^2} \right) ds = -\frac{1}{s} ds,$$

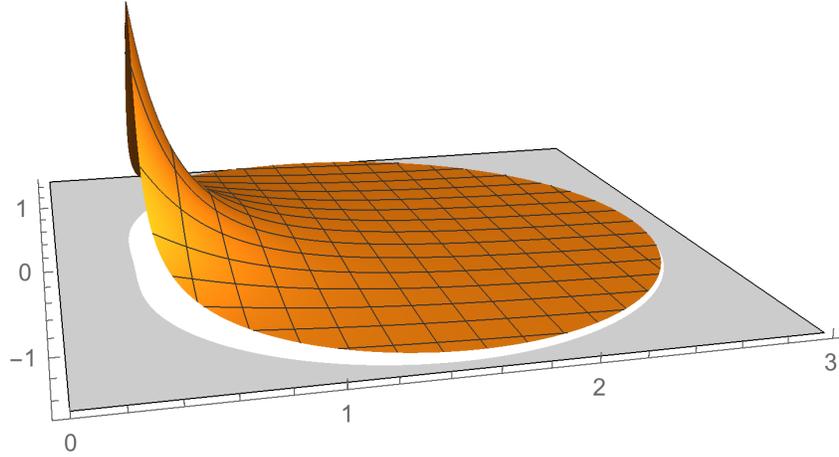
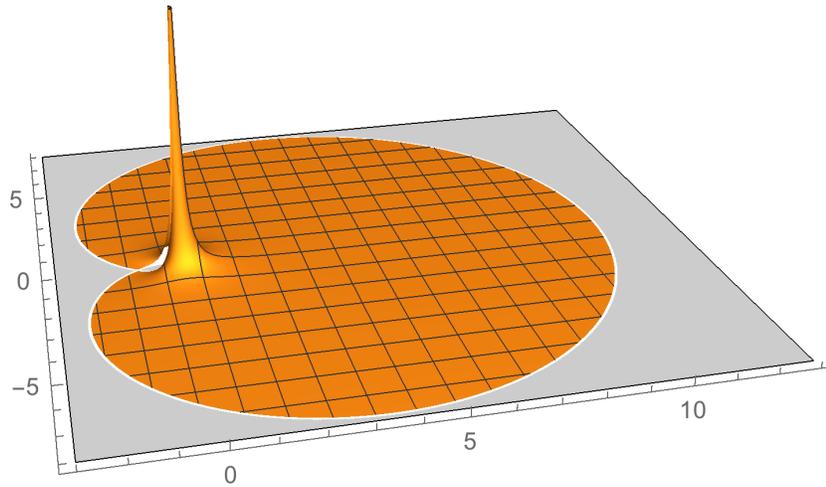
so that  $d\mu_{b_t} = w_t(\phi) \frac{1}{s} ds \, d\phi$ . Similarly, if  $\rho = \log r$ , then  $d\rho = (1/r) dr$ , so that  $d\mu_{b_t} = w_t(\theta) d\rho \, d\theta$ .  $\square$

Plots of  $w_t(\theta)$  are shown in Figure 6. Note that for  $t < 4$ , not all angles  $\theta$  actually occur in the domain  $\Sigma_t$ . Thus, for  $t < 4$ , the function  $w_t(\theta)$  is only defined for  $\theta$  in a certain interval  $(-\theta_{\max}(t), \theta_{\max}(t))$ —where, as shown in Section 4,  $\theta_{\max}(t) = \cos^{-1}(1 - t/2)$ . Plots of  $W_t$  for  $t = 1$  and  $t = 4$  are then shown in Figures 7, 8, and 9. Actually, when  $t = 1$ , the function  $w_t$  is almost constant (see Figure 19). Thus, the variation in  $W_t$  in Figure 7 comes almost entirely from the variation in the factor of  $1/r^2$  in (2.4).

We also observe that by Point 1 of Corollary 2.4, half the mass of  $\mu_{b_t}$  is contained in the unit disk and half in the complement of the unit disk. Thus, although the density  $W_t$  becomes large near the origin in, say, Figures 8 and 9, it is *not* correct to say that most of the mass of  $\mu_{b_t}$  is near the origin.

**2.2. A connection to free unitary Brownian motion.** It follows easily from Theorem 2.2 that the distribution of the argument of  $\lambda$  with respect to  $\mu_{b_t}$  has a density given by

$$a_t(\theta) = 2 \log[r_t(\theta)] w_t(\theta), \quad (2.8)$$

FIGURE 7. The density  $W_t$  with  $t = 1$ FIGURE 8. The density  $W_t$  for  $t = 4$ 

where, as in Theorem 2.2, we take  $r_t(\theta)$  to be the outer radius of the domain (with  $r_t(\theta) > 1$ ). After all, the Brown measure in the domain is computed in polar coordinates as  $(1/r^2)w_t(\theta)r dr d\theta$ . Integrating with respect to  $r$  from  $1/r_t(\theta)$  to  $r_t(\theta)$  then gives the claimed density for  $\theta$ .

Recall from Theorem 1.1 that the limiting eigenvalue distribution  $\nu_t$  for Brownian motion in the unitary group was determined by Biane. We now claim that the distribution in (2.8) is related to Biane's measure  $\nu_t$  by a natural change of variable. To each angle  $\theta$  arising in the region  $\Sigma_t$ , we associate another angle  $\phi$  by the formula

$$f_t(r_t(\theta)e^{i\theta}) = e^{i\phi}, \quad (2.9)$$

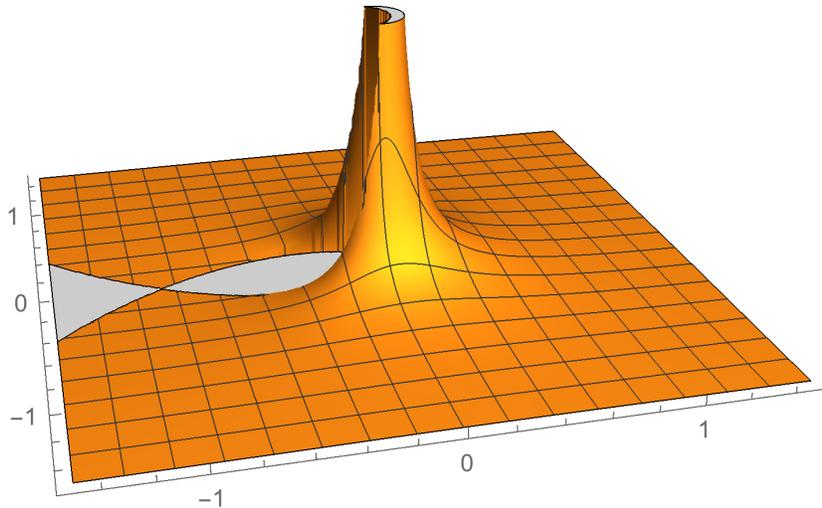


FIGURE 9. A detail of the density  $W_t$  for  $t = 4$

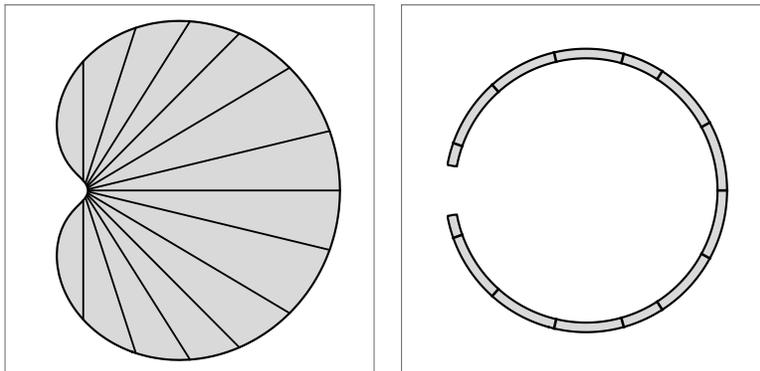


FIGURE 10. The map  $\Phi_t : \bar{\Sigma}_t \rightarrow S^1$  coincides with  $f_t$  on  $\partial\Sigma_t$  and is constant along each radial segment

where  $f_t$  is as in (1.6). (Recall that, by Definition 2.1, the boundary of  $\Sigma_t$  maps into the unit circle under  $f_t$ .) We then have the following remarkable direct connection between the Brown measure of  $b_t$  and Biane’s measure  $\nu_t$ .

**Proposition 2.5.** *If  $\theta$  is distributed according to the density in (2.8) and  $\phi$  is defined by (2.9), then  $\phi$  is distributed as Biane’s measure  $\nu_t$ .*

We may think of this result in a more geometric way, as follows. Define a map

$$\Phi_t : \bar{\Sigma}_t \rightarrow S^1$$

by requiring (a) that  $\Phi_t$  should agree with  $f_t$  on the boundary of  $\Sigma_t$ , and (b) that  $\Phi_t$  should be constant along each radial segment inside  $\Sigma_t$ , as in Figure 10. (This

specification makes sense because  $f_t$  has the same value at the two boundary points on each radial segment.) Explicitly,  $\Phi_t$  may be computed as

$$\Phi_t(\lambda) = f_t(r_t(\arg \lambda))e^{i \arg \lambda}.$$

Then Proposition 2.5 gives the following result, which may be summarized by saying that the distribution  $\nu_t$  of free unitary Brownian motion is a “shadow” of the Brown measure of  $b_t$ .

**Proposition 2.6.** *The push-forward of the Brown measure of  $b_t$  under the map  $\Phi_t$  is Biane’s measure  $\nu_t$  on  $S^1$ . Indeed, the Brown measure of  $b_t$  is the unique measure  $\mu$  on  $\bar{\Sigma}_t$  with the following two properties: (1) the push-forward of  $\mu$  by  $\Phi_t$  is  $\nu_t$  and (2)  $\mu$  is absolutely continuous with respect to Lebesgue measure with a density  $W$  having the form*

$$W(r, \theta) = \frac{1}{r^2}g(\theta)$$

in polar coordinates, for some continuous function  $g$ .

Now, the results of [5] and [33] already indicate a relationship between the free unitary Brownian motion  $u_t$  (whose spectral measure is  $\nu_t$ ) and the free multiplicative Brownian motion  $b_t$  (whose Brown measure we are studying in this paper). It is nevertheless striking to see such a direct relationship between  $\mu_{b_t}$  and  $\nu_t$ . Indeed, Proposition 2.6 precisely mirrors the relationship between the semicircle law and the circular law. If  $c_t$  is a circular random variable of variance  $t$ , and  $x_t$  is semicircular of variance  $t$ , then the distribution of  $x_t$  (the semicircle law on the interval  $[-2\sqrt{t}, 2\sqrt{t}]$ ) is the push-forward of the Brown measure of  $c_t$  (the uniform probability measure on the disk  $\bar{\mathbb{D}}(\sqrt{t})$  of radius  $\sqrt{t}$ ) under a similar “shadow map”: first project the disk onto its upper boundary circle via  $(x, y) \mapsto (x, \sqrt{t - x^2})$ , and then use the conformal map  $z \mapsto z + \frac{t}{z}$  from  $\mathbb{C} \setminus \bar{\mathbb{D}}(\sqrt{t})$  onto  $\mathbb{C} \setminus [-2\sqrt{t}, 2\sqrt{t}]$ . (The net result of these two operations is  $(x, y) \mapsto 2x$ .) Since, as described in the introduction,  $u_t$  and  $b_t$  are the “Lie group” versions of the “Lie algebra” operators  $x_t$  and  $c_t$ , it is pleasing that this shadow relationship between their Brown measures persists.

**2.3. The structure of the formula.** We now explain the significance of the two terms on the right-hand side of the formula (2.5) for  $w_t$ . Following the general construction of Brown measures in (1.3), the density of the Brown measure is computed as  $\frac{1}{4\pi}\Delta s_t(\lambda)$ , where  $\Delta$  is the Laplacian with respect to  $\lambda$  and where

$$s_t(\lambda) = \lim_{x \rightarrow 0^+} \tau[\log((b_t - \lambda)^*(b_t - \lambda) + x)].$$

It is then convenient to work in polar coordinates  $(r, \theta)$ . In these coordinates, we may write  $\Delta$  as

$$\Delta = \frac{1}{r^2} \left( \left( r \frac{\partial}{\partial r} \right)^2 + \frac{\partial^2}{\partial \theta^2} \right). \quad (2.10)$$

**Theorem 2.7.** *For each  $t > 0$ , the function  $s_t(\lambda)$  is real analytic for  $\lambda \in \Sigma_t$  and also for  $\lambda \in (\bar{\Sigma}_t)^c$ . At each boundary point,  $s_t(\lambda)$  and its first derivatives with respect to  $\lambda$  approach the same value from the inside of  $\Sigma_t$  as from the outside of  $\bar{\Sigma}_t$ . For  $\lambda$  inside  $\Sigma_t$ , we have*

$$\left( r \frac{\partial}{\partial r} \right)^2 s_t = \frac{2}{t}. \quad (2.11)$$

For  $\lambda$  inside  $\Sigma_t$ , we also have that  $\partial s_t/\partial\theta$  is independent of  $r$  with  $t$  and  $\theta$  fixed. Indeed,  $\partial s_t/\partial\theta(\lambda)$  is the unique function on  $\Sigma_t$  that is independent of  $r$  and agrees with the angular derivative of  $\log(|\lambda - 1|^2)$  as we approach  $\partial\Sigma_t$ .

The formula (2.11)—along with (2.10)—accounts for the first term on the right-hand side of (2.5). Then since the angular derivative of  $\log(|\lambda - 1|^2)$  is computable as

$$\frac{\partial}{\partial\theta} \log(|\lambda - 1|^2) = \frac{2 \operatorname{Im} \lambda}{|\lambda - 1|^2} = \frac{2r \sin \theta}{r^2 + 1 - 2 \cos \theta}$$

as in (2.6), we can recognize the second term on the right-hand side of (2.5) as the  $\theta$ -derivative of  $\partial s_t/\partial\theta$ . Thus, Theorem 2.7, together with the formula (2.10) for the Laplacian in polar coordinates, accounts for the formula (2.5) for  $w_t$ .

**2.4. Deriving the formula.** We now briefly indicate the method we will use to compute the Brown measure  $\mu_{b_t}$ . Following the general construction of the Brown measure in (1.3), we consider the function  $S$  defined by

$$S(t, \lambda, x) = \tau[\log((b_t - \lambda)^*(b_t - \lambda) + x)] \quad (2.12)$$

for  $\lambda \in \mathbb{C}$  and  $x > 0$ , where  $b_t$  is the free multiplicative Brownian motion and  $\tau$  is the trace in the von Neumann algebra in which  $b_t$  lives. It is easily verified that as  $x$  decreases with  $t$  and  $\lambda$  fixed,  $S(t, \lambda, x)$  also decreases. Hence, the limit

$$s_t(\lambda) = \lim_{x \rightarrow 0^+} S(t, \lambda, x)$$

exists, possibly with the value  $-\infty$ .

The general theory developed by Brown [9] shows that  $s_t(\lambda)$  is a subharmonic function of  $\lambda$  for each fixed  $t$ , so that the Laplacian (in the distribution sense) of  $s_t(\lambda)$  with respect to  $\lambda$  is a positive measure. If this measure happens to be absolutely continuous with respect to the Lebesgue measure, then the density  $W(t, \lambda)$  of the Brown measure is computed in terms of the value of  $s_t(\lambda)$ , as follows:

$$W(t, \lambda) = \frac{1}{4\pi} \Delta s_t(\lambda). \quad (2.13)$$

See also Chapter 11 in [40] and Section 2.3 in [33] for general information on Brown measures.

The first major step toward proving Theorem 2.2 is the following result.

**Theorem 2.8.** *The function  $S$  in (2.12) satisfies the following PDE:*

$$\frac{\partial S}{\partial t} = x \frac{\partial S}{\partial x} \left( 1 + (|\lambda|^2 - x) \frac{\partial S}{\partial x} - a \frac{\partial S}{\partial a} - b \frac{\partial S}{\partial b} \right), \quad \lambda = a + ib, \quad (2.14)$$

with the initial condition

$$S(0, \lambda, x) = \log(|\lambda - 1|^2 + x). \quad (2.15)$$

We emphasize that  $S(t, \lambda, x)$  is only defined for  $x > 0$ . Although, as we will see,  $\lim_{x \rightarrow 0} S(t, \lambda, x)$  is finite,  $\partial S/\partial x$  develops singularities in this limit. Thus, it is *not* correct to formally set  $x = 0$  in (2.14) to obtain  $\partial s_t/\partial t = 0$ . (Actually, it will turn out that  $s_t(\lambda)$  is independent of  $t$  for as long as  $\lambda$  remains outside  $\Sigma_t$ , but not after this time; see Section 7.2.)

After verifying this equation (Section 5), we will use the Hamilton–Jacobi formalism to analyze the solution (Section 6). In the remaining sections, we will then

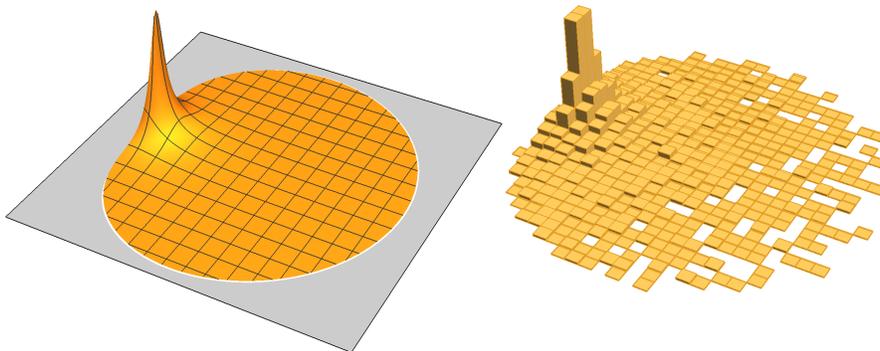


FIGURE 11. A plot of  $W_t$  with  $t = 2$  (left). A histogram of the eigenvalues of  $B_t^N$  with  $N = 2,000$  and  $t = 2$  (right) is shown for comparison

analyze the limit of the solution as  $x$  tends to zero and compute the Laplacian in (2.13).

By way of comparison, we mention that a similar PDE was used in Biane's paper [3]. There he studies the spectral measure  $\mu_t$  of  $x_0 + x_t$ , the free additive Brownian motion with a nonconstant initial distribution  $x_0$  freely independent from  $x_t$ . Biane studies the Cauchy transform  $G$  of  $\mu_t$ :

$$G(t, z) = \int_{\mathbb{R}} \frac{\mu_t(dx)}{z - x}, \quad \text{Im}(z) > 0,$$

and shows that  $G$  satisfies the complex inviscid Burger's equation

$$\frac{\partial G(t, z)}{\partial t} = -G(t, z) \frac{\partial G}{\partial z}. \quad (2.16)$$

The measure  $\mu_t$  may then be recovered, up to a constant, as  $\lim_{\varepsilon \rightarrow 0^+} \text{Im} G(t, x + i\varepsilon)$ .

In our paper, we similarly use a first-order, nonlinear PDE whose solution in a certain limit gives the desired measure. We note, however, that the PDE (2.16) is not actually the main source of information about  $\mu_t$  in [3]. By contrast, our analysis of the Brown measure of the free multiplicative Brownian motion  $b_t$  is based entirely on the PDE in Theorem 2.8.

### 3. COMPARISON WITH THE EIGENVALUE DISTRIBUTION OF $B_t^N$

As mentioned in Section 1.3, the Brown measure of the free multiplicative Brownian motion  $b_t$  is a natural candidate for the limiting empirical eigenvalue distribution of the Brownian motion  $B_t^N$  in  $\text{GL}(N; \mathbb{C})$ . We may express this idea formally as a conjecture.

**Conjecture 3.1.** *For all  $t > 0$ , the empirical eigenvalue distribution of  $B_t^N$  converges almost surely as  $N \rightarrow \infty$  to the Brown measure  $\mu_{b_t}$  of  $b_t$ .*

While natural, Conjecture 3.1 is technically difficult to approach. It is by now well known that the logarithmic singularity in the definition of the Brown measure can result in failure of convergence of the empirical distribution of eigenvalues to the Brown measure of the limit (in  $*$ -distribution) of the random matrix ensembles.

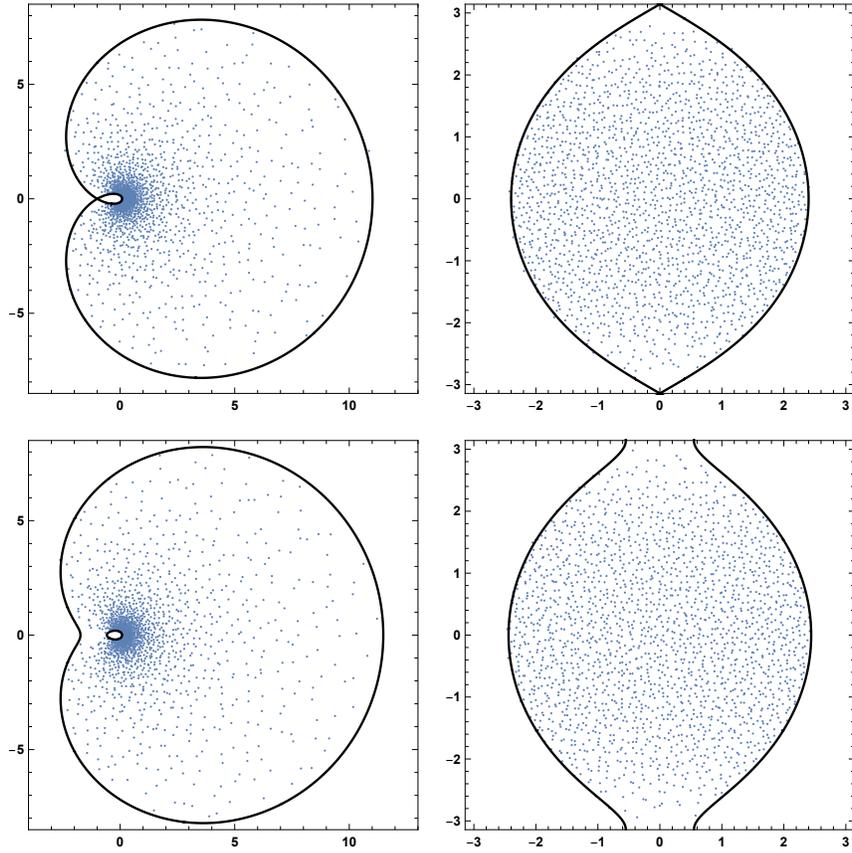


FIGURE 12. Simulation of the eigenvalues of  $B_t^N$  (left) and their logarithms (right) with  $N = 2,000$  and  $t = 4.1$ . The distribution of points on the right-hand side of the figure is approximately constant in the horizontal direction

Suppose, for example,  $T_N$  is an  $N \times N$  matrix with all 0 entries except for 1's just below the diagonal. Then all of  $T_N$ 's eigenvalues are 0, and hence the empirical eigenvalue distribution is a point-mass at 0 for each  $N$ . However, in  $*$ -distribution,  $T_N$  converges to a Haar unitary  $u$ , whose Brown measure is the uniform probability measure on the unit circle. (See, for example, Section 2.6 of [43].)

In [43], Śniady proved that convergence to the Brown measure is “generic”, in the sense that a small (vanishing in the limit) independent Gaussian perturbation of the original ensemble will always yield convergence to the Brown measure. (The required size of the perturbation was more recently explored in [27].) A main step in Śniady's proof is to show (in our language) that the empirical eigenvalue distribution of a complex Brownian motion  $Z_t$  on  $\mathfrak{gl}(N; \mathbb{C})$ , with any deterministic initial condition, converges to the appropriate Brown measure. That is to say, there is enough regular noise in such matrix diffusions to kill any pseudo-spectral discontinuities. It is natural to expect that the same should hold true for our geometric matrix diffusion  $B_t$ .

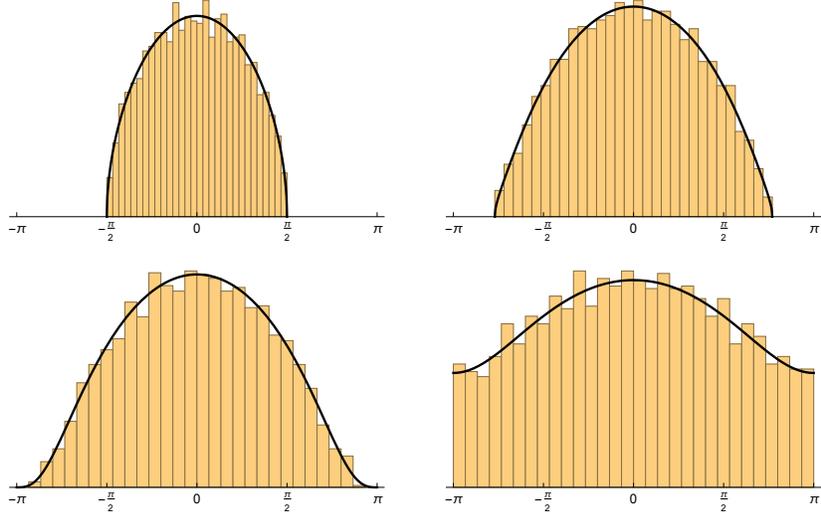


FIGURE 13. The density  $a_t(\theta)$  in (2.8) plotted against a histogram of the arguments of the eigenvalues of  $B_t^N$ , for  $N = 2,000$  and  $t = 2, 3.5, 4,$  and  $7$

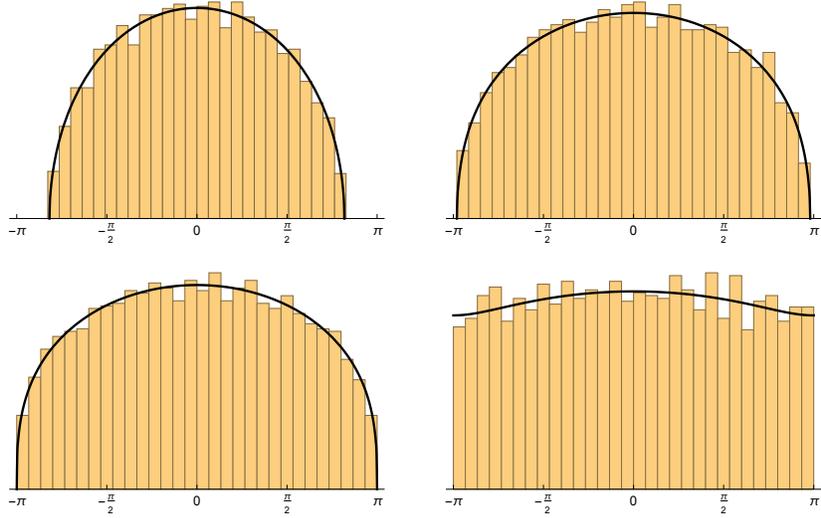


FIGURE 14. The density of Biane's measure  $\nu_t(\phi)$  plotted against a histogram of  $\{\Phi_t(\lambda_j)\}_{j=1}^N$ , for  $N = 2,000$  and  $t = 2, 3.5, 4,$  and  $7$

In support of Conjecture 3.1, we first note that for small  $t$ , the distribution of  $B_t^N$ —namely, the time- $t$  heat kernel measure on  $\mathrm{GL}(N; \mathbb{C})$ , based at the identity—is approximately Gaussian. Thus,  $B_t^N$  is distributed, for small  $t$ , similarly to the Ginibre ensemble, shifted by the identity and scaled by  $\sqrt{t}$ . Thus, Conjecture 3.1 leads us to expect that that  $\mu_{b_t}$  will be close, for small  $t$ , to the uniform probability

measure on a disk of radius  $\sqrt{t}$ . This expectation is confirmed by the asymptotics in Section 8.1.

We now offer several numerical tests of the conjecture. First, Figure 11 directly compares the density  $W_t$  with  $t = 2$  to the distribution of eigenvalues of  $B_t^N$  with  $t = 2$  and  $N = 2,000$ . Next, in light of Conjecture 3.1 and Point 2 of Corollary 2.4, we expect that the limiting distribution of the *logarithms* of the eigenvalues of  $B_t^N$  will be constant in the horizontal direction. This expectation is confirmed by simulations, as in the right-hand side of Figure 12.

Furthermore, Conjecture 3.1 predicts that the large- $N$  distribution of the arguments of the eigenvalues will be given by the density  $a_t$  in (2.8). Furthermore, Conjecture 3.1 and Proposition 2.6 predict that, if  $\{\lambda_j\}_{j=1}^N$  are the eigenvalues of  $B_t^N$ , then for large  $N$ , the empirical distribution of  $\{\Phi_t(\lambda_j)\}_{j=1}^N$  will be given by Biane's measure  $\nu_t$ . Both of these predictions are confirmed by simulations; see Figures 13 and 14.

#### 4. PROPERTIES OF $\Sigma_t$

We now verify some important properties of the regions  $\Sigma_t$  in Definition 2.1. Define

$$T(\lambda) = \begin{cases} |\lambda - 1|^2 \frac{\log(|\lambda|^2)}{|\lambda|^2 - 1}, & |\lambda| \neq 1 \\ |\lambda - 1|^2, & |\lambda| = 1 \end{cases}. \quad (4.1)$$

Note that the function

$$x \mapsto \frac{\log(x)}{x - 1}$$

has a removable singularity at  $x = 1$ , with a limiting value of 1 at  $x = 1$ . Thus,  $T(\lambda)$  is a real analytic function on all of  $\mathbb{C} \setminus \{0\}$ . Since, also,

$$\lim_{x \rightarrow 0} \frac{\log(x)}{x - 1} = +\infty,$$

we see that  $T(\lambda) \rightarrow +\infty$  as  $\lambda \rightarrow 0$ . By checking the three cases  $|\lambda| > 1$ ,  $|\lambda| = 1$ , and  $|\lambda| < 1$ , we may verify that  $T(\lambda) \geq 0$  for all  $\lambda$ , with equality only if  $\lambda = 1$ .

**Theorem 4.1.** *For all  $t > 0$ , the region  $\Sigma_t$  may be expressed as*

$$\Sigma_t = \{\lambda \in \mathbb{C} \mid T(\lambda) < t\}$$

*and the boundary of  $\Sigma_t$  may be expressed as*

$$\partial\Sigma_t = \{\lambda \in \mathbb{C} \mid T(\lambda) = t\}.$$

Thus, each fixed  $\lambda \in \mathbb{C}$  will be outside  $\bar{\Sigma}_t$  until  $t = T(\lambda)$  and will be inside  $\Sigma_t$  for all  $t > T(\lambda)$ . We may therefore say that  $T(\lambda)$  is *the time that the domain  $\Sigma_t$  gobbles up  $\lambda$* . See Figures 15 and 16.

**Theorem 4.2.** *For each  $t > 0$ , the region  $\Sigma_t$  has the following properties.*

- (1) *For  $t \leq 4$ , we have  $|\arg \lambda| \leq \cos^{-1}(1 - t/2)$  for all  $\lambda \in \bar{\Sigma}_t$ , with equality precisely for the points on the unit circle with  $\cos \theta = 1 - t/2$ .*
- (2) *Consider the ray from the origin with angle  $\theta$ ; if  $t \leq 4$ , assume  $|\theta| < \cos^{-1}(1 - t/2)$ . Then this ray intersects  $\Sigma_t$  precisely in an open interval of the form  $1/r_t(\theta) < r < r_t(\theta)$  for some  $r_t(\theta) > 1$ .*
- (3) *The boundary of  $\Sigma_t$  is smooth for all  $t > 0$  with  $t \neq 4$ . When  $t = 4$ , the boundary of  $\Sigma_t$  is smooth except at  $\lambda = -1$ , near which it looks like the transverse intersection of two smooth curves.*

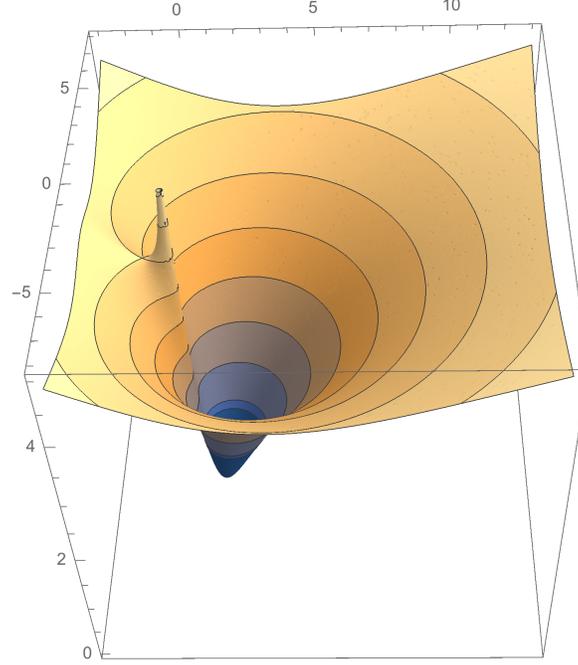


FIGURE 15. Plot of the function  $T(\lambda)$ , showing values between 0 and 5. The function has a global minimum at  $\lambda = 1$ , a saddle point at  $\lambda = -1$ , and a pole at  $\lambda = 0$

- (4) The region  $\Sigma_t$  is invariant under  $\lambda \mapsto 1/\lambda$  and under  $\lambda \mapsto \bar{\lambda}$ .
- (5) The region  $\Sigma_t$  coincides with the one defined by Biane in [5].

For Point 1, see Figure 17.

We now begin working toward the proofs of Theorems 4.1 and 4.2.

**Lemma 4.3.** For  $\lambda \in \mathbb{C}$  with  $|\lambda| \neq 1$ , we have  $|f_t(\lambda)| = 1$  if and only if  $T(\lambda) = t$ .

*Proof.* Since  $f_t(0) = 0$ , we must have  $\lambda \neq 0$  if  $|f_t(\lambda)|$  is going to equal 1. For nonzero  $\lambda$ , we compute that

$$\begin{aligned} \log(|f_t(\lambda)|) &= \log |\lambda| + \operatorname{Re} \left( \frac{t}{2} \frac{1 + \lambda}{1 - \lambda} \right) \\ &= \log |\lambda| + \frac{t}{2} \frac{1 - |\lambda|^2}{|\lambda - 1|^2}. \end{aligned} \tag{4.2}$$

Thus, for nonzero  $\lambda$ , the condition  $|f_t(\lambda)| = 1$  is equivalent to

$$0 = \log |\lambda| + \frac{t}{2} \frac{1 - |\lambda|^2}{|\lambda - 1|^2}.$$

When  $|\lambda| \neq 1$ , this condition simplifies to

$$t = |\lambda - 1|^2 \frac{\log(|\lambda|^2)}{|\lambda|^2 - 1} = T(\lambda),$$

as claimed.  $\square$

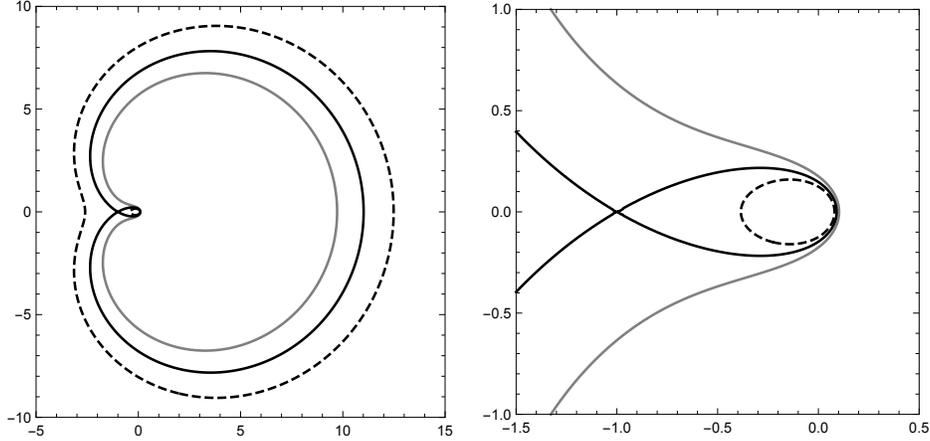


FIGURE 16. Level sets of the function  $T(\lambda)$  form the boundaries of the regions  $\Sigma_t$ . Shown for  $t = 3.7$  (gray),  $t = 4$  (black), and  $t = 4.3$  (dashed). The right-hand side of the figure gives a close-up view near  $\lambda = 0$ .

We now state some important properties of the function  $r_t$  occurring in the statement of Theorem 2.2; the proof is given on p. 23.

**Proposition 4.4.** *Consider a real number  $t > 0$  and an angle  $\theta \in (-\pi, \pi]$ , where if  $t \leq 4$ , we require  $|\theta| < \cos^{-1}(1 - t/2)$ . Then there exist exactly two radii  $r \neq 1$  for which  $|f_t(re^{i\theta})| = 1$ , and these radii have the form  $r = r_t(\theta)$  and  $r = 1/r_t(\theta)$  with  $r_t(\theta) > 1$ . Furthermore,  $r_t(\theta)$  depends analytically on  $\theta$  and if  $t \leq 4$ , then  $r_t(\theta) \rightarrow 1$  as  $\theta \rightarrow \pm \cos^{-1}(1 - t/2)$ .*

*If  $t \leq 4$  and  $\theta \in (-\pi, \pi]$  satisfies  $|\theta| \geq \cos^{-1}(1 - t/2)$ , then there are no radii  $r \neq 1$  with  $|f_t(r)| = 1$ .*

Using the proposition, we can now compute the sets  $F_t$  and  $E_t = \overline{F_t}$  that enter into the definition of  $\Sigma_t$ . (Recall (2.1) and (2.2).)

**Corollary 4.5.** *For  $t \leq 4$ , the set  $F_t$  consists of points of the form  $r_t(\theta)e^{i\theta}$  and  $(1/r_t(\theta))e^{i\theta}$  for  $-\cos^{-1}(1 - t/2) < \theta < \cos^{-1}(1 - t/2)$ . In this case, the closure of  $F_t$  consists of  $F_t$  together with the points  $e^{i\theta}$  on the unit circle with  $\cos \theta = 1 - t/2$ . There are two such points when  $t < 4$  and one such point when  $t = 4$ , namely  $-1$ .*

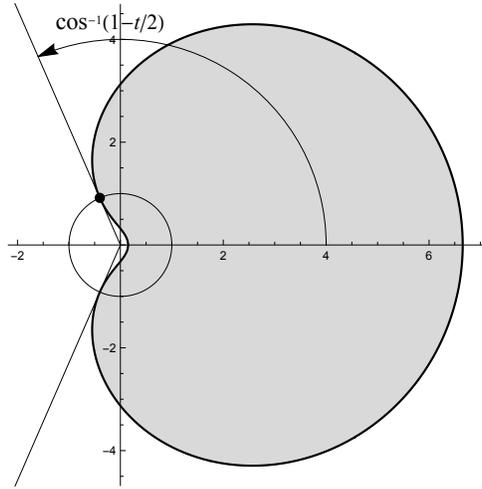
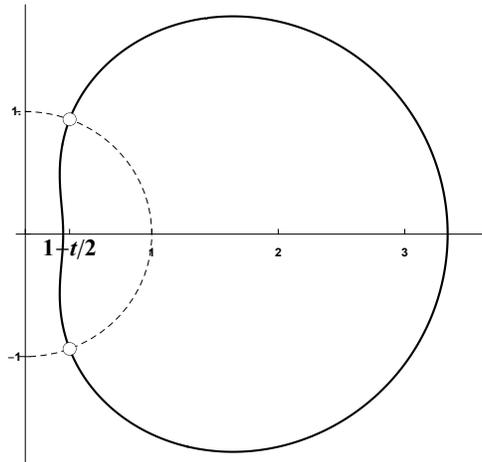
*For  $t > 4$ , the set  $F_t$  consists of points of the form  $r_t(\theta)e^{i\theta}$  and  $(1/r_t(\theta))e^{i\theta}$ , where  $\theta$  ranges over all possible angles, and this set is closed.*

See Figure 18.

We now set out to prove Proposition 4.4. In the proof, we will always rewrite the equation  $|f_t(\lambda)| = 1$ , for  $|\lambda| \neq 1$ , as  $T(\lambda) = t$  (Lemma 4.3).

**Lemma 4.6.** *Let us write the function  $T$  in (4.1) in polar coordinates. Then for each  $\theta$ , the function  $r \mapsto T(r, \theta)$  is strictly decreasing for  $0 < r < 1$  and strictly increasing for  $r > 1$ . For each  $\theta$ , the minimum value of  $T(r, \theta)$ , achieved at  $r = 1$ , is  $2(1 - \cos \theta)$ , and we have*

$$\lim_{r \rightarrow 0} T(r, \theta) = \lim_{r \rightarrow +\infty} T(r, \theta) = +\infty.$$

FIGURE 17. Illustration of Point 1 of Proposition 4.2, for  $t = 2.8$ FIGURE 18. The set  $F_t$  with  $t = 1.3$ , with the unit circle (dashed) shown for comparison.

*Proof.* We will show in Proposition 6.13 that the function  $T(\lambda)$  is the limit of another function  $t_*(\lambda_0, x_0)$  as  $x_0$  goes to zero. Explicitly, this amounts to saying that  $T(r, \theta) = g_\theta(\delta)$ , where  $g$  is defined in (6.61) and  $\delta = r + 1/r$ . Now,  $\delta$  is decreasing for  $0 < r < 1$  and increasing for  $r > 1$ . Thus, the claimed monotonicity of  $T$  follows if  $g_\theta(\delta)$  is an increasing function of  $\delta$  for each  $\theta$ , which we will show in the proof of Proposition 6.16.

For the convenience of the reader, we briefly outline how the argument goes in the context of the function  $T(r, \theta)$ . We note that

$$T(r, \theta) = (r^2 + 1 - 2r \cos \theta) \frac{\log(r^2)}{r^2 - 1},$$

where if we assign  $\log(r^2)/(r^2 - 1)$  the value 1 at  $r = 1$ , then  $T$  is analytic except at  $r = 0$ . We then compute that, after simplification,

$$\frac{\partial T}{\partial r} = 2 \left[ \frac{-2r + (1 + r^2) \cos \theta}{(r^2 - 1)^2} \right] \log(r^2) + \frac{r^2 + 1 - 2r \cos \theta}{r^2 - 1} \frac{2}{r}. \quad (4.3)$$

We then claim that for all  $\theta$ , we have  $\partial T/\partial r > 0$  for  $r > 1$  and  $\partial T/\partial r < 0$  for  $r < 1$ . Note that for each fixed  $r$ , the right-hand side of (4.3) depends linearly on  $\cos \theta$ . Thus, if, for a fixed  $r$ , if  $\partial T/\partial r$  is positive when  $\cos \theta = 1$  and when  $\cos \theta = -1$ , it will be positive for all  $\theta$ . Specifically, we may say that

$$\frac{\partial T}{\partial r}(r, \theta) \geq \min \left( \frac{\partial T}{\partial r}(r, 0), \frac{\partial T}{\partial r}(r, \pi) \right). \quad (4.4)$$

It is now an elementary (if slightly messy) computation to check that the right-hand side of (4.4) is strictly positive for all  $r > 1$ . A similar argument then shows that  $\partial T/\partial r$  is negative for all  $\theta$  and all  $0 < r < 1$ .

We conclude that for each  $\theta$ , the function  $r \mapsto T(r, \theta)$  is decreasing for  $0 < r < 1$  and increasing for  $r > 1$ . The minimum value therefore occurs at  $r = 1$ , and this value is the value of  $r^2 + 1 - 2 \cos \theta$  at  $r = 1$ , namely  $2(1 - \cos \theta)$ . Finally, we can easily see that for  $r$  approaching zero, we have

$$T(r, \theta) \sim -\log(r^2) \rightarrow +\infty$$

and for  $r$  approaching infinity, we have

$$T(r, \theta) \sim \log(r^2) \rightarrow +\infty,$$

as claimed.  $\square$

*Proof of Proposition 4.4.* The minimum value of  $T(r, \theta)$ , achieved at  $r = 1$ , is  $2 - 2 \cos \theta$ . This value is always less than  $t$ , as can be verified separately in the cases  $t > 4$  (all  $\theta$ ) and  $t \leq 4$  ( $|\theta| < \cos^{-1}(1 - t/2)$ ). Thus, Lemma 4.6 tells us that the equation  $T(r, \theta) = t$  has exactly one solution for  $r$  with  $0 < r < 1$  and exactly one solution for  $r > 1$ . Since, as is easily verified,  $T(1/r, \theta) = T(r, \theta)$ , the two solutions are reciprocals of each other, and we let  $r_t(\theta)$  denote the solution with  $r > 1$ . Since  $\partial T/\partial r$  is nonzero for all  $r \neq 1$ , the implicit function theorem tells us that  $r_t(\theta)$  depend analytically on  $\theta$ .

Now, if  $t \leq 4$  and  $\theta$  approaches  $\pm \cos^{-1}(1 - t/2)$ , the minimum value of  $2 - 2 \cos \theta$ —achieved at  $r = 1$ —approaches  $2 - 2(1 - t/2) = t$ . It should then be plausible that  $r_t(\theta)$  will approach  $r = 1$ . To make this claim rigorous, we need to show that  $T(r, \theta)$  increases rapidly enough as  $r$  increases from 1 that the  $T(r, \theta) = t$  is achieved close to  $r = 1$ . To that end, let  $g(r)$  denote the function on the right-hand side of (4.4), which is continuous everywhere and strictly positive for  $r > 1$ . Then for  $r > 1$ , we have

$$T(r, \theta) - (2 - 2 \cos \theta) \geq G(r) := \int_1^r g(s) ds. \quad (4.5)$$

Now,  $G(r)$  is continuous and strictly increasing for  $r \geq 0$ , with  $G(1) = 0$ . Thus,  $G$  it has a continuous inverse function satisfying  $G^{-1}(0) = 1$ .

For  $\varepsilon > 0$ , choose  $\delta > 0$  so that  $G^{-1}(R) < 1 + \varepsilon$  when  $0 < R < \delta$ . Then take  $\theta$  sufficiently close to  $\pm \cos^{-1}(1 - t/2)$  that  $2 - 2 \cos \theta$  is within  $\delta$  of  $t$ . Then

$$G^{-1}(t - (2 - 2 \cos \theta)) < 1 + \varepsilon,$$

which is to say that there is an  $R$  with  $1 < R < 1 + \varepsilon$  such that

$$\int_1^R g(s) ds = t - (2 - 2 \cos \theta).$$

From (4.5) we can then see that  $T(R, \theta) > t$ . Thus,  $r_t(\theta)$  will satisfy

$$1 < r_t(\theta) < R < 1 + \varepsilon.$$

We have therefore shown that  $r_t(\theta) \rightarrow 1$  as  $\theta \rightarrow \pm \cos^{-1}(1 - t/2)$ .

Finally, if  $t \leq 4$  and  $\theta \in (-\pi, \pi]$  satisfies  $|\theta| \geq \cos^{-1}(1 - t/2)$ , the minimum value of  $T(r, \theta)$ , achieved at  $r = 1$ , is  $2 - 2 \cos \theta \geq t$ . Thus, there are no values of  $r \neq 1$  where  $T(r, \theta) = t$ .  $\square$

We are now ready for the proofs of our main results about  $\Sigma_t$ .

*Proof of Theorem 4.1.* We first claim that the set  $E_t = \overline{F_t}$  is precisely the set where  $T(\lambda) = t$ . To see this, first note that  $F_t$  is, by Lemma 4.3, the set of  $\lambda$  with  $|\lambda| \neq 1$  where  $T(\lambda) = t$ . Then by Corollary 4.5, the closure of  $F_t$  is obtained by adding in the points on the unit circle (zero, one, or two such points, depending on  $t$ ) where  $\cos \theta = 1 - t/2$ . But these points are easily seen to be the points on the unit circle where  $T(\lambda) = t$ .

Using Corollary 4.5, we see that the complement of the set  $E_t = \{\lambda | T(\lambda) = t\}$  has two connected components when  $t < 4$  and three connected components when  $t \geq 4$ . Since  $T(1) = 0 < t$ , we have  $T(\lambda) < t$  on the entire connected component of  $E_t^c$  containing 1, which is, by definition, the region  $\Sigma_t$ . The remaining components of  $E_t^c$  are the unbounded component and (for  $t \geq 4$ ) the component containing 0. Since  $T(\lambda)$  tends to  $+\infty$  at zero and at infinity, we see that  $T(\lambda) > t$  on these regions, so that  $T(\lambda) < t$  precisely on  $\Sigma_t$ .

It is also clear from Corollary 4.5 that the boundary of the region  $\Sigma_t$  (i.e., the connected component of  $E_t^c$  containing 1) contains the entire set  $E_t = \{\lambda | T(\lambda) = t\}$ .  $\square$

*Proof of Theorem 4.2.* Point 1 follows easily from Corollary 4.5. For Point 2, we note that by Proposition 4.4, we have  $T(r, \theta) < t$  for  $1/r_t(\theta) < r < r_t(\theta)$ , and  $T(r, \theta) \geq t$  for  $0 < r \leq 1/r_t(\theta)$  and for  $r \geq r_t(\theta)$ . Thus, by Theorem 4.1, the ray with angle  $\theta$  intersects  $\Sigma_t$  precisely in the claimed interval.

For Point 3, we have already shown that  $\partial T / \partial r$  is nonzero except when  $r = 1$ . When  $r = 1$ , we know from (4.1) that

$$T(r, \theta) = |\lambda - 1|^2 = 2 - 2 \cos \theta.$$

Thus, when  $r = 1$ , we have  $\partial T / \partial \theta = 2 \sin \theta$ , which is nonzero except when  $\theta = 0$  or  $\theta = \pi$ . Thus, the gradient of  $T(\lambda)$  is nonzero except when  $\lambda = 0$  (where  $T(\lambda)$  is undefined), when  $\lambda = 1$ , and when  $\lambda = -1$ . Since 0 is never in  $\Sigma_t$  and 1 is always in  $\Sigma_t$ , the only possible singular point in the boundary of  $\Sigma_t$  is at  $\lambda = -1$ . Since  $T(r, \theta) = 2 - 2 \cos \pi = 4$  when  $r = 1$  and  $\theta = \pi$ , the point  $\lambda = -1$  belongs to the boundary of  $\Sigma_4$ .

Meanwhile, the Taylor expansion of  $T$  to second order at  $\lambda = -1$  is easily found to be  $T(\lambda) \approx 4 + (\operatorname{Re} \lambda + 1)^2/3 - (\operatorname{Im} \lambda)^2$ . By the Morse lemma, we can then make a smooth change of variables so that in the new coordinate system,

$$T(u, v) = 4 + u^2 - v^2 = 4 + (u + v)(u - v).$$

Thus, near  $\lambda = -1$ , the set  $T(\lambda) = 4$  is the union of the curves  $u + v = 0$  and  $u - v = 0$ .

The invariance of  $\Sigma_t$  under  $\lambda \mapsto 1/\lambda$  and under  $\lambda \mapsto \bar{\lambda}$  follows from the easily verified invariance of  $T(\lambda)$  under these transformations. Finally, we verify that the domain  $\Sigma_t$ , as we have defined it, coincides with the one originally introduced by Biane in [5]. Let us start with the case  $t < 4$ . According to the discussion at the bottom of p. 273 in [5], the boundary of Biane's domain  $\Sigma_t$  consists in this case of two analytic arcs. The interior of one arc lies in the open unit disk and the interior of the other arc lies in the complement of the closed unit disk, while the endpoints of both arcs lie on the unit circle. The first arc is then computed by applying a certain holomorphic function  $\chi(t, \cdot)$  to the support of Biane's measure  $\nu_t$  in the unit circle. Now,  $\chi(t, \cdot)$  satisfies  $f_t(\chi(t, z)) = z$  on the closed unit disk. (Combine the identity involving  $\kappa$  on p. 266 of [5] with the definition of  $\chi$  on p. 273.) We see that the interior of the first arc consists of points with  $|\lambda| \neq 1$  but  $|f_t(\lambda)| = 1$ . This arc must, therefore, coincide with the arc of points with radius  $1/r_t(\theta)$ . The second arc is obtained from the first by the map  $\lambda \mapsto 1/\lambda$  and therefore coincides with the points of radius  $r_t(\theta)$ . We can now see that the boundary of Biane's domain coincides with the boundary of the domain we have defined. A similar analysis applies to the cases  $t > 4$  and  $t = 4$ , using the description of the boundary of  $\Sigma_t$  in those cases at the top of p. 274 in [5].  $\square$

## 5. THE PDE FOR $S$

In this section, we will verify the PDE for  $S$  in Theorem 2.8. The claimed initial condition (2.15) holds because  $b_0 = 1$ . We now proceed to verify the equation (2.14) itself.

Let  $(c_t)_{t \geq 0}$  denote a free circular Brownian motion. The rules of free stochastic calculus, in "stochastic differential" form, are as follows; see [35, Lemma 2.5, Lemma 4.3]. If  $g_t$  and  $h_t$  are processes adapted to  $c_t$ , then

$$dc_t g_t dc_t^* = dc_t^* g_t dc_t = \tau(g_t) dt \quad (5.1)$$

$$dc_t g_t dc_t = dc_t^* g_t dc_t^* = 0 \quad (5.2)$$

$$dc_t dt = dc_t^* dt = 0 \quad (5.3)$$

$$\tau(g_t dc_t h_t) = \tau(g_t dc_t^* h_t) = 0. \quad (5.4)$$

In addition, we have the following Itô product rule: if  $a_t^1, \dots, a_t^n$  are processes adapted to  $c_t$ , then

$$d(a_t^1 \cdots a_t^n) = \sum_{j=1}^n (a_t^1 \cdots a_t^{j-1}) da_t^j (a_t^{j+1} \cdots a_t^n) \quad (5.5)$$

$$+ \sum_{1 \leq j < k \leq n} (a_t^1 \cdots a_t^{j-1}) da_t^j (a_t^{j+1} \cdots a_t^{k-1}) da_t^k (a_t^{k+1} \cdots a_t^n). \quad (5.6)$$

We let  $b_t$  be the free multiplicative Brownian motion, which satisfies the free stochastic differential equation

$$db_t = b_t dc_t, \quad b_0 = 1.$$

Throughout the rest of this section, we will use the notation

$$b_{t,\lambda} := b_t - \lambda.$$

**Lemma 5.1.** *We have*

$$\frac{\partial}{\partial t} \tau[(b_{t,\lambda}^* b_{t,\lambda})^n] = n \sum_{j=0}^{n-1} \tau[(b_{t,\lambda}^* b_{t,\lambda})^j] \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^*)^{n-j-1}]. \quad (5.7)$$

Note that second factor on the right-hand side of (5.7) has  $(b_{t,\lambda} b_{t,\lambda}^*)^{n-j}$ , with the adjoint on the *second* factor.

*Proof.* Note that  $db_{t,\lambda} = db_t = b_t dc_t$  and  $db_{t,\lambda}^* = db_t^* = dc_t^* b_t^*$ . We compute  $d\tau[(b_{t,\lambda}^* b_{t,\lambda})^n]$  by moving the  $d$  inside the trace and then applying the product rule in (5.5) and (5.6). By (5.4), the terms arising from (5.5) will not contribute. Furthermore, by (5.2), the only terms from (5.6) that contribute are those where one  $d$  goes on a factor of  $b_{t,\lambda}$  and one goes on a factor of  $b_{t,\lambda}^*$ .

By choosing all possible factors of  $b_{t,\lambda}$  and all possible factors of  $b_{t,\lambda}^*$ , we get  $n^2$  terms. In each term, after putting the  $d$  inside the trace, we can cyclically permute the factors until, say, the  $db_{t,\lambda}$  factor is at the end. There are then only  $n$  *distinct* terms that occur, each of which occurs  $n$  times. By (5.1), each distinct term is computed as

$$\begin{aligned} & \tau[(b_{t,\lambda}^* b_{t,\lambda})^j dc_t^* b_t^* b_{t,\lambda} (b_{t,\lambda}^* b_{t,\lambda})^{n-j-2} b_{t,\lambda}^* b_t dc_t] \\ &= \tau[b_t^* b_{t,\lambda} (b_{t,\lambda}^* b_{t,\lambda})^{n-j-2} b_{t,\lambda}^* b_t] \tau[(b_{t,\lambda}^* b_{t,\lambda})^j] dt \\ &= \tau[(b_{t,\lambda}^* b_{t,\lambda})^j] \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^*)^{n-j-1}] dt. \end{aligned}$$

(The reader who doubts the validity of using the cyclic invariance of the trace when some factors are differentials may compute each term by *first* using (5.1) and *then* using the cyclic invariance of the trace, with the same result.) Since each distinct term occurs  $n$  times, we obtain

$$d\tau[(b_{t,\lambda}^* b_{t,\lambda})^n] = n \sum_{l=0}^{n-1} \tau[(b_{t,\lambda}^* b_{t,\lambda})^j] \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^*)^{n-j-1}] dt$$

as claimed.  $\square$

**Lemma 5.2.** *The function  $S$  in (2.12) satisfies*

$$\frac{\partial S}{\partial t} = x \tau[(b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}] \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}]. \quad (5.8)$$

Of course, since  $b_t = b_{t,\lambda} + \lambda$ , we can rewrite (5.8) in a way that involves only  $b_{t,\lambda}$  and not  $b_t$ .

*Proof.* We note that the definition of  $S$  actually makes sense for all  $x \in \mathbb{C}$  with  $\operatorname{Re}(x) > 0$ , using the standard branch of the logarithm function. We note that for  $|x| > |z|$ , we have

$$\begin{aligned} \frac{1}{z+x} &= \frac{1}{x(1 - (-\frac{z}{x}))} \\ &= \frac{1}{x} \left[ 1 - \frac{z}{x} + \frac{z^2}{x^2} - \frac{z^3}{x^3} + \dots \right]. \end{aligned} \quad (5.9)$$

Integrating with respect to  $z$  gives

$$\log(z+x) = \log x + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left(\frac{z}{x}\right)^n.$$

Thus, for  $|x| > \|b_t^* b_t\|$ , we have

$$\tau[\log(b_{t,\lambda}^* b_{t,\lambda} + x)] = \log x + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n x^n} \tau[(b_{t,\lambda}^* b_{t,\lambda})^n]. \quad (5.10)$$

Assume for the moment that it is permissible to differentiate (5.10) term by term with respect to  $t$ . Then by Lemma 5.1, we have

$$\frac{\partial S}{\partial t} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{x^n} \sum_{j=0}^{n-1} \tau[(b_{t,\lambda}^* b_{t,\lambda})^j] \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^*)^{n-j-1}]. \quad (5.11)$$

Now, by [7, Proposition 3.2.3], the map  $t \mapsto b_t$  is continuous in the operator norm topology; in particular,  $\|b_t\|$  is a locally bounded function of  $t$ . From this observation, it is easy to see that the right-hand side of (5.11) converges locally uniformly in  $t$ . Thus, a standard result about interchange of limit and derivative (e.g., Theorem 7.17 in [41]) shows that the term-by-term differentiation is valid.

Now, in (5.11), we let  $k = j$  and  $l = n - j - 1$ , so that  $n = k + l + 1$ . Then  $k$  and  $l$  go from 0 to  $\infty$ , and we get

$$\frac{\partial S}{\partial t} = x \left( \frac{1}{x} \sum_{k=0}^{\infty} \frac{(-1)^k}{x^k} \tau[(b_{t,\lambda}^* b_{t,\lambda})^k] \right) \left( \frac{1}{x} \sum_{l=0}^{\infty} \frac{(-1)^l}{x^l} \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^*)^l] \right).$$

(We may check that the power of  $x$  in the denominator is  $k + l + 1 = n$  and that the power of  $-1$  is  $k + l = n - 1$ .) Thus, moving the sums inside the traces and using (5.9), we obtain the claimed form of  $\partial S/\partial t$ .

We have now established the claimed formula for  $\partial S/\partial t$  for  $x$  in the right half-plane, provided  $|x|$  is sufficiently large, depending on  $t$  and  $\lambda$ . Since, also,  $S(0, \lambda, x) = \log(|\lambda - 1|^2 + x)$ , we have, for sufficiently large  $|x|$ ,

$$S(t, \lambda, x) = \log(|\lambda - 1|^2 + x) + \int_0^t x \tau[(b_{s,\lambda}^* b_{s,\lambda} + x)^{-1}] \tau[b_s b_s^* (b_{s,\lambda} b_{s,\lambda}^*)^{-1}] ds. \quad (5.12)$$

We now claim that both sides of (5.12) are well-defined, holomorphic functions of  $x$ , for  $x$  in the right half-plane. This claim is easily established from the standard power-series representation of the inverse:

$$\begin{aligned} (A + x + h)^{-1} &= (A + x)^{-1} (1 + h(A + x)^{-1})^{-1} \\ &= (A + x)^{-1} \sum_{n=0}^{\infty} (-1)^n h^n (A + x)^{-n}, \end{aligned}$$

and a similar power-series representation of the logarithm. Thus, (5.12) actually holds for all  $x$  in the right half-plane. Differentiating with respect to  $t$  then establishes the claimed formula (5.8) for  $dS/dt$  for all  $x$  in the right half-plane.  $\square$

**Lemma 5.3.** *We have the following formulas for the derivatives of  $S$  with respect to  $x$  and  $\lambda$ :*

$$\begin{aligned} \frac{\partial S}{\partial x} &= \tau[(b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}] \\ \frac{\partial S}{\partial \lambda} &= -\tau[b_{t,\lambda}^* (b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}] \\ \frac{\partial S}{\partial \lambda} &= -\tau[b_{t,\lambda} (b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}]. \end{aligned}$$

*Proof.* We use the formula for the derivative of the trace of a logarithm (Lemma 1.1 in [9]):

$$\frac{d}{du}\tau[\log(f(u))] = \tau\left[f(u)^{-1}\frac{df}{du}\right].$$

(We emphasize that there is no such simple formula for the derivative of  $\log(f(u))$  without the trace, unless  $df/du$  commutes with  $f(u)$ .) The lemma easily follows from this formula.  $\square$

We are now ready for the verification of the differential equation for  $S$ .

*Proof of Theorem 2.8.* We note that

$$b_{t,\lambda}(b_{t,\lambda}^*b_{t,\lambda} + x) = (b_{t,\lambda}b_{t,\lambda}^* + x)b_{t,\lambda}.$$

Multiplying by  $(b_{t,\lambda}^*b_{t,\lambda} + x)^{-1}$  on the right and  $(b_{t,\lambda}b_{t,\lambda}^* + x)^{-1}$  on the left gives a useful identity:

$$(b_{t,\lambda}b_{t,\lambda}^* + x)^{-1}b_{t,\lambda} = b_{t,\lambda}(b_{t,\lambda}^*b_{t,\lambda} + x)^{-1}. \quad (5.13)$$

Replacing  $b_{t,\lambda}$  by its adjoint gives another version of the identity:

$$b_{t,\lambda}^*(b_{t,\lambda}b_{t,\lambda}^* + x)^{-1} = (b_{t,\lambda}^*b_{t,\lambda} + x)^{-1}b_{t,\lambda}^*. \quad (5.14)$$

Note that in both (5.13) and (5.14), both side have the same pattern of starred and unstarred variables, always with the two outer variables being the same and the middle one being different.

We also claim that

$$\tau[(b_{t,\lambda}^*b_{t,\lambda} + x)^{-1}] = \tau[(b_{t,\lambda}b_{t,\lambda}^* + x)^{-1}]. \quad (5.15)$$

To verify this identity for large  $x$ , we replace  $z$  by  $b_{t,\lambda}^*b_{t,\lambda}$  in the series (5.9) and note that by the cyclic invariance of the trace,

$$\tau[(b_{t,\lambda}^*b_{t,\lambda})^n] = \tau[(b_{t,\lambda}b_{t,\lambda}^*)^n].$$

The result for general  $x$  follows by an analyticity argument as in the proof of Lemma 5.2.

We start from the formula for  $\partial S/\partial t$  in Lemma 5.2. Noting that

$$\begin{aligned} b_t b_t^* &= (b_{t,\lambda} + \lambda)^*(b_{t,\lambda} + \lambda) \\ &= b_{t,\lambda} b_{t,\lambda}^* + \lambda b_{t,\lambda}^* + \bar{\lambda} b_{t,\lambda} + |\lambda|^2, \end{aligned}$$

we expand the second factor on the right-hand side of (5.8) as

$$\begin{aligned} \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}] &= \tau[b_{t,\lambda} b_{t,\lambda}^* (b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}] \\ &\quad + \lambda \tau[b_{t,\lambda}^* (b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}] \\ &\quad + \bar{\lambda} \tau[b_{t,\lambda} (b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}] \\ &\quad + |\lambda|^2 \tau[(b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}]. \end{aligned}$$

We then simplify the first term by writing  $b_{t,\lambda} b_{t,\lambda}^* = b_{t,\lambda} b_{t,\lambda}^* + x - x$ . In the middle two terms, we use (5.13), (5.14), and cyclic invariance of the trace. Using also (5.15), we get

$$\begin{aligned} \tau[b_t b_t^* (b_{t,\lambda} b_{t,\lambda}^* + x)^{-1}] &= 1 + (|\lambda|^2 - x) \tau[(b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}] \\ &\quad + \lambda \tau[b_{t,\lambda}^* (b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}] \\ &\quad + \bar{\lambda} \tau[b_{t,\lambda} (b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}]. \end{aligned} \quad (5.16)$$

Thus,

$$\frac{\partial S}{\partial t} = x\tau[(b_{t,\lambda}^* b_{t,\lambda} + x)^{-1}](\text{all the terms in (5.16)}). \quad (5.17)$$

All terms on the right-hand side of (5.17) are expressible using Lemma 5.3 in terms of derivatives of  $S$ , and the claimed differential equation follows.  $\square$

## 6. THE HAMILTON–JACOBI METHOD

**6.1. Setting up the method.** The equation (2.14) is a first-order, nonlinear PDE of Hamilton–Jacobi type. (The reader may consult, for example, Section 3.3 in the book of Evans [16], but we will give a brief self-contained account of the theory in the proof of Proposition 6.3.) We consider a Hamiltonian function obtained from the right-hand side of (2.14) by replacing each partial derivative with momentum variable, with an overall minus sign. Thus, we define

$$H(a, b, x, p_a, p_b, p_x) = -xp_x(1 + (a^2 + b^2)p_x - xp_x - ap_a - bp_b). \quad (6.1)$$

We then consider Hamilton’s equations for this Hamiltonian. That is to say, we consider this system of six coupled ODEs:

$$\begin{aligned} \frac{da}{dt} &= \frac{\partial H}{\partial p_a}; & \frac{db}{dt} &= \frac{\partial H}{\partial p_b}; & \frac{dx}{dt} &= \frac{\partial H}{\partial p_x}; \\ \frac{dp_a}{dt} &= -\frac{\partial H}{\partial a}; & \frac{dp_b}{dt} &= -\frac{\partial H}{\partial b}; & \frac{dp_x}{dt} &= -\frac{\partial H}{\partial x}. \end{aligned} \quad (6.2)$$

As convenient, we will let

$$\lambda(t) = a(t) + ib(t).$$

The initial conditions for  $a$ ,  $b$ , and  $x$  are arbitrary:

$$a(0) = a_0; \quad b(0) = b_0; \quad x(0) = x_0, \quad (6.3)$$

while those for  $p_a$ ,  $p_b$ , and  $p_x$  are determined by those for  $a$ ,  $b$ , and  $x$  as follows:

$$p_a(0) = 2(a_0 - 1)p_0; \quad p_b(0) = 2b_0p_0; \quad p_x(0) = p_0, \quad (6.4)$$

where

$$p_0 = \frac{1}{|\lambda_0 - 1|^2 + x_0} = \frac{1}{(a_0 - 1)^2 + b_0^2 + x_0}. \quad (6.5)$$

The motivation for (6.4) is that the momentum variables  $p_a$ ,  $p_b$ , and  $p_x$  will correspond to the derivatives of  $S$  along the curves  $(x(t), a(t), b(t))$ ; see (6.8). Thus, the initial momenta are simply the derivatives of the initial value (2.15) of  $S$ , evaluated at  $(a_0, b_0, x_0)$ .

For future reference, we record the value  $H_0$  of the Hamiltonian at time  $t = 0$ .

**Lemma 6.1.** *The value of the Hamiltonian at  $t = 0$  is*

$$H_0 = -x_0p_0^2. \quad (6.6)$$

*Proof.* Plugging  $t = 0$  into (6.1) and using (6.4) gives

$$H_0 = -x_0p_0(1 + (a_0^2 + b_0^2)p_0 - x_0p_0 - 2a_0(a_0 - 1)p_0 - 2b_0^2p_0),$$

which simplifies to

$$H_0 = -x_0p_0(1 - p_0(a_0^2 - 2a_0 + b_0^2 + x_0)).$$

But using the formula (6.5) for  $p_0$ , we see that  $a_0^2 - 2a_0 + b_0^2 + x_0$  equals  $1/p_0 - 1$ , from which (6.6) follows.  $\square$

The main result of this section is the following; the proof is given on p. 32.

**Theorem 6.2.** *Assume  $\lambda_0 \neq 0$  and  $x_0 > 0$ . Suppose a solution to the system (6.2) with initial conditions (6.3) and (6.4) exists with  $x(t) > 0$  for  $0 \leq t < T$ . Then we have*

$$\begin{aligned} S(t, \lambda(t), x(t)) &= \log(|\lambda_0 - 1|^2 + x_0) - \frac{x_0 t}{(|\lambda_0 - 1|^2 + x_0)^2} \\ &\quad + \log |\lambda(t)| - \log |\lambda_0| \end{aligned} \quad (6.7)$$

for all  $t \in [0, T)$ . Furthermore, the derivatives of  $S$  with respect to  $a$ ,  $b$ , and  $x$  satisfy

$$\begin{aligned} \frac{\partial S}{\partial x}(t, \lambda(t), x(t)) &= p_x(t); \\ \frac{\partial S}{\partial a}(t, \lambda(t), x(t)) &= p_a(t); \\ \frac{\partial S}{\partial b}(t, \lambda(t), x(t)) &= p_b(t). \end{aligned} \quad (6.8)$$

Note that  $S(t, \lambda, x)$  is only defined for  $x > 0$ . Thus, (6.7) and (6.8) only make sense as long as the solution to (6.2) exists with  $x(t) > 0$ .

Since our objective is to compute  $\Delta s_t(\lambda) = \partial^2 s_t / \partial a^2 + \partial^2 s_t / \partial^2 b^2$ , the formula (6.8) for the derivatives of  $S$  will ultimately be of as great importance as the formula (6.7) for  $S$  itself. We emphasize that we are not using the Hamilton–Jacobi method to *construct* a solution to (2.14); the function  $S(t, \lambda, x)$  is already defined in (2.12) in terms of free probability and is known (Theorem 2.8) to satisfy (2.14). Rather, we are using the Hamilton–Jacobi method to *analyze* a solution that is already known to exist.

We begin by briefly recapping the general form of the Hamilton–Jacobi method.

**Proposition 6.3.** *Fix a function  $H(\mathbf{x}, \mathbf{p})$  and consider a function  $S(t, \mathbf{x})$  satisfying*

$$\frac{\partial S}{\partial t} = -H(\mathbf{x}, \nabla_{\mathbf{x}} S). \quad (6.9)$$

Suppose the pair  $(\mathbf{x}(t), \mathbf{p}(t))$  satisfies Hamilton’s equations:

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}(\mathbf{x}(t), \mathbf{p}(t)); \quad \frac{dp_j}{dt} = -\frac{\partial H}{\partial x_j}(\mathbf{x}(t), \mathbf{p}(t))$$

with initial conditions

$$\mathbf{x}(0) = \mathbf{x}_0; \quad \mathbf{p}(0) = (\nabla_{\mathbf{x}} S)(0, \mathbf{x}_0). \quad (6.10)$$

Then we have

$$S(t, \mathbf{x}(t)) = S(0, \mathbf{x}_0) - H(\mathbf{x}_0, \mathbf{p}_0) t + \int_0^t \mathbf{p}(s) \cdot \frac{d\mathbf{x}}{ds} ds \quad (6.11)$$

and

$$(\nabla_{\mathbf{x}} S)(t, \mathbf{x}(t)) = \mathbf{p}(t). \quad (6.12)$$

Again, we are not trying to construct solutions to (6.9), but rather to analyze a solution that is already assumed to exist.

*Proof.* Take an arbitrary (for the moment) smooth curve  $\mathbf{x}(t)$  and note that

$$\begin{aligned} \frac{d}{dt}S(t, \mathbf{x}(t)) &= \frac{\partial S}{\partial t}(t, \mathbf{x}(t)) + \frac{\partial S}{\partial x_j}(t, \mathbf{x}(t)) \frac{dx_j}{dt} \\ &= -H(\mathbf{x}(t), (\nabla_{\mathbf{x}}S)(t, \mathbf{x}(t))) + (\nabla_{\mathbf{x}}S)(t, \mathbf{x}(t)) \cdot \frac{d\mathbf{x}}{dt}, \end{aligned} \quad (6.13)$$

where we use the Einstein summation convention. Let us use the notation

$$\mathbf{p}(t) = (\nabla_{\mathbf{x}}S)(t, \mathbf{x}(t));$$

that is  $p_j(t) = \partial S / \partial x_j(t, \mathbf{x}(t))$ . Then (6.13) may be rewritten as

$$\frac{d}{dt}S(t, \mathbf{x}(t)) = -H(\mathbf{x}(t), \mathbf{p}(t)) + \mathbf{p}(t) \cdot \frac{d\mathbf{x}}{dt}. \quad (6.14)$$

If we can choose  $\mathbf{x}(t)$  so that  $\mathbf{p}(t)$  is somehow computable, then the right-hand side of (6.14) would be known and we could integrate to get  $S(t, \mathbf{x}(t))$ .

To see how we might be able to compute  $\mathbf{p}(t)$ , we try differentiating:

$$\begin{aligned} \frac{dp_j}{dt} &= \frac{d}{dt} \frac{\partial S}{\partial x_j}(t, \mathbf{x}(t)) \\ &= \frac{\partial^2 S}{\partial t \partial x_j}(t, \mathbf{x}(t)) + \frac{\partial^2 S}{\partial x_k \partial x_j}(t, \mathbf{x}(t)) \frac{dx_k}{dt}. \end{aligned} \quad (6.15)$$

Now, from (6.9), we have

$$\begin{aligned} \frac{\partial^2 S}{\partial t \partial x_j} &= \frac{\partial^2 S}{\partial x_j \partial t} \\ &= -\frac{\partial}{\partial x_j} H(\mathbf{x}, \nabla_{\mathbf{x}}S) \\ &= -\frac{\partial H}{\partial x_j}(\mathbf{x}, \nabla_{\mathbf{x}}S) - \frac{\partial H}{\partial p_k}(\mathbf{x}, \nabla_{\mathbf{x}}S) \frac{\partial^2 S}{\partial x_j \partial x_k}. \end{aligned}$$

Thus, (6.15) becomes (suppressing the dependence on the path)

$$\frac{dp_j}{dt} = -\frac{\partial H}{\partial x_j} + \left( \frac{dx_k}{dt} - \frac{\partial H}{\partial p_k} \right) \frac{\partial^2 S}{\partial x_k \partial x_j}. \quad (6.16)$$

If we now take  $\mathbf{x}(t)$  to satisfy

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}, \quad (6.17)$$

the second term on the right-hand side of (6.16) vanishes, and we find that  $\mathbf{p}(t)$  satisfies

$$\frac{dp_j}{dt} = -\frac{\partial H}{\partial x_j}. \quad (6.18)$$

With this choice of  $\mathbf{x}(t)$ , (6.14) becomes

$$\frac{d}{dt}S(t, \mathbf{x}(t)) = -H(\mathbf{x}_0, \mathbf{p}_0) + \mathbf{p}(t) \cdot \frac{d\mathbf{x}}{dt}, \quad (6.19)$$

because  $H$  is constant along the solutions to Hamilton's equations.

Note that not all solutions  $(\mathbf{x}(t), \mathbf{p}(t))$  to Hamilton's equations (6.17) and (6.18) will arise by the above method. After all, we are assuming that  $\mathbf{p}(t) = (\nabla_{\mathbf{x}}S)(t, \mathbf{x}(t))$ , from which it follows that the initial conditions  $(\mathbf{x}_0, \mathbf{p}_0)$  will be of the form in (6.10).

On the other hand, suppose we take a pair  $(\mathbf{x}_0, \mathbf{p}_0)$  as in (6.10). Let us then take  $\mathbf{x}(t)$  to be the solution to

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}(\mathbf{x}(t), (\nabla_{\mathbf{x}} S)(t, \mathbf{x}(t))), \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (6.20)$$

where since  $S$  is a fixed, “known” function, this ODE for  $\mathbf{x}(t)$  will have unique solutions for as long as they exist. If we set  $\mathbf{p}(t) = (\nabla_{\mathbf{x}} S)(t, \mathbf{x}(t))$ , then  $\mathbf{p}(0) = \mathbf{p}_0$  as in (6.10) and (6.20) says that the pair  $(\mathbf{x}(t), \mathbf{p}(t))$  satisfies the first of Hamilton’s equations. Applying (6.16) with this choice of  $\mathbf{x}(t)$  shows that the pair  $(\mathbf{x}(t), \mathbf{p}(t))$  also satisfies the second of Hamilton’s equations. Thus,  $(\mathbf{x}(t), \mathbf{p}(t))$  must be the *unique* solution to Hamilton’s equations with the given initial condition  $(\mathbf{x}_0, \mathbf{p}_0)$ .

We conclude that for *any* solution to Hamilton’s equations with initial conditions of the form (6.10), the formula (6.14) holds. Since, also,  $H$  is constant along solutions to Hamilton’s equations, we may replace  $H(\mathbf{x}(t), \mathbf{p}(t))$  by  $H(\mathbf{x}_0, \mathbf{p}_0)$  in (6.14), at which point, integration with respect to  $t$  gives (6.11). Finally, (6.12) holds by the definition of  $\mathbf{p}(t)$ .  $\square$

We are now ready for the proof of Theorem 6.2.

*Proof of Theorem 6.2.* The PDE (2.14) is of the type in (6.9), with  $H$  given by (6.1). The initial conditions (6.4) are obtained by differentiating the initial condition  $S(0, \lambda, x) = \log(|\lambda - 1|^2 + x)$ .

We let  $\mathbf{x}(t) = (a(t), b(t), x(t))$  and  $\mathbf{p}(t) = (p_a(t), p_b(t), p_x(t))$ . For the case of the Hamiltonian (6.1), a simple computation shows that

$$\begin{aligned} \mathbf{p} \cdot \frac{d\mathbf{x}}{dt} &= \mathbf{p} \cdot \nabla_{\mathbf{p}} H \\ &= 2H + xp_x \\ &= 2H_0 + xp_x. \end{aligned}$$

Thus, the general formula (6.11) becomes, in this case,

$$S(t, \mathbf{x}(t)) = S(0, \mathbf{x}_0) + H(\mathbf{x}_0, \mathbf{p}_0) t + \int_0^t x(s)p_x(s) ds. \quad (6.21)$$

But we also may compute that

$$\begin{aligned} \frac{d}{dt} \log(\sqrt{a^2 + b^2}) &= \frac{1}{2} \frac{1}{a^2 + b^2} (a\dot{a} + b\dot{b}) \\ &= \frac{1}{2} \frac{1}{a^2 + b^2} \left( a \frac{\partial H}{\partial p_a} + b \frac{\partial H}{\partial p_b} \right) \\ &= xp_x. \end{aligned}$$

Thus,

$$\int_0^t x(s)p_x(s) ds = \log |\lambda(t)| - \log |\lambda_0|. \quad (6.22)$$

If we now plug in the value of  $S(0, \mathbf{x}_0) = S(0, \lambda_0, x_0)$  and use Lemma 6.1 along with the definition (6.5) of  $p_0$ , we obtain (6.7). Finally, (6.8) is just the general formula (6.12), applied to the case at hand.  $\square$

**6.2. Constants of motion.** We now identify several constants of motion for the system (6.2), from which various useful formulas can be derived. Throughout the section, we assume we have a solution to (6.2) with the initial conditions (6.3) and (6.4), defined on a time-interval of the form  $0 \leq t < T$ . We continue the notation  $\lambda(t) = a(t) + ib(t)$ .

**Proposition 6.4.** *Along any solution of (6.2), following quantities remain constant:*

- (1) *The Hamiltonian  $H$ ,*
- (2) *The “angular momentum” in the  $(a, b)$  variables, namely  $ap_b - bp_a$ , and*
- (3) *The argument of  $\lambda$ , assuming  $\lambda_0 \neq 0$ .*

*Proof.* For any system of the form (6.2), the Hamiltonian  $H$  itself is a constant of motion, as may be verified easily from the equations. The conservation of the angular momentum is a consequence of the invariance of  $H$  under simultaneous rotations of  $(a, b)$  and  $(p_a, p_b)$ ; see Proposition 2.30 and Conclusion 2.31 in [30]. This result can also be verified by direct computation from (6.2).

Finally, note from (6.22) that if  $\lambda_0 \neq 0$ , then  $\log |\lambda(t)|$  remains finite as long as the solution to (6.2) exists, so that  $\lambda(t)$  cannot pass through the origin. We then compute that

$$\frac{d}{dt} \tan(\arg \lambda(t)) = \frac{d}{dt} \frac{b}{a} = \frac{\dot{b}a - b\dot{a}}{a^2} = \frac{xp_x b a - bxp_x a}{a^2} = 0.$$

(If  $a = 0$ , we instead compute the time-derivative of  $\cot(\arg \lambda)$ , which also equals zero.)  $\square$

**Proposition 6.5.** *The Hamiltonian  $H$  in (6.1) is invariant under the one-parameter group of symplectic linear transformations given by*

$$(a, b, x, p_a, p_b, p_x) \mapsto (e^{\sigma/2}a, e^{\sigma/2}b, e^{\sigma}x, e^{-\sigma/2}p_a, e^{-\sigma/2}p_b, e^{-\sigma}p_x), \quad (6.23)$$

with  $\sigma$  varying over  $\mathbb{R}$ . Thus, the generator of this family of transformations, namely,

$$\Psi := xp_x + \frac{1}{2}(ap_a + bp_b) \quad (6.24)$$

is a constant of motion for the system (6.2). The constant  $\Psi$  may be computed in terms of  $x_0$  and  $\lambda_0$  as

$$\Psi = p_0(a_0(a_0 - 1) + b_0^2 + x_0) \quad (6.25)$$

where  $p_0$  is as in (6.5).

*Proof.* The claimed invariance of  $H$  is easily checked from the formula (6.1). One can easily check that  $\Psi$  is the generator of this family. That is to say, if we replace  $H$  by  $\Psi$  in (6.2), the solution is given by the map in (6.23). Thus, by a simple general result,  $\Psi$  will be a constant of motion; see Conclusion 2.31 in [30]. Of course, one can also check by direct computation that the function in (6.24) is constant along solutions to (6.2). The expression (6.25) then follows easily from the initial conditions in (6.4).  $\square$

**Proposition 6.6.** *For all  $t$ , we have*

$$x(t)p_x(t)^2 = x_0p_0^2e^{-Ct}, \quad (6.26)$$

where  $C = 2\Psi - 1$  and  $\Psi$  is as in (6.24). The constant  $C$  in (6.26) may be computed in terms of  $x_0$  and  $\lambda_0$  as

$$C = p_0(|\lambda_0|^2 - 1 + x_0) = \frac{|\lambda_0|^2 - 1 + x_0}{|\lambda_0 - 1|^2 + x_0}. \quad (6.27)$$

*Proof.* We compute that

$$\begin{aligned} \dot{x} &= \frac{\partial H}{\partial p_x} = \frac{H}{p_x} - xp_x(a^2 + b^2 - x) \\ \dot{p}_x &= -\frac{\partial H}{\partial x} = -\frac{H}{x} - xp_x^2 \end{aligned} \quad (6.28)$$

and then that

$$\begin{aligned} \frac{d}{dt}(xp_x^2) &= \dot{x}p_x^2 + 2xp_x\dot{p}_x \\ &= p_x H - xp_x^3(a^2 + b^2 - x) - 2Hp_x - 2x^2p_x^3 \end{aligned}$$

This result simplifies to

$$\begin{aligned} \frac{d}{dt}(xp_x^2) &= xp_x^2 \left[ 1 - 2 \left( xp_x + \frac{1}{2}(ap_a + bp_b) \right) \right] \\ &= -xp_x^2(2\Psi - 1). \end{aligned}$$

The unique solution to this equation is (6.26). The expression (6.27) is obtained by evaluating  $\Psi$  at  $t = 0$ , using the initial conditions (6.4), and simplifying.  $\square$

We now make an important application of preceding results.

**Theorem 6.7.** *Suppose a solution to (6.2) exists with  $x(t) > 0$  for  $0 \leq t < t_*$ , but that  $\lim_{t \rightarrow t_*} x(t) = 0$ . Then*

$$\lim_{t \rightarrow t_*} \log |\lambda(t)| = \frac{Ct_*}{2}, \quad (6.29)$$

where  $C = 2\Psi - 1$  is as in Proposition 6.6. Furthermore, we have

$$\lim_{t \rightarrow t_*} (ap_a + bp_b) = \lim_{t \rightarrow t_*} \frac{2 \log |\lambda(t)|}{t} + 1. \quad (6.30)$$

Equation (6.30) is a key step in the derivation of our main result; see Section 7.1. We will write (6.29) in a more explicit way in Proposition 6.12, after the time  $t_*$  has been determined. We note also from Proposition 6.6 that since  $x(t)$  approaches zero as  $t$  approaches  $t_*$ , then  $p_x(t)$  must be blowing up, so that  $x(t)p_x(t)^2$  can remain positive in this limit.

*Proof.* Using the constant of motion  $\Psi$  in (6.24), we can rewrite the Hamiltonian  $H$  as

$$H = -xp_x(1 + (a^2 + b^2)p_x - 2\Psi + xp_x). \quad (6.31)$$

Now, by assumption, the variable  $x$  approaches zero as  $t$  approaches  $t_*$ . Furthermore, by Proposition 6.6,  $xp_x^2$  remains finite in this limit, so that  $xp_x = \sqrt{x}\sqrt{xp_x^2}$  tends to zero. Thus, in the  $t \rightarrow t_*$  limit, the  $xp_x$  terms in (6.31) vanish while  $xp_x^2$  remains finite, leaving us with

$$H = -\lim_{t \rightarrow t_*} xp_x^2(a^2 + b^2).$$

Since  $H$  is a constant of motion, we may write this result as

$$\lim_{t \rightarrow t_*} (a^2 + b^2) = - \lim_{t \rightarrow t_*} \frac{H_0}{xp_{x^2}} = \lim_{t \rightarrow t_*} \frac{x_0 p_0^2}{xp_x^2} = e^{Ct_*} \quad (6.32)$$

where we have used Lemma 6.1 in the second equality and Proposition 6.6 in the third. The formula (6.29) follows.

Meanwhile, as  $t$  approaches  $t_*$ , the  $xp_x$  term in the formula (6.24) for  $\Psi$  vanishes and we find, using (6.29), that

$$\lim_{t \rightarrow t_*} (ap_a + bp_b) = 2\Psi = C + 1 = \lim_{t \rightarrow t_*} \frac{2 \log |\lambda(t)|}{t} + 1,$$

as claimed in (6.30), where we have used (6.29) in the last equality.  $\square$

**6.3. Solving the equations.** We now solve the system (6.2) subject to the initial conditions (6.3) and (6.4). The formula in Proposition 6.6 for  $x(t)p_x(t)^2$  will be a key tool. Although we are mainly interested in the case  $x_0 > 0$ , we will need in Section 7.2 to allow  $x_0$  to be slightly negative.

We begin by with the following elementary lemma.

**Lemma 6.8.** *Consider a number  $a^2 \in \mathbb{R}$  and let  $a$  be either of the two square roots of  $a^2$ . Then the solution to the equation*

$$\dot{y} = y^2 - a^2 \quad (6.33)$$

subject to the initial condition  $y(0) = y_0 > 0$  is

$$y(t) = \frac{y_0 \cosh(at) - a \sinh(at)}{\cosh(at) - y_0 \frac{\sinh(at)}{a}} \quad (6.34)$$

If  $a^2 \geq y_0^2$ , the solution exists for all  $t > 0$ . If  $a^2 < y_0^2$ , then  $y(t)$  is a strictly increasing function of  $t$  until the first positive time  $t_*$  at which the solution blows up. This time is given by

$$t_* = \frac{1}{a} \tanh^{-1} \left( \frac{a}{y_0} \right) \quad (6.35)$$

$$= \frac{1}{2a} \log \left( \frac{1 + a/y_0}{1 - a/y_0} \right). \quad (6.36)$$

Here, we use the principal branch of the inverse hyperbolic tangent, with branch cuts  $(-\infty, -1]$  and  $[1, \infty)$  on the real axes, which corresponds to using the principal branch of the logarithm. When  $a = 0$ , we interpret the right-hand side of (6.35) or (6.36) as having its limiting value as  $a$  approaches zero, namely  $1/y_0$ .

In passing from (6.35) to (6.36), we have used the standard formula for the inverse hyperbolic tangent,

$$\tanh^{-1}(x) = \frac{1}{2} \log \left( \frac{1+x}{1-x} \right). \quad (6.37)$$

In (6.34), we interpret  $\sinh(at)/a$  as having the value  $t$  when  $a = 0$ . If  $a^2 < 0$ , so that  $a$  is pure imaginary, one can rewrite the solution in terms of ordinary trigonometric functions, using the identities  $\cosh(i\alpha) = \cos \alpha$  and  $\sinh(i\alpha) = i \sin \alpha$ . For each fixed  $t$ , the solution is an even analytic function of  $a$  and therefore an analytic function of  $a^2$ .

*Proof.* If  $a$  is nonzero and real, we may integrate (6.33) to obtain

$$\begin{aligned} t &= \frac{1}{2a} \int_0^t \left( \frac{1}{y(\tau) - a} - \frac{1}{y(\tau) + a} \right) \dot{y}(\tau) \, d\tau \\ &= \frac{1}{2a} \log \left( \frac{y(\tau) - a}{y(\tau) + a} \right) \Big|_{\tau=0}^t \\ &= \frac{1}{2a} \log \left( \frac{y(t) - a y_0 + a}{y(t) + a y_0 - a} \right). \end{aligned}$$

It is then straightforward to solve for  $y(t)$  and simplify to obtain (6.34). Similar computations give the result when  $a$  is zero (recalling that we interpret  $\sinh(at)/a$  as equaling  $t$  when  $a = 0$ ) and when  $a$  is nonzero and pure imaginary. Alternatively, one may check by direct computation that the function on the right-hand side of (6.34) satisfies the equation (6.33) for all  $a \in \mathbb{C}$ .

Now, if  $a^2 \geq y_0^2 > 0$ , the denominator in (6.34) is easily seen to be nonzero for all  $t$  and there is no singularity. If  $a^2$  is positive but less than  $y_0^2$ , the denominator remains positive until it becomes zero when  $\tanh(at) = a/y_0$ . If  $a^2$  is negative, so that  $a = i\alpha$  for some nonzero  $\alpha \in \mathbb{R}$ , we write the solution using ordinary trigonometric functions as

$$y(t) = y_0 \frac{\cos(\alpha t) + \frac{\alpha}{y_0} \sin(\alpha t)}{\cos(\alpha t) - \frac{y_0}{\alpha} \sin(\alpha t)}. \quad (6.38)$$

The denominator in (6.38) becomes zero at  $\alpha t = \tan^{-1}(\alpha/y_0) < \pi/2$ . Finally, if  $a^2 = 0$ , the solution is  $y(t) = y_0/(1 - y_0 t)$ , which blows up at  $t = 1/y_0$ .

It is then not hard to check that for all cases with  $a^2 < y_0^2$ , the blow-up time can be computed as  $t_* = \frac{1}{a} \tanh^{-1}(a/y_0)$ , where we use the principal branch of the inverse hyperbolic tangent, with branch cuts  $(-\infty, -1]$  and  $[1, \infty)$  on the real axis. (At  $a = 0$  we have a removable singularity with a value of  $1/y_0$ .) This recipe corresponds to using the principal branch of the logarithm in the last expression in (6.36).  $\square$

We now apply Lemma 6.8 to compute the  $p_x$ -component of the solution to (6.2). We use the following notations, some of which have been introduced previously:

$$p_0 = \frac{1}{|\lambda_0 - 1|^2 + x_0} \quad (6.39)$$

$$\delta = \frac{|\lambda_0|^2 + 1 + x_0}{|\lambda_0|} \quad (6.40)$$

$$C = 2\Psi - 1 = p_0(|\lambda_0|^2 - 1 + x_0) \quad (6.41)$$

$$y_0 = p_0 + \frac{C}{2} = \frac{1}{2} p_0 |\lambda_0| \delta \quad (6.42)$$

$$a^2 = C^2/4 + x_0 p_0^2. \quad (6.43)$$

We now make the following standing assumptions:

$$\begin{aligned} \lambda_0 &\neq 0 \\ p_0 &> 0 \\ \delta &> 0. \end{aligned} \quad (6.44)$$

We note that under these assumptions,  $y_0$  is positive. Furthermore, we may compute that

$$a = \frac{1}{2}p_0 |\lambda_0| \sqrt{\delta^2 - 4}. \quad (6.45)$$

from which we obtain

$$\frac{a^2}{y_0^2} = \frac{\delta^2 - 4}{\delta^2} < 1, \quad (6.46)$$

so that  $a^2 < y_0^2$ . Now, the assumptions  $p_0 > 0$  and  $\delta > 0$  can be written as  $x_0 > -|\lambda_0 - 1|^2$  and  $x_0 > -(1 + |\lambda_0|^2)$ . Thus, for  $\lambda_0 \neq 0$ , the assumptions (6.44) are always satisfied if  $x_0 > 0$ . Furthermore, except when  $\lambda_0 = 1$ , some negative values of  $x_0$  are allowed.

**Proposition 6.9.** *Under the assumptions (6.44), the  $p_x$ -component of the solution to (6.2) subject to the initial conditions (6.3) and (6.4) is given by*

$$p_x(t) = p_0 \frac{\cosh(at) + \frac{2|\lambda_0| - \delta}{\sqrt{\delta^2 - 4}} \sinh(at)}{\cosh(at) - \frac{\delta}{\sqrt{\delta^2 - 4}} \sinh(at)} e^{-Ct} \quad (6.47)$$

for as long as the solution to the system (6.2) exists. Here we write  $a$  as in (6.45) and we use the same choice of  $\sqrt{\delta^2 - 4}$  in the computation of  $a$  as in the two times  $\sqrt{\delta^2 - 4}$  appears explicitly in (6.47). If  $\delta = 2$ , we interpret  $\sinh(at)/\sqrt{\delta^2 - 4}$  as equaling  $\frac{1}{2}p_0 |\lambda_0| t$ .

If  $x_0 \geq 0$ , the numerator in the fraction on the right-hand side of (6.47) is positive for all  $t$ . Hence when  $x_0 \geq 0$ , we see that  $p_x(t)$  is positive for as long as the solution exists and  $1/p_x(t)$  extends to a real-analytic function of  $t$  defined for all  $t \in \mathbb{R}$ .

The first time  $t_*(\lambda_0, x_0)$  at which the expression on the right-hand side of (6.47) blows up is

$$t_*(\lambda_0, x_0) = \frac{2(\delta - 2 \cos \theta_0)}{\sqrt{\delta^2 - 4}} \tanh^{-1} \left( \frac{\sqrt{\delta^2 - 4}}{\delta} \right) \quad (6.48)$$

$$= \frac{\delta - 2 \cos \theta_0}{\sqrt{\delta^2 - 4}} \log \left( \frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}} \right), \quad (6.49)$$

where  $\theta_0 = \arg \lambda_0$  and  $\sqrt{\delta^2 - 4}$  is either of the two square roots of  $\delta^2 - 4$ . The principal branch of the inverse hyperbolic tangent should be used in (6.48), with branch cuts  $(-\infty, -1]$  and  $[1, \infty)$  on the real axis, which corresponds to using the principal branch of the logarithm in (6.49). When  $\delta = 2$ , we interpret  $t_*(\lambda_0, x_0)$  as having its limiting value as  $\delta$  approaches 2, namely  $\delta - 2 \cos \theta_0$ .

Note that the expression

$$\frac{1}{a} \tanh^{-1} \left( \frac{a}{b} \right)$$

is an even function of  $a$  with  $b$  fixed, with a removable singularity at  $a = 0$ . This expression is therefore an analytic function of  $a^2$  near the origin. In particular, the value of  $t_*(\lambda_0, x_0)$  does not depend on the choice of square root of  $\delta^2 - 4$ .

*Proof of Proposition 6.9.* We assume at first that  $x_0 \neq 0$ . We recall from Proposition 6.6 that  $x(t)p_x(t)^2$  is equal to  $x_0 p_0^2 e^{-Ct}$ , which is never zero, since we assume  $x_0$  is nonzero and  $p_0$  is positive. Thus, as long as the solution to the system (6.2)

exists, both  $x(t)$  and  $p_x(t)$  must be nonzero—and must have the same signs they had at  $t = 0$ . Using (6.28) and the fact that  $H$  is a constant of motion, we obtain

$$\dot{p}_x(t) = \frac{x_0 p_0^2}{x(t)} - x_0 p_0^2 e^{-Ct}$$

But  $x_0 p_0^2/x(t) = p_x(t)^2 e^{Ct}$  and we obtain

$$\dot{p}_x(t) = p_x(t)^2 e^{Ct} - x_0 p_0^2 e^{-Ct}.$$

Then if  $y(t) = e^{Ct} p_x(t) + C/2$ , we find that  $y$  satisfies (6.33). Thus, we obtain  $p_x(t) = (y(t) - C/2)e^{-Ct}$ , where  $y(t)$  is as in (6.34), which simplifies to the claimed formula for  $p_x$ . The same formula holds for  $x_0 = 0$ , by the continuous dependence of the solutions on initial conditions. (It is also possible to solve the system (6.2) with  $x_0 = 0$  by postulating that  $x(t)$  is identically zero and working out the equations for the other variables.)

In this paragraph only, we assume  $x_0 \geq 0$ . Then  $a^2 \geq 0$ , with  $a = 0$  occurring only if  $x_0 = 0$  and  $|\lambda_0| = 1$ , so that  $\delta = 2$ . In that case, the numerator on the right-hand side of (6.47) is identically equal to 1. If  $a^2 > 0$ , then the numerator will always be positive provided that

$$\left( \frac{2|\lambda_0| - \delta}{\sqrt{\delta^2 - 4}} \right)^2 \leq 1,$$

which is equivalent to

$$(\delta^2 - 4) - (2|\lambda_0| - \delta)^2 \geq 0.$$

But a computation shows that

$$(\delta^2 - 4) - (2|\lambda_0| - \delta)^2 = 4x_0, \quad (6.50)$$

and we are assuming  $x_0 \geq 0$ . Now, since the numerator in (6.47) is always positive, we conclude that  $p_x$  remains positive until it blows up.

For any value of  $x_0$ , the blow-up time for the function on the right-hand side of (6.47) is computed by plugging the expression (6.46) for  $a/y_0$  into the formula (6.36), giving

$$\begin{aligned} t_*(\lambda_0, x_0) &= \frac{1}{y_0} \frac{1}{a/y_0} \tanh^{-1} \left( \frac{a}{y_0} \right) \\ &= \frac{2}{p_0 |\lambda_0| \delta} \frac{\delta}{\sqrt{\delta^2 - 4}} \tanh^{-1} \left( \frac{\sqrt{\delta^2 - 4}}{\delta} \right). \end{aligned}$$

After computing that

$$\frac{1}{p_0 |\lambda_0|} = \frac{|\lambda_0 - 1|^2 + x_0}{|\lambda_0|} = \delta - 2 \cos \theta_0,$$

we obtain the claimed formula (6.48) for  $t_*(\lambda_0, x_0)$ .  $\square$

**Remark 6.10.** *If  $x_0 < 0$ , then numerator on the right-hand side of (6.47) can become zero. The time  $\sigma$  at which this happens is computed using (6.45) and (6.50) as*

$$\sigma = \frac{2}{p_0 |\lambda_0| \sqrt{\delta^2 - 4}} \tanh^{-1} \left( - \left( 1 + \frac{4x_0}{(2|\lambda_0| - \delta)^2} \right)^{1/2} \right).$$

*By considering separately the cases  $|\lambda_0| \neq 1$  and  $|\lambda_0| = 1$ , we can verify that  $\sigma$  tends to infinity, locally uniformly in  $\lambda_0$ , as  $x_0$  tends to zero from below. Thus,*

for small negative values of  $x_0$ , the function on the right-hand side of (6.47) will remain positive until the time  $t_*(\lambda_0, x_0)$  at which it blows up.

We now show that the whole system (6.2) has a solution up to the time at which the function on the right-hand side of (6.47) blows up.

**Proposition 6.11.** *Assume that  $x_0$  and  $\lambda_0$  satisfy the assumptions (6.44). Assume further that if  $x_0 < 0$ , then  $|x_0|$  is sufficiently small that  $p_x$  remains positive until it blows up, as in Remark 6.10. Then the solution to the system (6.2) exists up to the time  $t_*(\lambda_0, x_0)$  in Proposition 6.9.*

For any  $x_0$ , we have

$$\lim_{t \rightarrow t_*(\lambda_0, x_0)} x(t) = 0. \quad (6.51)$$

If  $x_0 = 0$ , the solution has  $x(t) \equiv 0$  and  $\lambda(t) \equiv \lambda_0$ .

*Proof.* Let  $T$  be the maximum time such that the solution to (6.2) exists on  $[0, T)$ . We now compute formulas for the solution on this interval. Recall from Proposition 6.9 that if  $x_0 \geq 0$ , then  $p_x(t)$  remains positive for as long as the solution exists; by Remark 6.10, the same assertion holds if  $x_0$  is small and negative.

Now, since  $x p_x^2 = x_0 p_0^2 e^{-Ct}$ , we see that

$$x(t) = \frac{1}{p_x(t)^2} x_0 p_0^2 e^{-Ct}. \quad (6.52)$$

Since  $p_x(t)$  remains positive until it blows up,  $x(t)$  remains bounded until time  $t_*(\lambda_0, x_0)$ , at which time  $x(t)$  approaches zero, as claimed in (6.51). We recall from Proposition 6.4 that the argument of  $\lambda(t)$  remains constant. Then as in shown in (6.22), we have

$$\log |\lambda(t)| = \log |\lambda_0| + \int_0^t x(s) p_x(s) ds. \quad (6.53)$$

Finally,

$$\frac{dp_a}{dt} = -\frac{\partial H}{\partial a} = -2axp_x^2 + xp_x p_a \quad (6.54)$$

which is a first-order, linear equation for  $p_a$ , which can be solved using an integrating factor. A similar calculation applies to  $p_b$ .

Suppose now that the existence time  $T$  of the whole system were smaller than the time  $t_*(\lambda_0, x_0)$  at which the right-hand side of (6.47) blows up. Then from the formulas (6.52), (6.53), and (6.54), we see that all functions involved would remain bounded up to time  $T$ . But then by a standard result,  $T$  could not actually be the maximal time. The solution to the system (6.2) must therefore exist all the way up to time  $t_*(\lambda_0, x_0)$ .

Finally, we note that when  $x_0 = 0$ , (6.52) gives  $x(t) \equiv 0$  and (6.53) gives  $|\lambda(t)| \equiv |\lambda_0|$ . Since also the argument of  $\lambda(t)$  is constant, we see that  $\lambda(t) \equiv \lambda_0$ .  $\square$

**6.4. More about the lifetime of the solution.** In light of Propositions 6.9 and 6.11, the lifetime of the solution to the system (6.2) is  $t_*(\lambda_0, x_0)$ , as computed in (6.48) or (6.49). In this subsection, we (1) analyze the behavior of  $\log |\lambda(t)|$  as  $t$  approaches  $t_*(\lambda_0, x_0)$ , (2) analyze the behavior of  $t_*(\lambda_0, x_0)$  as  $x_0$  approaches zero, and (3) show that  $t_*(\lambda_0, x_0)$  is an increasing function of  $x_0$  with  $\lambda_0$  fixed.

**Proposition 6.12.** *Assume that  $x_0$  and  $\lambda_0$  satisfy the assumptions (6.44). Then*

$$\lim_{t \rightarrow t_*(\lambda_0, x_0)} \log |\lambda(t)| = \frac{\delta - 2/|\lambda_0|}{2\sqrt{\delta^2 - 4}} \log \left( \frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}} \right), \quad (6.55)$$

where  $\delta$  is as in (6.40).

Notice that there is a strong similarity between the formula (6.49) for  $t_*(\lambda_0, x_0)$  and the expression on the right-hand side of (6.55).

*Proof.* By (6.29) in Theorem 6.7, we have  $\lim_{t \rightarrow t_*(\lambda_0, x_0)} \log |\lambda(t)| = Ct_*(\lambda_0, x_0)/2$ , where by (6.27),

$$C = \frac{(|\lambda_0|^2 - 1 + x_0)/|\lambda_0|}{(|\lambda_0 - 1|^2 + x_0)/|\lambda_0|} = \frac{\delta - 2/|\lambda_0|}{\delta - 2 \cos \theta_0}.$$

From this result and the second expression (6.49) for  $t_*(\lambda_0, x_0)$ , (6.55) follows easily.  $\square$

**Proposition 6.13.** *If  $t_*(\lambda_0, x_0)$  is defined by (6.49), then for all nonzero  $\lambda_0$  we have*

$$t_*(\lambda_0, 0) = T(\lambda_0), \quad (6.56)$$

where the function  $T$  is defined in (4.1). Furthermore, when  $x_0 = 0$ , we have

$$\lim_{t \rightarrow t_*(\lambda_0, x_0)} \log |\lambda(t)| = \log |\lambda_0|. \quad (6.57)$$

Recall that the formula for  $t_*(\lambda_0, x_0)$  is defined under the standing assumptions in (6.44). Note that for all  $\lambda_0 \neq 0$ , the value  $x_0 = 0$  satisfies these assumptions.

Since  $\log(x)/(x-1) \rightarrow 1$  as  $x \rightarrow 1$ , we see that  $t_*(\lambda_0, 0)$  is a continuous function of  $\lambda_0 \in \mathbb{C}^*$ . Comparing the formula for  $t_*(\lambda_0, 0)$  to Theorem 4.1, we have the following consequence.

**Corollary 6.14.** *For  $\lambda_0 \in \Sigma_t$ , we have  $t_*(\lambda_0, 0) < t$ , while for  $\lambda_0 \in \partial\Sigma_t$ , we have  $t_*(\lambda_0, 0) = t$ , and for  $\lambda_0 \notin \bar{\Sigma}_t$ , we have  $t_*(\lambda_0, 0) > t$ .*

*Proof of Proposition 6.13.* In the limit as  $x_0 \rightarrow 0$ , we have

$$\delta = \frac{|\lambda_0|^2 + 1}{|\lambda_0|},$$

and

$$\delta^2 - 4 = \left( \frac{|\lambda_0|^2 - 1}{|\lambda_0|} \right)^2,$$

so that

$$\sqrt{\delta^2 - 4} = \pm \frac{|\lambda_0|^2 - 1}{|\lambda_0|}. \quad (6.58)$$

In the case  $|\lambda_0| = 1$ , the limiting value of  $\delta$  is 2. Making use of the elementary limit

$$\lim_{\delta \rightarrow 2^+} \frac{1}{\sqrt{\delta^2 - 4}} \log \left( \frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}} \right) = 1. \quad (6.59)$$

Thus, using (6.49), we obtain in this case,

$$\lim_{x_0 \rightarrow 0} t_*(\lambda_0, x_0) = 2 - 2 \cos \theta_0 = |\lambda_0 - 1|^2, \quad |\lambda_0| = 1,$$

which agrees with the value of  $T(\lambda_0)$  when  $|\lambda_0| = 1$ .

In the case  $|\lambda_0| \neq 1$ , we note that the quantity  $(1/b) \log((a+b)/(a-b))$  is an even function of  $b$  with  $a$  fixed. We may therefore choose the plus sign on the right-hand side of (6.58), regardless of the sign of  $|\lambda_0|^2 - 1$ . We then obtain, using (6.49),

$$\begin{aligned} \lim_{x_0 \rightarrow 0} t_*(\lambda_0, x_0) &= \frac{(|\lambda_0|^2 + 1)/|\lambda_0| - 2 \cos \theta_0}{(|\lambda_0|^2 - 1)/|\lambda_0|} \log \left( \frac{2|\lambda_0|^2/|\lambda_0|}{2/|\lambda_0|} \right) \\ &= \frac{|\lambda_0|^2 + 1 - 2|\lambda_0| \cos \theta_0}{|\lambda_0|^2 - 1} \log(|\lambda_0|^2) \\ &= T(\lambda_0). \end{aligned} \tag{6.60}$$

A similar calculation, beginning from (6.55), establishes (6.57).  $\square$

**Remark 6.15.** *If we began with (6.48) instead of (6.49), we would obtain by similar reasoning*

$$t_*(\lambda_0, 0) = \frac{2|\lambda_0 - 1|^2}{|\lambda_0|^2 - 1} \tanh^{-1} \left( \frac{|\lambda_0|^2 - 1}{|\lambda_0|^2 + 1} \right).$$

Using (6.37), this expression is easily seen to agree with  $T(\lambda_0)$  but is more transparent in its behavior at  $|\lambda_0| = 1$ .

**Proposition 6.16.** *For each  $\lambda_0$ , the function  $t_*(\lambda_0, x_0)$  is a strictly increasing function of  $x_0$  for  $x_0 \geq 0$ , and*

$$\lim_{x_0 \rightarrow +\infty} t_*(\lambda_0, x_0) = +\infty.$$

*Proof.* We note that the quantity  $\delta$  in (6.40) is an increasing function of  $x_0$  with  $\lambda_0$  fixed, with  $\delta$  tending to infinity as  $x_0$  tends to infinity. We note also that if  $x_0 \geq 0$ , then

$$\delta \geq |\lambda_0| + \frac{1}{|\lambda_0|} \geq 2.$$

It therefore suffices to show that for each angle  $\theta_0$ , the function

$$g_{\theta_0}(\delta) := \frac{\delta - 2 \cos \theta_0}{\sqrt{\delta^2 - 4}} \log \left( \frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}} \right), \tag{6.61}$$

is strictly increasing, non-negative, continuous function of  $\delta$  for  $\delta \geq 2$  that tends to  $+\infty$  as  $\delta$  tends to infinity. Here when  $\delta = 2$ , we interpret  $g_{\theta_0}(\delta)$  as having the value  $2 - 2 \cos \theta_0$ , in accordance with the limit (6.59).

Throughout the proof, we use the notation

$$\gamma = \sqrt{\delta^2 - 4}.$$

We note that

$$\lim_{\delta \rightarrow \infty} \frac{\delta - 2 \cos \theta_0}{\gamma} = 1.$$

Meanwhile, for large  $\delta$ , we have

$$\begin{aligned}\delta - \gamma &= \delta \left(1 - \sqrt{1 - 4/\delta^2}\right) \\ &= \delta \left(1 - \left(1 - \frac{1}{2} \frac{4}{\delta^2} + O\left(\frac{1}{\delta^3}\right)\right)\right) \\ &= \frac{2}{\delta} + O\left(\frac{1}{\delta^3}\right),\end{aligned}$$

whereas

$$\delta + \gamma = 2\delta + O\left(\frac{1}{\delta}\right).$$

Thus,  $g_{\theta_0}(\delta)$  grows like  $\log(\delta^2)$  as  $\delta \rightarrow \infty$ .

Our definition of  $g_{\theta_0}(\delta)$  for  $\delta = 2$ , together with (6.59), shows that  $g_{\theta_0}$  is non-negative and continuous there. To show that  $g_{\theta_0}$  is an increasing function of  $\delta$ , we show that  $\partial g_{\theta_0}/\partial \delta$  is positive for  $\delta > 2$ . The derivative is computed, after simplification, as

$$\frac{\partial g_{\theta_0}}{\partial \delta} = \frac{2}{\gamma^3} \left( (\delta - 2 \cos \theta_0) \gamma + (\delta \cos \theta_0 - 2) \log \left( \frac{\delta + \gamma}{\delta - \gamma} \right) \right).$$

Since this expression depends linearly on  $\cos \theta_0$  with  $\delta$  fixed, if it is positive when  $\cos \theta_0 = 1$  and also when  $\cos \theta_0 = -1$ , it will be positive always. Thus, it suffices to verify the positivity of the functions

$$(\delta - 2) \left( \gamma + \log \left( \frac{\delta + \gamma}{\delta - \gamma} \right) \right) \quad (6.62)$$

and

$$(\delta + 2) \left( \gamma - \log \left( \frac{\delta + \gamma}{\delta - \gamma} \right) \right). \quad (6.63)$$

Now, (6.62) is clearly positive for all  $\delta > 2$ . Meanwhile, a computation shows that

$$\frac{d}{d\delta} \left( \gamma - \log \left( \frac{\delta + \gamma}{\delta - \gamma} \right) \right) = \frac{\delta - 2}{\gamma} > 0$$

and

$$\lim_{\delta \rightarrow 2^+} \left( \gamma - \log \left( \frac{\delta + \gamma}{\delta - \gamma} \right) \right) = 0,$$

from which we conclude that (6.63) is also positive for all  $\delta > 2$ .  $\square$

**6.5. Surjectivity.** In Section 7.3, we will compute  $s_t(\lambda) := \lim_{x \rightarrow 0^+} S(t, \lambda, x)$  for  $\lambda$  in  $\Sigma_t$ . We will do so by evaluating  $S$  (and its derivatives) along curves of the form  $(t, \lambda(t), x(t))$  and then the taking the limit as we approach the time  $t_*$  when  $x(t)$  becomes zero. For this method to be successful, we need the following result, whose proof appears on p. 44.

**Theorem 6.17.** *Fix  $t > 0$ . Then for all  $\lambda \in \Sigma_t$ , there exists a unique  $\lambda_0 \in \mathbb{C}$  and  $x_0 > 0$  such that the solution to (6.2) with these initial conditions exists on  $[0, t)$  with  $\lim_{u \rightarrow t^-} x(u) = 0$  and  $\lim_{u \rightarrow t^-} \lambda(u) = \lambda$ . For all  $\lambda \in \Sigma_t$ , the corresponding  $\lambda_0$  also belongs to  $\Sigma_t$ .*

Define functions  $\Lambda_0^t : \Sigma_t \rightarrow \Sigma_t$  and  $X_0^t : \Sigma_t \rightarrow (0, \infty)$  by letting  $\Lambda_0^t(\lambda)$  and  $X_0^t(\lambda)$  be the corresponding values of  $\lambda_0$  and  $x_0$ , respectively. Then  $\Lambda_0^t$  and  $X_0^t$  extend to continuous maps of  $\bar{\Sigma}_t$  into  $\bar{\Sigma}_t$  and  $[0, \infty)$ , respectively, with the continuous extensions satisfying  $\Lambda_t(\lambda) = \lambda$  and  $X_0^t(\lambda) = 0$  for  $\lambda \in \partial \Sigma_t$ .

We first recall that we have shown (Proposition 6.16) that the lifetime of the path to be a strictly increasing function of  $x_0 \geq 0$  with  $\lambda_0$  fixed. If  $\lambda_0$  is outside  $\Sigma_t$ , then by Theorem 4.1 and Proposition 6.13, the lifetime at least  $t$ , even at  $x_0 = 0$ . (That is to say,  $T(\lambda_0) = t_*(\lambda_0, 0) \geq t$  for  $\lambda_0$  outside  $\Sigma_t$ .) Thus, for  $\lambda_0$  outside  $\Sigma_t$ , the lifetime cannot equal  $t$  for  $x_0 > 0$ . On the other hand, if  $\lambda_0 \in \Sigma_t$ , then  $t_*(\lambda_0, 0) < t$  and Proposition 6.16 tells us that there is a unique  $x_0 > 0$  with  $t_*(\lambda_0, x_0) = t$ .

**Lemma 6.18.** *Fix  $t > 0$ . Define maps*

$$\begin{aligned} x_0^t : \Sigma_t &\rightarrow [0, \infty) \\ \lambda_t : \Sigma_t &\rightarrow \mathbb{C} \setminus \{0\} \end{aligned}$$

as follows. For  $\lambda_0 \in \Sigma_t$ , we let  $x_0^t(\lambda_0)$  denote the unique positive value of  $x_0$  for which  $t_*(\lambda_0, x_0) = t$ . Then we set

$$\lambda_t(\lambda_0) = \lim_{u \rightarrow t^-} \lambda(u),$$

where  $\lambda(\cdot)$  is computed with initial conditions  $\lambda(0) = \lambda_0$  and  $x(0) = x_0^t(\lambda_0)$ . Then both  $x_0^t$  and  $\lambda_t$  extend continuously from  $\Sigma_t$  to  $\bar{\Sigma}_t$ , with the extended maps satisfying  $x_0^t(\lambda_0) = 0$  and  $\lambda_t(\lambda_0) = \lambda_0$  for  $\lambda_0 \in \partial\Sigma_t$ . The extended map  $\lambda_t$  is a homeomorphism of  $\bar{\Sigma}_t$  to itself.

We note that the desired function  $\Lambda_0^t$  in Theorem 6.17 is the inverse function to  $\lambda_t$  and that  $X_0^t(\lambda) = x_0^t(\lambda_t^{-1}(\lambda))$ .

Recall from Proposition 6.4 that the argument of  $\lambda(t)$  is constant. By the formula (6.29) in Theorem 6.7 together with the expression (6.27) for the constant  $C$ , we can write

$$\lambda_t(\lambda_0) = \frac{\lambda_0}{|\lambda_0|} e^{Ct/2} = \frac{\lambda_0}{|\lambda_0|} \exp\left(\frac{t}{2} \frac{|\lambda_0|^2 - 1 + x_0^t(\lambda_0)}{|\lambda_0 - 1|^2 + x_0^t(\lambda_0)}\right), \quad (6.64)$$

where we have used that  $t_*(\lambda_0, x_0^t(\lambda_0)) = t$ . As noted in the proof of Proposition 6.12, this formula can also be written as

$$\lambda_t(\lambda_0) = \frac{\lambda_0}{|\lambda_0|} \exp\left(\frac{\delta - 2/|\lambda_0|}{2\sqrt{\delta^2 - 4}} \log\left(\frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}}\right)\right), \quad (6.65)$$

where  $\delta = (|\lambda_0|^2 + 1 + x_0^t(\lambda_0))/|\lambda_0|$ .

*Proof.* We start by trying to compute the function  $x_0^t$ , which we will do by finding the correct value of  $\delta$  and then solving for  $x_0^t$ . Recall that the lifetime  $t_*(\lambda_0, x_0)$  is computed as  $g_{\theta_0}(\delta)$ , where  $\delta$  is as in (6.40) and  $g_{\theta_0}$  is as in (6.61). As we have computed in (6.60), we have

$$g_{\theta_0}\left(\frac{r_0^2 + 1}{r_0}\right) = T(r_0 e^{i\theta_0}).$$

Assume, then, that the ray with angle  $\theta_0$  intersects  $\Sigma_t$  and let  $r_t(\theta_0)$  be the outer (for definiteness) radius at which this ray intersects the boundary of  $\Sigma_t$ . Then Theorem 4.1 tells us that  $T(r_t(\theta_0)e^{i\theta_0}) = t$ , and we conclude that

$$g_{\theta_0}\left(\frac{r_t(\theta_0)^2 + 1}{r_t(\theta_0)}\right) = t. \quad (6.66)$$

Consider, then, some  $\lambda_0 \in \Sigma_t$  with  $\arg(\lambda_0) = \theta_0$ . By the formula (6.49), to find  $x_0$  with  $t_*(\lambda_0, x_0) = t$ , we first find  $\delta$  so that  $g_{\theta_0}(\delta) = t$ . (Note that the value of

$\delta$  depends only on the argument of  $\lambda_0$ .) We then adjust  $x_0$  so that  $(|\lambda_0|^2 + x_0 + 1)/|\lambda_0| = \delta$ . Since the correct value of  $\delta$  is given in (6.66), this means that we should choose  $x_0$  so that

$$\frac{|\lambda_0|^2 + x_0 + 1}{|\lambda_0|} = \frac{r_t(\theta_0)^2 + 1}{r_t(\theta_0)}.$$

We can solve this relation for  $x_0$  to obtain

$$x_0^t(\lambda_0) = |\lambda_0| \left( \frac{r_t(\arg \lambda_0)^2 + 1}{r_t(\arg \lambda_0)} - \frac{|\lambda_0|^2 + 1}{|\lambda_0|} \right). \quad (6.67)$$

Now, we have shown that  $r_t(\theta)$  is continuous for the full range of angles  $\theta$  occurring in  $\bar{\Sigma}_t$ . Since 0 is not in  $\bar{\Sigma}_t$ , we can then see that the formula (6.67) is well defined and continuous on all of  $\bar{\Sigma}_t$ . For  $\lambda_0 \in \partial\Sigma_t$ , we have that  $|\lambda_0|$  equals  $r_t(\arg \lambda_0)$  or  $1/r_t(\arg \lambda_0)$ , so that  $x_0^t(\lambda_0)$  equals zero.

Now, the point 0 is always outside  $\bar{\Sigma}_t$ , while the point 1 is always in  $\Sigma_t$  and therefore not on the boundary of  $\Sigma_t$ . Thus, since  $x_0^t$  is continuous on  $\bar{\Sigma}_t$  and zero precisely on the boundary, we see from (6.64) that  $\lambda_t$  is continuous on  $\bar{\Sigma}_t$ . Furthermore, on  $\partial\Sigma_t$ , we compute  $\lambda_t(\lambda_0)$  by putting  $x_0^t(\lambda_0) = 0$  in (6.64). Suppose now that  $\lambda_0$  is in  $\partial\Sigma_t$ . Then  $x_0^t(\lambda_0) = 0$  and, by Theorem 4.1, the function  $T(\lambda_0)$  in (4.1) has the value  $t$ , so that

$$\frac{t}{2} \frac{|\lambda_0|^2 - 1}{|\lambda_0 - 1|^2} = \log(|\lambda_0|).$$

Thus, from (6.64), we see that  $\lambda_t(\lambda_0) = \lambda_0$ .

Consider an angle  $\theta_0$  for which the ray  $\text{Ray}(\theta_0)$  with angle  $\theta_0$  intersects  $\bar{\Sigma}_t$  and let  $\delta$  be chosen so that  $g_{\theta_0}(\delta) = t$ , noting again that the value of  $\delta$  depends only on  $\theta_0 = \arg \lambda_0$ . We now observe from (6.65) that  $|\lambda_t(\lambda_0)|$  is a strictly increasing function of  $|\lambda_0|$  with  $\delta$  fixed. Thus,  $\lambda_t$  is a strictly increasing function of the interval  $\text{Ray}(\theta_0) \cap \bar{\Sigma}_t$  into  $\text{Ray}(\theta_0)$  that fixes the endpoints. Thus, actually,  $\lambda_t$  maps this interval bijectively into itself. Since this holds for all  $\theta_0$ , we conclude that  $\lambda_t$  maps  $\bar{\Sigma}_t$  bijectively into itself. The continuity of the inverse then holds because  $\lambda_t$  is continuous and  $\bar{\Sigma}_t$  is compact.  $\square$

*Proof of Theorem 6.17.* We have noted before the statement of Lemma 6.18 that if the desired pair  $(x_0, \lambda_0)$  exists,  $\lambda_0$  must be in  $\Sigma_t$ . The lemma then tells us that a unique pair  $(x_0, \lambda_0)$  exists with  $\lambda_0 \in \Sigma_t$ . We compute  $\Lambda_0^t(\lambda)$  as  $\lambda_t^{-1}(\lambda)$  and  $X_0^t(\lambda)$  as  $x_0^t(\lambda_t^{-1}(\lambda))$ , both of which extend continuously to  $\bar{\Sigma}_t$ . For  $\lambda \in \partial\Sigma_t$ , we have  $\lambda_t^{-1}(\lambda) = \lambda$  and  $x_0^t(\lambda_t^{-1}(\lambda)) = x_0^t(\lambda) = 0$ .  $\square$

## 7. LETTING $x$ TEND TO ZERO

**7.1. Outline.** Our goal is to compute the Laplacian with respect to  $\lambda$  of the function  $s_t(\lambda) := \lim_{x \rightarrow 0^+} S(t, \lambda, x)$ , using the Hamilton–Jacobi method of Theorem 6.2. We want the curve  $x(\cdot)$  occurring in (6.7) and (6.8) to approach zero at time  $t$ ; a simple way we might try to accomplish this is to let the initial condition  $x_0$  approach zero. Suppose, then, that  $x_0$  is very small. Using various formulas from Section 6.3, we then find that *for as long as the solution to the system (6.2) exists*, the whole curve  $x(\cdot)$  will be small and the whole curve  $\lambda(\cdot)$  will be approximately constant. Thus, by taking  $x_0 \approx 0$  and  $\lambda_0 \approx \lambda$ , we obtain a curve with  $x(t) \approx 0$  and

$\lambda(t) \approx \lambda$ . We may then hope to compute  $s_t(\lambda)$  by letting  $\lambda_0$  and  $\lambda(t)$  approach  $\lambda$  and  $x_0$  approach zero in the Hamilton–Jacobi formula (6.7), with the result that

$$s_t(\lambda) = \log(|\lambda - 1|^2). \quad (7.1)$$

It is essential to note, however, that this approach is only valid if the solution to system (6.2) exists up to time  $t$ . Corollary 6.14 tells us that for  $x_0 \approx 0$ , the solution will exist beyond time  $t$  provided  $\lambda$  is outside  $\bar{\Sigma}_t$ . Thus, we expect that for  $\lambda$  outside  $\bar{\Sigma}_t$ , the function  $s_t$  will be given by (7.1) and therefore that  $\Delta s_t$  will be zero. (The function  $\log(|\lambda - 1|^2)$  is harmonic except at the point  $\lambda = 1$ , which is always inside  $\Sigma_t$ .)

To analyze  $s_t(\lambda)$  for  $\lambda$  inside  $\Sigma_t$ , we first make use of the surjectivity result in Theorem 6.17. The theorem says that for each  $t > 0$  and  $\lambda \in \Sigma_t$ , there exist  $x_0 > 0$  and  $\lambda_0 \in \Sigma_t$  such that  $x(u)$  approaches 0 and  $\lambda(u)$  approaches  $\lambda$  as  $u$  approaches  $t$ . We then use the formula (6.30) in Theorem 6.7. In light of the second Hamilton–Jacobi formula (6.8), we can write (6.30) as

$$\begin{aligned} \lim_{u \rightarrow t} \left( a \frac{\partial S}{\partial a} + b \frac{\partial S}{\partial b} \right) (u, \lambda(u), x(u)) &= \lim_{u \rightarrow t} \frac{2 \log |\lambda(t)|}{t} + 1 \\ &= \frac{2 \log |\lambda|}{t} + 1. \end{aligned} \quad (7.2)$$

Once we have established enough regularity in the function  $S(t, \lambda, x)$  near  $x = 0$ , we will be able to identify the left-hand side of (7.2) with the corresponding derivative of  $s_t$ , giving the following explicit formula for one of the derivatives of  $s_t$ :

$$\left( a \frac{\partial s_t}{\partial a} + b \frac{\partial s_t}{\partial b} \right) (\lambda) = \frac{2 \log |\lambda|}{t} + 1. \quad (7.3)$$

We now compute in logarithmic polar coordinates, with  $\rho = \log |\lambda|$  and  $\theta = \arg \lambda$ . We may recognize the left-hand side of (7.3) as the derivative of  $s_t$  with respect to  $\rho$ , giving

$$\frac{\partial s_t}{\partial \rho} = \frac{2\rho}{t} + 1 \quad (7.4)$$

for points inside  $\Sigma_t$ . Remarkably,  $\partial s_t / \partial \rho$  is independent of  $\theta$ ! Thus,

$$\frac{\partial}{\partial \rho} \frac{\partial s_t}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{\partial s_t}{\partial \rho} = 0,$$

meaning that  $\partial s_t / \partial \theta$  is independent of  $\rho$ .

Now, we will show in Section 7.4 that the first derivatives of  $s_t$  have the same value as we approach a point  $\lambda \in \partial \Sigma_t$  from the inside as when we approach  $\lambda$  from the outside. We can therefore give a complete description of the function  $\partial s_t / \partial \theta$  on  $\Sigma_t$  as follows. It is the unique function on  $\Sigma_t$  that is independent of  $\rho$  (or, equivalently, independent of  $r = |\lambda|$ ) and whose boundary values agree

$$\frac{\partial}{\partial \theta} \log(|\lambda - 1|^2) = \frac{2b}{|\lambda - 1|^2} = \frac{2r \sin \theta}{r^2 + 1 - 2r \cos \theta}. \quad (7.5)$$

Since the points on the outer boundary of  $\Sigma_t$  have the polar form  $(r_t(\theta), \theta)$ , we conclude that

$$\frac{\partial s_t}{\partial \theta} = \frac{2r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta}.$$

From this result, the expression (7.4), and the formula for the Laplacian in logarithmic polar coordinates, we obtain

$$\begin{aligned}\Delta s_t(\lambda) &= \frac{1}{|\lambda|^2} \left( \frac{\partial^2 s_t}{\partial \rho^2} + \frac{\partial^2 s_t}{\partial \theta^2} \right) \\ &= \frac{1}{|\lambda|^2} \left( \frac{2}{t} + \frac{\partial}{\partial \theta} \frac{2r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta} \right)\end{aligned}$$

for points inside  $\Sigma_t$ , accounting for the formula in Theorem 2.2.

We now briefly discuss what is needed to make the preceding arguments rigorous. If  $\lambda$  is outside  $\bar{\Sigma}_t$  and  $x$  is small and positive, we need to know that we can find a  $\lambda_0$  close to  $\lambda$  and a small, positive  $x_0$  such that with these initial conditions,  $x(t) = x$  and  $\lambda(t) = \lambda$ . To show this, we apply the inverse function theorem to the map  $U_t(\lambda_0, x_0) := (\lambda(t), x(t))$  in a neighborhood of the point  $(\lambda_0, x_0) = (\lambda, 0)$ .

For  $\lambda$  inside  $\Sigma_t$ , we need to know first that  $S(t, \lambda, x)$  is continuous—in all three variables—up to  $x = 0$ . After all,  $s_t(\lambda)$  is defined letting  $x$  tend to zero in the expression  $S(t, \lambda, x)$ , with  $t$  and  $\lambda$  fixed. But the Hamilton–Jacobi formula (6.7) gives a formula for  $S(u, \lambda(u), x(u))$ , in which the first two variables in  $S$  are not remaining constant. Furthermore, we want to apply also the Hamilton–Jacobi formula (6.8) for the derivatives of  $S$ , which means we need also continuity of the derivatives of  $S$  with respect to  $\lambda$  up to  $x = 0$ . Using another inverse function theorem argument, we will show that after making the change of variable  $z = \sqrt{x}$ , the function  $S$  will extend smoothly up to  $x = z = 0$ , from which the needed regularity will follow.

We use the following notation throughout the section.

**Notation 7.1.** *We will let*

$$x(t; \lambda_0, x_0)$$

*denote the  $x$ -component of the solution to (6.2) with  $\lambda(0) = \lambda_0$  and  $x(0) = x_0$  (and with initial values of the momenta given by (6.4)), and similarly for the other components of the solution.*

**7.2. Outside  $\bar{\Sigma}_t$ .** The goal of this subsection is to prove the following result.

**Theorem 7.2.** *Fix a pair  $(t, \lambda)$  with  $\lambda$  outside  $\bar{\Sigma}_t$ . Then*

$$s_t(\lambda) := \lim_{x \rightarrow 0^+} S(t, \lambda, x) = \log(|\lambda - 1|^2). \quad (7.6)$$

*Thus,*

$$\Delta s_t(\lambda) = 0$$

*whenever  $\lambda$  is outside  $\bar{\Sigma}_t$ .*

As we have discussed in Section 7.1, the idea is that for  $\lambda$  outside  $\bar{\Sigma}_t$  and  $x$  small and positive, we should try to find a  $\lambda_0$  close to  $\lambda$  and a small, positive  $x_0$  such that  $x(u)$  and  $\lambda(u)$  will approach 0 and  $\lambda$ , respectively, as  $u$  approaches  $t$ . To that end, we define, for each  $t > 0$ , a map  $U_t$  from an open subset of  $\mathbb{R} \times \mathbb{C}$  into  $\mathbb{R} \times \mathbb{C}$  by

$$U_t(\lambda_0, x_0) = (\lambda(t; \lambda_0, x_0), x(t; \lambda_0, x_0)).$$

We wish to evaluate the derivative of this map at the point  $(\lambda_0, x_0) = (\lambda, 0)$ . For this idea to make sense,  $\lambda(t; \lambda_0, x_0)$  and  $x(t; \lambda_0, x_0)$  must be defined in a neighborhood of  $(\lambda, 0)$ ; it is for this reason that we have allowed  $x_0$  to be negative in Section 6.3.

The domain of  $U_t$  consists of pairs  $(x_0, \lambda_0)$  such that (1) the assumptions (6.44) are satisfied; (2) the function  $p_x(\cdot)$  remains positive until it blows up, as in Remark 6.10; and (3) we have  $t_*(\lambda_0, x_0) > t$ . We note that these conditions allow  $x_0$  to be slightly negative and that all the results of Section 6.3 hold under these conditions. We also note that by Proposition 6.11, if  $x_0 = 0$ , then  $x(t) \equiv 0$  and  $\lambda(t) \equiv \lambda_0$ ; thus,

$$U_t(\lambda_0, 0) = (\lambda_0, 0). \quad (7.7)$$

We now fix a pair  $(t, \lambda)$  with  $\lambda$  outside of  $\bar{\Sigma}_t$  (so that  $\lambda \neq 1$ ). By Corollary 6.14, we then have  $t_*(\lambda, 0) > t$ .

**Lemma 7.3.** *The Jacobian of  $U_t$  at  $(\lambda, 0)$  has the form*

$$U'_t(\lambda, 0) = \begin{pmatrix} I_{2 \times 2} & \frac{\partial \lambda}{\partial x_0}(t; \lambda, 0) \\ 0 & \frac{\partial x}{\partial x_0}(t; \lambda, 0) \end{pmatrix} \quad (7.8)$$

with  $\partial x / \partial x_0(t; \lambda, 0) > 0$ . In particular, the inverse function theorem applies at  $(\lambda, 0)$ .

*Proof.* The claimed form of the second column of  $U'_t(\lambda, 0)$  follows immediately from (7.7). We then compute from (6.52) that

$$\begin{aligned} \frac{\partial x(t; \lambda_0, x_0)}{\partial x_0}(0, \lambda_0) &= \frac{1}{p_x(t)^2} p_0^2 e^{-Ct} + x_0 \frac{\partial}{\partial x_0} \left[ \frac{1}{p_x(t)^2} p_0^2 e^{-Ct} \right] \Big|_{x_0=0} \\ &= \frac{1}{p_x(t)^2} p_0^2 e^{-Ct}, \end{aligned} \quad (7.9)$$

which is positive.  $\square$

*Proof of Theorem 7.2.* We note that the inverse of the matrix in (7.8) will have a positive entry in the bottom right corner, meaning that  $U_t^{-1}$  has the property that  $\partial x_0 / \partial x > 0$ . It follows that the  $x_0$ -component of  $U_t^{-1}(\lambda, x)$  will be positive for  $x$  small and positive. In that case, the solution to the system (6.2) will have  $x(u) > 0$  up to the blow-up time. The blow-up time, in turn, exceeds  $t$  for all points in the domain of  $U_t$ .

We may, therefore, apply the Hamilton–Jacobi formula (6.7), which we write as follows. We let HJ denote the right-hand side of the Hamilton–Jacobi formula (6.7):

$$\begin{aligned} \text{HJ}(t, \lambda_0, x_0) &= \log(|\lambda_0 - 1|^2 + x_0) - \frac{x_0 t}{(|\lambda_0 - 1|^2 + x_0)^2} \\ &\quad + \log |\lambda(t; \lambda_0, x_0)| - \log |\lambda_0| \end{aligned} \quad (7.10)$$

and we then have

$$S(t, \lambda(t; \lambda_0, x_0), x(t; \lambda_0, x_0)) = \text{HJ}(t, \lambda_0, x_0). \quad (7.11)$$

If  $x$  is small and positive, we therefore obtain

$$S(t, \lambda, x) = \text{HJ}(t, U_t^{-1}(\lambda, x)),$$

where we note that by definition  $\lambda(t; U_t^{-1}(\lambda, x)) = \lambda$ .

Now, in the limit  $x \rightarrow 0^+$  with  $\lambda$  fixed, the inverse function theorem tells us that  $U_t^{-1}(x, \lambda) \rightarrow (0, \lambda)$ . Thus, the limit (7.6) may be computed by putting  $\lambda(t; \lambda_0, x_0) = \lambda$  in (7.11) and letting  $x_0$  tend to zero and  $\lambda_0$  tend to  $\lambda$ . This process gives (7.6).

Finally, when  $\lambda_0 = 0$ , we can use continuous dependence of the solutions on the initial conditions. The formula for  $p_x(t)$  in Proposition 6.9 has a limit as  $|\lambda_0|$  tends

to zero, so that  $\delta$  tends to  $+\infty$ . From (6.46), we find that  $a^2 = y_0^2$ , so that from (6.34),  $y(t) \equiv y_0$ . We then obtain

$$p_x(t) = e^{-Ct}p_0,$$

which remains nonsingular for all  $t$ . We can then continue to use the formula (6.52) for  $x(t)$ . Furthermore, by exponentiating (6.53) and letting  $|\lambda_0|$  tend to zero, we find that  $\lambda(t) \equiv 0$ . We then continue to use the remaining formulas in the proof of Proposition 6.11 and find that the solution to the system exists for all time.

When  $\lambda_0 = 0$ , we apply the Hamilton–Jacobi formula in the form (6.21), which is to say that we replace the last two terms in (6.7) by  $\int_0^t x(s)p_x(s) ds$ . We then compute as in (7.9) that the derivative of  $x(t; 0, x_0)$  with respect to  $x_0$  is positive at  $x_0 = 0$ . Thus, by the inverse function theorem, for small positive  $x$ , we can find a small positive  $x_0$  that gives  $x(t; 0, x_0) = x$ . We then apply (6.21) with  $\lambda_0 = 0$  and  $\lambda(t) = 0$ , and let  $x$  tend to zero, which means that  $x_0$  also tends to zero. As  $x_0$  tends to zero, the function

$$x(s)p_x(s) = \frac{x(s)p_x(s)^2}{p_x(s)} = \frac{x_0 p_0^2 e^{-Cs}}{p_x(s)}$$

tends to zero uniformly and we obtain (7.6).  $\square$

**7.3. Inside  $\Sigma_t$ .** In this subsection, we establish the needed regularity of  $S(t, \lambda, x)$  as  $x$  tends to zero, for  $\lambda$  in  $\Sigma_t$ . This result, whose proof is on p. 51, together with Theorem 6.7, will allow us to understand the structure of  $s_t$  and its derivatives on  $\Sigma_t$ .

**Theorem 7.4.** *Define*

$$\tilde{S}(t, \lambda, z) = S(t, \lambda, z^2), \quad z > 0.$$

*Fix a pair  $(\sigma, \mu)$  with  $\mu$  in  $\Sigma_\sigma$ . Then  $\tilde{S}(t, \lambda, z)$ , initially defined for  $z > 0$ , extends to a real-analytic function in a neighborhood of  $(\sigma, \mu, 0)$  inside  $\mathbb{R} \times \mathbb{C} \times \mathbb{R}$ .*

We emphasize that the analytically extended  $\tilde{S}$  does not satisfy the identity  $\tilde{S}(t, \lambda, z) = S(t, \lambda, z^2)$ . Indeed, since  $\sqrt{x(t)}p_x(t)$  is always bounded away from zero (Proposition 6.6), the second Hamilton–Jacobi formula (6.8) tells us that  $\partial\tilde{S}/\partial z(t, \lambda, z) = 2\sqrt{x}\partial S/\partial x(t, \lambda, z^2)$  has a nonzero limit as  $z$  tends to zero, ruling out a smooth extension that is even in  $z$ .

**Corollary 7.5.** *Fix a pair  $(\sigma, \mu)$  with  $\mu$  in  $\Sigma_\sigma$ . Then the functions*

$$S(t, \lambda, x), \quad \frac{\partial S}{\partial a}(t, \lambda, x), \quad \frac{\partial S}{\partial b}(t, \lambda, x), \quad \sqrt{x}\frac{\partial S}{\partial x}(t, \lambda, x) \quad (7.12)$$

*all have extensions that are continuous in all three variables to the set of  $(t, \lambda, x)$  with  $\lambda \in \Sigma_t$  and  $x \geq 0$ . Furthermore, for each  $t > 0$ , the function  $s_t$  is infinitely differentiable on  $\Sigma_t$ , and its derivatives with respect to  $a$  and  $b$  agree with the  $x \rightarrow 0^+$  limit of  $\partial S/\partial a$  and  $\partial S/\partial b$ . If we let  $t_*$  be short for  $t_*(\lambda_0, x_0)$ , then for all  $\lambda_0$  and  $x_0 > 0$ , we have*

$$\begin{aligned} s_t(t_*, \lambda(t_*; \lambda_0, x_0)) &= \log(|\lambda_0 - 1|^2 + x_0) - \frac{x_0 t_*}{(|\lambda_0 - 1|^2 + x_0)^2} \\ &\quad + \log |\lambda(t_*; \lambda_0, x_0)| - \log |\lambda_0| \end{aligned}$$

and

$$\begin{aligned}\frac{\partial s_t}{\partial a}(t_*, \lambda(t_*; \lambda_0, x_0)) &= \lim_{t \rightarrow t_*} p_a(t) \\ \frac{\partial s_t}{\partial b}(t_*, \lambda(t_*; \lambda_0, x_0)) &= \lim_{t \rightarrow t_*} p_b(t).\end{aligned}\quad (7.13)$$

*Proof.* We note that the four functions in (7.12) may be computed as

$$\tilde{S}(t, \lambda, \sqrt{x}), \quad \frac{\partial \tilde{S}}{\partial a}(t, \lambda, \sqrt{x}), \quad \frac{\partial \tilde{S}}{\partial b}(t, \lambda, \sqrt{x}), \quad \frac{1}{2} \frac{\partial \tilde{S}}{\partial z}(t, \lambda, \sqrt{x}),$$

respectively, and that  $S(t, \lambda, 0) = \tilde{S}(t, \lambda, 0)$ . The first claim then follows from Theorem 7.4. Now that the continuity of  $S$  and its derivatives has been established, we may let  $t$  approach  $t_*(\lambda_0, x_0)$  in the Hamilton–Jacobi formulas (6.7) and (6.8) to obtain the second claim.  $\square$

**Corollary 7.6.** *Let us write  $\lambda \in \Sigma_t$  in logarithmic polar coordinates, with  $\rho = \log |\lambda|$  and  $\theta = \arg \lambda$ . Then for each pair  $(t, \lambda)$  with  $\lambda \in \Sigma_t$ , we have*

$$\frac{\partial s_t}{\partial \rho}(t, \lambda) = \frac{2\rho}{t} + 1. \quad (7.14)$$

Furthermore,  $\partial s_t / \partial \theta$  is independent of  $\rho$ ; that is,

$$\frac{\partial s_t}{\partial \theta} = m_t(\theta),$$

for some smooth function  $m_t$ . Thus,

$$\frac{\partial^2 s_t}{\partial \rho^2} + \frac{\partial^2 s_t}{\partial \theta^2} = \frac{2}{t} + \frac{\partial}{\partial \theta} m_t(\theta) \quad (7.15)$$

for some smooth function  $m_t$ , and

$$\left( \frac{\partial^2}{\partial a^2} + \frac{\partial^2}{\partial b^2} \right) s_t(\lambda) = \frac{1}{|\lambda|^2} \left( \frac{2}{t} + \frac{\partial}{\partial \theta} m_t(\theta) \right). \quad (7.16)$$

In Section 7.4, we will obtain a formula for the function  $m_t(\theta)$  appearing in Corollary 7.6.

*Proof.* The derivative  $\partial / \partial \rho$  may be computed in ordinary polar coordinates as  $r \partial / \partial r$  or in rectangular coordinates as  $a \partial / \partial a + b \partial / \partial b$ . It then follows from the Hamilton–Jacobi formula (6.8) that

$$\left( a \frac{\partial S}{\partial a} + b \frac{\partial S}{\partial b} \right) (t, \lambda(t), x(t)) = a(t) p_a(t) + b(t) p_b(t).$$

Now, for each pair  $(t, \lambda)$  with  $\lambda \in \Sigma_t$ , Theorem 6.17 tells us that we can find  $(\lambda_0, x_0)$  so that

$$\lim_{u \rightarrow t} x(u) = 0; \quad \lim_{u \rightarrow t} \lambda(u) = \lambda.$$

In light of (7.13), the formula (7.14) then follows from the formula (6.30) in Theorem 6.7.

Now,  $\partial s_t / \partial \rho$  is manifestly independent of  $\theta$ . Since, by Corollary 7.5,  $s_t$  is an analytic, hence  $C^2$ , function on  $\Sigma_t$ , we conclude that

$$\frac{\partial}{\partial \rho} \frac{\partial s_t}{\partial \theta} = \frac{\partial}{\partial \theta} \frac{\partial s_t}{\partial \rho} = 0,$$

showing that  $\partial s_t / \partial \theta$  is independent of  $\rho$ . The formula (7.15) then follows by differentiating (7.14) with respect to  $\rho$ . Finally, if we use the standard formula for the Laplacian in polar coordinates,

$$\Delta = \frac{1}{r^2} \left( \left( r \frac{\partial}{\partial r} \right)^2 + \frac{\partial^2}{\partial \theta^2} \right) = \frac{1}{r^2} \left( \frac{\partial^2}{\partial \rho^2} + \frac{\partial^2}{\partial \theta^2} \right).$$

we obtain (7.16) from (7.15).  $\square$

We now begin preparations for the proof of Theorem 7.4. Recall from Proposition 6.6 that  $x(t)p_x(t)^2 = x_0 p_0^2 e^{-Ct}$ , where  $C = 2\Psi - 1$  is a constant computed from  $x_0$  and  $\lambda_0$  as in (6.27). Recall also from Proposition 6.9 that for  $x_0 > 0$ , the function  $1/p_x(t)$  extends to real analytic function of  $t$  defined for all  $t \in \mathbb{R}$ . We then define, for  $x_0 > 0$ ,

$$z(t; \lambda_0, x_0) = \sqrt{x_0} p_0 e^{-Ct/2} \frac{1}{p_x(t; \lambda_0, x_0)} \quad (7.17)$$

for all  $t \in \mathbb{R}$ . For  $t < t_*(\lambda_0, x_0)$ , the function  $z(t; \lambda_0, x_0)$  is positive and satisfies

$$z(t; \lambda_0, x_0)^2 = x(t; \lambda_0, x_0),$$

while for  $t = t_*(\lambda_0, x_0)$ , we have  $z(t; \lambda_0, x_0) = 0$  and for  $t > t_*(\lambda_0, x_0)$ , the function  $z(t; \lambda_0, x_0)$  is negative.

Furthermore, using (6.53) and Point 3 of Proposition 6.4, we see that

$$\lambda(t; \lambda_0, x_0) = \lambda_0 e^{\int_0^t x(s)p_x(s) ds},$$

where by Proposition 6.6, we have

$$x(s)p_x(s) = \frac{x(s)p_x(s)^2}{p_x(s)} = \frac{x_0 p_0^2 e^{-Cs}}{p_x(s)}.$$

Since  $1/p_x(s)$  extends to an analytic function of  $s \in \mathbb{R}$ , we see that  $\lambda(t)$  extends to an analytic function of  $t \in \mathbb{R}$ . We may therefore define a map

$$V(t, \lambda_0, x_0) := (t, \lambda(t; \lambda_0, x_0), z(t; \lambda_0, x_0)),$$

for all  $t \in \mathbb{R}$ ,  $\lambda_0 \in \mathbb{C}$ , and  $x_0 > 0$ .

**Proposition 7.7.** *Suppose  $(t, \lambda_0, x_0)$  has the property that  $\lambda_0 \neq 0$  and  $t_*(\lambda_0, x_0) = t$ , so that  $z(t; \lambda_0, x_0) = 0$ . Then the Jacobian matrix of  $V$  at  $(t, \lambda_0, x_0)$  is invertible.*

*Proof.* We make some convenient changes of variables. First, we replace  $(t, \lambda_0, x_0)$  with  $(t, \lambda_0, \delta)$ , where  $\delta$  is as in (6.40). This change has a smooth inverse, since we can recover  $x_0$  from  $\delta$  as

$$x_0 = |\lambda_0| \delta - |\lambda_0|^2 - 1.$$

Then we write  $\lambda_0$  in terms of its polar coordinates,  $(r_0, \theta_0)$ . Finally, we write  $\lambda(t; \lambda_0, x_0)$  in logarithmic polar coordinates,

$$\rho(t; \lambda_0, x_0) := \log |\lambda(t; \lambda_0, x_0)|; \quad \theta(t; \lambda_0, x_0) := \arg(\lambda(t; \lambda_0, x_0)),$$

where by Point 3 of Proposition 6.4,  $\theta(t; \lambda_0, x_0) = \theta_0$ .

Thus, to prove the proposition, it suffices to verify that the Jacobian matrix of the map

$$W(t, \theta_0, r_0, \delta) := (t, \theta_0, \rho(t; \lambda_0, x_0), z(t; \lambda_0, x_0))$$

is invertible. We observe that this Jacobian has the form

$$W' = \begin{pmatrix} I_{2 \times 2} & 0 \\ * & K \end{pmatrix},$$

where

$$K = \begin{pmatrix} \frac{\partial \rho}{\partial r_0} & \frac{\partial \rho}{\partial \delta} \\ \frac{\partial z}{\partial r_0} & \frac{\partial z}{\partial \delta} \end{pmatrix}.$$

Now, by Proposition 6.9, the lifetime is independent of  $r_0$  with  $\delta$  and  $\theta_0$  fixed. Thus, if we start at a point with  $t_*(\lambda_0, x_0) = t$  and vary  $r_0$ , the lifetime will remain equal to  $t$  and  $z(t; x_0, \lambda_0)$  will remain equal to 0. Thus, at the point in question,  $\partial z / \partial r_0 = 0$ . Meanwhile, Proposition 6.12 gives a formula for the value of  $\rho(t; \lambda_0, x_0)$  at  $t = t_*(\lambda_0, x_0)$ , from which we can easily see that  $\partial \rho / \partial r_0 > 0$ . It therefore remains only to verify that  $\partial z / \partial \delta$  is nonzero.

Now,  $z(t_*(\lambda_0, x_0); \lambda_0, x_0) = 0$ . If we differentiate this relation with respect to  $x_0$  with  $\lambda_0$  fixed, we find that

$$\frac{\partial z}{\partial x_0} = -\frac{\partial z}{\partial t} \frac{\partial t_*(\lambda_0, x_0)}{\partial x_0}. \quad (7.18)$$

The derivative  $\partial t_*/\partial x_0$  may be computed as

$$\frac{\partial t_*(\lambda_0, x_0)}{\partial x_0} = \frac{\partial g_{\theta_0}(\delta)}{\partial \delta} \frac{\partial \delta}{\partial x_0},$$

where  $g_{\theta_0}$  is as in (6.61). But the proof of Proposition 6.16 shows that  $\partial g_{\theta_0} / \partial \delta > 0$  for all  $\delta > 2$ , while from the formula (6.40) for  $\delta$ , we see that  $\partial \delta / \partial x_0 > 0$ . (Note also that  $\delta > 2$  whenever  $x_0 > 0$ .) Thus,  $\partial t_*(\lambda_0, x_0) / \partial x_0 > 0$ .

Meanwhile, from (7.17) and (6.47), we have

$$z(t) = \frac{\sqrt{x_0} e^{Ct/2}}{\cosh(at) + \frac{2|\lambda_0| - \delta}{\sqrt{\delta^2 - 4}} \sinh(at)} \left( \cosh(at) - \frac{\delta}{\sqrt{\delta^2 - 4}} \sinh(at) \right).$$

If we differentiate with respect to  $t$  and evaluate at the time  $t_*$  when the last factor is zero, the product rule gives

$$\begin{aligned} & \left. \frac{\partial z(t; \lambda_0, x_0)}{\partial t} \right|_{t=t_*(\lambda_0, x_0)} \\ &= 0 + \frac{\sqrt{x_0} e^{Ct/2}}{\cosh(at) + \frac{2|\lambda_0| - \delta}{\sqrt{\delta^2 - 4}} \sinh(at)} a \left( \sinh(at) - \frac{\delta}{\sqrt{\delta^2 - 4}} \cosh(at) \right), \end{aligned}$$

which is negative because  $\delta / \sqrt{\delta^2 - 4} > 1$  and the denominator is positive (Proposition 6.9). Thus, from (7.18), we conclude that  $\partial z / \partial x_0 > 0$ .  $\square$

We are now ready for the proof of the main result of this section.

*Proof of Theorem 7.4.* By (7.11), we have

$$\begin{aligned} \tilde{S}(t; \lambda(t; \lambda_0, x_0), z(t; \lambda_0, x_0)) &= S(t; \lambda(t; \lambda_0, x_0), x(t; \lambda_0, x_0)) \\ &= \text{HJ}(t, \lambda_0, x_0), \end{aligned} \quad (7.19)$$

where HJ is as in (7.10), whenever  $t_*(\lambda_0, x_0) > \sigma$ . Fix a point  $(\sigma, \mu)$  with  $\mu \in \Sigma_\sigma$ . Then by Theorem 6.17, we can find a pair  $(\lambda_0, x_0)$  with  $t_*(\lambda_0, x_0) = t$ —so that  $z(t; \lambda_0, x_0) = 0$ —and  $\lambda(t; \lambda_0, x_0) = \lambda$ . We now construct a local inverse  $V^{-1}$  to  $V$  around the point  $V(t, \lambda_0, x_0) = (t, \lambda, 0)$ .

For any triple  $(t, \lambda, z)$  in the domain of  $V^{-1}$ , we write  $V^{-1}(t, \lambda, z)$  as  $(t, \lambda_0, x_0)$ . We note that if  $z > 0$  then  $t_*(\lambda_0, x_0)$  must be greater than  $t$ , because if we had

$t_*(\lambda_0, x_0) \leq t$ , then  $z(t; \lambda_0, x_0) = z$  would be zero or negative. Thus, we may apply (7.19) at  $(t, \lambda_0, x_0) = V^{-1}(t, \lambda, z)$  to obtain

$$\tilde{S}(t, \lambda, z) = \text{HJ}(V^{-1}(t, \lambda, z)), \quad (7.20)$$

whenever  $(t, \lambda, z)$  is in the domain of  $V^{-1}$  and  $z > 0$ .

Recall now that  $\lambda(t; \lambda_0, x_0)$  extends to an analytic function of  $t \in \mathbb{R}$ . Thus, the function HJ in (7.10) extends to a smooth function of  $t \in \mathbb{R}$ ,  $\lambda_0 \in \mathbb{C} \setminus \{0\}$ , and  $x_0 > 0$ , defined even if  $t_*(\lambda_0, x_0) < t$ . Therefore, the right-hand side of (7.20) provides the claimed smooth extension of  $\tilde{S}$ .  $\square$

**7.4. Near the boundary of  $\Sigma_t$ .** We start by considering what is happening right on the boundary of  $\Sigma_t$ .

**Remark 7.8.** *Neither the method of Section 7.2 nor the method of Section 7.3 allows us to compute the value of  $s_t(\lambda)$  for  $\lambda$  in the boundary of  $\Sigma_t$ . Although we expect that this value will be  $\log(|\lambda - 1|^2)$ , the question is irrelevant to the computation of the Brown measure. After all, we are supposed to consider  $\Delta s_t$  computed in the distribution sense, that is, the distribution whose value on a test function  $\psi$  is*

$$\int_{\mathbb{C}} s_t(\lambda) \Delta \psi(\lambda) d^2 \lambda. \quad (7.21)$$

The value of (7.21) is unaffected by the value of  $s_t(\lambda)$  for  $\lambda$  in  $\partial \Sigma_t$ , which is a set of measure zero in  $\mathbb{C}$ .

It is nevertheless essential to understand the behavior of  $s_t(\lambda)$  as  $\lambda$  approaches the boundary of  $\Sigma_t$ .

**Definition 7.9.** *We say that a function  $f : \mathbb{C} \rightarrow \mathbb{R}$  is **analytic up to the boundary from inside**  $\Sigma_t$  if the following conditions hold. First,  $f$  is real analytic on  $\Sigma_t$ . Second, for each  $\lambda \in \partial \Sigma_t$ , we can find an open set  $U$  containing  $\lambda$  and a real analytic function  $g$  on  $U$  such that  $g$  agrees with  $f$  on  $U \cap \Sigma_t$ . We may similarly define what it means for  $f$  to be **analytic up to the boundary from outside**  $\Sigma_t$ .*

**Proposition 7.10.** *For each  $t > 0$ , the function  $s_t$  is analytic up to the boundary from inside  $\Sigma_t$  and analytic up to the boundary from outside  $\Sigma_t$ .*

Note that the proposition is not claiming that  $s_t$  is an analytic function on all of  $\mathbb{C}$ . Indeed, our main results tell us that  $\frac{1}{4\pi} \Delta s_t(\lambda)$  is identically zero for  $\lambda$  outside  $\Sigma_t$  but approaches a typically nonzero value as  $\lambda$  approaches a boundary point from the inside. As we approach from the inside a boundary point with polar coordinates  $(r, \theta)$ , the limiting value of  $\frac{1}{4\pi} \Delta s_t(\lambda)$  is  $w_t(\theta)/r^2$ . This quantity certainly cannot always be zero, or the Brown measure of  $b_t$  would be identically zero. Actually, we will see in Section 8.1 that  $w_t(\theta)$  is strictly positive except when  $t = 4$  and  $\theta = \pi$ .

*Proof.* We have shown that  $s_t(\lambda) = \log(|\lambda - 1|^2)$  for  $\lambda$  in  $(\overline{\Sigma_t})^c$ . Since  $1 \in \Sigma_t$ , we see that  $s_t$  is analytic from the outside of  $\Sigma_t$ .

To address the analyticity from the inside, first note that by applying (7.20) with  $z = 0$ , we have

$$s_t(\lambda) = S(t, \lambda, 0) = \tilde{S}(t, \lambda, 0) = \text{HJ}(V^{-1}(t, \lambda, 0)),$$

where HJ is as in (7.10). But if  $x_0^t : \Sigma_t \rightarrow \mathbb{R}$  and  $\lambda_t : \Sigma_t \rightarrow \mathbb{C}$  are as in Lemma 6.18, then we can see that

$$V^{-1}(t, \lambda, 0) = (t, \lambda_t^{-1}(\lambda), x_0^t(\lambda_t^{-1}(\lambda)))$$

and we conclude that

$$s_t(\lambda) = \text{HJ}(t, \lambda_t^{-1}(\lambda), x_0^t(\lambda_t^{-1}(\lambda))). \quad (7.22)$$

We now claim that the function  $x_0^t(\lambda_0)$ , initially defined for  $\lambda_0 \in \bar{\Sigma}_t$ , extends to an analytic function in a neighborhood of  $\bar{\Sigma}_t$ . For  $t \geq 4$ , we can simply use the formula (6.67) for all nonzero  $\lambda_0$ . For  $t < 4$ , however, the formula (6.67) becomes undefined in a neighborhood of a point where  $\partial\Sigma_t$  intersects the unit circle.

Nevertheless, we can make a general argument as follows. To compute  $x_0^t(\lambda_0)$ , we solve the equation  $t_*(\lambda_0, x_0) = t$  for  $x_0$  as a function of  $\lambda_0$ . To do this, we first solve the equation  $g_{\theta_0}(\delta) = t$  for  $\delta_{\theta_0, t}$  and then solve for  $x_0$  in terms of  $\delta$  as  $x_0 = |\lambda_0|\delta - |\lambda_0|^2 - 1$ . Now, we know from the proof of Proposition 6.16 that  $g_{\theta_0}(\delta) = t$  has a solution when  $|\theta_0| \leq \theta_{\max}(t) = \cos^{-1}(1 - t/2)$ , with the solution being  $\delta = 2$  when  $\theta_0 = \pm\theta_{\max}(t)$ . We can also verify that  $\partial g_{\theta_0}/\partial\delta > 0$  for all  $\delta \geq 2$ . This was verified for  $\delta > 2$  in the proof of Proposition 6.16. To see that the result holds even when  $\delta = 2$ , it suffices to verify that the expressions in (6.62) and (6.63) have positive limits as  $\delta \rightarrow 2^+$ . We omit this verification and simply note that the limits have the values 1 and  $1/3$ , respectively. It then follows from the implicit function theorem that (1) the solution  $\delta_{\theta_0, t}$  continues to exist (with  $\delta < 2$ ) for  $|\theta_0|$  slightly larger than  $\theta_{\max}(t)$ , and (2) the solution  $\delta_{\theta_0, t}$  depends analytically on  $\theta_0$ . Then, the expression

$$x_0^t(\lambda_0) = |\lambda_0|\delta_{\theta_0, t} - |\lambda_0|^2 - 1$$

makes sense and is analytic for all nonzero  $\lambda_0$  with  $|\arg \lambda_0| < \theta_{\max}(t) + \varepsilon_t$ . We note that in this expression,  $x_0^t(\lambda_0)$  can be negative—for example if  $|\lambda_0| = 1$  and  $\arg \lambda_0 > \theta_{\max}(t)$ .

We now consider the function  $\lambda_t$ , defined as

$$\lambda_t(\lambda_0) = \lambda(t; \lambda_0, x_0^t(\lambda_0)),$$

and we recall that  $\lambda_t(\lambda_0) = \lambda_0$  for  $\lambda_0 \in \partial\Sigma_t$ . Although  $\lambda_t$  was initially defined for  $\lambda_0$  in  $\bar{\Sigma}_t$ , it has an analytic extension to a neighborhood of  $\bar{\Sigma}_t$ , namely the set of  $\lambda_0$  in the domain of the extended function  $x_0^t$  for which the pair  $(\lambda_0, x_0)$  satisfy the assumptions in (6.44). We now claim that the derivative of  $\lambda_t(\lambda_0)$  is invertible at each point in its domain. We use polar coordinates in both domain and range. Since  $\arg(\lambda_t(\lambda_0)) = \arg \lambda_0$ , the derivative will have the form

$$\lambda_t'(\lambda_0) = \begin{pmatrix} \frac{\partial|\lambda_t|}{\partial r} & \frac{\partial \arg \lambda_t}{\partial \theta} \\ 0 & 1 \end{pmatrix},$$

and it therefore suffices to check that  $\partial|\lambda_t|/\partial r$  is nonzero. To see this, we use the formula (6.65), where  $\delta = \delta_{\theta_0, t}$  as in the previous paragraph. Since  $\delta$  is independent of  $|\lambda_0|$  with  $t$  and  $\arg \lambda_0$  fixed, we can easily verify from (6.65) that  $\partial|\lambda_t|/\partial r > 0$ .

Now, we have already established that  $s_t$  is analytic in the interior of  $\Sigma_t$ . Consider, then, a point  $\lambda$  in  $\partial\Sigma_t$ , so that  $\lambda_t(\lambda) = \lambda$ . Since  $\lambda_t'(\lambda)$  is invertible, it has an analytic local inverse  $\lambda_t^{-1}$  defined near  $\lambda$ . Then the formula (7.22) gives an analytic extension of  $s_t$  to a neighborhood of  $\lambda$ .  $\square$

**Proposition 7.11.** *Fix a point  $\mu$  on the boundary of  $\Sigma_t$ . Then the functions*

$$s_t(\lambda), \quad \frac{\partial s_t}{\partial a}(\lambda), \quad \frac{\partial s_t}{\partial b}(\lambda)$$

*all approach the same value when  $\lambda$  approaches  $\mu$  from inside  $\Sigma_t$  as when  $\lambda$  approaches  $\mu$  from outside  $\bar{\Sigma}_t$ .*

*Proof.* We begin by considering  $s_t$  itself. The limit as  $\lambda$  approaches  $\mu$  from the inside may be computed by using (7.22). By Lemma 6.18, as  $\lambda$  approaches  $\mu$  from the inside,  $\lambda_t^{-1}(\lambda)$  approaches  $\lambda_t^{-1}(\mu) = \mu$ , and  $x_0^t(\lambda_t^{-1}(\lambda))$  approaches 0. Thus, the limiting value of  $s_t$  from the inside is

$$\text{HJ}(t, \mu, 0) = \log(|\mu - 1|^2),$$

where HJ is given by (7.10) and where we have used that  $\lambda(t; \mu, 0) = \mu$ . (See the last part of Proposition 6.11.) Since  $s_t(\lambda) = \log(|\lambda - 1|^2)$  outside  $\bar{\Sigma}_t$ , the limit of  $s_t$  from the outside agrees with the limit from the inside.

Next we consider the derivatives, which we compute in logarithmic polar coordinates  $\rho = \log |\lambda|$  and  $\theta = \arg \lambda$ . By (7.14), we have

$$\frac{\partial s_t}{\partial \rho}(\lambda) = \left( a \frac{\partial s_t}{\partial a} + b \frac{\partial s_t}{\partial b} \right) (\lambda) = \frac{\log(|\lambda|^2)}{t} + 1$$

for  $\lambda \in \Sigma_t$ . Letting  $\lambda$  approach  $\mu$  from the inside gives the value  $\log(|\mu|^2)/t + 1$ . Since  $\mu$  is on the boundary of  $\Sigma_t$ , Theorem 4.1 says that  $T(\mu) = t$ , so that

$$\begin{aligned} \frac{\log |\mu|^2}{t} + 1 &= \frac{|\mu|^2 - 1}{|\mu - 1|^2} + 1 \\ &= \frac{2(|\mu|^2 - \operatorname{Re} \mu)}{|\mu - 1|^2}. \end{aligned}$$

Taking the corresponding derivative of the “outside” function  $\log(|\lambda - 1|^2)$  and letting  $\lambda$  tend to  $\mu$  from the outside gives the same result.

Finally, we recall from Proposition 6.4 that  $ap_b - bp_a$  is a constant of motion. Thus, by the second Hamilton–Jacobi formula (6.8) and the initial conditions (6.4), we have

$$\begin{aligned} a \frac{\partial s_t}{\partial b}(u, \lambda(u), x(u)) - b \frac{\partial s_t}{\partial a}(u, \lambda(u), x(u)) &= a_0 p_{b,0} - b_0 p_{a,0} \\ &= (2a_0 b_0 - 2b_0(a_0 - 1))p_0 \\ &= \frac{2b_0}{|\lambda_0 - 1|^2 + x_0}. \end{aligned}$$

If we choose  $x_0$  and  $\lambda_0$  so that  $t_*(\lambda_0, x_0) = t$  we can use the regularity result in Corollary 7.5 to let  $u$  tend to  $t$ . This gives

$$a \frac{\partial s_t}{\partial b}(\lambda) - b \frac{\partial s_t}{\partial a}(\lambda) = \frac{2b_0}{|\lambda_0 - 1|^2 + x_0},$$

where now  $\lambda_0 = \lambda_t^{-1}(\lambda)$  and  $x_0 = x_0^t(\lambda_t^{-1}(\lambda))$ . As  $\lambda$  approaches  $\mu$ , Theorem 6.17 says that the value of  $\lambda_0$  approaches  $\mu$  and  $x_0$  approaches 0, so we get

$$\lim_{\lambda \rightarrow \mu^{\text{inside}}} \left( a \frac{\partial s_t}{\partial b}(\lambda) - b \frac{\partial s_t}{\partial a}(\lambda) \right) = \frac{2 \operatorname{Im} \mu}{|\mu - 1|^2}.$$

Taking the corresponding derivative of  $\log(|\lambda - 1|^2)$  and letting  $\lambda$  tend to  $\mu$  from the outside gives the same result.  $\square$

**7.5. Proof of the main result.** In this subsection, we prove our first main result, Theorem 2.2. Proposition 2.3 will then be proved in Section 8.1, while Propositions 2.5 and 2.6 will be proved in Section 8.2.

**Proposition 7.12.** *For each fixed  $t$ , the restriction to  $\Sigma_t$  of the function*

$$\frac{\partial s_t}{\partial \theta}(t, \lambda)$$

*is the unique function that on  $\Sigma_t$  that (1) extends continuously to the boundary, (2) agrees with the  $\theta$ -derivative of  $\log(|\lambda - 1|^2)$  on the boundary, and (3) is independent of  $r = |\lambda|$ . Thus, the function  $m_t$  in Corollary 7.6 is given by*

$$m_t(\theta) = \frac{2r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta}, \quad (7.23)$$

where  $r_t(\theta)$  is the outer radius of the domain  $\Sigma_t$  (Figure 3).

*Proof.* We have already established in Corollary 7.6 that  $\partial s_t / \partial \theta$  is independent of  $\rho$  (or equivalently, of  $r$ ) in  $\Sigma_t$ . Then Propositions 7.10 and 7.11 tell us that  $\partial s_t / \partial \theta$  is continuous up to the boundary and agrees there with the angular derivative of  $\log(|\lambda - 1|^2)$ . Thus, to compute  $\partial s_t / \partial \theta$  at a point in  $\Sigma_t$ , we travel along a radial segment (in either direction) until we hit the boundary at radius  $r_t(\theta)$  or  $1/r_t(\theta)$ . We then evaluate the angular derivative of  $\log(|\lambda - 1|^2)$ , as in (7.5), giving the claimed expression for  $\partial s_t / \partial \theta = m_t(\theta)$ .  $\square$

**Proposition 7.13.** *For each  $t > 0$ , the distributional Laplacian of  $s_t(\lambda)$  with respect to  $\lambda$  may be computed as follows. Take the pointwise Laplacian of  $s_t$  outside  $\bar{\Sigma}_t$  (giving zero), take the pointwise Laplacian of  $s_t$  inside  $\Sigma_t$  (giving the expression (7.16) in Corollary 7.6) and ignore the boundary of  $\Sigma_t$ .*

*Proof.* Since, by Proposition 7.10,  $s_t$  is analytic up to the boundary of  $\Sigma_t$  from the inside, Green's second identity says that

$$\begin{aligned} \int_{\Sigma_t} s_t(\lambda) \Delta \psi(\lambda) d^2 \lambda &= \int_{\Sigma_t} (\Delta s_t(\lambda)) \psi(\lambda) d^2 \lambda \\ &\quad + \int_{\partial \Sigma_t} (s_t(\lambda) \nabla \psi(\lambda) - \psi(\lambda) \nabla s_t(\lambda)) \cdot \hat{n} dS, \end{aligned}$$

for any test function  $\psi$ , where in the last integral, the limiting value of  $\nabla s_t$  from the inside should be used. This identity holds because the boundary of  $\Sigma_t$  is smooth for  $t \neq 4$  and piecewise smooth when  $t = 4$  (Point 3 of Theorem 4.2). We also have similar formula for the integral over the complement of  $\bar{\Sigma}_t$ , provided that  $\psi$  is compactly supported, but with the direction of the unit normal reversed. Proposition 7.11 then tells us that the boundary terms in the two integrals cancel, giving

$$\int_{\mathbb{C}} s_t(\lambda) \Delta \chi(\lambda) d^2 \lambda = \int_{(\bar{\Sigma}_t)^c} (\Delta s_t(\lambda)) \chi(\lambda) d^2 \lambda + \int_{\Sigma_t} (\Delta s_t(\lambda)) \chi(\lambda) d^2 \lambda, \quad (7.24)$$

where the integral over  $(\bar{\Sigma}_t)^c$  is actually zero, since  $\Delta s_t(\lambda) = 0$  there. The formula (7.24) says that the distributional Laplacian of  $s_t$  may be computed by taking the ordinary, pointwise Laplacian in  $\bar{\Sigma}_t$  and in  $\Sigma_t$  and ignoring the boundary of  $\Sigma_t$ .  $\square$

We now have all the ingredients for a proof of Theorem 2.2.

*Proof of Theorem 2.2.* Proposition 7.13 tells us that we can compute the distributional Laplacian of  $s_t$  separately inside  $\Sigma_t$  and outside  $\bar{\Sigma}_t$ , ignoring the boundary. Theorem 7.2 tells us that the Laplacian outside  $\bar{\Sigma}_t$  is zero. Corollary 7.6 gives us the form of  $\Delta s_t$  inside  $\Sigma_t$ , while Proposition 2.6 identifies the function  $m_t$  appearing in Corollary 7.6. The claimed formula for the Brown measure therefore holds.  $\square$

## 8. FURTHER PROPERTIES OF THE BROWN MEASURE

**8.1. The formula for  $\omega$ .** In this subsection, we derive the formula for  $w_t$  given in Proposition 2.3 in terms of the density  $\omega$ . Throughout, we will write the function  $T$  in (4.1) in polar coordinates as

$$T(r, \theta) = (r^2 + 1 - 2r \cos \theta) \frac{\log(r^2)}{r^2 - 1}. \quad (8.1)$$

We start with a simple rewriting of the expression for  $w_t$  in Theorem 2.2.

**Lemma 8.1.** *The density  $w_t(\theta)$  in Theorem 2.2 may also be written as*

$$w_t(\theta) = \frac{1}{2\pi t} \left( 1 + \frac{\partial}{\partial \theta} [h(r_t(\theta)) \sin \theta] \right), \quad (8.2)$$

where

$$h(r) = r \frac{\log(r^2)}{r^2 - 1}.$$

*Proof.* We start by noting that the point with polar coordinates  $(r_t(\theta), \theta)$  is on the boundary of  $\Sigma_t$ . Thus, by Theorem 4.1, we have  $T(r_t(\theta), \theta) = t$ , from which we obtain

$$\frac{1}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta} = \frac{1 \log(r_t(\theta))}{t r_t(\theta)^2 - 1}.$$

Thus, we may write

$$\frac{2r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta} = \frac{2}{t} h(r_t(\theta)) \sin \theta,$$

from which the claimed formula follows easily from the expression in Theorem 2.2.  $\square$

We now formulate the main result of this subsection, whose proof is on p. 58.

**Theorem 8.2.** *Consider the function  $\omega(r, \theta)$  defined in (2.7). Although the right-hand side of (2.7) is indeterminate at  $r = 1$ , the function  $\omega$  has a smooth extension to all  $r > 0$  and all  $\theta$ . The function  $w_t(\theta)$  in Theorem 2.2 can then be expressed as*

$$w_t(\theta) = \frac{1}{2\pi t} \omega(r_t(\theta), \theta).$$

*The function  $\omega$  has the following properties.*

- (1) *We have  $\omega(1/r, \theta) = \omega(r, \theta)$  for all  $r > 0$  and all  $\theta$ .*
- (2) *When  $r = 1$ , we have*

$$\omega(1, \theta) = 3 \frac{1 + \cos \theta}{2 + \cos \theta}.$$

*In particular,  $\omega(1, 0) = 2$  and  $\omega(1, \pi) = 0$ .*

- (3) *The density  $\omega(r, \theta)$  is strictly positive except when  $r = 1$  and  $\theta = \pm\pi$ . Furthermore,  $\omega(r, \theta) \leq 2$  with equality precisely when  $r = 1$  and  $\theta = 0$ .*

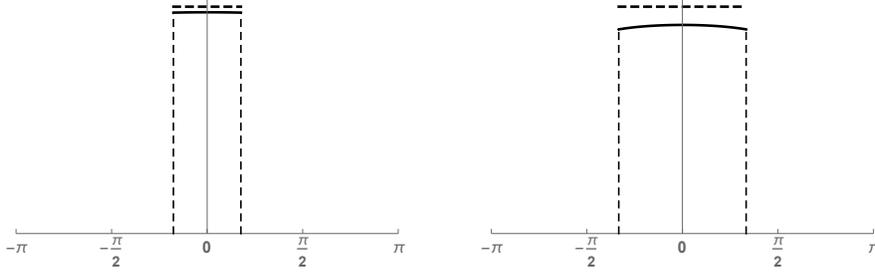


FIGURE 19. Plots of  $w_t(\theta)$  (black) and  $1/(\pi t)$  (dashed) for  $t = 0.3$  and 1

(4) We have

$$\lim_{r \rightarrow 0} \omega(r, \theta) = 1,$$

where the limit is uniform in  $\theta$ .

We now derive consequences for  $w_t$ . For  $t \leq 4$ , the density  $w_t(\theta)$  is only defined for  $-\theta_{\max}(t) < \theta < \theta_{\max}(t)$ , where  $\theta_{\max}(t) = \cos^{-1}(1 - t/2)$ , while for  $t > 4$ , the density  $w_t(\theta)$  is defined for all  $\theta$ . (Recall Theorem 4.2.)

**Corollary 8.3** (Positivity). *If  $t > 4$ , then  $w_t(\theta)$  is strictly positive for all  $\theta$ . If  $t < 4$ , then  $w_t(\theta)$  is strictly positive for  $-\theta_{\max}(t) < \theta < \theta_{\max}(t)$  and the limit as  $\theta$  approaches  $\pm\theta_{\max}(t)$  of  $w_t(\theta)$  is strictly positive. Finally, if  $t = 4$ , then  $w_t(\theta)$  is strictly positive for  $-\pi < \theta < \pi$ , but  $\lim_{\theta \rightarrow \pm\pi} w_t(\theta) = 0$ .*

*Proof.* The only time  $\omega(r, \theta)$  equals zero is when  $r = 1$  and  $\theta = \pm\pi$ . When  $t > 4$ , the function  $r_t(\theta)$  is continuous and greater than 1 for all  $\theta$ , so that  $w_t(\theta)$  is strictly positive in this case. When  $t \leq 4$ , we know from Proposition 4.4 that  $r_t(\theta)$  is greater than 1 for  $|\theta| < \theta_{\max}(t)$  and approaches 1 when  $\theta$  approaches  $\pm\theta_{\max}(t)$ . Thus,  $w_t(\theta) = \omega(r_t(\theta), \theta)$  is strictly positive for  $|\theta| < \theta_{\max}(t)$ . When  $t < 4$ , we have  $\theta_{\max}(t) = \cos^{-1}(1 - t/2) < \pi$  and the limiting value of  $w_t(\theta)$ —namely  $\omega(1, \theta_{\max})$ —will be positive. Finally, when  $t = 4$ , we have  $\theta_{\max}(t) = \pi$  and the limiting value of  $w_t(\theta)$  is  $\omega(1, \pi) = 0$ .  $\square$

**Corollary 8.4** (Asymptotics). *The density  $w_t(\theta)$  has the property that*

$$w_t(\theta) \sim \frac{1}{\pi t}$$

*for small  $t$ . More precisely, for all sufficiently small  $t$  and all  $\theta \in (-\theta_{\max}(t), \theta_{\max}(t))$ , the quantity  $\pi t w_t(\theta)$  is close to 1. Furthermore,*

$$w_t(\theta) \sim \frac{1}{2\pi t}$$

*for large  $t$ . More precisely, for all sufficiently large  $t$  and all  $\theta$ , the quantity  $2\pi t w_t(\theta)$  is close to 1.*

See Figures 19 and 20. The small- and large- $t$  behavior of the region  $\Sigma_t$  can also be determined using the behavior of the function  $T(\lambda)$  near  $\lambda = 1$  (small  $t$ ) and near  $\lambda = 0$  (large  $t$ ), together with the invariance of the region under  $\lambda \mapsto 1/\lambda$ . For small  $t$ , the region resembles a disk of radius  $\sqrt{t}$  around 1, while for large  $t$ ,

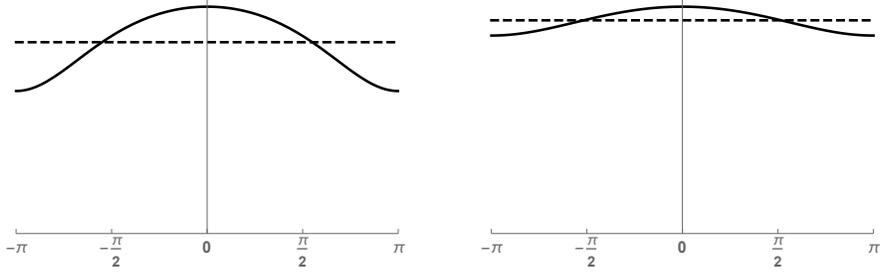


FIGURE 20. Plots of  $w_t(\theta)$  (black) and  $1/(2\pi t)$  (dashed) for  $t = 7$  and 10

the region resembles an annulus with inner radius  $e^{-t/2}$  and outer radius  $e^{t/2}$ . In particular, the expected behavior of the Brown measure for small  $t$  can be observed: it resembles the uniform probability measure on a disk of radius  $\sqrt{t}$  centered at 1.

*Proof.* When  $t$  is small, the entire boundary of  $\Sigma_t$  will be close to  $\lambda = 1$ , since this is the only point where  $T(\lambda) = 0$ . Furthermore, when  $t$  is small,  $\theta_{\max}(t) = \cos^{-1}(1 - t/2)$  is close to zero. When  $t$  is small, therefore, the quantity

$$\pi t w_t(\theta) = \frac{1}{2} \omega(r_t(\theta), \theta)$$

will be close to  $\omega(1, 0)/2 = 1$  for all  $\theta \in (-\theta_{\max}(t), \theta_{\max}(t))$ , by Point 2 of Theorem 8.2.

When  $t$  is large (in particular, greater than 4), the inner boundary of the domain will be close to  $\lambda = 0$ , since this is the only point in the unit disk where  $T(\lambda)$  is large. Thus, for large  $t$ , the inner radius  $1/r_t(\theta)$  of the domain will be uniformly small, and therefore

$$2\pi t w_t(\theta) = \omega(r_t(\theta), \theta) = \omega(1/r_t(\theta), \theta)$$

will be uniformly close to 1, by Point 4 of Theorem 8.2.  $\square$

*Proof of Theorem 8.2.* We note that the function  $T$  in (8.1) can be written as

$$T(r, \theta) = \left( r + \frac{1}{r} - 2 \cos \theta \right) h(r),$$

so that

$$\begin{aligned} \frac{\partial T}{\partial r} &= \left( 1 - \frac{1}{r^2} \right) h(r) + \left( r + \frac{1}{r} - 2 \cos \theta \right) h'(r); \\ \frac{\partial T}{\partial \theta} &= 2 \sin \theta h(r). \end{aligned}$$

Applying implicit differentiation to the identity  $T(r_t(\theta), \theta) = t$  then gives

$$\frac{dr_t(\theta)}{d\theta} = -\frac{\partial T / \partial \theta}{\partial T / \partial r}. \quad (8.3)$$

By the chain rule and (8.3),  $\frac{d}{d\theta}[h(r_t(\theta)) \sin \theta] = q(r_t(\theta), \theta)$ , where

$$\begin{aligned} q(r, \theta) &= h(r) \cos \theta - h'(r) \sin \theta \frac{\partial T / \partial \theta}{\partial T / \partial r} \\ &= h(r) \cos \theta - \frac{2h'(r) \sin^2 \theta h(r)}{\left(1 - \frac{1}{r^2}\right) h(r) + \left(r + \frac{1}{r} - 2 \cos \theta\right) h'(r)}. \end{aligned} \quad (8.4)$$

After computing that

$$h'(r) = \frac{2}{r^2 - 1} - \frac{r^2 + 1}{r(r^2 - 1)} h(r),$$

it is a straightforward but tedious exercise to simplify (8.4) and obtain the claimed formula (2.7).

Since  $h(1/r) = h(r)$ , we may readily verify Point (1); both numerator and denominator in the fraction on the right-hand side of (2.7) change by a factor of  $1/r^2$  when  $r$  is replaced by  $1/r$ .

To understand the behavior of  $\omega$  at  $r = 1$ , we need to understand the function  $h$  better. We may easily calculate that  $h$  has a removable singularity at  $r = 1$  with  $h(1) = 1$ ,  $h'(1) = 0$ , and  $h''(1) = -1/3$ . We also claim that  $h$  satisfies  $0 < h(r) \leq 1$ , with  $h(r) = 1$  only at  $r = 1$ . To verify the claim, we first compute that  $\lim_{r \rightarrow 0} h(r) = 0$  and that

$$h'(r) = \frac{2(r^2 - 1) + (r^2 + 1) \log(1/r^2)}{(r^2 - 1)^2}.$$

Using the Taylor expansion of logarithm, we may then compute that

$$h'(r) = \frac{1}{(r^2 - 1)^2} \sum_{k=3}^{\infty} \left( \frac{2}{k} - \frac{1}{k+1} \right) (1 - r^2)^k > 0$$

for  $0 < r < 1$ . Thus,  $h(r)$  increases from 0 to 1 on  $[0, 1]$ .

We now write  $h$  in the form

$$h(r) = 1 - c(r)(r - 1)^2 \quad (8.5)$$

for some analytic function  $c(r)$ , with  $c(1) = 1/6$ . The minus sign in (8.5) is convenient because  $h$  has a strict global maximum at 1, which means  $c(r)$  is strictly positive everywhere.

Now, since  $h(1) = 1$ , the fraction on the right-hand side of (2.7) is of  $0/0$  form when  $r = 1$ . To rectify this situation, we observe that  $\alpha$  and  $\beta$  may be written as

$$\alpha(r) = (r - 1)^2 [1 + 2rc(r)]; \quad \beta(r) = (r - 1)^2 [1 - (r^2 + 1)c(r)].$$

Thus, we can take a factor of  $(r - 1)^2$  out of numerator and denominator to obtain

$$\omega(r, \theta) = 1 + h(r) \frac{\tilde{\alpha}(r) \cos \theta + \tilde{\beta}(r)}{\tilde{\beta}(r) \cos \theta + \tilde{\alpha}(r)}, \quad (8.6)$$

where  $\tilde{\alpha}(r) = 1 + 2rc(r)$  and  $\tilde{\beta}(r) = 1 - (r^2 + 1)c(r)$ . This expression is no longer of  $0/0$  form at  $r = 1$ . Indeed, since  $h(1) = 1$  and  $c(1) = 1/6$ , we may easily verify the claimed formula for  $\omega(1, \theta)$  in Point 2 of the theorem. We will shortly verify that the denominator in the fraction on the right-hand side of (8.6) is positive for all  $r > 0$  and all  $\theta$ , from which the claimed smooth extension of  $\omega$  follows.

To verify the claimed positivity of  $\omega$ , we first observe that  $\tilde{\beta}(r)z + \tilde{\alpha}(r)$  is positive when  $z = 1$  (with a value of  $2 - (r - 1)^2 c(r) = 1 + h(r)$ ) and also positive when

$z = -1$  (with a value of  $(r+1)^2 c(r)$ ), and hence positive for all  $-1 \leq z \leq 1$ . Thus, the denominator in the fraction on the right-hand side of (8.6) is never zero. We then compute that

$$\frac{d}{dz} \frac{\tilde{\alpha}(r)z + \tilde{\beta}(r)}{\tilde{\beta}(r)z + \tilde{\alpha}(r)} = \frac{\tilde{\alpha}(r)^2 - \tilde{\beta}(r)^2}{(\tilde{\beta}(r)z + \tilde{\alpha}(r))^2} = \frac{(r+1)^2 h(r)}{(\tilde{\beta}(r)z + \tilde{\alpha}(r))^2} > 0$$

for all  $r$  and  $\theta$ . Thus,  $(\tilde{\alpha}(r)z + \tilde{\beta}(r))/(\tilde{\beta}(r)z + \tilde{\alpha}(r))$  increases from  $-1$  to  $1$  as  $z$  increases from  $-1$  to  $1$ . Since  $h(r)$  is positive, we conclude that

$$1 - h(r) \leq \omega(r, \theta) \leq 1 + h(r)$$

for all  $r$  and  $\theta$ , with equality when  $\cos \theta = -1$  in the first case and when  $\cos \theta = 1$  in the second case. Since  $h(r) \leq 1$  with equality only at  $r = 1$ , we see that  $\omega(r, \theta)$  is positive except when  $r = 1$  and  $\cos \theta = -1$ . Similarly,  $\omega(r, \theta) \leq 2$  with equality only if  $r = 1$  and  $\cos \theta = 0$ .

Finally, from the definition (8.5) and the fact that  $\lim_{r \rightarrow 0} h(r) = 0$ , we find that  $\lim_{r \rightarrow 0} c(r) = 1$ . Thus, as  $r \rightarrow 0$ , we have  $\tilde{\alpha}(r) \rightarrow 1$  and  $\tilde{\beta}(r) \rightarrow 0$ . In this limit, the fraction on the right-hand side of (8.6) converges uniformly to  $\cos \theta$ , while  $h(r)$  tends to zero, giving Point 4.  $\square$

**8.2. The connection to free unitary Brownian motion.** Recall from Theorem 1.1 that the spectral measure  $\nu_t$  of the free unitary Brownian motion  $u_t$  was computed by Biane. In this subsection, we prove Proposition 2.5, which connects the Brown measure of  $b_t$  to Biane's measure  $\nu_t$ . The support of  $\nu_t$  is a proper subset of the unit circle for  $t < 4$  and the entire unit circle for  $t \geq 4$ . For  $t < 4$ , the support of  $\nu_t$  consists of points with angles  $\phi$  satisfying  $|\phi| \leq \phi_{\max}(t)$ , where

$$\phi_{\max}(t) = \frac{1}{2} \sqrt{t(4-t)} + \cos^{-1}(1-t/2).$$

Recall the definition in (1.6) of the function  $f_t$ . Then  $f_t$  maps the boundary of  $\Sigma_t$  into the unit circle. (This is true by the definition (2.1) for points in  $\partial\Sigma_t$  outside the unit circle and follows by continuity for points in  $\partial\Sigma_t$  in the unit circle.) Indeed, let the *outer boundary* of  $\Sigma_t$ , denoted  $\partial\Sigma_t^{\text{out}}$ , be the portion of  $\partial\Sigma_t$  outside the open unit disk. Then  $f_t$  is a homeomorphism of  $\partial\Sigma_t^{\text{out}}$  to the support of  $\nu_t$ :

$$f_t : \partial\Sigma_t^{\text{out}} \leftrightarrow \text{supp}(\nu_t). \quad (8.7)$$

In particular, for  $t < 4$ , let us define

$$\theta_{\max}(t) = \cos^{-1}(1-t/2),$$

so that the two points in  $\partial\Sigma_t \cap S^1$  have angles  $\pm\theta_{\max}(t)$  (Theorem 4.2). Then

$$f_t(e^{i\theta_{\max}(t)}) = e^{i\phi_{\max}(t)},$$

as may be verified by direct computation from the definition of  $f_t$ . (Use the formula (8.11) below with  $r = 1$  and  $\cos \theta = 1 - t/2$ .)

We now describe the map (8.7) more concretely. We denote by  $\lambda_t(\theta)$  the point at angle  $\theta$  in  $\partial\Sigma_t^{\text{out}}$ :

$$\lambda_t(\theta) = r_t(\theta)e^{i\theta},$$

where for  $t < 4$ , we require  $|\theta| \leq \theta_{\max}(t)$ . Then the map in (8.7) can be thought of as a map of  $\theta$  to  $\phi$  determined by the relation

$$f_t(\lambda_t(\theta)) = e^{i\phi}. \quad (8.8)$$

We now observe a close relationship between the density  $w_t(\theta)$  in Theorem 2.2 and the map in (8.8).

**Proposition 8.5.** *Let  $\phi$  and  $\theta$  be related as in (8.8), where if  $t < 4$ , we require  $|\phi| \leq \phi_{\max}(t)$  and  $|\theta| \leq \theta_{\max}(t)$ . Then the density  $w_t$  in Theorem 2.2 may be computed as*

$$w_t(\theta) = \frac{1}{2\pi t} \frac{d\phi}{d\theta}. \quad (8.9)$$

We may also write this formula as a logarithmic derivative of  $f_t$  along the outer boundary of  $\Sigma_t$ :

$$w_t(\theta) = \frac{1}{2\pi t} \frac{1}{i} \frac{\frac{d}{d\theta} f_t(\lambda_t(\theta))}{f_t(\lambda_t(\theta))}. \quad (8.10)$$

*Proof.* We compute that

$$\operatorname{Im} \left( \frac{1 + \lambda}{1 - \lambda} \right) = \frac{2 \operatorname{Im} \lambda}{|\lambda - 1|^2} = \frac{2r \sin \theta}{r^2 + 1 - 2r \cos \theta}.$$

Thus, using the definition (1.6) of  $f_t$ , we find that

$$\arg(f_t(\lambda)) = \arg \lambda + \arg e^{\frac{t}{2} \frac{1+\lambda}{1-\lambda}} = \theta + t \frac{r \sin \theta}{r^2 + 1 - 2r \cos \theta}. \quad (8.11)$$

Evaluating this expression at the point  $\lambda_t(\theta)$  gives

$$\begin{aligned} \phi &= \arg(f_t(\lambda_t(\theta))) \\ &= \theta + t \frac{r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta}. \end{aligned} \quad (8.12)$$

(Strictly speaking,  $\phi$  and  $\theta$  are only defined “mod  $2\pi$ ,” but for any local continuous version of  $\theta$ , the last expression in (8.12) gives a local continuous version of  $\phi$ .)

Thus,

$$d\phi = \left( 1 + t \frac{d}{d\theta} \frac{r_t(\theta) \sin \theta}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta} \right) d\theta$$

and the formula (8.9) follows easily by recalling the definition (2.5) of  $w_t$ . The expression (8.10) is then obtained by noting that  $\phi = \frac{1}{i} \log f_t(\lambda_t(\theta))$ .  $\square$

**Proposition 8.6.** *Biane’s measure  $\nu_t$  may be computed as*

$$d\nu_t(\phi) = \frac{r_t(\theta)^2 - 1}{r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta} \frac{d\phi}{2\pi} \quad (8.13)$$

or as

$$d\nu_t(\phi) = \frac{\log(r_t(\theta))}{\pi t} d\phi. \quad (8.14)$$

Here, as usual,  $r_t(\theta)$  is the outer radius of the domain  $\Sigma_t$  and  $\theta$  is viewed as a function of  $\phi$  by inverting the relationship (8.8). When  $t < 4$ , the formula should be used only for  $|\phi| \leq \phi_{\max}(t)$ .

*Proof.* We make use of the expression for  $\nu_t$  in Theorem 1.1. If  $\phi$  is in the interior of the support of  $\nu_t$ , then  $\chi_t(e^{i\phi})$  is in the open unit disk, so that the density of  $\nu_t$  is nonzero at this point. Now, since  $\chi_t$  is an inverse function to  $f_t$  we see that  $\chi_t(e^{i\phi})$  is (for  $\phi$  in the interior of the support of  $\nu_t$ ) the unique point  $\lambda$  with  $|\lambda| < 1$  for which  $f_t(\lambda) = e^{i\phi}$ . Thus,

$$d\nu_t(\phi) = \frac{1 - 1/r_t(\theta)^2}{1 + 1/r_t(\theta)^2 - 2 \cos \theta / r_t(\theta)} \frac{d\phi}{2\pi},$$

which reduces to (8.13). Finally, since  $T(\lambda) = t$  on  $\partial\Sigma_t$  (Theorem 4.1), we have

$$(r_t(\theta)^2 + 1 - 2r_t(\theta) \cos \theta) \frac{\log(r_t(\theta)^2)}{r_t(\theta)^2 - 1} = t, \quad (8.15)$$

which allows us to obtain (8.14) from (8.13).  $\square$

We are now ready for the proof of Proposition 2.5.

*Proof of Proposition 2.5.* The distribution of  $\arg \lambda$  with respect to the Brown measure of  $b_t$  is given in (2.8) as  $2 \log(r_t(\theta)) w_t(\theta) d\theta$ , which we write using Proposition 8.5 and Proposition 8.6 as

$$\begin{aligned} 2 \log(r_t(\theta)) w_t(\theta) d\theta &= 2 \log(r_t(\theta)) \frac{1}{2\pi t} \frac{d\phi}{d\theta} d\theta \\ &= \log(r_t(\theta)) \frac{d\phi}{\pi t} \\ &= d\nu_t(\phi), \end{aligned}$$

as claimed.  $\square$

*Proof of Proposition 2.6.* The value  $\Phi_t(\lambda)$  is computed by first taking the argument of  $\lambda$  to obtain  $\theta$  and then applying the map in (8.8) to obtain  $\phi$ . Thus, the first result is just a restatement of Proposition 2.5. For the uniqueness claim, suppose a measure  $\mu$  on  $\Sigma_t$  has the form

$$d\mu(\lambda) = \frac{1}{r^2} g(\theta) r dr d\theta.$$

Then the distribution of the argument  $\theta$  of  $\lambda$  will be, by integrating out the radial variable,  $2 \log(r_t(\theta)) g(\theta) d\theta$ . The distribution of  $\phi$  will then be

$$2 \log(r_t(\theta)) g(\theta) \frac{d\theta}{d\phi} d\phi = 2 \log(r_t(\theta)) g(\theta) \frac{1}{2\pi t w_t(\theta)} d\phi.$$

The only way this can reduce to Biane's measure as computed in (8.14) is if  $g$  coincides with  $w_t$ .  $\square$

## REFERENCES

- [1] M. Anshelevich and A. N. Sengupta, Quantum free Yang-Mills on the plane, *J. Geom. Phys.* **62** (2012), 330–343.
- [2] Z. D. Bai, Circular law, *Ann. Probab.* **25** (1997), 494–529.
- [3] P. Biane, On the free convolution with a semi-circular distribution, *Indiana Univ. Math. J.* **46** (1997), 705–718.
- [4] P. Biane, “Free Brownian motion, free stochastic calculus and random matrices.” In *Free Probability Theory* (Waterloo, ON, 1995), 1–19. Fields Institute Communications 12. Providence, RI: American Mathematical Society, 1997.
- [5] P. Biane, Segal–Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems, *J. Funct. Anal.*, **144** (1997), 232–286.
- [6] P. Biane, Processes with free increments, *Math. Z.* **227** (1998), 143–174.
- [7] P. Biane and R. Speicher, Stochastic calculus with respect to free Brownian motion and analysis on Wigner space, *Probab. Theory Related Fields* **112** (1998), 373–409.
- [8] P. Biane and R. Speicher, Free diffusions, free entropy and free Fisher information, *Ann. Inst. H. Poincaré Probab. Statist.* **37** (2001), 581–606.
- [9] Brown, L. G. Lidskiĭ's theorem in the type II case. In *Geometric methods in operator algebras* (Kyoto, 1983), 1–35, Pitman Res. Notes Math. Ser., 123, Longman Sci. Tech., Harlow, 1986.
- [10] G. Cébron, Free convolution operators and free Hall transform, *J. Funct. Anal.* **265** (2013), 2645–2708.

- [11] A. Dahlqvist and J. Norris, Yang–Mills measure and the master field on the sphere, preprint arXiv:1703.10578 [math.PR].
- [12] B. K. Driver,  $YM_2$ : continuum expectations, lattice convergence, and lassos, *Comm. Math. Phys.* **123** (1989), 575–616.
- [13] B. K. Driver, B. C. Hall, and T. Kemp, The large- $N$  limit of the Segal–Bargmann transform on  $\mathbb{U}_N$ , *J. Funct. Anal.* **265** (2013), 2585–2644.
- [14] B. K. Driver, B. C. Hall, and T. Kemp, Three proofs of the Makeenko–Migdal equation for Yang–Mills theory on the plane, *Commun. Math. Phys.* 351 (2017), **2** (2017), 741–74.
- [15] B. K. Driver, F. Gabriel, B. C. Hall, and T. Kemp, The Makeenko–Migdal equation for Yang–Mills theory on compact surfaces, *Commun. Math. Phys.* 352 (2017), **3** (2017), 967–978.
- [16] L. C. Evans, Partial differential equations. Second edition. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 2010. xxii+749 pp.
- [17] B. Fuglede and R. V. Kadison, On determinants and a property of the trace in finite factors, *Proc. Nat. Acad. Sci. U. S. A.* **37** (1951), 425–431.
- [18] B. Fuglede and R. V. Kadison, Determinant theory in finite factors, *Ann. of Math. (2)* **55** (1952), 520–530.
- [19] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, *J. Math. Phys.* **6** (1965), 440–449.
- [20] V. L. Girko, The circular law. (Russian) *Teor. Veroyatnost. i Primenen.* **29** (1984), 669–679.
- [21] F. Götze and A. Tikhomirov, The circular law for random matrices, *Ann. Probab.* **38** (2010), 1444–1491.
- [22] D. Gross, Two dimensional QCD as a string theory, *Nucl. Phys. B* **400** (1993), 181–208.
- [23] D. Gross and W. Taylor, Two-dimensional QCD is a string theory, *Nucl. Phys. B* **400** (1993), 161–180.
- [24] L. Gross, C. King, and A. N. Sengupta, Two-dimensional Yang–Mills theory via stochastic differential equations, *Ann. Physics* **194** (1989), 65–112.
- [25] L. Gross and P. Malliavin, Hall’s transform and the Segal–Bargmann map. In Itô’s stochastic calculus and probability theory (N. Ikeda, S. Watanabe, M. Fukushima and H. Kunita, Eds.), 73–116, Springer, 1996.
- [26] E. Gudowska-Nowak, R. A. Janik, J. Jurkiewicz, and M. A. Nowak, Infinite products of large random matrices and matrix-valued diffusion, *Nuclear Phys. B* **670** (2003), 479–507.
- [27] O. Feldheim, E. Paquette, and O. Zeitouni, Regularization of non-normal matrices by Gaussian noise, *Int. Math. Res. Not. IMRN* **18** (2015), 8724–8751.
- [28] B. C. Hall, The Segal–Bargmann “coherent state” transform for compact Lie groups. *J. Funct. Anal.* **122** (1994), 103–151.
- [29] B. C. Hall, Harmonic analysis with respect to heat kernel measure, *Bull. Amer. Math. Soc. (N.S.)* **38** (2001), 43–78.
- [30] B. C. Hall, Quantum theory for mathematicians. Graduate Texts in Mathematics, **267**. Springer, New York, 2013. xvi+554 pp.
- [31] B. C. Hall, The Segal–Bargmann transform for unitary groups in the large- $N$  limit, preprint arXiv:1308.0615 [math.RT].
- [32] B. C. Hall, The large- $N$  limit for two-dimensional Yang–Mills theory, *Comm. Math. Phys.* **363** (2018), 789–828.
- [33] B. C. Hall and T. Kemp, Brown measure support and the free multiplicative Brownian motion, preprint arXiv:1810.00153 [math.FA].
- [34] C. Ho, The two-parameter free unitary Segal–Bargmann transform and its Biane–Gross–Malliavin identification, *J. Funct. Anal.* **271** (2016), 3765–3817.
- [35] T. Kemp, The large- $N$  limits of Brownian motions on  $\mathbb{GL}_N$ , *Int. Math. Res. Not.*, (2016), 4012–4057.
- [36] T. Kemp, Heat kernel empirical laws on  $\mathbb{U}_N$  and  $\mathbb{GL}_N$ . *J. Theoret. Probab.* **2** (2017), 397–451.
- [37] T. Kemp, I. Nourdin, G. Peccati, and R. Speicher, Wigner chaos and the fourth moment. *Ann. Probab.* **40** (2012), 1577–1635.
- [38] T. Lévy, The master field on the plane, *Astérisque* No. 388 (2017), ix+201 pp.
- [39] R. Lohmayer, H. Neuberger, and T. Wettig, Possible large- $N$  transitions for complex Wilson loop matrices, *J. High Energy Phys.* 2008, no. 11, 053, 44 pp.
- [40] J. A. Mingo and R. Speicher, Free probability and random matrices. Fields Institute Monographs, 35. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.

- [41] W. Rudin, Principles of mathematical analysis. Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976.
- [42] I. M. Singer, On the master field in two dimensions, In: Functional analysis on the eve of the 21st century, Vol. 1 (New Brunswick, NJ, 1993), 263–281, Progr. Math., 131, Birkhauser Boston, Boston, MA, 1995.
- [43] P. Śniady, Random regularization of Brown spectral measure, *J. Funct. Anal.* **193** (2002), 291–313.
- [44] T. Tao and V. Vu, Random matrices: universality of ESDs and the circular law. With an appendix by Manjunath Krishnapur. *Ann. Probab.* **38** (2010), 2023–2065.
- [45] D. V. Voiculescu, Limit laws for random matrices and free products. *Invent. Math.* **104** (1991), 201–220.
- [46] E. Wigner, Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)* **62** (1955), 548–564.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA SAN DIEGO, LA JOLLA, CA 92093,  
USA

*E-mail address:* `bdriver@ucsd.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NOTRE DAME, NOTRE DAME, IN 46556, USA

*E-mail address:* `bhall@nd.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA SAN DIEGO, LA JOLLA, CA 92093,  
USA

*E-mail address:* `tkemp@math.ucsd.edu`