Probability Theory and Related Fields (2022) 184:209–273
https://doi.org/10.1007/s00440-022-01142-z

The Brown measure of the free multiplicative Brownian motion

Bruce K. Driver[1] · Brian Hall[2] · Todd Kemp[1]

Received: 7 January 2021 / Revised: 1 April 2022 / Accepted: 23 April 2022 /
Published online: 7 October 2022

Abstract

The free multiplicative Brownian motion b_t is the large-N limit of the Brownian motion on $\mathsf{GL}(N;\mathbb{C})$, 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 Σ_t that appeared in the work of Biane. In the present paper, we compute the Brown measure completely. It has a continuous density W_t on $\overline{\Sigma}_t$, which is strictly positive and real analytic on Σ_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 Σ_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 the circular and semicircular laws. We develop several new methods, based on stochastic differential equations and PDE, to prove these results.

Hall's research supported in part by a grant from the Simons Foundation.
Kemp's research supported in part by NSF CAREER Award DMS-1254807, NSF Award DMS-1800733, and NSF Award DMS-2055340.

✉ Todd Kemp
tkemp@ucsd.edu

Bruce K. Driver
bdriver@ucsd.edu

Brian Hall
bhall@nd.edu

1 Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA

2 Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556, USA

Mathematics Subject Classification 60B20 · 46L54 · 35F21

Contents

1 Introduction

1.1 The Brown measure

Let $(\mathcal{A}, \tau)$ be a tracial von Neumann algebra: a von Neumann algebra $\mathcal{A}$ together with a faithful, normal, tracial state $\tau : \mathcal{A} \to \mathbb{C}$. Such algebras frequently arise from large-N limits of random matrix models, with τ playing the role of the normalized trace for matrices. For an element a of $\mathcal{A}$, the notion of the empirical eigenvalue distribution of a matrix is then played by the Brown measure of a, defined as follows [5]. We let

$$s_a(\lambda) = \tau[\log((a - \lambda)^*(a - \lambda))], \tag{1.1}$$

which is defined as a finite real number for almost every λ. (The quantity $s_a(\lambda)$ is twice the logarithm of the Fuglede–Kadison determinant [9, 10] of $a - \lambda$.) Then s_a is a subharmonic function and the Brown measure μ_a of a is defined in terms of the distributional Laplacian of s_a:

$$\mu_a = \frac{1}{4\pi} \Delta s_a(\lambda).$$

The definition of the Brown measure is the operator-algebra counterpart to Girko's method [11] of computing the empirical eigenvalue distribution of a random matrix.

By regularizing the right-hand side of (1.1), one can construct the Brown measure μ_a as a weak limit,

$$d\mu_a(\lambda) = \lim_{\varepsilon\to 0^+} \frac{1}{4\pi}\Delta_\lambda \tau[\log[(a-\lambda)^*(a-\lambda)+\varepsilon)]\,d\lambda, \tag{1.2}$$

where Δ_λ is the Laplacian with respect to λ and $d\lambda$ is the Lebesgue measure on the plane (see [23, Section 11.5] and [18, Eq. (2.11)]). Here, the positive parameter ε regularizes the logarithm, so that $\tau[\log[(a-\lambda)^*(a-\lambda)+\varepsilon)]$ is a smooth function of $\lambda \in \mathbb{C}$.

In general, μ_a is a probability measure supported on the spectrum of a. If a happens to be a normal operator, μ_a coincides with the law (or distribution) of a. That is to say, if a is normal, $\mu_a(E) = \tau(P_a(E))$, where P_a is the projection-valued measure associated to a by the spectral theorem.

1.2 The free unitary and multiplicative Brownian motions

Let σ_t be a free semicircular Brownian motion (e.g., [4, Section 1.1]) and let c_t be a free circular Brownian motion, which may be constructed as $c_t = (x_t + iy_t)/\sqrt{2}$, where x_t and y_t are freely independent semicircular Brownian motions. These are the large-N limits, in the sense of $*$-distribution, of Brownian motions in the space of Hermitian $N \times N$ matrices and in the space of all $N \times N$ matrices, respectively. We then introduce the free unitary Brownian motion u_t and the free multiplicative Brownian motion b_t given by the free stochastic differential equations

$$\begin{aligned} du_t &= iu_t\,d\sigma_t - \tfrac{1}{2}u_t\,dt \\ db_t &= b_t\,dc_t, \end{aligned}$$

both starting at 1. These processes are the large-N limits of Brownian motions in the unitary group and in the general linear group, respectively. (For the free unitary Brownian motion, this limiting result is due to Biane, while for the free multiplicative Brownian motion it was conjectured by Biane [3] and proved by the third author [20]).

Biane also computed the law ν_t of u_t. We now record this result, since it relates directly to the results of the present paper (Sect. 2.2). Let f_t denote the holomorphic function on $\mathbb{C}\backslash\{1\}$ defined by

$$f_t(\lambda) = \lambda e^{\frac{t}{2}\frac{1+\lambda}{1-\lambda}}. \tag{1.3}$$

Then f_t has a holomorphic inverse χ_t in the open unit disk, and χ_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 [2, 3]). *The spectral measure ν_t of the free unitary Brownian motion u_t is supported in the arc*

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

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

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

See, for example, p. 275 in [3]. In the present paper, we compute the Brown measure μ_{b_t} of the free *multiplicative* Brownian motion and show a direct relationship between μ_{b_t} and the law ν_t of the free *unitary* Brownian motion u_t (Sect. 2.2).

1.3 The Brown measure of b_t

The main result of this paper is a formula for the Brown measure μ_{b_t} of the free multiplicative Brownian motion b_t.

A previous result [18] of the second and third authors showed that the *support* of μ_{b_t} is contained in closure of a certain region Σ_t introduced by Biane in [3]; see Fig. 1 and Definition 2.1. (We reprove that result in the present paper by a different method; see Theorem 6.2 in Sect. 6.2). Nonrigorous results on the support of the Brown measure were also obtained in the physics literature by Gudowska-Nowak et al. [13] and then by Lohmayer et al. [22]. None of the results mentioned in this paragraph say anything about the actual Brown measure itself—only about its support.

We conjecture that the Brown measure of b_t coincides with the limiting eigenvalue distribution of the corresponding Brownian motion B_t^N in the general linear group (see Fig. 2). Proving such results is, however, well known to be a difficult problem, which we do not address here.

Since the first version of this paper appeared on the arXiv, four subsequent works have appeared that use the techniques developed here to analyze Brown measures of other operators. First, work of Ho and Zhong [19] has extended the results of the present paper to the case of a free multiplicative Brownian motion with an arbitrary unitary initial condition. This means that they compute the Brown measure of ub_t, where u is a unitary element freely independent of b_t. Ho and Zhong also compute the Brown measure of $x_0 + c_t$, where c_t is a free circular Brownian motion and x_0 is a self-adjoint element freely independent of c_t. Second, Demni and Hamdi [7] have computed the support of the Brown measure of $u_t P$, where u_t is the free unitary Brownian motion and P is a projection freely independent of u_t. Third, Hall and Ho [16] have computed the Brown measure of $x_0 + i\sigma_t$, where σ_t is the free semicircular Brownian motion and x_0 is a self-adjoint element freely independent of x_t. Last, Hall and Ho [17] have computed the Brown of $ub_{s,\tau}$, where $b_{s,\tau}$ is a family of multiplicative Brownian motions allowing different diffusion rates in the Hermitian and skew-Hermitian directions and where u is freely independent of $b_{s,\tau}$.

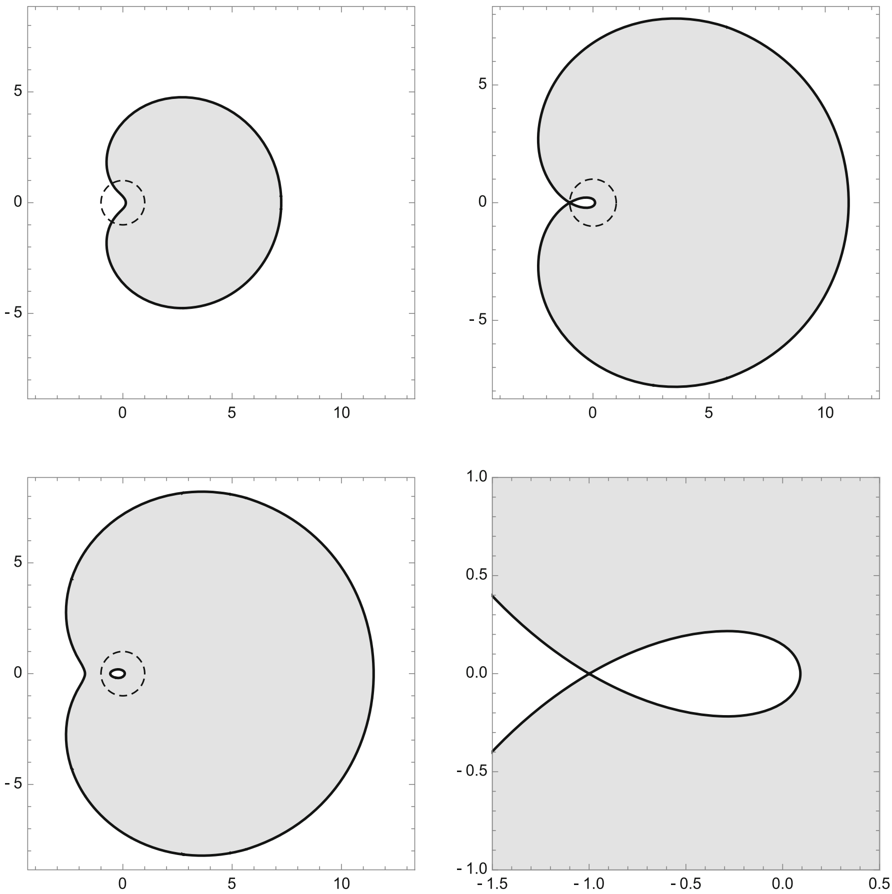

Fig. 1 The first three images show Σ_t for $t = 3$, $t = 4$, and $t = 4.1$, with the unit circle indicated for comparison. The last image shows a detail of the $t = 4$ case

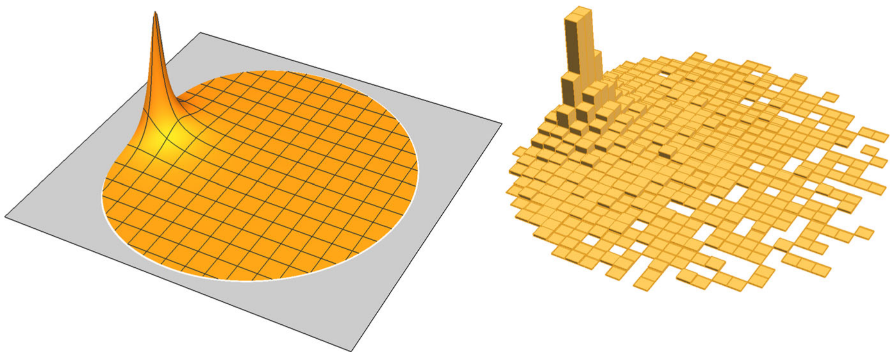

Fig. 2 The Brown measure μ_{b_t} (left) and the eigenvalues of a simulation of the corresponding Brownian motion B_t^N (right), for $t = 1$ and $N = 2000$

The reader may also consult the expository article [15] by the second author, which provides a nontechnical introduction to the techniques used in the present paper. See also the paper [12] of Grela, Nowak, and Tarnowski that explains the PDE method from a physical perspective.

2 Statement of main results

2.1 A formula for the Brown measure

To state our main result, we need to briefly describe the regions Σ_t. For each $t > 0$, consider the holomorphic function f_t on $\mathbb{C}\backslash\{1\}$ defined by (1.3). 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

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

and

$$E_t = \overline{F}_t. \tag{2.2}$$

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

We will show (Theorem 3.1) that Σ_t may also be characterized as

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

where $T(\lambda) = |\lambda - 1|^2 \log(|\lambda|^2)/(|\lambda|^2 - 1)$. Each region Σ_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 θ, if this ray intersects Σ_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 Figs. 3 and 4). See Sect. 3 for more information.

We are now ready to state our main result.

Theorem 2.2 *For all $t > 0$, the Brown measure μ_{b_t} of b_t is absolutely continuous with respect to the Lebesgue measure on the plane and $\mu_{b_t}(\Sigma_t) = 1$. In Σ_t, the density W_t of μ_{b_t} with respect to the Lebesgue measure is strictly positive and real analytic, with the following form in polar coordinates:*

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

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

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

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

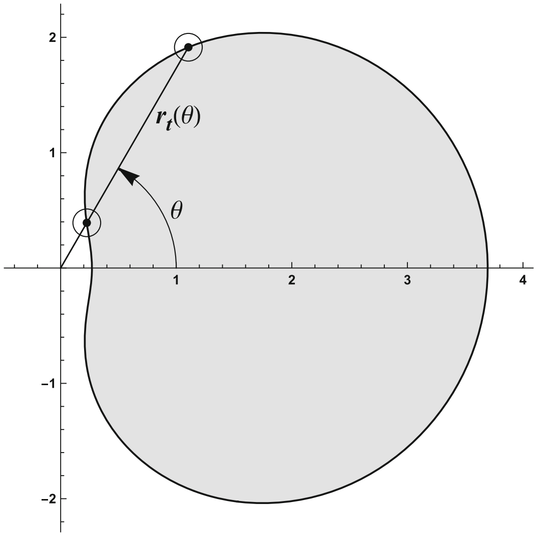

Fig. 3 We let $r_t(\theta)$ denote the larger of the two radii where the ray with angle θ intersects $\partial\Sigma_t$. Shown for $t = 1.5$

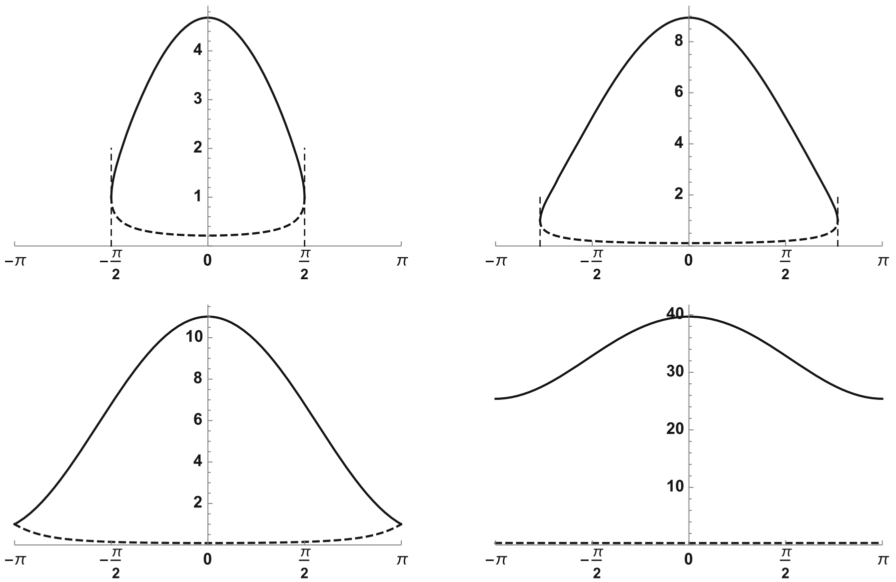

Fig. 4 Graphs of $r_t(\theta)$ (black) and $1/r_t(\theta)$ (dashed) for $t = 2$, 3.5, 4, and 7

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

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

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 Σ_t. According to Proposition 7.5, w_t can also be computed as a logarithmic derivative along the boundary of Σ_t of the function f_t in (1.3).

It follows from (2.3) that the function T equals t on the boundary of Σ_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 θ. We may then use this computation to rewrite (2.5) in a form that no longer involves a derivative with respect to θ, as follows.

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

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

Here

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

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 Σ_t and then parametrize the boundary by the angle θ; see Fig. 5. Using Proposition 2.3, we can derive small- and large-t asymptotics of $w_t(\theta)$ as follows:

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

See Sect. 7 for details.

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

Corollary 2.4 *The Brown measure μ_{b_t} of b_t has the following properties.*

(1) *μ_{b_t} is invariant under the maps $\lambda \mapsto 1/\lambda$ and $\lambda \mapsto \bar{\lambda}$.*

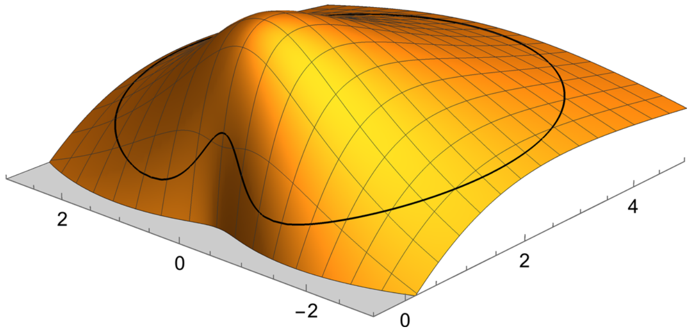

Fig. 5 The function $w_t(\theta)$ is computed by evaluating ω on the boundary of Σ_t and parametrizing the boundary by the angle θ. Shown for $t = 2$

(2) *Let Ξ_t denote the image of $\Sigma_t \backslash (-\infty, 0)$ under the complex logarithm map, using the standard branch cut along the negative real axis. We write points $z \in \Xi_t$ as (ρ, θ). Then for points in Ξ_t, the pushforward of μ_{b_t} by the logarithm map has density $\omega_t(\rho, \theta)$ given by*

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

independent of ρ.

Proof As we have stated above, the region Σ_t is invariant under the maps $\lambda \mapsto 1/\lambda$ and $\lambda \mapsto \bar{\lambda}$. The invariance of μ_{b_t} under $\lambda \mapsto \bar{\lambda}$ follows from the fact that w_t is even. Now, we may compute μ_{b_t} in polar coordinates as

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

where $\log r$ and θ are the real and imaginary parts of the complex logarithm of $\lambda = re^{i\theta}$, as claimed in Point 2. The invariance of the measure under $\lambda \mapsto 1/\lambda$, that is, under $(r, \theta) \mapsto (1/r, -\theta)$ is now evident. □

Plots of $w_t(\theta)$ are shown in Fig. 6. Note that for $t < 4$, not all angles θ actually occur in the domain Σ_t. Thus, for $t < 4$, the function $w_t(\theta)$ is only defined for θ in a certain interval $(-\theta_{\max}(t), \theta_{\max}(t))$—where, as shown in Sect. 3, $\theta_{\max}(t) = \cos^{-1}(1 - t/2)$. Plots of W_t for $t = 1$ and $t = 4$ are then shown in Fig. 7. Actually, when $t = 1$, the function w_t is almost constant. Thus, the variation in W_t in the top part of Fig. 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 μ_{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 when $t = 4$, it is *not* correct to say that most of the mass of μ_{b_t} is near the origin.

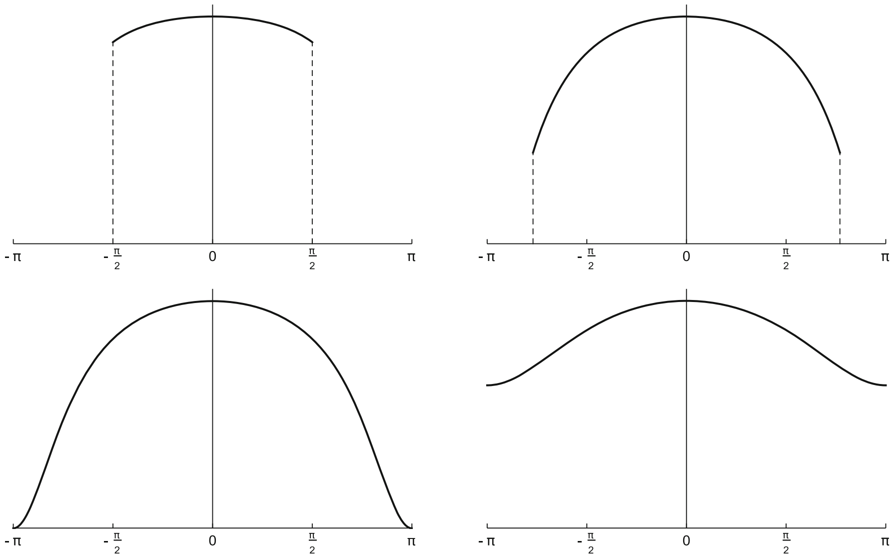

Fig. 6 Plots of $w_t(\theta)$ for $t = 2$, 3.5, 4 and 7

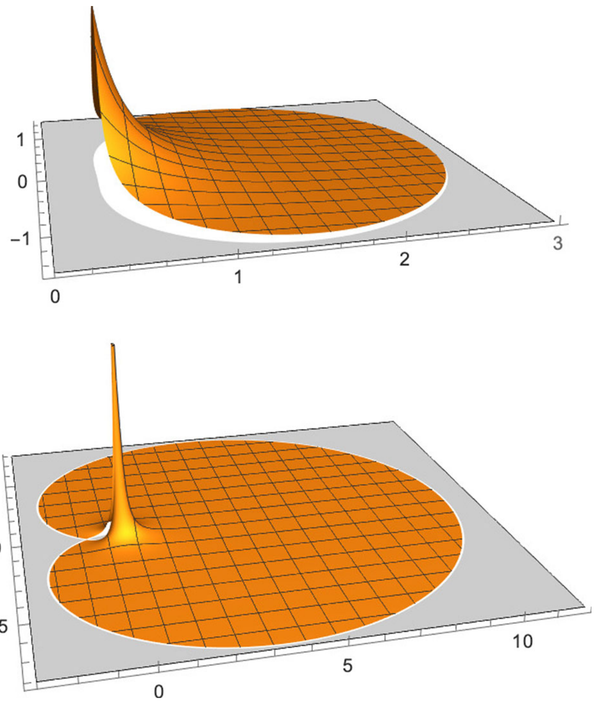

Fig. 7 The density W_t with $t = 1$ (top) and $t = 4$ (bottom)

2.2 A connection to free unitary Brownian motion

It follows easily from Theorem 2.2 that the distribution of the argument of λ with respect to μ_{b_t} has a density given by

$$a_t(\theta) = 2\log[r_t(\theta)]w_t(\theta), \tag{2.8}$$

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 θ.

Recall from Theorem 1.1 that the limiting eigenvalue distribution ν_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 ν_t by a natural change of variables. To each angle θ arising in the region Σ_t, we associate another angle ϕ by the formula

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

where f_t is as in (1.3). (Recall that, by Definition 2.1, the boundary of Σ_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 ν_t.

Proposition 2.5 *If θ is distributed according to the density in (2.8) and ϕ is defined by (2.9), then ϕ is distributed as Biane's measure ν_t.*

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

$$\Phi_t : \overline{\Sigma}_t \to S^1$$

by requiring (a) that Φ_t should agree with f_t on the boundary of Σ_t, and (b) that Φ_t should be constant along each radial segment inside $\overline{\Sigma}_t$, as in Fig. 8. (This specification makes sense because f_t has the same value at the two boundary points on each radial segment). Explicitly, Φ_t may 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 ν_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 Φ_t is Biane's measure ν_t on S^1. Indeed, the Brown measure of b_t is the* unique *measure μ on $\overline{\Sigma}_t$ with the following two properties: (1) the push-forward of μ by Φ_t is ν_t and (2) μ 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.

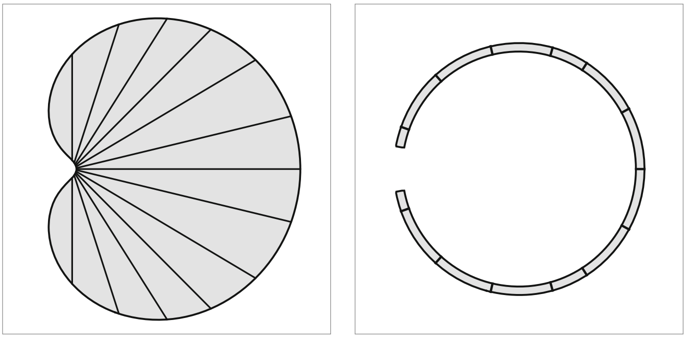

Fig. 8 The map $\Phi_t : \overline{\Sigma}_t \to S^1$ coincides with f_t on $\partial\Sigma_t$ and maps each radial segment in Σ_t to a single point in S^1

Now, the results of [3, 18] already suggest a relationship between the free unitary Brownian motion u_t (whose law is ν_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 μ_{b_t} and ν_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 $\overline{\mathbb{D}}(\sqrt{t})$ 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 $\mathbb{C}\backslash\overline{\mathbb{D}}(\sqrt{t})$ onto $\mathbb{C}\backslash[-2\sqrt{t}, 2\sqrt{t}]$. The net result of these two operations is the map $(x, y) \mapsto 2x$, and the push-forward of the uniform measure on $\overline{\mathbb{D}}(\sqrt{t})$ under this map is the semicircular measure on $[-2\sqrt{t}, 2\sqrt{t}]$.

2.3 Deriving the formula

We now briefly indicate the method we will use to compute the Brown measure μ_{b_t}. Following the general construction of the Brown measure in (1.2), we consider the function S defined by

$$S(t, \lambda, \varepsilon) = \tau[\log((b_t - \lambda)^*(b_t - \lambda) + \varepsilon)] \tag{2.10}$$

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

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

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

The general theory developed by Brown [5] shows that $s_t(\lambda)$ is a subharmonic function of λ for each fixed t, so that the Laplacian (in the distribution sense) of $s_t(\lambda)$ with respect to λ 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:

$$W(t,\lambda) = \frac{1}{4\pi}\Delta s_t(\lambda). \tag{2.11}$$

See also Chapter 11 in [23] and Section 2.3 in [18] for general information on Brown measures.

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

Theorem 2.7 *The function S in (2.10) satisfies the following PDE:*

$$\frac{\partial S}{\partial t} = \varepsilon\frac{\partial S}{\partial \varepsilon}\left(1 + (|\lambda|^2 - \varepsilon)\frac{\partial S}{\partial \varepsilon} - a\frac{\partial S}{\partial a} - b\frac{\partial S}{\partial b}\right), \quad \lambda = a + ib, \tag{2.12}$$

with the initial condition

$$S(0,\lambda,\varepsilon) = \log(|\lambda - 1|^2 + \varepsilon). \tag{2.13}$$

We emphasize that $S(t,\lambda,\varepsilon)$ is only defined for $\varepsilon > 0$. Although, as we will see, $\lim_{\varepsilon\to 0^+} S(t,\lambda,\varepsilon)$ is finite, $\partial S/\partial\varepsilon$ develops singularities in this limit. Thus, it is *not* correct to formally set $\varepsilon = 0$ in (2.12) 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 λ remains outside $\overline{\Sigma}_t$, but not after this time; see Sect. 6.2).

After verifying this equation (Sect. 4), we will use the Hamilton–Jacobi formalism to analyze the solution (Sect. 5). In the remaining sections, we will then analyze the limit of the solution as ε tends to zero and compute the Laplacian in (2.11). The expository article [15] of the second author provides an introduction to the methods used in the present paper.

By way of comparison, we mention that a similar PDE appeared in Biane's paper [1]. There he studies the spectral measure μ_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 μ_t:

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

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

$$\frac{\partial G(t,z)}{\partial t} = -G(t,z)\frac{\partial G}{\partial z}. \tag{2.14}$$

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

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 (2.14) is not actually the main source of information about μ_t in [1]. By contrast, our analysis of the Brown measure of the free multiplicative Brownian motion b_t is based entirely on the PDE in Theorem 2.7.

Finally, we mention that, for the case of the circular Brownian motion c_t, a PDE similar to the one in Theorem 2.12 appeared in work of Burda et al. [6, Equation (9)].

3 Properties of Σ_t

We now verify some important properties of the regions Σ_t in Definition 2.1. Define

$$T(\lambda) = |\lambda - 1|^2 \frac{\log(|\lambda|^2)}{|\lambda|^2 - 1}. \tag{3.1}$$

Since the function

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

has a removable singularity at $x = 1$ with a limiting value of 1, we interpret $T(\lambda)$ as equaling $|\lambda - 1|^2$ when $|\lambda|^2 = 1$. Then $T(\lambda)$ is a real analytic function on all of $\mathbb{C}\backslash\{0\}$. Since, also,

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

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

Theorem 3.1 *For all $t > 0$, the region Σ_t may be expressed as*

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

and the boundary of Σ_t may be expressed as

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

Thus, each fixed $\lambda \in \mathbb{C}$ will be outside $\overline{\Sigma}_t$ until $t = T(\lambda)$ and will be inside Σ_t for all $t > T(\lambda)$. We may therefore say that $T(\lambda)$ is *the time that the domain Σ_t gobbles up λ*. See Figs. 9 and 10.

Theorem 3.2 *For each $t > 0$, the region Σ_t has the following properties.*

(1) *For $t \leq 4$, we have $|\arg\lambda| \leq \cos^{-1}(1 - t/2)$ for all $\lambda \in \overline{\Sigma}_t$, with equality precisely for the points on the unit circle with $\cos\theta = 1 - t/2$.*

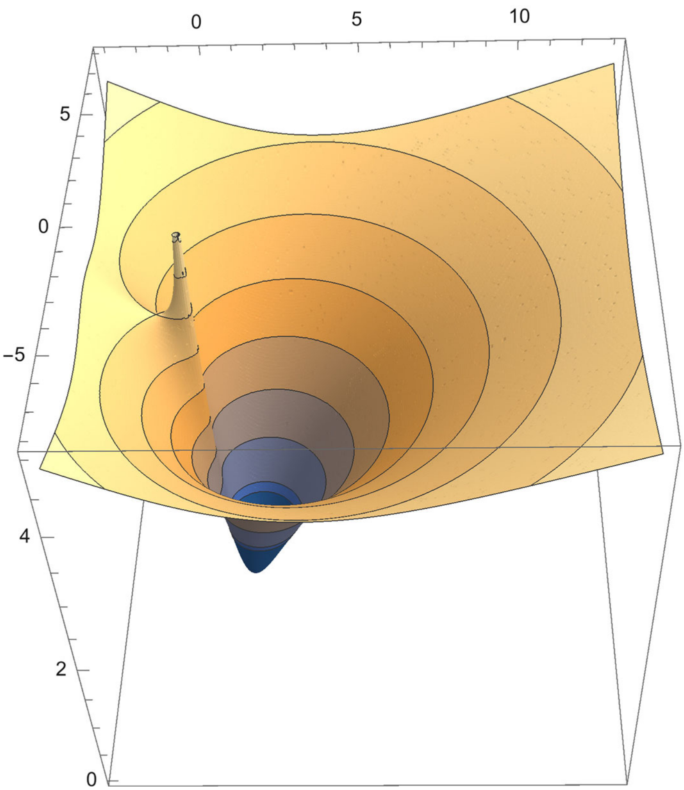

Fig. 9 Plot of the function $T(\lambda)$, showing values between 0 and 5. The function has a global minimum at $\lambda = 1$, a saddle point at $\lambda = -1$, and a pole at $\lambda = 0$

(2) *Consider the ray from the origin with angle θ; if $t \leq 4$, assume $|\theta| < \cos^{-1}(1 - t/2)$. Then this ray intersects Σ_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 Σ_t is smooth for all $t > 0$ with $t \neq 4$. When $t = 4$, the boundary of Σ_t is smooth except at $\lambda = -1$, near which it looks like the transverse intersection of two smooth curves.*
(4) *The region Σ_t is invariant under $\lambda \mapsto 1/\lambda$ and under $\lambda \mapsto \bar{\lambda}$.*
(5) *The region Σ_t coincides with the one defined by Biane in* [3].

We now begin working toward the proofs of Theorems 3.1 and 3.2.

Lemma 3.3 *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 λ, we compute that

$$\log(|f_t(\lambda)|) = \log|\lambda| + \operatorname{Re}\left(\frac{t}{2}\frac{1+\lambda}{1-\lambda}\right)$$

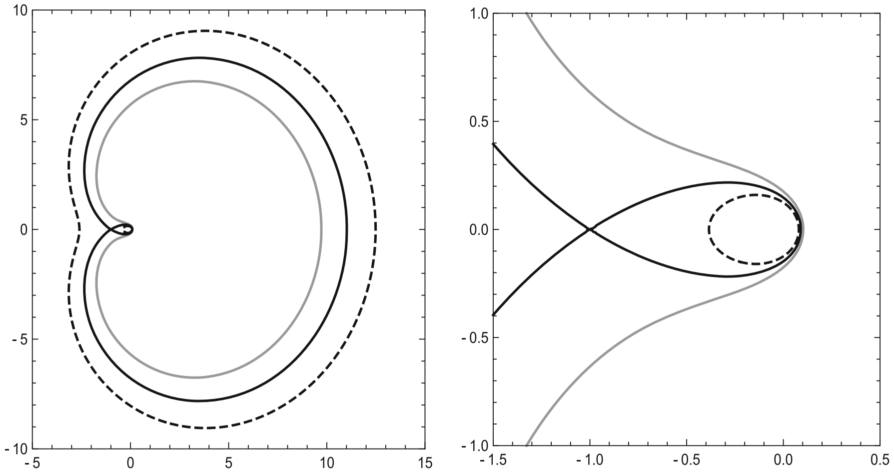

Fig. 10 Level sets of the function $T(\lambda)$ form the boundaries of the regions Σ_t. Shown for $t = 3.7$ (gray), $t = 4$ (black), and $t = 4.3$ (dashed). The right-hand side of the figure gives a close-up view near $\lambda = 0$

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

Thus, for nonzero λ, 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. □

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. 15.

Proposition 3.4 *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 $\left|f_t(re^{i\theta})\right| = 1$, 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 θ and if $t \leq 4$, then $r_t(\theta) \to 1$ as $\theta \to \pm\cos^{-1}(1-t/2)$.*

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 Σ_t. (Recall (2.1) and (2.2)).

Corollary 3.5 *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 θ ranges over all possible angles, and this set is closed.

We now set out to prove Proposition 3.4. In the proof, we will always rewrite the equation $|f_t(\lambda)| = 1$, for $|\lambda| \neq 1$, as $T(\lambda) = t$ (Lemma 3.3).

Lemma 3.6 *Let us write the function T in (3.1) in polar coordinates. Then for each θ, the function $r \mapsto T(r,\theta)$ is strictly decreasing for $0 < r < 1$ and strictly increasing for $r > 1$. For each θ, the minimum value of $T(r,\theta)$, achieved at $r = 1$, is $2(1-\cos\theta)$, and we have*

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

Proof We will show in Proposition 5.13 that the function $T(\lambda)$ is the limit of another function $t_*(\lambda_0, \varepsilon_0)$ as ε_0 goes to zero. Explicitly, this amounts to saying that $T(r,\theta) = g_\theta(\delta)$, where g is defined in (5.58) and $\delta = r + 1/r$. Now, δ is decreasing for $0 < r < 1$ and increasing for $r > 1$. Thus, the claimed monotonicity of T follows if $g_\theta(\delta)$ an increasing function δ for each θ, which we will show in the proof of Proposition 5.16.

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,

$$\frac{\partial T}{\partial r} = 2\left[\frac{-2r + (1+r^2)\cos\theta}{(r^2-1)^2}\right]\log(r^2) + \frac{r^2 + 1 - 2r\cos\theta}{r^2 - 1}\frac{2}{r}. \tag{3.2}$$

We then claim that for all θ, 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 (3.2) 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 θ. Specifically, we may say that

$$\frac{\partial T}{\partial r}(r,\theta) \geq \min\left(\frac{\partial T}{\partial r}(r,0), \frac{\partial T}{\partial r}(r,\pi)\right). \tag{3.3}$$

It is now an elementary (if slightly messy) computation to check that the right-hand side of (3.3) is strictly positive for all $r > 1$. A similar argument then shows that $\partial T/\partial r$ is negative for all θ and all $0 < r < 1$.

We conclude that for each θ, 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

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

and for r approaching infinity, we have

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

as claimed. $\square$

Proof of Proposition 3.4 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 θ) and $t \leq 4$ ($|\theta| < \cos^{-1}(1 - t/2)$). Thus, Lemma 3.6 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 θ.

Now, if $t \leq 4$ and θ 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 (3.3), which is continuous everywhere and strictly positive for $r > 1$. Then for $r > 1$, we have

$$T(r, \theta) - (2 - 2\cos\theta) \geq G(r) := \int_1^r g(s)\, ds. \tag{3.4}$$

Now, $G(r)$ is continuous and strictly increasing for $r \geq 1$, 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 θ sufficiently close to $\pm\cos^{-1}(1 - t/2)$ that $2 - 2\cos\theta$ is within δ 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 (3.4) 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 $r_t(\theta) \to 1$ as $\theta \to \pm\cos^{-1}(1 - t/2)$.

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 $T(r, \theta) = t$. □

We are now ready for the proofs of our main results about Σ_t.

Proof of Theorem 3.1 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 3.3, the set of λ with $|\lambda| \neq 1$ where $T(\lambda) = t$. Then by Corollary 3.5, 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 3.5, 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 Σ_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 Σ_t.

It is also clear from Corollary 3.5 that the boundary of the region Σ_t (i.e., the connected component of E_t^c containing 1) contains the entire set $E_t = \{\lambda | T(\lambda) = t\}$. □

Proof of Theorem 3.2 Point 1 follows easily from Corollary 3.5. For Point 2, we note that by Proposition 3.4, 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 3.1, the ray with angle θ intersects Σ_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 (3.1) 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 Σ_t and 1 is always in Σ_t, the only possible singular point in the boundary of Σ_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 Σ_4.

Meanwhile, the Taylor expansion of T to second order at $\lambda = -1$ is easily found to be $T(\lambda) \approx 4 + (\mathrm{Re}\,\lambda + 1)^2/3 - (\mathrm{Im}\,\lambda)^2$. 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 Σ_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 Σ_t, as we have defined it, coincides with the one originally introduced by Biane [3]. Let us start with the case $t < 4$. According to the discussion at the bottom of p. 273 in [3], the boundary of Biane's domain Σ_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 ν_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 κ on p. 266 of [3] with the definition of χ 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 Σ_t in those cases at the top of p. 274 in [3]. □

4 The PDE for *S*

In this section, we will verify the PDE for S in Theorem 2.7. The claimed initial condition (2.13) holds because $b_0 = 1$. We now proceed to verify the Eq. (2.12) itself. 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 use the notation

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

Lemma 4.1 *The function S in (2.10) satisfies*

$$\frac{\partial S}{\partial t} = \varepsilon\tau\left[\left(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon\right)^{-1}\right]\tau\left[b_t b_t^*\left(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon\right)^{-1}\right]. \tag{4.1}$$

Proof The basic tools for computing with SDEs involving the free circular Brownian motion c_t are the free Itô formulas, which may be stated informally as

$$\begin{aligned} \tau[g_t\, dc_t] = \tau[g_t\, dc_t^*] = 0 \\ dc_t^*\, g_t\, dc_t = dc_t\, g_t\, dc_t^* = \tau[g_t], \end{aligned} \tag{4.2}$$

for a continuous adapted process g_t. Free stochastic calculus was developed by Biane and Speicher [4] and extended by Kümmerer and Speicher [21]. We will specifically use the free stochastic product rule and free functional Itô formula developed by

Nikitopoulos [24]. (An earlier version of this paper, available on the arXiv, used a power series argument in place of Nikitopoulos's result).

For each $\lambda \in \mathbb{C}$, define a self-adjoint element m_t by

$$m_t = b_{t,\lambda}^* b_{t,\lambda}.$$

Then by the free stochastic product rule [24, Thm. 3.2.5] and the free SDE for b_t, we have

$$\begin{aligned} dm_t &= dc_t^* \, b_t^* b_{t,\lambda} + b_{t,\lambda}^* b_t \, dc_t + dc_t^* \, b_t^* b_t \, dc_t \\ &= dc_t^* \, b_t^* b_{t,\lambda} + b_{t,\lambda}^* b_t \, dc_t + \tau[b_t^* b_t] \, dt. \end{aligned} \tag{4.3}$$

Then by the free functional Itô formula [24, Thm. 3.5.3], we have

$$d\tau[\log(m_t + \varepsilon)] = \tau[R \, dm_t] - \frac{1}{2}\tau[R \, dm_t \; R \, dm_t], \tag{4.4}$$

where

$$R = (m_t + \varepsilon)^{-1}.$$

Noting that $S = \tau[\log(m_t + \varepsilon)]$, and using (4.2) and (4.3), (4.4) becomes

$$\frac{\partial S}{\partial t} = \tau[R]\tau[b_t^* b_t] - \tau[R]\tau[b_{t,\lambda}^* b_t b_t^* b_{t,\lambda} R], \tag{4.5}$$

Equation (4.5) is actually just of Eq. (33) in Example 3.5.5 of [24] with $n = 1$,

$$a_t^1 = b_{t,\lambda}^* b_t; \quad b_t^1 = 1; \quad c_t^1 = 1; \quad d_t^1 = b_t^* b_{t,\lambda}; \quad l_t = \tau[b_t^* b_t]1.$$

To compute further, we note that

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

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

$$\left(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon\right)^{-1} b_{t,\lambda} = b_{t,\lambda} \left(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon\right)^{-1}. \tag{4.6}$$

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

$$b_{t,\lambda}^* (b_{t,\lambda} b_{t,\lambda}^* + \varepsilon)^{-1} = (b_{t,\lambda}^* b_{t,\lambda} + \varepsilon)^{-1} b_{t,\lambda}^*. \tag{4.7}$$

We also claim that

$$\tau[(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon)^{-1}] = \tau[(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon)^{-1}]. \tag{4.8}$$

This result can be verified for large ε by expanding both sides in powers of $1/\varepsilon$ and checking the identity term by term. The result for general ε then follows by analyticity of both sides in ε.

We now use (4.7) to show that

$$\begin{aligned} b_{t,\lambda} R b_{t,\lambda}^* &= b_{t,\lambda} b_{t,\lambda}^* \left(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon\right)^{-1} \\ &= \left(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon - \varepsilon\right)\left(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon\right)^{-1} \\ &= 1 - \varepsilon \left(b_{t,\lambda} b_{t,\lambda}^* + \varepsilon\right)^{-1}. \end{aligned}$$

Thus, (4.5) becomes the claimed formula (4.1) for $\partial S/\partial t$. □

Lemma 4.2 *We have the following formulas for the derivatives of S with respect to ε and λ:*

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

Proof The lemma follows easily from the formula for the derivative of the trace of a logarithm (Lemma 1.1 in [5]):

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

(We emphasize that there is no such simple formula for the derivative of $\log(f(u))$ without the trace, unless df/du commutes with $f(u)$). □

We are now ready for the proof of our main result.

Proof of Theorem 2.7 We start from the formula for $\partial S/\partial t$ in Lemma 4.1. 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 (4.1) as

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

We then simplify the first term by writing $b_{t,\lambda}b_{t,\lambda}^* = b_{t,\lambda}b_{t,\lambda}^* + \varepsilon - \varepsilon$. In the middle two terms, we use (4.6), (4.7), and cyclic invariance of the trace. Using also (4.8), we get

$$\begin{aligned}\tau[b_t b_t^*(b_{t,\lambda}b_\lambda^* + \varepsilon)^{-1}] = 1 &+ (|\lambda|^2 - \varepsilon)\tau[(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon)^{-1}] \\ &+ \lambda\tau[b_{t,\lambda}^*(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon)^{-1}] \\ &+ \bar{\lambda}\tau[b_{t,\lambda}(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon)^{-1}]. \end{aligned} \tag{4.9}$$

Thus,

$$\frac{\partial S}{\partial t} = \varepsilon\tau[(b_{t,\lambda}^* b_{t,\lambda} + \varepsilon)^{-1}](\text{all the terms in } (4.9)). \tag{4.10}$$

All terms on the right-hand side of (4.10) are expressible using Lemma 4.2 in terms of derivatives of S, and the claimed differential equation follows. □

5 The Hamilton–Jacobi method

5.1 Setting up the method

The Eq. (2.12) is a first-order, nonlinear PDE of Hamilton–Jacobi type. (The reader may consult, for example, Section 3.3 in the book of Evans [8], but we will give a brief self-contained account of the theory in the proof of Proposition 5.3). We consider a Hamiltonian function obtained from the right-hand side of (2.12) by replacing each partial derivative with momentum variable, with an overall minus sign. Thus, we define

$$H(a, b, \varepsilon, p_a, p_b, p_\varepsilon) = -\varepsilon p_\varepsilon(1 + (a^2 + b^2)p_\varepsilon - \varepsilon p_\varepsilon - ap_a - bp_b). \tag{5.1}$$

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

$$\begin{aligned} \frac{da}{dt} &= \frac{\partial H}{\partial p_a}; & \frac{db}{dt} &= \frac{\partial H}{\partial p_b}; & \frac{d\varepsilon}{dt} &= \frac{\partial H}{\partial p_\varepsilon}; \\ \frac{dp_a}{dt} &= -\frac{\partial H}{\partial a}; & \frac{dp_b}{dt} &= -\frac{\partial H}{\partial b}; & \frac{dp_\varepsilon}{dt} &= -\frac{\partial H}{\partial \varepsilon}. \end{aligned} \tag{5.2}$$

As convenient, we will let

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

The initial conditions for a, b, and ε are arbitrary:

$$a(0) = a_0; \quad b(0) = b_0; \quad \varepsilon(0) = \varepsilon_0, \tag{5.3}$$

while those for p_a, p_b, and p_ε are determined by those for a, b, and ε as follows:

$$p_a(0) = 2(a_0 - 1)p_0; \quad p_b(0) = 2b_0 p_0; \quad p_\varepsilon(0) = p_0, \tag{5.4}$$

where

$$p_0 = \frac{1}{|\lambda_0 - 1|^2 + \varepsilon_0} = \frac{1}{(a_0 - 1)^2 + b_0^2 + \varepsilon_0}. \tag{5.5}$$

The motivation for (5.4) is that the momentum variables p_a, p_b, and p_ε will correspond to the derivatives of S along the curves $(a(t), b(t), \varepsilon(t))$; see (5.8). Thus, the initial momenta are simply the derivatives of the initial value (2.13) of S, evaluated at $(a_0, b_0, \varepsilon_0)$.

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

Lemma 5.1 *The value of the Hamiltonian at $t = 0$ is*

$$H_0 = -\varepsilon_0 p_0^2. \tag{5.6}$$

Proof Plugging $t = 0$ into (5.1) and using (5.4) gives

$$H_0 = -\varepsilon_0 p_0 \left(1 + \left(a_0^2 + b_0^2\right) p_0 - \varepsilon_0 p_0 - 2a_0(a_0 - 1)p_0 - 2b_0^2 p_0\right),$$

which simplifies to

$$H_0 = -\varepsilon_0 p_0 \left(1 - p_0 \left(a_0^2 - 2a_0 + b_0^2 + \varepsilon_0\right)\right).$$

But using the formula (5.5) for p_0, we see that $a_0^2 - 2a_0 + b_0^2 + \varepsilon_0$ equals $1/p_0 - 1$, from which (5.6) follows. □

The main result of this section is the following; the proof is given on p. 22.

Theorem 5.2 *Assume $\lambda_0 \neq 0$ and $\varepsilon_0 > 0$. Suppose a solution to the system (5.2) with initial conditions (5.3) and (5.4) exists with $\varepsilon(t) > 0$ for $0 \le t < T$. Then we have*

$$\begin{aligned} S(t, \lambda(t), \varepsilon(t)) = \log(|\lambda_0 - 1|^2 + \varepsilon_0) - \frac{\varepsilon_0 t}{(|\lambda_0 - 1|^2 + \varepsilon_0)^2} \\ + \log|\lambda(t)| - \log|\lambda_0| \end{aligned} \tag{5.7}$$

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

$$\begin{aligned} \frac{\partial S}{\partial \varepsilon}(t, \lambda(t), \varepsilon(t)) &= p_\varepsilon(t); \\ \frac{\partial S}{\partial a}(t, \lambda(t), \varepsilon(t)) &= p_a(t); \\ \frac{\partial S}{\partial b}(t, \lambda(t), \varepsilon(t)) &= p_b(t). \end{aligned} \tag{5.8}$$

Note that $S(t, \lambda, \varepsilon)$ is only defined for $\varepsilon > 0$. Thus, (5.7) and (5.8) only make sense as long as the solution to (5.2) exists with $\varepsilon(t) > 0$.

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 (5.8) for the derivatives of S will ultimately be of as great importance as the formula (5.7) for S itself. We emphasize that we are not using the Hamilton–Jacobi method to *construct* a solution to (2.12); the function $S(t, \lambda, \varepsilon)$ is already defined in (2.10) in terms of free probability and is known (Theorem 2.7) to satisfy (2.12). 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 5.3 *Fix an open set $U \subset \mathbb{R}^n$, a time-interval $[0, T]$, and a function $H(\mathbf{x}, \mathbf{p})$. Consider a function $S(t, \mathbf{x})$ satisfying*

$$\frac{\partial S}{\partial t} = -H(\mathbf{x}, \nabla_{\mathbf{x}} S), \quad \mathbf{x} \in U,\ t \in [0, T]. \tag{5.9}$$

Consider a pair $(\mathbf{x}(t), \mathbf{p}(t))$ with $\mathbf{x}(t) \in U$, $\mathbf{p}(t) \in \mathbb{R}^n$, and $t \in [0, T_1]$ with $T_1 \leq T$. Assume this pair 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

$$\mathbf{x}(0) = \mathbf{x}_0; \quad \mathbf{p}(0) = (\nabla_{\mathbf{x}} S)(0, \mathbf{x}_0). \tag{5.10}$$

Then we have

$$S(t, \mathbf{x}(t)) = S(0, \mathbf{x}_0) - H(\mathbf{x}_0, \mathbf{p}_0)\, t + \int_0^t \mathbf{p}(s) \cdot \frac{d\mathbf{x}}{ds}\, ds \tag{5.11}$$

and

$$(\nabla_{\mathbf{x}} S)(t, \mathbf{x}(t)) = \mathbf{p}(t). \tag{5.12}$$

Again, we are not trying to construct solutions to (5.9), 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

$$\begin{aligned} \frac{d}{dt} S(t, \mathbf{x}(t)) &= \frac{\partial S}{\partial t}(t, \mathbf{x}(t)) + \frac{\partial S}{\partial x_j}(t, \mathbf{x}(t)) \frac{dx_j}{dt} \\ &= -H(\mathbf{x}(t), (\nabla_{\mathbf{x}} S)(t, \mathbf{x}(t))) + (\nabla_{\mathbf{x}} S)(t, \mathbf{x}(t)) \cdot \frac{d\mathbf{x}}{dt}, \end{aligned} \tag{5.13}$$

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 (5.13) may be rewritten as

$$\frac{d}{dt} S(t, \mathbf{x}(t)) = -H(\mathbf{x}(t), \mathbf{p}(t)) + \mathbf{p}(t) \cdot \frac{d\mathbf{x}}{dt}. \tag{5.14}$$

If we can choose $\mathbf{x}(t)$ so that $\mathbf{p}(t)$ is somehow computable, then the right-hand side of (5.14) 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:

$$\begin{aligned}\frac{dp_j}{dt} &= \frac{d}{dt} \frac{\partial S}{\partial x_j}(t, \mathbf{x}(t)) \\ &= \frac{\partial^2 S}{\partial t \partial x_j}(t, \mathbf{x}(t)) + \frac{\partial^2 S}{\partial x_k \partial x_j}(t, \mathbf{x}(t)) \frac{dx_k}{dt}.\end{aligned} \tag{5.15}$$

Now, from (5.9), 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, (5.15) becomes (suppressing the dependence on the path)

$$\frac{dp_j}{dt} = -\frac{\partial H}{\partial x_j} + \left(\frac{dx_k}{dt} - \frac{\partial H}{\partial p_k} \right) \frac{\partial^2 S}{\partial x_k \partial x_j}. \tag{5.16}$$

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

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}, \tag{5.17}$$

the second term on the right-hand side of (5.16) vanishes, and we find that $\mathbf{p}(t)$ satisfies

$$\frac{dp_j}{dt} = -\frac{\partial H}{\partial x_j}. \tag{5.18}$$

With this choice of $\mathbf{x}(t)$, (5.14) becomes

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

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 Eqs. (5.17) and (5.18) 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 (5.10).

On the other hand, suppose we take a pair $(\mathbf{x}_0,\mathbf{p}_0)$ as in (5.10). Let us then take $\mathbf{x}(t)$ to be the solution to

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}(\mathbf{x}(t),(\nabla_{\mathbf{x}}S)(t,\mathbf{x}(t))),\quad \mathbf{x}(0)=\mathbf{x}_0, \tag{5.19}$$

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 (5.10) and (5.19) says that the pair $(\mathbf{x}(t),\mathbf{p}(t))$ satisfies the first of Hamilton's equations. Applying (5.16) 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 (5.10), the formula (5.14) 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 (5.14), at which point, integration with respect to t gives (5.11). Finally, (5.12) holds by the definition of $\mathbf{p}(t)$. □

We are now ready for the proof of Theorem 5.2.

Proof of Theorem 5.2 We apply Proposition 5.3 with $n=3$ and the open set U consisting of triples (a,b,ε) with $\varepsilon>0$. The PDE (2.12) is of the type in (5.9), with H given by (5.1). The initial conditions (5.4) are obtained by differentiating the initial condition $S(0,\lambda,\varepsilon)=\log(|\lambda-1|^2+\varepsilon)$.

We let $\mathbf{x}(t)=(a(t),b(t),\varepsilon(t))$ and $\mathbf{p}(t)=(p_a(t),p_b(t),p_\varepsilon(t))$. For the case of the Hamiltonian (5.1), a simple computation shows that

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

Thus, the general formula (5.11) becomes, in this case,

$$S(t,\mathbf{x}(t)) = S(0,\mathbf{x}_0) + H(\mathbf{x}_0,\mathbf{p}_0)\,t + \int_0^t \varepsilon(s)p_\varepsilon(s)\,ds. \tag{5.20}$$

But we also may compute that

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

Thus,

$$\int_0^t \varepsilon(s)p_\varepsilon(s)\,ds = \log|\lambda(t)| - \log|\lambda_0|. \tag{5.21}$$

If we now plug in the value of $S(0,\mathbf{x}_0) = S(0,\lambda_0,\varepsilon_0)$ and use Lemma 5.1 along with the definition (5.5) of p_0, we obtain (5.7). Finally, (5.8) is just the general formula (5.12), applied to the case at hand. □

5.2 Constants of motion

We now identify several constants of motion for the system (5.2), from which various useful formulas can be derived. Throughout the section, we assume we have a solution to (5.2) with the initial conditions (5.3) and (5.4), defined on a time-interval of the form $0 \le t < T$. We continue the notation $\lambda(t) = a(t) + ib(t)$.

Proposition 5.4 *Along any solution of* (5.2)*, 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 λ, assuming $\lambda_0 \neq 0$.*

Proof For any system of the form (5.2), 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 [14]. This result can also be verified by direct computation from (5.2).

Finally, note from (5.21) that if $\lambda_0 \neq 0$, then $\log|\lambda(t)|$ remains finite as long as the solution to (5.2) exists, so that $\lambda(t)$ cannot pass through the origin. We then compute that

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

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

Proposition 5.5 *The Hamiltonian H in* (5.1) *in invariant under the one-parameter group of symplectic linear transformations given by*

$$(a,b,\varepsilon,p_a,p_b,p_\varepsilon) \mapsto (e^{\sigma/2}a, e^{\sigma/2}b, e^{\sigma}\varepsilon, e^{-\sigma/2}p_a, e^{-\sigma/2}p_b, e^{-\sigma}p_\varepsilon), \tag{5.22}$$

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

$$\Psi := \varepsilon p_\varepsilon + \frac{1}{2}(ap_a + bp_b) \tag{5.23}$$

is a constant of motion for the system (5.2). The constant Ψ may be computed in terms of ε_0 and λ_0 as

$$\Psi = p_0(a_0(a_0 - 1) + b_0^2 + \varepsilon_0) \tag{5.24}$$

where p_0 is as in (5.5).

Proof The claimed invariance of H is easily checked from the formula (5.1). One can easily check that Ψ is the generator of this family. That is to say, if we replace H by Ψ in (5.2), the solution is given by the map in (5.22). Thus, by a simple general result, Ψ will be a constant of motion; see Conclusion 2.31 in [14]. Of course, one can also check by direct computation that the function in (5.23) is constant along solutions to (5.2). The expression (5.24) then follows easily from the initial conditions in (5.4). □

Proposition 5.6 *For all t, we have*

$$\varepsilon(t)p_\varepsilon(t)^2 = \varepsilon_0 p_0^2 e^{-Ct}, \tag{5.25}$$

where $C = 2\Psi - 1$ and Ψ is as in (5.23). The constant C in (5.25) may be computed in terms of ε_0 and λ_0 as

$$C = p_0(|\lambda_0|^2 - 1 + \varepsilon_0) = \frac{|\lambda_0|^2 - 1 + \varepsilon_0}{|\lambda_0 - 1|^2 + \varepsilon_0}. \tag{5.26}$$

Proof We compute that

$$\begin{aligned}\dot{\varepsilon} &= \frac{\partial H}{\partial p_\varepsilon} = \frac{H}{p_\varepsilon} - \varepsilon p_\varepsilon(a^2 + b^2 - \varepsilon) \\ \dot{p}_\varepsilon &= -\frac{\partial H}{\partial \varepsilon} = -\frac{H}{\varepsilon} - \varepsilon p_\varepsilon^2\end{aligned} \tag{5.27}$$

and then that

$$\begin{aligned}\frac{d}{dt}(\varepsilon p_\varepsilon^2) &= \dot{\varepsilon} p_\varepsilon^2 + 2\varepsilon p_\varepsilon \dot{p}_\varepsilon \\ &= p_\varepsilon H - \varepsilon p_\varepsilon^3(a^2 + b^2 - \varepsilon) - 2Hp_\varepsilon - 2\varepsilon^2 p_\varepsilon^3\end{aligned}$$

This result simplifies to

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

The unique solution to this equation is (5.25). The expression (5.26) is obtained by evaluating Ψ at $t = 0$, using the initial conditions (5.4), and simplifying. □

We now make an important application of preceding results.

Theorem 5.7 *Suppose a solution to* (5.2) *exists with* $\varepsilon(t) > 0$ *for* $0 \le t < t_*$*, but that* $\lim_{t\to t^*} \varepsilon(t) = 0$*. Then*

$$\lim_{t\to t_*} \log|\lambda(t)| = \frac{Ct_*}{2}, \tag{5.28}$$

where $C = 2\Psi - 1$ *is as in Proposition* 5.6*. Furthermore, we have*

$$\lim_{t\to t_*} (ap_a + bp_b) = \lim_{t\to t_*} \frac{2\log|\lambda(t)|}{t} + 1. \tag{5.29}$$

Equation (5.29) is a key step in the derivation of our main result; see Sect. 6.1. We will write (5.28) in a more explicit way in Proposition 5.12, after the time t_* has been determined. We note also from Proposition 5.6 that since $\varepsilon(t)$ approaches zero as t approaches t_*, then $p_\varepsilon(t)$ must be blowing up, so that $\varepsilon(t)p_\varepsilon(t)^2$ can remain positive in this limit.

Proof Using the constant of motion Ψ in (5.23), we can rewrite the Hamiltonian H as

$$H = -\varepsilon p_\varepsilon(1 + (a^2 + b^2)p_\varepsilon - 2\Psi + \varepsilon p_\varepsilon). \tag{5.30}$$

Now, by assumption, the variable ε approaches zero as t approaches t_*. Furthermore, by Proposition 5.6, $\varepsilon p_\varepsilon^2$ remains finite in this limit, so that $\varepsilon p_\varepsilon = \sqrt{\varepsilon}\sqrt{\varepsilon p_\varepsilon^2}$ tends to zero. Thus, in the $t \to t_*$ limit, the $\varepsilon p_\varepsilon$ terms in (5.30) vanish while $\varepsilon p_\varepsilon^2$ remains finite, leaving us with

$$H = -\lim_{t\to t*} \varepsilon p_\varepsilon^2(a^2 + b^2).$$

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

$$\lim_{t\to t_*} (a^2 + b^2) = -\lim_{t\to t_*} \frac{H_0}{\varepsilon p_{\varepsilon^2}} = \lim_{t\to t_*} \frac{\varepsilon_0 p_0^2}{\varepsilon p_\varepsilon^2} = e^{Ct_*}$$

where we have used Lemma 5.1 in the second equality and Proposition 5.6 in the third. The formula (5.28) follows.

Meanwhile, as t approaches t_*, the $\varepsilon p_\varepsilon$ term in the formula (5.23) for Ψ vanishes and we find, using (5.28), that

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

as claimed in (5.29), where we have used (5.28) in the last equality. □

5.3 Solving the equations

We now solve the system (5.2) subject to the initial conditions (5.3) and (5.4). The formula in Proposition 5.6 for $\varepsilon(t)p_\varepsilon(t)^2$ will be a key tool. Although we are mainly interested in the case $\varepsilon_0 > 0$, we will need in Sect. 6.2 to allow ε_0 to be slightly negative.

We begin by with the following elementary lemma.

Lemma 5.8 *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*

$$\dot{y} = y^2 - a^2 \tag{5.31}$$

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

$$y(t) = \frac{y_0 \cosh(at) - a\sinh(at)}{\cosh(at) - y_0 \frac{\sinh(at)}{a}} \tag{5.32}$$

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

$$\begin{aligned} t_* &= \frac{1}{a}\tanh^{-1}\left(\frac{a}{y_0}\right) && (5.33)\\ &= \frac{1}{2a}\log\left(\frac{1 + a/y_0}{1 - a/y_0}\right). && (5.34) \end{aligned}$$

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 (5.33) or (5.34) as having its limiting value as a approaches zero, namely $1/y_0$.

In passing from (5.33) to (5.34), we have used the standard formula for the inverse hyperbolic tangent,

$$\tanh^{-1}(x) = \frac{1}{2}\log\left(\frac{1+x}{1-x}\right). \tag{5.35}$$

In (5.32), 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 (5.31) to obtain

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

It is then straightforward to solve for $y(t)$ and simplify to obtain (5.32). 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 (5.32) satisfies the Eq. (5.31) for all $a \in \mathbb{C}$.

Now, if $a^2 \geq y_0^2 > 0$, the denominator in (5.32) 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

$$y(t) = y_0\frac{\cos(\alpha t) + \frac{\alpha}{y_0}\sin(\alpha t)}{\cos(\alpha t) - \frac{y_0}{\alpha}\sin(\alpha t)}. \tag{5.36}$$

The denominator in (5.36) 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 (5.34). □

We now apply Lemma 5.8 to compute the p_ε-component of the solution to (5.2). We use the following notations, some of which have been introduced previously:

$$p_0 = \frac{1}{|\lambda_0 - 1|^2 + \varepsilon_0} \tag{5.37}$$

$$\delta = \frac{|\lambda_0|^2 + 1 + \varepsilon_0}{|\lambda_0|} \tag{5.38}$$

$$C = 2\Psi - 1 = p_0(|\lambda_0|^2 - 1 + \varepsilon_0) \tag{5.39}$$

$$y_0 = p_0 + \frac{C}{2} = \frac{1}{2}p_0\,|\lambda_0|\,\delta \tag{5.40}$$

$$a^2 = C^2/4 + \varepsilon_0 p_0^2. \tag{5.41}$$

We now make the following standing assumptions:

$$\begin{aligned} \lambda_0 &\neq 0 \\ p_0 &> 0 \\ \delta &> 0. \end{aligned} \tag{5.42}$$

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

$$a = \frac{1}{2} p_0 \, |\lambda_0| \sqrt{\delta^2 - 4}. \tag{5.43}$$

from which we obtain

$$\frac{a^2}{y_0^2} = \frac{\delta^2 - 4}{\delta^2} < 1, \tag{5.44}$$

so that $a^2 < y_0^2$. Now, the assumptions $p_0 > 0$ and $\delta > 0$ can be written as $\varepsilon_0 > -|\lambda_0 - 1|^2$ and $\varepsilon_0 > -(1 + |\lambda_0|^2)$. Thus, for $\lambda_0 \neq 0$, the assumptions (5.42) are always satisfied if $\varepsilon_0 > 0$. Furthermore, except when $\lambda_0 = 1$, some negative values of ε_0 are allowed.

Proposition 5.9 *Under the assumptions* (5.42)*, the p_ε-component of the solution to* (5.2) *subject to the initial conditions* (5.3) *and* (5.4) *is given by*

$$p_\varepsilon(t) = p_0 \frac{\cosh(at) + \frac{2|\lambda_0| - \delta}{\sqrt{\delta^2 - 4}} \sinh(at)}{\cosh(at) - \frac{\delta}{\sqrt{\delta^2 - 4}} \sinh(at)} e^{-Ct} \tag{5.45}$$

for as long as the solution to the system (5.2) *exists. Here we write a as in* (5.43) *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* (5.45)*. If $\delta = 2$, we interpret $\sinh(at)/\sqrt{\delta^2 - 4}$ as equaling $\frac{1}{2} p_0 \, |\lambda_0| \, t$.*

If $\varepsilon_0 \geq 0$, the numerator in the fraction on the right-hand side of (5.45) *is positive for all t. Hence when $\varepsilon_0 \geq 0$, we see that $p_\varepsilon(t)$ is positive for as long as the solution exists and $1/p_\varepsilon(t)$ extends to a real-analytic function of t defined for all $t \in \mathbb{R}$.*

The first time $t_*(\lambda_0, \varepsilon_0)$ at which the expression on the right-hand side of (5.45) *blows up is*

$$t_*(\lambda_0, \varepsilon_0) = \frac{2(\delta - 2\cos\theta_0)}{\sqrt{\delta^2 - 4}} \tanh^{-1}\left(\frac{\sqrt{\delta^2 - 4}}{\delta}\right) \tag{5.46}$$

$$= \frac{\delta - 2\cos\theta_0}{\sqrt{\delta^2 - 4}} \log\left(\frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}}\right), \tag{5.47}$$

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* (5.46)*, with branch cuts* $(-\infty, -1]$ *and* $[1, \infty)$ *on the real axis, which corresponds to using the principal branch of the logarithm in* (5.47)*. When* $\delta = 2$*, we interpret* $t_*(\lambda_0, \varepsilon_0)$ *as having its limiting value as* δ *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 $t_*(\lambda_0, \varepsilon_0)$ does not depend on the choice of square root of $\delta^2 - 4$.

Proof of Proposition 5.9 We assume at first that $\varepsilon_0 \neq 0$. We recall from Proposition 5.6 that $\varepsilon(t) p_\varepsilon(t)^2$ is equal to $\varepsilon_0 p_0^2 e^{-Ct}$, which is never zero, since we assume ε_0 is nonzero and p_0 is positive. Thus, as long as the solution to the system (5.2) exists, both $\varepsilon(t)$ and $p_\varepsilon(t)$ must be nonzero—and must have the same signs they had at $t = 0$. Using (5.27) and the fact that H is a constant of motion, we obtain

$$\dot{p}_\varepsilon(t) = \frac{\varepsilon_0 p_0^2}{\varepsilon(t)} - \varepsilon_0 p_0^2 e^{-Ct}$$

But $\varepsilon_0 p_0^2/\varepsilon(t) = p_\varepsilon(t)^2 e^{Ct}$ and we obtain

$$\dot{p}_\varepsilon(t) = p_\varepsilon(t)^2 e^{Ct} - \varepsilon_0 p_0^2 e^{-Ct}.$$

Then if $y(t) = e^{Ct} p_\varepsilon(t) + C/2$, we find that y satisfies (5.31). Thus, we obtain $p_\varepsilon(t) = (y(t) - C/2)e^{-Ct}$, where $y(t)$ is as in (5.32), which simplifies to the claimed formula for p_ε. The same formula holds for $\varepsilon_0 = 0$, by the continuous dependence of the solutions on initial conditions. (It is also possible to solve the system (5.2) with $\varepsilon_0 = 0$ by postulating that $\varepsilon(t)$ is identically zero and working out the equations for the other variables).

In this paragraph only, we assume $\varepsilon_0 \geq 0$. Then $a^2 \geq 0$, with $a = 0$ occurring only if $\varepsilon_0 = 0$ and $|\lambda_0| = 1$, so that $\delta = 2$. In that case, the numerator on the right-hand side of (5.45) 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

$$(\delta^2 - 4) - (2\,|\lambda_0| - \delta)^2 = 4\varepsilon_0, \tag{5.48}$$

and we are assuming $\varepsilon_0 \geq 0$. Now, since the numerator in (5.45) is always positive, we conclude that p_ε remains positive until it blows up.

For any value of ε_0, the blow-up time for the function on the right-hand side of (5.45) is computed by plugging the expression (5.44) for a/y_0 into the formula (5.34), giving

$$\begin{aligned} t_*(\lambda_0, \varepsilon_0) &= \frac{1}{y_0}\frac{1}{a/y_0}\tanh^{-1}\left(\frac{a}{y_0}\right) \\ &= \frac{2}{p_0\,|\lambda_0|\,\delta}\frac{\delta}{\sqrt{\delta^2-4}}\tanh^{-1}\left(\frac{\sqrt{\delta^2-4}}{\delta}\right). \end{aligned}$$

After computing that

$$\frac{1}{p_0\,|\lambda_0|} = \frac{|\lambda_0 - 1|^2 + \varepsilon_0}{|\lambda_0|} = \delta - 2\cos\theta_0,$$

we obtain the claimed formula (5.46) for $t_*(\lambda_0, \varepsilon_0)$. □

Remark 5.10 If $\varepsilon_0 < 0$, then numerator on the right-hand side of (5.45) can become zero. The time σ at which this happens is computed using (5.43) and (5.48) as

$$\sigma = \frac{2}{p_0\,|\lambda_0|\,\sqrt{\delta^2-4}}\tanh^{-1}\left(-\left(1 + \frac{4\varepsilon_0}{(2\,|\lambda_0| - \delta)^2}\right)^{1/2}\right).$$

By considering separately the cases $|\lambda_0| \neq 1$ and $|\lambda_0| = 1$, we can verify that σ tends to infinity, locally uniformly in λ_0, as ε_0 tends to zero from below. Thus, for small negative values of ε_0, the function on the right-hand side of (5.45) will remain positive until the time $t_*(\lambda_0, \varepsilon_0)$ at which it blows up.

We now show that the whole system (5.2) has a solution up to the time at which the function on the right-hand side of (5.45) blows up.

Proposition 5.11 *Assume that ε_0 and λ_0 satisfy the assumptions (5.42). Assume further that if $\varepsilon_0 < 0$, then $|\varepsilon_0|$ is sufficiently small that p_ε remains positive until it blows up, as in Remark 5.10. Then the solution to the system (5.2) exists up to the time $t_*(\lambda_0, \varepsilon_0)$ in Proposition 5.9.*

For any ε_0, we have

$$\lim_{t \to t_*(\lambda_0, \varepsilon_0)} \varepsilon(t) = 0. \tag{5.49}$$

If $\varepsilon_0 = 0$, the solution has $\varepsilon(t) \equiv 0$ and $\lambda(t) \equiv \lambda_0$.

Proof Let T be the maximum time such that the solution to (5.2) exists on $[0, T)$. We now compute formulas for the solution on this interval. Recall from Proposition 5.9 that if $\varepsilon_0 \geq 0$, then $p_\varepsilon(t)$ remains positive for as long as the solution exists; by Remark 5.10, the same assertion holds if ε_0 is small and negative.

Now, since $\varepsilon p_\varepsilon^2 = \varepsilon_0 p_0^2 e^{-Ct}$, we see that

$$\varepsilon(t) = \frac{1}{p_\varepsilon(t)^2} \varepsilon_0 p_0^2 e^{-Ct}. \tag{5.50}$$

Since $p_\varepsilon(t)$ remains positive until it blows up, $\varepsilon(t)$ remains bounded until time $t_*(\lambda_0, \varepsilon_0)$, at which time $\varepsilon(t)$ approaches zero, as claimed in (5.49). We recall from Proposition 5.4 that the argument of $\lambda(t)$ remains constant. Then as in shown in (5.21), we have

$$\log |\lambda(t)| = \log |\lambda_0| + \int_0^t \varepsilon(s) p_\varepsilon(s)\, ds. \tag{5.51}$$

Finally,

$$\frac{dp_a}{dt} = -\frac{\partial H}{\partial a} = -2a\varepsilon p_\varepsilon^2 + \varepsilon p_\varepsilon p_a \tag{5.52}$$

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 $t_*(\lambda_0, \varepsilon_0)$ at which the right-hand side of (5.45) blows up. Then from the formulas (5.50), (5.51), and (5.52), 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 (5.2) must therefore exist all the way up to time $t_*(\lambda_0, \varepsilon_0)$.

Finally, we note that when $\varepsilon_0 = 0$, (5.50) gives $\varepsilon(t) \equiv 0$ and (5.51) gives $|\lambda(t)| \equiv |\lambda_0|$. Since also the argument of $\lambda(t)$ is constant, we see that $\lambda(t) \equiv \lambda_0$. □

5.4 More about the lifetime of the solution

In light of Propositions 5.9 and 5.11, the lifetime of the solution to the system (5.2) is $t_*(\lambda_0, \varepsilon_0)$, as computed in (5.46) or (5.47). In this subsection, we (1) analyze the behavior of $\log |\lambda(t)|$ as t approaches $t_*(\lambda_0, \varepsilon_0)$, (2) analyze the behavior of $t_*(\lambda_0, \varepsilon_0)$ as ε_0 approaches zero, and (3) show that $t_*(\lambda_0, \varepsilon_0)$ is an increasing function of ε_0 with λ_0 fixed.

Proposition 5.12 *Assume that ε_0 and λ_0 satisfy the assumptions* (5.42). *Then*

$$\lim_{t \to t_*(\lambda_0, \varepsilon_0)} \log |\lambda(t)| = \frac{\delta - 2/|\lambda_0|}{2\sqrt{\delta^2 - 4}} \log \left(\frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}} \right), \tag{5.53}$$

where δ is as in (5.38).

Notice that there is a strong similarity between the formula (5.47) for $t_*(\lambda_0, \varepsilon_0)$ and the expression on the right-hand side of (5.53).

Proof By (5.28) in Theorem 5.7, we have $\lim_{t\to t_*(\lambda_0,\varepsilon_0)} \log|\lambda(t)| = Ct_*(\lambda_0, \varepsilon_0)/2$, where by (5.26),

$$C = \frac{(|\lambda_0|^2 - 1 + \varepsilon_0)/|\lambda_0|}{(|\lambda_0 - 1|^2 + \varepsilon_0)/|\lambda_0|} = \frac{\delta - 2/|\lambda_0|}{\delta - 2\cos\theta_0}.$$

From this result and the second expression (5.47) for $t_*(\lambda_0, \varepsilon_0)$, (5.53) follows easily. □

Proposition 5.13 *If $t_*(\lambda_0, \varepsilon_0)$ is defined by (5.47), then for all nonzero λ_0 we have*

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

where the function T is defined in (3.1). Furthermore, when $\varepsilon_0 = 0$, we have

$$\lim_{t\to t_*(\lambda_0,\varepsilon_0)} \log|\lambda(t)| = \log|\lambda_0|. \tag{5.54}$$

Recall that the formula for $t_*(\lambda_0, \varepsilon_0)$ is defined under the standing assumptions in (5.42). Note that for all $\lambda_0 \neq 0$, the value $\varepsilon_0 = 0$ satisfies these assumptions.

Since $\log(x)/(x - 1) \to 1$ as $x \to 1$, 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 3.1, we have the following consequence.

Corollary 5.14 *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 $\lambda_0 \notin \overline{\Sigma}_t$, we have $t_*(\lambda_0, 0) > t$.*

Proof of Proposition 5.13 In the limit as $\varepsilon_0 \to 0$, 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

$$\sqrt{\delta^2 - 4} = \pm\frac{|\lambda_0|^2 - 1}{|\lambda_0|}. \tag{5.55}$$

In the case $|\lambda_0| = 1$, the limiting value of δ is 2. We then make use of the elementary limit

$$\lim_{\delta\to 2^+} \frac{1}{\sqrt{\delta^2 - 4}} \log\left(\frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}}\right) = 1. \tag{5.56}$$

Thus, using (5.47), we obtain in this case,

$$\lim_{\varepsilon_0 \to 0} t_*(\lambda_0, \varepsilon_0) = 2 - 2\cos\theta_0 = |\lambda_0 - 1|^2, \quad |\lambda_0| = 1,$$

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 (5.55), regardless of the sign of $|\lambda_0|^2 - 1$. We then obtain, using (5.47),

$$\begin{aligned}\lim_{\varepsilon_0 \to 0} t_*(\lambda_0, \varepsilon_0) &= \frac{(|\lambda_0|^2 + 1)/\,|\lambda_0| - 2\cos\theta_0}{(|\lambda_0|^2 - 1)/\,|\lambda_0|} \log\left(\frac{2\,|\lambda_0|^2\,/\,|\lambda_0|}{2/\,|\lambda_0|}\right) \\ &= \frac{|\lambda_0|^2 + 1 - 2\,|\lambda_0|\cos\theta_0}{|\lambda_0|^2 - 1} \log(|\lambda_0|^2) \\ &= T(\lambda_0). \end{aligned} \tag{5.57}$$

A similar calculation, beginning from (5.53), establishes (5.54). □

Remark 5.15 If we began with (5.46) instead of (5.47), 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 (5.35), this expression is easily seen to agree with $T(\lambda_0)$ but is more transparent in its behavior at $|\lambda_0| = 1$.

Proposition 5.16 *For each λ_0, the function $t_*(\lambda_0, \varepsilon_0)$ is a strictly increasing function of ε_0 for $\varepsilon_0 \geq 0$, and*

$$\lim_{\varepsilon_0 \to +\infty} t_*(\lambda_0, \varepsilon_0) = +\infty.$$

Proof We note that the quantity δ in (5.38) is an increasing function of ε_0 with λ_0 fixed, with δ tending to infinity as ε_0 tends to infinity. We note also that if $\varepsilon_0 \geq 0$, then

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

It therefore suffices to show that for each angle θ_0, the function

$$g_{\theta_0}(\delta) := \frac{\delta - 2\cos\theta_0}{\sqrt{\delta^2 - 4}} \log\left(\frac{\delta + \sqrt{\delta^2 - 4}}{\delta - \sqrt{\delta^2 - 4}}\right), \tag{5.58}$$

is strictly increasing, non-negative, continuous function of δ for $\delta \geq 2$ that tends to $+\infty$ as δ 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 (5.56).

Throughout the proof, we use the notation

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

We note that

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

Meanwhile, for large δ, 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 $\delta \to \infty$.

Our definition of $g_{\theta_0}(\delta)$ for $\delta = 2$, together with (5.56), shows that g_{θ_0} is non-negative and continuous there. To show that g_{θ_0} is an increasing function of δ, 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 δ 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

$$(\delta - 2)\left(\gamma + \log\left(\frac{\delta+\gamma}{\delta-\gamma}\right)\right) \tag{5.59}$$

and

$$(\delta + 2)\left(\gamma - \log\left(\frac{\delta+\gamma}{\delta-\gamma}\right)\right). \tag{5.60}$$

Now, (5.59) 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

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

from which we conclude that (5.60) is also positive for all $\delta > 2$. $\square$

5.5 Surjectivity

In Sect. 6.3, we will compute $s_t(\lambda) := \lim_{\varepsilon\to 0^+} S(t,\lambda,\varepsilon)$ for λ in Σ_t. We will do so by evaluating S (and its derivatives) along curves of the form $(t,\lambda(t),\varepsilon(t))$ and then the taking the limit as we approach the time t_* when $\varepsilon(t)$ becomes zero. For this method to be successful, we need the following result, whose proof appears on p. 34.

Theorem 5.17 *Fix $t > 0$. Then for all $\lambda \in \Sigma_t$, there exists a unique $\lambda_0 \in \mathbb{C}$ and $\varepsilon_0 > 0$ such that the solution to (5.2) with these initial conditions exists on $[0,t)$ with $\lim_{u\to t^-}\varepsilon(u) = 0$ and $\lim_{u\to t^-}\lambda(u) = \lambda$. For all $\lambda \in \Sigma_t$, the corresponding λ_0 also belongs to Σ_t.*

Define functions $\Lambda_0^t : \Sigma_t \to \Sigma_t$ and $E_0^t : \Sigma_t \to (0,\infty)$ by letting $\Lambda_0^t(\lambda)$ and $E_0^t(\lambda)$ be the corresponding values of λ_0 and ε_0, respectively. Then Λ_0^t and E_0^t extend to continuous maps of $\overline{\Sigma}_t$ into $\overline{\Sigma}_t$ and $[0,\infty)$, respectively, with the continuous extensions satisfying $\Lambda_t(\lambda) = \lambda$ and $E_0^t(\lambda) = 0$ for $\lambda \in \partial\Sigma_t$.

We first recall that we have shown (Proposition 5.16) that the lifetime of the path to be a strictly increasing function of $\varepsilon_0 \geq 0$ with λ_0 fixed. If λ_0 is outside Σ_t, then by Theorem 3.1 and Proposition 5.13, the lifetime is at least t, even at $\varepsilon_0 = 0$. (That is to say, $T(\lambda_0) = t_*(\lambda_0, 0) \geq t$ for λ_0 outside Σ_t.) Thus, for λ_0 outside Σ_t, the lifetime cannot equal t for $\varepsilon_0 > 0$. On the other hand, if $\lambda_0 \in \Sigma_t$, then $t_*(\lambda_0, 0) < t$ and Proposition 5.16 tells us that there is a unique $\varepsilon_0 > 0$ with $t_*(\lambda_0,\varepsilon_0) = t$.

Lemma 5.18 *Fix $t > 0$. Define maps*

$$\begin{aligned}\varepsilon_0^t &: \Sigma_t \to [0,\infty)\\ \lambda_t &: \Sigma_t \to \mathbb{C}\backslash\{0\}\end{aligned}$$

as follows. For $\lambda_0 \in \Sigma_t$, we let $\varepsilon_0^t(\lambda_0)$ denote the unique positive value of ε_0 for which $t_*(\lambda_0,\varepsilon_0) = t$. Then we set

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

where $\lambda(\cdot)$ is computed with initial conditions $\lambda(0) = \lambda_0$ and $\varepsilon(0) = \varepsilon_0^t(\lambda_0)$. Then both ε_0^t and λ_t extend continuously from Σ_t to $\overline{\Sigma}_t$, with the extended maps satisfying

$\varepsilon_0^t(\lambda_0) = 0$ and $\lambda_t(\lambda_0) = \lambda_0$ for $\lambda_0 \in \partial\Sigma_t$. The extended map λ_t is a homeomorphism of $\overline{\Sigma}_t$ to itself.

We note that the desired function Λ_0^t in Theorem 5.17 is the inverse function to λ_t and that $E_0^t(\lambda) = \varepsilon_0^t(\lambda_t^{-1}(\lambda))$.

Recall from Proposition 5.4 that the argument of $\lambda(t)$ is constant. By the formula (5.28) in Theorem 5.7 together with the expression (5.26) for the constant C, we can write

$$\lambda_t(\lambda_0) = \frac{\lambda_0}{|\lambda_0|}e^{Ct/2} = \frac{\lambda_0}{|\lambda_0|}\exp\left(\frac{t}{2}\frac{|\lambda_0|^2 - 1 + \varepsilon_0^t(\lambda_0)}{|\lambda_0 - 1|^2 + \varepsilon_0^t(\lambda_0)}\right), \tag{5.61}$$

where we have used that $t_*(\lambda_0, \varepsilon_0^t(\lambda_0)) = t$. As noted in the proof of Proposition 5.12, this formula can also be written as

$$\lambda_t(\lambda_0) = \frac{\lambda_0}{|\lambda_0|}\exp\left(\frac{\delta - 2/|\lambda_0|}{2\sqrt{\delta^2-4}}\log\left(\frac{\delta + \sqrt{\delta^2-4}}{\delta - \sqrt{\delta^2-4}}\right)\right), \tag{5.62}$$

where $\delta = (|\lambda_0|^2 + 1 + \varepsilon_0^t(\lambda_0))/|\lambda_0|$.

Proof We start by trying to compute the function ε_0^t, which we will do by finding the correct value of δ and then solving for ε_0^t. Recall that the lifetime $t_*(\lambda_0, \varepsilon_0)$ is computed as $g_{\theta_0}(\delta)$, where δ is as in (5.38) and g_{θ_0} is as in (5.58). As we have computed in (5.57), 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 θ_0 intersects Σ_t and let $r_t(\theta_0)$ be the outer (for definiteness) radius at which this ray intersects the boundary of Σ_t. Then Theorem 3.1 tells us that $T(r_t(\theta_0)e^{i\theta_0}) = t$, and we conclude that

$$g_{\theta_0}\left(\frac{r_t(\theta_0)^2+1}{r_t(\theta_0)}\right) = t. \tag{5.63}$$

Consider, then, some $\lambda_0 \in \Sigma_t$ with $\arg(\lambda_0) = \theta_0$. By the formula (5.47), to find ε_0 with $t_*(\lambda_0, \varepsilon_0) = t$, we first find δ so that $g_{\theta_0}(\delta) = t$. (Note that the value of δ depends only on the argument of λ_0.) We then adjust ε_0 so that $(|\lambda_0|^2 + \varepsilon_0 + 1)/|\lambda_0| = \delta$. Since the correct value of δ is given in (5.63), this means that we should choose ε_0 so that

$$\frac{|\lambda_0|^2 + \varepsilon_0 + 1}{|\lambda_0|} = \frac{r_t(\theta_0)^2 + 1}{r_t(\theta_0)}.$$

We can solve this relation for ε_0 to obtain

$$\varepsilon_0^t(\lambda_0) = |\lambda_0| \left(\frac{r_t(\arg \lambda_0)^2 + 1}{r_t(\arg \lambda_0)} - \frac{|\lambda_0|^2 + 1}{|\lambda_0|} \right). \tag{5.64}$$

Now, we have shown that $r_t(\theta)$ is continuous for the full range of angles θ occurring in $\overline{\Sigma}_t$. Since 0 is not in $\overline{\Sigma}_t$, we can then see that the formula (5.64) is well defined and continuous on all of $\overline{\Sigma}_t$. 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 $\varepsilon_0^t(\lambda_0)$ equals zero.

Now, the point 0 is always outside $\overline{\Sigma}_t$, while the point 1 is always in Σ_t and therefore not on the boundary of Σ_t. Thus, since ε_0^t is continuous on $\overline{\Sigma}_t$ and zero precisely on the boundary, we see from (5.61) that λ_t is continuous on $\overline{\Sigma}_t$. Furthermore, on $\partial\Sigma_t$, we compute $\lambda_t(\lambda_0)$ by putting $\varepsilon_0^t(\lambda_0) = 0$ in (5.61). Suppose now that λ_0 is in $\partial\Sigma_t$. Then $\varepsilon_0^t(\lambda_0) = 0$ and, by Theorem 3.1, the function $T(\lambda_0)$ in (3.1) has the value t, so that

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

Thus, from (5.61), we see that $\lambda_t(\lambda_0) = \lambda_0$.

Consider an angle θ_0 for which the ray $\mathrm{Ray}(\theta_0)$ with angle θ_0 intersects $\overline{\Sigma}_t$ and let δ be chosen so that $g_{\theta_0}(\delta) = t$, noting again that the value of δ depends only on $\theta_0 = \arg \lambda_0$. We now observe from (5.62) that $|\lambda_t(\lambda_0)|$ is a strictly increasing function of $|\lambda_0|$ with δ fixed. Thus, λ_t is a strictly increasing function of the interval $\mathrm{Ray}(\theta_0) \cap \overline{\Sigma}_t$ into $\mathrm{Ray}(\theta_0)$ that fixes the endpoints. Thus, actually, λ_t maps this interval bijectively into itself. Since this holds for all θ_0, we conclude that λ_t maps $\overline{\Sigma}_t$ bijectively into itself. The continuity of the inverse then holds because λ_t is continuous and $\overline{\Sigma}_t$ is compact. $\square$

Proof of Theorem 5.17 We have noted before the statement of Lemma 5.18 that if the desired pair $(\lambda_0, \varepsilon_0)$ exists, λ_0 must be in Σ_t. The lemma then tells us that a unique pair $(\lambda_0, \varepsilon_0)$ exists with $\lambda_0 \in \Sigma_t$. We compute $\Lambda_0^t(\lambda)$ as $\lambda_t^{-1}(\lambda)$ and $E_0^t(\lambda)$ as $\varepsilon_0^t(\lambda_t^{-1}(\lambda))$, both of which extend continuously to $\overline{\Sigma}_t$. For $\lambda \in \partial\Sigma_t$, we have $\lambda_t^{-1}(\lambda) = \lambda$ and $\varepsilon_0^t(\lambda_t^{-1}(\lambda)) = \varepsilon_0^t(\lambda) = 0$. $\square$

6 Letting ε tend to zero

6.1 Outline

Our goal is to compute the Laplacian with respect to λ of the function $s_t(\lambda) := \lim_{\varepsilon \to 0^+} S(t, \lambda, \varepsilon)$, using the Hamilton–Jacobi method of Theorem 5.2. We want the curve $\varepsilon(\cdot)$ occurring in (5.7) and (5.8) to approach zero at time t; a simple way we might try to accomplish this is to let the initial condition ε_0 approach zero. Suppose, then, that ε_0 is very small. Using various formulas from Sect. 5.3, we then find that *for as long as the solution to the system* (5.2) *exists*, the whole curve $\varepsilon(\cdot)$ will be small and the whole curve $\lambda(\cdot)$ will be approximately constant. Thus, by taking $\varepsilon_0 \approx 0$ and $\lambda_0 \approx \lambda$, we obtain a curve with $\varepsilon(t) \approx 0$ and $\lambda(t) \approx \lambda$. We may then hope to compute $s_t(\lambda)$ by letting λ_0 and $\lambda(t)$ approach λ and ε_0 approach zero in the Hamilton–Jacobi formula (5.7), with the result that

$$s_t(\lambda) = \log(|\lambda - 1|^2). \tag{6.1}$$

It is essential to note, however, that this approach is only valid if the solution to system (5.2) exists up to time t. Corollary 5.14 tells us that for $\varepsilon_0 \approx 0$, the solution will exist beyond time t provided λ is outside $\overline{\Sigma}_t$. Thus, we expect that for λ outside $\overline{\Sigma}_t$, the function s_t will be given by (6.1) and therefore that Δs_t will be zero. (The function $\log(|\lambda - 1|^2)$ is harmonic except at the point $\lambda = 1$, which is always inside Σ_t.)

To analyze $s_t(\lambda)$ for λ inside Σ_t, we first make use of the surjectivity result in Theorem 5.17. The theorem says that for each $t > 0$ and $\lambda \in \Sigma_t$, there exist $\varepsilon_0 > 0$ and $\lambda_0 \in \Sigma_t$ such that $\varepsilon(u)$ approaches 0 and $\lambda(u)$ approaches λ as u approaches t. We then use the formula (5.29) in Theorem 5.7. In light of the second Hamilton–Jacobi formula (5.8), we can write (5.29) as

$$\begin{aligned}\lim_{u \to t} \left(a\frac{\partial S}{\partial a} + b\frac{\partial S}{\partial b} \right)(u, \lambda(u), \varepsilon(u)) &= \lim_{u \to t} \frac{2 \log |\lambda(t)|}{t} + 1 \\ &= \frac{2 \log |\lambda|}{t} + 1. \end{aligned} \tag{6.2}$$

Once we have established enough regularity in the function $S(t, \lambda, \varepsilon)$ near $\varepsilon = 0$, we will be able to identify the left-hand side of (6.2) with the corresponding derivative of s_t, giving the following explicit formula for one of the derivatives of s_t:

$$\left(a\frac{\partial s_t}{\partial a} + b\frac{\partial s_t}{\partial b} \right)(\lambda) = \frac{2 \log |\lambda|}{t} + 1. \tag{6.3}$$

We now compute in logarithmic polar coordinates, with $\rho = \log |\lambda|$ and $\theta = \arg \lambda$. We may recognize the left-hand side of (6.3) as the derivative of s_t with respect to ρ, giving

$$\frac{\partial s_t}{\partial \rho} = \frac{2\rho}{t} + 1 \tag{6.4}$$

for points inside Σ_t. Remarkably, $\partial s_t / \partial \rho$ is independent of θ! 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 ρ.

Now, we will show in Sect. 6.4 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 λ from the outside. We can therefore give a complete description of the function $\partial s_t/\partial\theta$ on Σ_t as follows. It is the unique function on Σ_t that is independent of ρ (or, equivalently, independent of $r = |\lambda|$) and whose boundary values agree

$$\frac{\partial}{\partial\theta}\log(|\lambda-1|^2) = \frac{2b}{|\lambda-1|^2} = \frac{2r\sin\theta}{r^2+1-2r\cos\theta}. \tag{6.5}$$

Since the points on the outer boundary of Σ_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 (6.4), 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 Σ_t, accounting for the formula in Theorem 2.2.

We now briefly discuss what is needed to make the preceding arguments rigorous. If λ is outside $\overline{\Sigma}_t$ and ε is small and positive, we need to know that we can find a λ_0 close to λ and a small, positive ε_0 such that with these initial conditions, $\varepsilon(t) = \varepsilon$ and $\lambda(t) = \lambda$. To show this, we apply the inverse function theorem to the map $U_t(\lambda_0, \varepsilon_0) := (\lambda(t), \varepsilon(t))$ in a neighborhood of the point $(\lambda_0, \varepsilon_0) = (\lambda, 0)$.

For λ inside Σ_t, we need to know first that $S(t, \lambda, \varepsilon)$ is continuous—in all three variables—up to $\varepsilon = 0$. After all, $s_t(\lambda)$ is defined letting ε tend to zero in the expression $S(t, \lambda, \varepsilon)$, *with t and λ fixed*. But the Hamilton–Jacobi formula (5.7) gives a formula for $S(u, \lambda(u), \varepsilon(u))$, in which the first two variables in S are not remaining constant. Furthermore, we want to apply also the Hamilton–Jacobi formula (5.8) for the derivatives of S, which means we need also continuity of the derivatives of S with respect to λ up to $\varepsilon = 0$. Using another inverse function theorem argument, we will show that after making the change of variable $z = \sqrt{\varepsilon}$, the function S will extend smoothly up to $\varepsilon = z = 0$, from which the needed regularity will follow.

We use the following notation throughout the section.

Notation 6.1 *We will let*

$$\varepsilon(t; \lambda_0, \varepsilon_0)$$

denote the ε-component of the solution to (5.2) with $\lambda(0) = \lambda_0$ and $\varepsilon(0) = \varepsilon_0$ (and with initial values of the momenta given by (5.4)), and similarly for the other components of the solution.

6.2 Outside $\overline{\Sigma}_t$

The goal of this subsection is to prove the following result.

Theorem 6.2 *Fix a pair* (t, λ) *with* λ *outside* $\overline{\Sigma}_t$*. Then*

$$s_t(\lambda) := \lim_{\varepsilon \to 0^+} S(t, \lambda, \varepsilon) = \log(|\lambda - 1|^2). \tag{6.6}$$

Thus,

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

whenever λ *is outside* $\overline{\Sigma}_t$*.*

As we have discussed in Sect. 6.1, the idea is that for λ outside $\overline{\Sigma}_t$ and ε small and positive, we should try to find a λ_0 close to λ and a small, positive ε_0 such that $\varepsilon(u)$ and $\lambda(u)$ will approach 0 and λ, 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

$$U_t(\lambda_0, \varepsilon_0) = (\lambda(t; \lambda_0, \varepsilon_0), \varepsilon(t; \lambda_0, \varepsilon_0)).$$

We wish to evaluate the derivative of this map at the point $(\lambda_0, \varepsilon_0) = (\lambda, 0)$. For this idea to make sense, $\lambda(t; \lambda_0, \varepsilon_0)$ and $\varepsilon(t; \lambda_0, \varepsilon_0)$ must be defined in a neighborhood of $(\lambda, 0)$; it is for this reason that we have allowed ε_0 to be negative in Sect. 5.3.

The domain of U_t consists of pairs $(\lambda_0, \varepsilon_0)$ such that (1) the assumptions (5.42) are satisfied; (2) the function $p_\varepsilon(\cdot)$ remains positive until it blows up, as in Remark 5.10; and (3) we have $t_*(\lambda_0, \varepsilon_0) > t$. We note that these conditions allow ε_0 to be slightly negative and that all the results of Sect. 5.3 hold under these conditions. We also note that by Proposition 5.11, if $\varepsilon_0 = 0$, then $\varepsilon(t) \equiv 0$ and $\lambda(t) \equiv \lambda_0$; thus,

$$U_t(\lambda_0, 0) = (\lambda_0, 0). \tag{6.7}$$

We now fix a pair (t, λ) with λ outside of $\overline{\Sigma}_t$ (so that $\lambda \neq 1$). By Corollary 5.14, we then have $t_*(\lambda, 0) > t$.

Lemma 6.3 *The Jacobian of* U_t *at* $(\lambda, 0)$ *has the form*

$$U_t'(\lambda, 0) = \begin{pmatrix} I_{2\times 2} & \frac{\partial \lambda}{\partial \varepsilon_0}(t; \lambda, 0) \\ 0 & \frac{\partial \varepsilon}{\partial \varepsilon_0}(t; \lambda, 0) \end{pmatrix} \tag{6.8}$$

with $\partial\varepsilon/\partial\varepsilon_0(t; \lambda, 0) > 0$*. 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 (6.7). We then compute from (5.50) that

$$\begin{aligned}\frac{\partial \varepsilon(t;\lambda_0,\varepsilon_0)}{\partial \varepsilon_0}(0,\lambda_0) &= \frac{1}{p_\varepsilon(t)^2}p_0^2 e^{-Ct} + \varepsilon_0 \frac{\partial}{\partial \varepsilon_0}\left[\frac{1}{p_\varepsilon(t)^2}p_0^2 e^{-Ct}\right]\Bigg|_{\varepsilon_0=0} \\ &= \frac{1}{p_\varepsilon(t)^2}p_0^2 e^{-Ct},\end{aligned} \tag{6.9}$$

which is positive. □

Proof of Theorem 6.2 We note that the inverse of the matrix in (6.8) will have a positive entry in the bottom right corner, meaning that U_t^{-1} has the property that $\partial\varepsilon_0/\partial\varepsilon > 0$. It follows that the ε_0-component of $U_t^{-1}(\lambda,\varepsilon)$ will be positive for ε small and positive. In that case, the solution to the system (5.2) will have $\varepsilon(u) > 0$ 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 (5.7), which we write as follows. We let HJ denote the right-hand side of the Hamilton–Jacobi formula (5.7):

$$\begin{aligned}\mathrm{HJ}(t,\lambda_0,\varepsilon_0) &= \log(|\lambda_0-1|^2+\varepsilon_0) - \frac{\varepsilon_0 t}{(|\lambda_0-1|^2+\varepsilon_0)^2} \\ &\quad + \log|\lambda(t;\lambda_0,\varepsilon_0)| - \log|\lambda_0|\end{aligned} \tag{6.10}$$

and we then have

$$S(t,\lambda(t;\lambda_0,\varepsilon_0),\varepsilon(t;\lambda_0,\varepsilon_0)) = \mathrm{HJ}(t,\lambda_0,\varepsilon_0). \tag{6.11}$$

If ε is small and positive, we therefore obtain

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

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

Now, in the limit $\varepsilon \to 0^+$ with λ fixed, the inverse function theorem tells us that $U_t^{-1}(\varepsilon,\lambda) \to (0,\lambda)$. Thus, the limit (6.6) may be computed by putting $\lambda(t;\lambda_0,\varepsilon_0) = \lambda$ in (6.11) and letting ε_0 tend to zero and λ_0 tend to λ. This process gives (6.6).

Finally, when $\lambda_0 = 0$, we can use continuous dependence of the solutions on the initial conditions. The formula for $p_\varepsilon(t)$ in Proposition 5.9 has a limit as $|\lambda_0|$ tends to zero, so that δ tends to $+\infty$. From (5.44), we find that $a^2 = y_0^2$, so that from (5.32), $y(t) \equiv y_0$. We then obtain

$$p_\varepsilon(t) = e^{-Ct}p_0,$$

which remains nonsingular for all t. We can then continue to use the formula (5.50) for $\varepsilon(t)$. Furthermore, by exponentiating (5.51) 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 5.11 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 (5.20), which is to say that we replace the last two terms in (5.7) by $\int_0^t \varepsilon(s)p_\varepsilon(s)\,ds$. We then compute as in (6.9) that the derivative of $\varepsilon(t;0,\varepsilon_0)$ with respect to ε_0 is positive at $\varepsilon_0 = 0$. Thus, by the inverse function theorem, for small positive ε, we can find a small positive ε_0

that gives $\varepsilon(t; 0, \varepsilon_0) = \varepsilon$. We then apply (5.20) with $\lambda_0 = 0$ and $\lambda(t) = 0$, and let ε tend to zero, which means that ε_0 also tends to zero. As ε_0 tends to zero, the function

$$\varepsilon(s)p_\varepsilon(s) = \frac{\varepsilon(s)p_\varepsilon(s)^2}{p_\varepsilon(s)} = \frac{\varepsilon_0 p_0^2 e^{-Cs}}{p_\varepsilon(s)}$$

tends to zero uniformly and we obtain (6.6). □

6.3 Inside Σ_t

In this subsection, we establish the needed regularity of $S(t, \lambda, \varepsilon)$ as ε tends to zero, for λ in Σ_t. This result, whose proof is on p. 41, together with Theorem 5.7, will allow us to understand the structure of s_t and its derivatives on Σ_t.

Theorem 6.4 *Define*

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

Fix a pair (σ, μ) *with* μ *in* Σ_σ. *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 $\sqrt{\varepsilon(t)}p_\varepsilon(t)$ is always bounded away from zero (Proposition 5.6), the second Hamilton–Jacobi formula (5.8) tells us that $\partial\tilde{S}/\partial z(t, \lambda, z) = 2\sqrt{\varepsilon}\partial S/\partial\varepsilon(t, \lambda, z^2)$ has a nonzero limit as z tends to zero, ruling out a smooth extension that is even in z.

Corollary 6.5 *Fix a pair* (σ, μ) *with* μ *in* Σ_σ. *Then the functions*

$$S(t, \lambda, \varepsilon), \quad \frac{\partial S}{\partial a}(t, \lambda, \varepsilon), \quad \frac{\partial S}{\partial b}(t, \lambda, \varepsilon), \quad \sqrt{\varepsilon}\frac{\partial S}{\partial \varepsilon}(t, \lambda, \varepsilon) \tag{6.12}$$

all have extensions that are continuous in all three variables to the set of $(t, \lambda, \varepsilon)$ *with* $\lambda \in \Sigma_t$ *and* $\varepsilon \geq 0$. *Furthermore, for each* $t > 0$, *the function* s_t *is infinitely differentiable on* Σ_t, *and its derivatives with respect to a and b agree with the* $\varepsilon \to 0^+$ *limit of* $\partial S/\partial a$ *and* $\partial S/\partial b$. *If we let* t_* *be short for* $t_*(\lambda_0, \varepsilon_0)$, *then for all* λ_0 *and* $\varepsilon_0 > 0$, *we have*

$$s_t(t_*, \lambda(t_*; \lambda_0, \varepsilon_0)) = \log(|\lambda_0 - 1|^2 + \varepsilon_0) - \frac{\varepsilon_0 t_*}{(|\lambda_0 - 1|^2 + \varepsilon_0)^2} + \log|\lambda(t_*; \lambda_0, \varepsilon_0)| - \log|\lambda_0|$$

and

$$\begin{aligned} \frac{\partial s_t}{\partial a}(t_*, \lambda(t_*; \lambda_0, \varepsilon_0)) &= \lim_{t \to t_*} p_a(t) \\ \frac{\partial s_t}{\partial b}(t_*, \lambda(t_*; \lambda_0, \varepsilon_0)) &= \lim_{t \to t_*} p_b(t). \end{aligned} \tag{6.13}$$

Proof We note that the four functions in (6.12) may be computed as

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

respectively, and that $S(t,\lambda,0)=\tilde{S}(t,\lambda,0)$. The first claim then follows from Theorem 6.4. Now that the continuity of S and its derivatives has been established, we may let t approach $t_*(\lambda_0,\varepsilon_0)$ in the Hamilton–Jacobi formulas (5.7) and (5.8) to obtain the second claim. □

Corollary 6.6 *Let us write $\lambda\in\Sigma_t$ in logarithmic polar coordinates, with $\rho=\log|\lambda|$ and $\theta=\arg\lambda$. Then for each pair (t,λ) with $\lambda\in\Sigma_t$, we have*

$$\frac{\partial s_t}{\partial\rho}(t,\lambda)=\frac{2\rho}{t}+1. \tag{6.14}$$

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

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

for some smooth function m_t. Thus,

$$\frac{\partial^2 s_t}{\partial\rho^2}+\frac{\partial^2 s_t}{\partial\theta^2}=\frac{2}{t}+\frac{\partial}{\partial\theta}m_t(\theta) \tag{6.15}$$

for some smooth function m_t, and

$$\left(\frac{\partial^2}{\partial a^2}+\frac{\partial^2}{\partial b^2}\right)s_t(\lambda)=\frac{1}{|\lambda|^2}\left(\frac{2}{t}+\frac{\partial}{\partial\theta}m_t(\theta)\right). \tag{6.16}$$

In Sect. 6.4, we will obtain a formula for the function $m_t(\theta)$ appearing in Corollary 6.6.

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 (5.8) that

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

Now, for each pair (t,λ) with $\lambda\in\Sigma_t$, Theorem 5.17 tells us that we can find $(\lambda_0,\varepsilon_0)$ so that

$$\lim_{u\to t}\varepsilon(u)=0;\quad \lim_{u\to t}\lambda(u)=\lambda.$$

In light of (6.13), the formula (6.14) then follows from the formula (5.29) in Theorem 5.7.

Now, $\partial s_t/\partial\rho$ is manifestly independent of θ. Since, by Corollary 6.5, s_t is an analytic, hence C^2, function on Σ_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 ρ. The formula (6.15) then follows by differentiating (6.14) with respect to ρ. 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 (6.16) from (6.15). □

We now begin preparations for the proof of Theorem 6.4. Recall from Proposition 5.6 that $\varepsilon(t)p_\varepsilon(t)^2=\varepsilon_0p_0^2e^{-Ct}$, where $C=2\Psi-1$ is a constant computed from ε_0 and λ_0 as in (5.26). Recall also from Proposition 5.9 that for $\varepsilon_0>0$, the function $1/p_\varepsilon(t)$ extends to real analytic function of t defined for all $t\in\mathbb{R}$. We then define, for $\varepsilon_0>0$,

$$z(t;\lambda_0,\varepsilon_0)=\sqrt{\varepsilon_0}p_0e^{-Ct/2}\frac{1}{p_\varepsilon(t;\lambda_0,\varepsilon_0)} \tag{6.17}$$

for all $t\in\mathbb{R}$. For $t<t_*(\lambda_0,\varepsilon_0)$, the function $z(t;\lambda_0,\varepsilon_0)$ is positive and satisfies

$$z(t;\lambda_0,\varepsilon_0)^2=\varepsilon(t;\lambda_0,\varepsilon_0),$$

while for $t=t_*(\lambda_0,\varepsilon_0)$, we have $z(t;\lambda_0,\varepsilon_0)=0$ and for $t>t_*(\lambda_0,\varepsilon_0)$, the function $z(t;\lambda_0,\varepsilon_0)$ is negative.

Furthermore, using (5.51) and Point 3 of Proposition 5.4, we see that

$$\lambda(t;\lambda_0,\varepsilon_0)=\lambda_0e^{\int_0^t\varepsilon(s)p_\varepsilon(s)\,ds},$$

where by Proposition 5.6, we have

$$\varepsilon(s)p_\varepsilon(s)=\frac{\varepsilon(s)p_\varepsilon(s)^2}{p_\varepsilon(s)}=\frac{\varepsilon_0p_0^2e^{-Cs}}{p_\varepsilon(s)}.$$

Since $1/p_\varepsilon(s)$ 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

$$V(t,\lambda_0,\varepsilon_0):=(t,\lambda(t;\lambda_0,\varepsilon_0),z(t;\lambda_0,\varepsilon_0)),$$

for all $t\in\mathbb{R}$, $\lambda_0\in\mathbb{C}$, and $\varepsilon_0>0$.

Proposition 6.7 *Suppose $(t, \lambda_0, \varepsilon_0)$ has the property that $\lambda_0 \neq 0$ and $t_*(\lambda_0, \varepsilon_0) = t$, so that $z(t; \lambda_0, \varepsilon_0) = 0$. Then the Jacobian matrix of V at $(t, \lambda_0, \varepsilon_0)$ is invertible.*

Proof We make some convenient changes of variables. First, we replace $(t, \lambda_0, \varepsilon_0)$ with (t, λ_0, δ), where δ is as in (5.38). This change has a smooth inverse, since we can recover ε_0 from δ as

$$\varepsilon_0 = |\lambda_0|\, \delta - |\lambda_0|^2 - 1.$$

Then we write λ_0 in terms of its polar coordinates, (r_0, θ_0). Finally, we write $\lambda(t; \lambda_0, \varepsilon_0)$ in logarithmic polar coordinates,

$$\rho(t; \lambda_0, \varepsilon_0) := \log |\lambda_0(t; \lambda_0, \varepsilon_0)|\,; \quad \theta(t; \lambda_0, \varepsilon_0) := \arg(\lambda(t; \varepsilon_0, \lambda_0)),$$

where by Point 3 of Proposition 5.4, $\theta(t; \lambda_0, \varepsilon_0) = \theta_0$.

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

$$W(t, \theta_0, r_0, \delta) := (t, \theta_0, \rho(t; \lambda_0, \varepsilon_0), z(t; \lambda_0, \varepsilon_0))$$

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 5.9, the lifetime is independent of r_0 with δ and θ_0 fixed. Thus, if we start at a point with $t_*(\lambda_0, \varepsilon_0) = t$ and vary r_0, the lifetime will remain equal to t and $z(t; \varepsilon_0, \lambda_0)$ will remain equal to 0. Thus, at the point in question, $\partial z/\partial r_0 = 0$. Meanwhile, Proposition 5.12 gives a formula for the value of $\rho(t; \lambda_0, \varepsilon_0)$ at $t = t_*(\lambda_0, \varepsilon_0)$, 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, $z(t_*(\lambda_0, \varepsilon_0); \lambda_0, \varepsilon_0) = 0$. If we differentiate this relation with respect to ε_0 with λ_0 fixed, we find that

$$\frac{\partial z}{\partial \varepsilon_0} = -\frac{\partial z}{\partial t} \frac{\partial t_*(\lambda_0, \varepsilon_0)}{\partial \varepsilon_0}. \tag{6.18}$$

The derivative $\partial t_*/\partial \varepsilon_0$ may be computed as

$$\frac{\partial t_*(\lambda_0, \varepsilon_0)}{\partial \varepsilon_0} = \frac{\partial g_{\theta_0}(\delta)}{\partial \delta} \frac{\partial \delta}{\partial \varepsilon_0},$$

where g_{θ_0} is as in (5.58). But the proof of Proposition 5.16 shows that $\partial g_\theta/\partial\delta > 0$ for all $\delta > 2$, while from the formula (5.38) for δ, we see that $\partial\delta/\partial\varepsilon_0 > 0$. (Note also that $\delta > 2$ whenever $\varepsilon_0 > 0$.) Thus, $\partial t_*(\lambda_0, \varepsilon_0)/\partial\varepsilon_0 > 0$.

Meanwhile, from (6.17) and (5.45), we have

$$z(t) = \frac{\sqrt{\varepsilon_0}e^{Ct/2}}{\cosh(at) + \frac{2|\lambda_0|-\delta}{\sqrt{\delta^2-4}}\sinh(at)}\left(\cosh(at) - \frac{\delta}{\sqrt{\delta^2-4}}\sinh(at)\right).$$

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

$$\begin{aligned}&\left.\frac{\partial z(t;\lambda_0,\varepsilon_0)}{\partial t}\right|_{t=t_*(\lambda_0,\varepsilon_0)}\\ &\quad = 0 + \frac{\sqrt{\varepsilon_0}e^{Ct/2}}{\cosh(at) + \frac{2|\lambda_0|-\delta}{\sqrt{\delta^2-4}}\sinh(at)}a\left(\sinh(at) - \frac{\delta}{\sqrt{\delta^2-4}}\cosh(at)\right),\end{aligned}$$

which is negative because $\delta/\sqrt{\delta^2-4} > 1$ and the denominator is positive (Proposition 5.9). Thus, from (6.18), we conclude that $\partial z/\partial\varepsilon_0 > 0$. □

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

Proof of Theorem 6.4 By (6.11), we have

$$\begin{aligned}\tilde{S}(t;\lambda(t;\lambda_0,\varepsilon_0), z(t;\lambda_0,\varepsilon_0)) &= S(t;\lambda(t;\lambda_0,\varepsilon_0), \varepsilon(t;\lambda_0,\varepsilon_0))\\ &= \mathrm{HJ}(t,\lambda_0,\varepsilon_0),\end{aligned} \tag{6.19}$$

where HJ is as in (6.10), whenever $t_*(\lambda_0,\varepsilon_0) > \sigma$. Fix a point (σ,μ) with $\mu \in \Sigma_\sigma$. Then by Theorem 5.17, we can find a pair $(\lambda_0,\varepsilon_0)$ with $t_*(\lambda_0,\varepsilon_0) = t$—so that $z(t;\lambda_0,\varepsilon_0) = 0$—and $\lambda(t;\lambda_0,\varepsilon_0) = \lambda$. We now construct a local inverse V^{-1} to V around the point $V(t,\lambda_0,\varepsilon_0) = (t,\lambda,0)$.

For any triple (t,λ,z) in the domain of V^{-1}, we write $V^{-1}(t,\lambda,z)$ as $(t,\lambda_0,\varepsilon_0)$. We note that if $z > 0$ then $t_*(\lambda_0,\varepsilon_0)$ must be greater than t, because if we had $t_*(\lambda_0,\varepsilon_0) \le t$, then $z(t;\lambda_0,\varepsilon_0) = z$ would be zero or negative. Thus, we may apply (6.19) at $(t,\lambda_0,\varepsilon_0) = V^{-1}(t,\lambda,z)$ to obtain

$$\tilde{S}(t,\lambda,z) = \mathrm{HJ}(V^{-1}(t,\lambda,z)), \tag{6.20}$$

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

Recall now that $\lambda(t;\lambda_0,\varepsilon_0)$ extends to an analytic function of $t \in \mathbb{R}$. Thus, the function HJ in (6.10) extends to a smooth function of $t \in \mathbb{R}$, $\lambda_0 \in \mathbb{C}\backslash\{0\}$, and $\varepsilon_0 > 0$, defined even if $t_*(\lambda_0,\varepsilon_0) < t$. Therefore, the right-hand side of (6.20) provides the claimed smooth extension of $\tilde{S}$. □

6.4 Near the boundary of Σ_t

We start by considering what is happening right on the boundary of Σ_t.

Remark 6.8 Neither the method of Sect. 6.2 nor the method of Sect. 6.3 allows us to compute the value of $s_t(\lambda)$ for λ in the boundary of Σ_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 Δs_t *computed in the distribution sense*, that is, the distribution whose value on a test function ψ is

$$\int_{\mathbb{C}} s_t(\lambda)\Delta\psi(\lambda)\, d^2\lambda. \tag{6.21}$$

The value of (6.21) is unaffected by the value of $s_t(\lambda)$ for λ 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 λ *approaches* the boundary of Σ_t.

Definition 6.9 We say that a function $f : \mathbb{C} \to \mathbb{R}$ is analytic up to the boundary from inside Σ_t if the following conditions hold. First, f is real analytic on Σ_t. Second, for each $\lambda \in \partial\Sigma_t$, we can find an open set U containing λ 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 Σ_t.

Proposition 6.10 *For each $t > 0$, the function s_t is analytic up to the boundary from inside Σ_t and analytic up to the boundary from outside Σ_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 λ outside Σ_t but approaches a typically nonzero value as λ approaches a boundary point from the inside. As we approach from the inside a boundary point with polar coordinates (r, θ), 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 Sect. 7.1 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 λ in $(\overline{\Sigma}_t)^c$. Since $1 \in \Sigma_t$, we see that s_t is analytic from the outside of Σ_t.

To address the analyticity from the inside, first note that by applying (6.20) with $z = 0$, we have

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

where HJ is as in (6.10). But if $\varepsilon_0^t : \Sigma_t \to \mathbb{R}$ and $\lambda_t : \Sigma_t \to \mathbb{C}$ are as in Lemma 5.18, then we can see that

$$V^{-1}(t, \lambda, 0) = (t, \lambda_t^{-1}(\lambda), \varepsilon_0^t(\lambda_t^{-1}(\lambda)))$$

and we conclude that

$$s_t(\lambda) = \mathrm{HJ}(t, \lambda_t^{-1}(\lambda), \varepsilon_0^t(\lambda_t^{-1}(\lambda))). \tag{6.22}$$

We now claim that the function $\varepsilon_0^t(\lambda_0)$, initially defined for $\lambda_0 \in \overline{\Sigma}_t$, extends to an analytic function in a neighborhood of $\overline{\Sigma}_t$. For $t \geq 4$, we can simply use the formula (5.64) for all nonzero λ_0. For $t < 4$, however, the formula (5.64) 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 $\varepsilon_0^t(\lambda_0)$, we solve the equation $t_*(\lambda_0, \varepsilon_0) = t$ for ε_0 as a function of λ_0. To do this, we first solve the equation $g_{\theta_0}(\delta) = t$ for $\delta_{\theta_0,t}$ and then solve for ε_0 in terms of δ as $\varepsilon_0 = |\lambda_0|\,\delta - |\lambda_0|^2 - 1$. Now, we know from the proof of Proposition 5.16 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 5.16. To see that the result holds even when $\delta = 2$, it suffices to verify that the expressions in (5.59) and (5.60) have positive limits as $\delta \to 2^+$. 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 θ_0. Then, the expression

$$\varepsilon_0^t(\lambda_0) = |\lambda_0|\,\delta_{\theta_0,t} - |\lambda_0|^2 - 1$$

makes sense and is analytic for all nonzero λ_0 with $|\arg\lambda_0| < \theta_{\max}(t) + \alpha_t$, for some positive quantity α_t. We note that in this expression, $\varepsilon_0^t(\lambda_0)$ can be negative—for example if $|\lambda_0| = 1$ and $\arg\lambda_0 > \theta_{\max}(t)$.

We now consider the function λ_t, defined as

$$\lambda_t(\lambda_0) = \lambda(t; \lambda_0, \varepsilon_0^t(\lambda_0)),$$

and we recall that $\lambda_t(\lambda_0) = \lambda_0$ for $\lambda_0 \in \partial\Sigma_t$. Although λ_t was initially defined for λ_0 in $\overline{\Sigma}_t$, it has an analytic extension to a neighborhood of $\overline{\Sigma}_t$, namely the set of λ_0 in the domain of the extended function ε_0^t for which the pair $(\lambda_0, \varepsilon_0)$ satisfy the assumptions in (5.42). 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 (5.62), where $\delta = \delta_{\theta_0,t}$ as in the previous paragraph. Since δ is independent of $|\lambda_0|$ with t and $\arg\lambda_0$ fixed, we can easily verify from (5.62) that $\partial\,|\lambda_t|\,/\partial r > 0$.

Now, we have already established that s_t is analytic in the interior of Σ_t. Consider, then, a point λ in $\partial\Sigma_t$, so that $\lambda_t(\lambda) = \lambda$. Since $\lambda_t'(\lambda)$ is invertible, it has a analytic local inverse λ_t^{-1} defined near λ. Then the formula (6.22) gives an analytic extension of s_t to a neighborhood of λ. □

Proposition 6.11 *Fix a point μ on the boundary of Σ_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 λ approaches μ from inside Σ_t as when λ approaches μ from outside $\overline{\Sigma}_t$.

Proof We begin by considering s_t itself. The limit as λ approaches μ from the inside may be computed by using (6.22). By Lemma 5.18, as λ approaches μ from the inside, $\lambda_t^{-1}(\lambda)$ approaches $\lambda_t^{-1}(\mu) = \mu$, and $\varepsilon_0^t(\lambda_t^{-1}(\lambda))$ approaches 0. Thus, the limiting value of s_t from the inside is

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

where HJ is given by (6.10) and were we have used that $\lambda(t; \mu, 0) = \mu$. (See the last part of Proposition 5.11). Since $s_t(\lambda) = \log(|\lambda - 1|^2)$ outside $\overline{\Sigma}_t$, 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 (6.14), 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 λ approach μ from the inside gives the value $\log(|\mu|^2)/t + 1$. Since μ is on the boundary of Σ_t, Theorem 3.1 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 λ tend to μ from the outside gives the same result.

Finally, we recall from Proposition 5.4 that $ap_b - bp_a$ is a constant of motion. Thus, by the second Hamilton–Jacobi formula (5.8) and the initial conditions (5.4), we have

$$\begin{aligned}a\frac{\partial s_t}{\partial b}(u, \lambda(u), \varepsilon(u)) - b\frac{\partial s_t}{\partial a}(u, \lambda(u), \varepsilon(u)) &= a_0 p_{b,0} - b_0 p_{a,0} \\ &= (2a_0 b_0 - 2b_0(a_0 - 1))p_0 \\ &= \frac{2b_0}{|\lambda_0 - 1|^2 + \varepsilon_0}.\end{aligned}$$

If we choose ε_0 and λ_0 so that $t_*(\lambda_0, \varepsilon_0) = t$ we can use the regularity result in Corollary 6.5 to let u tend to t. This gives

$$a\frac{\partial s_t}{\partial b}(\lambda) - b\frac{\partial s_t}{\partial a}(\lambda) = \frac{2b_0}{|\lambda_0 - 1|^2 + \varepsilon_0},$$

where now $\lambda_0 = \lambda_t^{-1}(\lambda)$ and $\varepsilon_0 = \varepsilon_0^t(\lambda_t^{-1}(\lambda))$. As λ approaches μ, Theorem 5.17 says that the value of λ_0 approaches μ and ε_0 approaches 0, so we get

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

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

6.5 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 Sect. 7.1, while Propositions 2.5 and 2.6 will be proved in Sect. 7.2.

Proposition 6.12 *For each fixed t, the restriction to Σ_t of the function*

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

is the unique function that on Σ_t that (1) extends continuously to the boundary, (2) agrees with the θ-derivative of $\log(|\lambda - 1|^2)$ *on the boundary, and (3) is independent of $r = |\lambda|$. Thus, the function m_t in Corollary 6.6 is given by*

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

where $r_t(\theta)$ is the outer radius of the domain Σ_t (Fig. 3).

Proof We have already established in Corollary 6.6 that $\partial s_t/\partial\theta$ is independent of ρ (or equivalently, of r) in Σ_t. Then Propositions 6.10 and 6.11 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 Σ_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 (6.5), giving the claimed expression for $\partial s_t/\partial\theta = m_t(\theta)$. $\square$

Proposition 6.13 *For each $t > 0$, the distributional Laplacian of $s_t(\lambda)$ with respect to λ may be computed as follows. Take the pointwise Laplacian of s_t outside $\overline{\Sigma_t}$ (giving zero), take the pointwise Laplacian of s_t inside Σ_t (giving the expression* (6.16) *in Corollary* 6.6*) and ignore the boundary of Σ_t.*

Proof Since, by Proposition 6.10, s_t is analytic up to the boundary of Σ_t from the inside, Green's second identity says that

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

for any test function ψ, where in the last integral, the limiting value of ∇s_t from the inside should be used. This identity holds because the boundary of Σ_t is smooth for $t \neq 4$ and piecewise smooth when $t = 4$ (Point 3 of Theorem 3.2). We also have similar formula for the integral over the complement of $\overline{\Sigma}_t$, provided that ψ is compactly supported, but with the direction of the unit normal reversed. Proposition 6.11 then tells us that the boundary terms in the two integrals cancel, giving

$$\int_{\mathbb{C}} s_t(\lambda)\Delta\chi(\lambda)\, d^2\lambda = \int_{(\overline{\Sigma}_t)^c} (\Delta s_t(\lambda))\chi(\lambda)\, d^2\lambda + \int_{\Sigma_t} (\Delta s_t(\lambda))\chi(\lambda)\, d^2\lambda, \tag{6.23}$$

where the integral over $(\overline{\Sigma}_t)^c$ is actually zero, since $\Delta s_t(\lambda) = 0$ there. The formula (6.23) says that the distributional Laplacian of s_t may be computed by taking the ordinary, pointwise Laplacian in $\overline{\Sigma}_t$ and in Σ_t and ignoring the boundary of Σ_t. □

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

Proof of Theorem 2.2 Proposition 6.13 tells us that we can compute the distributional Laplacian of s_t separately inside Σ_t and outside $\overline{\Sigma}_t$, ignoring the boundary. Theorem 6.2 tells us that the Laplacian outside $\overline{\Sigma}_t$ is zero. Corollary 6.6 gives us the form of Δs_t inside Σ_t, while Proposition 2.6 identifies the function m_t appearing in Corollary 6.6. The claimed formula for the Brown measure therefore holds. □

7 Further properties of the Brown measure

7.1 The formula for ω

In this subsection, we derive the formula for w_t given in Proposition 2.3 in terms of the density ω. Throughout, we will write the function T in (3.1) in polar coordinates as

$$T(r,\theta) = (r^2 + 1 - 2r\cos\theta)\frac{\log(r^2)}{r^2 - 1}. \tag{7.1}$$

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

Lemma 7.1 *The density $w_t(\theta)$ in Theorem 2.2 may also be written as*

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

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 Σ_t. Thus, by Theorem 3.1, we have $T(r_t(\theta), \theta) = t$, from which we obtain

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

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. □

We now formulate the main result of this subsection, whose proof is on p. 47.

Theorem 7.2 *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 ω has a smooth extension to all $r > 0$ and all θ. 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 ω has the following properties.

(1) *We have $\omega(1/r, \theta) = \omega(r, \theta)$ for all $r > 0$ and all θ.*
(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 $\theta = 0$.*
(4) *We have*

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

where the limit is uniform in θ.

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 θ. (Recall Theorem 3.2).

Corollary 7.3 (Positivity). *If $t > 4$, then $w_t(\theta)$ is strictly positive for all θ. If $t < 4$, then $w_t(\theta)$ is strictly positive for $-\theta_{\max}(t) < \theta < \theta_{\max}(t)$ and the limit as θ 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 $\lim_{\theta\to\pm\pi} w_t(\theta) = 0$.*

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 and greater than 1 for all θ, so that $w_t(\theta)$ is strictly positive in this case. When $t \leq 4$, we know from Proposition 3.4 that $r_t(\theta)$ is greater than 1 for $|\theta| < \theta_{\max}(t)$ and approaches 1 when θ 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 $\omega(1,\pi) = 0$. □

Corollary 7.4 (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 θ, the quantity $2\pi t w_t(\theta)$ is close to 1.

The small- and large-t behavior of the region Σ_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, 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 Σ_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 7.2.

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 7.2. □

Proof of Theorem 7.2 We note that the function T in (7.1) 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

$$\frac{dr_t(\theta)}{d\theta} = -\frac{\partial T/\partial\theta}{\partial T/\partial r}. \tag{7.2}$$

By the chain rule and (7.2), $\frac{d}{d\theta}[h(r_t(\theta))\sin\theta] = q(r_t(\theta), \theta)$, where

$$\begin{aligned}q(r, \theta) &= h(r)\cos\theta - h'(r)\sin\theta\frac{\partial T/\partial\theta}{\partial T/\partial r}\\ &= h(r)\cos\theta - \frac{2h'(r)\sin^2\theta\ h(r)}{\left(1 - \frac{1}{r^2}\right) h(r) + \left(r + \frac{1}{r} - 2\cos\theta\right) h'(r)}.\end{aligned} \tag{7.3}$$

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 (7.3) 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 ω 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 $\lim_{r\to 0} h(r) = 0$ 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

$$h(r) = 1 - c(r)(r-1)^2 \tag{7.4}$$

for some analytic function $c(r)$, with $c(1) = 1/6$. The minus sign in (7.4) 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 α and β 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

$$\omega(r,\theta) = 1 + h(r)\frac{\tilde{\alpha}(r)\cos\theta + \tilde{\beta}(r)}{\tilde{\beta}(r)\cos\theta + \tilde{\alpha}(r)}, \tag{7.5}$$

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 (7.5) is positive for all $r > 0$ and all θ, from which the claimed smooth extension of ω follows.

To verify the claimed positivity of ω, we first observe that $\tilde{\beta}(r)z + \tilde{\alpha}(r)$ is positive when $z = 1$ (with a value of $2 - (r-1)^2c(r) = 1 + h(r)$) and also positive when $z = -1$ (with a value of $(r+1)^2c(r)$), and hence positive for all $-1 \leq z \leq 1$. Thus, the denominator in the fraction on the right-hand side of (7.5) is never zero. We then compute that

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

for all r and θ. 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 θ, 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 $\cos\theta = 1$.

Finally, from the definition (7.4) and the fact that $\lim_{r\to 0} h(r) = 0$, we find that $\lim_{r\to 0} c(r) = 1$. Thus, as $r \to 0$, we have $\tilde{\alpha}(r) \to 1$ and $\tilde{\beta}(r) \to 0$. In this limit, the fraction on the right-hand side of (7.5) converges uniformly to $\cos\theta$, while $h(r)$ tends to zero, giving Point 4. $\square$

7.2 The connection to free unitary Brownian motion

Recall from Theorem 1.1 that the spectral measure ν_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 ν_t. The support of ν_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 ν_t consists of points with angles ϕ 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 (1.3) of the function f_t. Then f_t maps the boundary of Σ_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 Σ_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 ν_t:

$$f_t : \partial\Sigma_t^{\text{out}} \leftrightarrow \text{supp}(\nu_t). \tag{7.6}$$

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 3.2). 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 (7.10) below with $r = 1$ and $\cos\theta = 1 - t/2$.)

We now describe the map (7.6) more concretely. We denote by $\lambda_t(\theta)$ the point at angle θ in $\partial\Sigma_t^{\mathrm{out}}$:

$$\lambda_t(\theta) = r_t(\theta)e^{i\theta},$$

where for $t < 4$, we require $|\theta| \leq \theta_{\max}(t)$. Then the map in (7.6) can be thought of as a map of θ to ϕ determined by the relation

$$f_t\left(\lambda_t(\theta)\right) = e^{i\phi}. \tag{7.7}$$

We now observe a close relationship between the density $w_t(\theta)$ in Theorem 2.2 and the map in (7.7).

Proposition 7.5 *Let ϕ and θ be related as in (7.7), 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*

$$w_t(\theta) = \frac{1}{2\pi t}\frac{d\phi}{d\theta}. \tag{7.8}$$

We may also write this formula as a logarithmic derivative of f_t along the outer boundary of Σ_t:

$$w_t(\theta) = \frac{1}{2\pi t}\frac{1}{i}\frac{\frac{d}{d\theta}f_t(\lambda_t(\theta))}{f_t(\lambda_t(\theta))}. \tag{7.9}$$

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 (1.3) of f_t, we find that

$$\arg(f_t(\lambda)) = \arg\lambda + \arg e^{\frac{t}{2}\frac{1+\lambda}{1-\lambda}} = \theta + t\frac{r\sin\theta}{r^2+1-2r\cos\theta}. \tag{7.10}$$

Evaluating this expression at the point $\lambda_t(\theta)$ gives

$$\begin{aligned}\phi &= \arg(f_t(\lambda_t(\theta))) \\ &= \theta + t\frac{r_t(\theta)\sin\theta}{r_t(\theta)^2+1-2r_t(\theta)\cos\theta}.\end{aligned} \tag{7.11}$$

(Strictly speaking, ϕ and θ are only defined "mod 2π," but for any local continuous version of θ, the last expression in (7.11) gives a local continuous version of ϕ.) 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 (7.8) follows easily by recalling the definition (2.5) of w_t. The expression (7.9) is then obtained by noting that $\phi = \frac{1}{i}\log f_t(\lambda_t(\theta))$. □

Proposition 7.6 *Biane's measure ν_t may be computed as*

$$d\nu_t(\phi) = \frac{r_t(\theta)^2 - 1}{r_t(\theta)^2 + 1 - 2r_t(\theta)\cos\theta}\frac{d\phi}{2\pi} \tag{7.12}$$

or as

$$d\nu_t(\phi) = \frac{\log(r_t(\theta))}{\pi t}d\phi. \tag{7.13}$$

Here, as usual, $r_t(\theta)$ is the outer radius of the domain Σ_t and θ is viewed as a function of ϕ by inverting the relationship (7.7). When $t < 4$, the formula should be used only for $|\phi| \leq \phi_{\max}(t)$.

Proof We make use of the expression for ν_t in Theorem 1.1. If ϕ is in the interior of the support of ν_t, then $\chi_t(e^{i\phi})$ is in the open unit disk, so that the density of ν_t is nonzero at this point. Now, since χ_t is an inverse function to f_t we see that $\chi_t(e^{i\phi})$ is (for ϕ in the interior of the support of ν_t) the unique point λ 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 (7.12). Finally, since $T(\lambda) = t$ on $\partial\Sigma_t$ (Theorem 3.1), we have

$$(r_t(\theta)^2 + 1 - 2r_t(\theta)\cos\theta)\frac{\log(r_t(\theta)^2)}{r_t(\theta)^2 - 1} = t, \tag{7.14}$$

which allows us to obtain (7.13) from (7.12). □

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 7.5 and Proposition 7.6 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. □

Proof of Proposition 2.6 The value $\Phi_t(\lambda)$ is computed by first taking the argument of λ to obtain θ and then applying the map in (7.7) to obtain ϕ. Thus, the first result is just a restatement of Proposition 2.5. For the uniqueness claim, suppose a measure μ on Σ_t has the form

$$d\mu(\lambda) = \frac{1}{r^2} g(\theta)\, r\, dr\, d\theta.$$

Then the distribution of the argument θ of λ will be, by integrating out the radial variable, $2\log(r_t(\theta))g(\theta)\, d\theta$. The distribution of ϕ 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 (7.13) is if g coincides with w_t. □

Acknowledgements The authors thank Ching-Wei Ho and Maciej Nowak for useful discussions. We also thank the referee for helping us to streamline the presentation. The figures were generated in Mathematica and the Mathematica notebook file is available online at https://www.notebookarchive.org/driver-hall-kemp-plots-and-simulations--2019-05-bjnnq2h/.

References

1. Biane, P.: On the free convolution with a semi-circular distribution. Indiana Univ. Math. J. **46**, 705–718 (1997)
2. Biane, P.: Free Brownian motion, free stochastic calculus and random matrices. In: Free Probability Theory (Waterloo, ON, 1995). Fields Institute Communications 12, pp. 1–19. American Mathematical Society, Providence, RI (1997)
3. Biane, P.: Segal–Bargmann transform, functional calculus on matrix spaces and the theory of semi-circular and circular systems. J. Funct. Anal. **144**, 232–286 (1997)
4. Biane, P., Speicher, R.: Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. Probab. Theory Relat. Fields **112**, 373–409 (1998)
5. Brown, L.G.: Lidskiĭ's theorem in the type II case. In: Geometric Methods in Operator Algebras (Kyoto. 1983), Pitman Research Notes in Mathematics Series, vol. 123, pp. 1–35. Longman Scientific and Technical, Harlow (1986)
6. Burda, Z., Grela, J., Nowak, M.A., Tarnowski, W., Warchoł, P.: Dysonian dynamics of the Ginibre ensemble. Phys. Rev. Lett. **113**, 104102 (2014)
7. Demni, N., Hamdi, T.: Support of the Brown measure of the product of a free unitary Brownian motion by a free self-adjoint projection. J. Funct. Anal. **282**, 109362 (2022)
8. Evans, L.C.: Partial differential equations. In: Graduate Studies in Mathematics, vol. 19, 2nd edn. American Mathematical Society, Providence, RI (2010)

9. Fuglede, B., Kadison, R.V.: On determinants and a property of the trace in finite factors. Proc. Natl. Acad. Sci. USA **37**, 425–431 (1951)
10. Fuglede, B., Kadison, R.V.: Determinant theory in finite factors. Ann. Math. **55**, 520–530 (1952)
11. Girko, V.L.: The circular law. Teor. Veroyatnost. i Primenen. **29**, 669–679 (1984). (**Russian**)
12. Grela, J., Nowak, M.A., Tarnowski, W.: Eikonal formulation of large dynamical random matrix models. Phys. Rev E **104**, 054111 (2021)
13. Gudowska-Nowak, E., Janik, R.A., Jurkiewicz, J., Nowak, M.A.: Infinite products of large random matrices and matrix-valued diffusion. Nucl. Phys. B **670**, 479–507 (2003)
14. Hall, B.C.: Quantum theory for mathematicians. In: Graduate Texts in Mathematics, **267**. Springer, New York (2013)
15. Hall, B.C.: PDE methods in random matrix theory. In: Harmonic Analysis and Applications, pp. 77–124. Springer (2021)
16. Hall, B.C., Ho, C.-W.: The Brown measure of the sum of a self-adjoint element and an imaginary multiple of a semicircular element. Lett. Math. Phys. **112**, 19 (2022)
17. Hall, B.C., Ho, C.-W.: The Brown measure of a family of free multiplicative Brownian motions. arXiv:2104.07859 [math.PR]
18. Hall, B.C., Kemp, T.: Brown measure support and the free multiplicative Brownian motion. Adv. Math. **355**, 106771 (2019)
19. Ho, C.-W., Zhong, P.: Brown measures of free circular and multiplicative Brownian motions with self-adjoint and unitary initial conditions. J. Eur. Math. Soc., to appear
20. Kemp, T.: The large-N limits of Brownian motions on $\mathbb{GL}_N$. Int. Math. Res. Not. **2016**, 4012–4057 (2016)
21. Kümmerer, B., Speicher, R.: Stochastic integration on the Cuntz algebra O_∞. J. Funct. Anal. **103**, 372–408 (1992)
22. Lohmayer, R., Neuberger, H., Wettig, T.: Possible large-N transitions for complex Wilson loop matrices. J. High Energy Phys. **2008**(11), 053 (2008)
23. Mingo, J.A., Speicher, R.: Free probability and random matrices. In: Fields Institute Monographs, vol. 35. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON (2017)
24. Nikitopoulos, E.A.: Itô's formula for noncommutative C^2 functions of free Itô processes with respect to circular Brownian motion. arXiv:2011.08493 [math.OA]