Liberation of Projections

Benoît Collins[*]
University of Ottawa
& CNRS Lyon I & AIMR Tohoku
585 King Edward, Ottawa, ON K1N 6N5
bcollins@uottawa.ca

Todd Kemp[†]
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112
tkemp@math.ucsd.edu

January 14, 2014

Abstract

We study the liberation process for projections: $(p, q) \mapsto (p_t, q) = (u_t p u_t^*, q)$ where u_t is a free unitary Brownian motion freely independent from $\{p, q\}$. Its action on the operator-valued angle qp_tq between the projections induces a flow on the corresponding spectral measures μ_t; we prove that the Cauchy transform of the measure satisfies a holomorphic PDE. We develop a theory of subordination for the boundary values of this PDE, and use it to show that the spectral measure μ_t possesses a piecewise analytic density for any $t > 0$ and any initial projections of trace $\frac{1}{2}$. We us this to prove the Unification Conjecture for free entropy and information in this trace $\frac{1}{2}$ setting.

Contents

[*]Supported in part by an NSERC discovery grant and Ontario's Early Research Award.

[†]Supported in part by NSF grant DMS-1001894, NSF CAREER award DMS-1254807, and the Hellman Fellowship.

1 Introduction

Let V and W be subspaces of a the finite-finite dimensional complex space $\mathbb{C}^d$. From elementary linear algebra, it follows that $\dim V \cap W \geq \max\{\dim V + \dim W - d, 0\}$. In fact, this inequality is *almost surely* equality:

$$\dim(V \cap W) = \max\{\dim V + \dim W - d, 0\} \;\; a.s. \tag{1.1}$$

The *almost surely* can be interpreted in a number of ways: for example, it can be taken with respect to any reasonable probability measure on the (product) Grassmannian manifold of subspaces. We will discuss a more probabilistic interpretation shortly.

When two subspaces satisfy the equality of Equation 1.1, they are said to be in general position. For convenience, we may rewrite this relation as follows. Let P, Q be the orthogonal projections onto V and W respectively, and let $P \wedge Q$ denote the orthogonal projection onto $V \cap W$. The dimension of a subspace is the trace of the projection onto it, so the subspaces are in general position if and only if $\mathrm{Tr}\,(P \wedge Q) = \max\{\mathrm{Tr}\,P + \mathrm{Tr}\,Q - d, 0\}$. Normalizing, letting $\mathrm{tr}\; = \frac{1}{d}\mathrm{Tr}$, two projections (i.e. subspaces) are in general position if and only if

$$\mathrm{tr}\,(P \wedge Q) = \max\{\mathrm{tr}\,P + \mathrm{tr}\,Q - 1, 0\}. \tag{1.2}$$

In this language, a good way to express the genericity of general position is as follows: let P, Q be as above, and let U be a *random unitary matrix*. If $\tilde{P} = UPU^*$ (i.e. the projection onto the rotation of the image of P by U), then $\tilde{P}$ and Q are in general position almost surely. This statement is valid for all reasonable notions of *random unitary matrix*; for example, it holds true if U is sampled from some measure that has a continuous, strictly positive density with respect to the Haar measure. (Indeed, the subset of the Grassmannian product where general position fails to hold is a subvariety of lower dimension, and so it is easy to see that any reasonable measure will assign it probability 0.) Since any neighborhood of the identity has positive measure, it follows that rotations arbitrarily close to the identity produce projections in general position, agreeing with our intuition. Note: the same result applies as well even if P, Q are *random* projections, provided that the random unitary U is independent from $\{P, Q\}$.

An important flow of random unitaries is given by the unitary Brownian motion. The group $\mathbb{U}_d$ of unitary $d\times d$ matrices is a compact Lie group, and so its left-invariant Riemannian metric (given by the Hilbert-Schmidt norm on the Lie algebra $\mathfrak{u}_d$) gives rise to a heat kernel, which generates a Markov process: Brownian motion U_t. This stochastic process can be constructed from the standard Brownian motion W_t on $\mathfrak{u}_d$: it satisfies the (Stratonovich) stochastic differential equation $dU_t = U_t \circ dW_t$. The Lie algebra $\mathfrak{u}_d$ consists of skew-Hermitian matrices; in random matrix theory, it is more conventional to consider Brownian motion taking values in Hermitian matrices, so set $X_t = -iW_t$. Making this substitution, and converting the SDE to Itô form, gives

$$dU_t = iU_t\,dX_t - \frac{1}{2}U_t\,dt. \tag{1.3}$$

(Note: this equation corresponds to the normalization $\mathbb{E}\,\mathrm{tr}\,(X_t^2) = t$ where tr is the normalized trace.) The distribution of U_t is the heat kernel at time t; on any Lie group, this has a strictly positive smooth density for any $t > 0$, cf. [29]. Hence, the above discussion proves the following.

Proposition 1.1. *Let P, Q be orthogonal projections on $\mathbb{C}^d$. Let U_t be the Brownian motion on the unitary group $\mathbb{U}_d$, and set $P_t = U_tPU_t^*$. Then for each $t > 0$, P_t and Q are almost surely in general position.*

Beyond the question of general position (relating only to dimension), it is interesting to consider the relative position of the two subspaces, and how it evolves under the Brownian conjugation. The principle angles between P_t and Q are encoded in the *operator-valued angle* QP_tQ (whose eigenvalues are trigonometric polynomials in said angles).

The purpose of the present paper is to address the nature of the flow of the operator-valued angle in an *infinite-dimensional* context. Let $\mathscr{A}$ be a II_1-factor – a von Neumann subalgebra of the bounded operators on some Hilbert space $\mathcal{H}$, with a unique tracial state τ. (For intuition, one may think of a limit as $d \to \infty$ of the algebra of $d \times d$ *random matrices*. We will make the notion of such a limit precise in Section 1.1; the limit can always be identified as living in a free group factor.) Let $p, q \in \mathscr{A}$ be projections, and let $p \wedge q$ be the projection onto $p\mathcal{H} \cap q\mathcal{H}$. Then $p \wedge q$ is in $\mathscr{A}$, since $\mathscr{A}$ is weakly closed. As in the finite dimensional setting, say that p, q are in general position if

$$\tau(p \wedge q) = \max\{\tau(p) + \tau(q) - 1, 0\}. \tag{1.4}$$

Instead of attempting to make sense of a random operator drawn from the unfathomably large group of unitaries on $\mathcal{H}$, we use the tools of *free probability* (Section 1.1) to proceed. We may assume that $\mathscr{A}$ is rich enough to possess a *free unitary Brownian motion* u_t (Section 1.3) free from $\{p, q\}$, by enlarging $\mathscr{A}$ if necessary. The operator-valued process u_t can be thought of as a limit of the Brownian motions U_t in $\mathbb{U}_d$ as $d \to \infty$ (cf. [3]). In fact, the analogue of Proposition 1.1 holds true in this context.

Proposition 1.2. *Let p, q be projections in a II_1-factor $(\mathscr{A}, \tau)$, and let u_t be a free unitary Brownian motion in $\mathscr{A}$, freely independent from $\{p, q\}$. Set $p_t = u_t p u_t^*$. Then for each $t > 0$, p_t and q are in general position.*

Proof. In [31, Lemma 12.5], it was proved that, for any two projections $\tilde{p}, q$, if the algebras $A = \mathbb{C}\langle \tilde{p} \rangle$ and $B = \mathbb{C}\langle q \rangle$ possess an $L^1(\tau)$ liberation gradient $j(A : B)$, then p, q are in general position. Letting $\tilde{p} = upu^*$ for some unitary free from $\{p, q\}$, it was proved in [31, Proposition 8.7] that this liberation gradient exists (in fact in $L^2(\tau) \subset L^1(\tau)$) provided that the law of u possesses an L^3-density with respect to the Haar measure on the circle. In fact, the density of u_t is continuous and bounded, with a density that is real analytic on the set where it is positive, cf. [3] (refined in [31, Corollary 1.7]). This proves the proposition. □

Remark 1.3. This argument was not known initially to the authors of the present paper; it was also unknown to the authors of [2]. Indeed, [2, Thm. 8.2] gave an alternate proof of Proposition 1.2; unfortunately, this proof was flawed (as will be discussed in Section 1.4). The present paper arose, in part, as an attempt to address this flaw.

The process $t \mapsto (p_t, q)$ (as t ranges through $[0, \infty)$) is known as the *free liberation* of the initial pair (p, q). It was introduced in [31] as a technical tool for the analysis of free entropy and free Fisher information. As $t \to \infty$, the free unitary Brownian motion u_t tends (in the weak sense) to a Haar unitary operator. (This is the infinite-dimensional version of the statement that the heat kernel measure on the unitary group flows towards the Haar measure.) Thus, the pair (p_t, q) tends towards (upu^*, q) where u is a Haar unitary free from $\{p, q\}$. The two operators upu^* and q are therefore free (cf. [22]), and so the spectral measure μ of the self-adjoint operator $q^{1/2}upu^*q^{1/2} = qp_tq$ is given by the free multiplicative convolution of the spectral measures of upu^* and q separately (cf. [33]). Let $\tau(p) = \alpha$ and $\tau(q) = \beta$. Then $\mu_{upu^*} = \mu_p = (1-\alpha)\delta_0 + \alpha\delta_1$ while $\mu_q = (1-\beta)\delta_0 + \beta\delta_1$. This convolution was calculated explicitly in [33, Ex. 3.6.7], via the Cauchy transform (cf. Section 1.1): setting $\mu = \mu_p \boxtimes \mu_q$, the authors show that

$$G_\mu(z) = \int_0^1 \frac{\mu(dx)}{z - x} = \frac{z + \alpha + \beta - 2 - \sqrt{z^2 - 2(\alpha + \beta - 2\alpha\beta)z + (\alpha - \beta)^2}}{2z(z-1)}, \quad z \in \mathbb{C}_+. \tag{1.5}$$

The Stieltjes inversion formula (again see Section 1.1) then shows that

$$\mu = (1 - \min\{\alpha, \beta\})\delta_0 + \max\{\alpha + \beta - 1, 0\}\delta_1 + \frac{\sqrt{(r_+ - x)(x - r_-)}}{2\pi x(1-x)} \mathbb{1}_{[r_-, r_+]}\, dx, \tag{1.6}$$

where $r_\pm = \alpha + \beta - 2\alpha\beta \pm 2\sqrt{\alpha\beta(1-\alpha)(1-\beta)}$ are the roots of the quadratic polynomial in the radical in Equation 1.5. It should be noted that this measure is precisely the limit of the empirical eigenvalue distribution of a Jacobi Ensemble in random matrix theory, which has recently been much studied in part due to its applications to MANOVA (multivariate analysis of variance) problems in statistics, cf. [8, 9, 10, 11, 12, 14].

The general position rank shows up as the mass of the spectral measure μ concentrated at 1; this is no accident. By a theorem of von Neumann (cf. [34]), $p \wedge q$ is the weak $\lim_{n\to\infty}(pq)^n$, and so

$$\tau(p \wedge q) = \lim_{n\to\infty} \tau[(pq)^n] = \lim_{n\to\infty} \tau[(qpq)^n] = \lim_{n\to\infty} \int_0^1 x^n \, \mu_{qpq}(dx) \tag{1.7}$$

which is precisely equal to $\mu_{qpq}\{1\}$. Proposition 1.2 shows that this point mass at 1 is structural: it is present in the law of qp_tq for all $t > 0$.

In the present paper, we study the law $\mu_t = \mu_{qp_tq}$ for all $t \geq 0$. As the calculation above shows, in the limit as $t \to \infty$, the operator-valued angle qp_tq flows towards a universal form determined only by $\tau(p)$ and $\tau(q)$. For any $t > 0$, the law μ_t completely determines the structure of the von Neumann algebra generated by p_t and q, cf. [17, 18, 25]. Properties of this measure μ_t are important for the analysis of free entropy in [18]. The main theorems of this paper relate to the analysis of this flow. Since μ (the weak limit of μ_t as $t \to \infty$) has a (structural) mass of $1 - \min\{\alpha, \beta\}$ at 0, it is sensible to remove this static singularity from the dynamics. The first major theorem of this paper shows that this produces a smooth flow equation for the Cauchy transform of μ_t; and the Cauchy transform itself is analytic in both z *and* t.

Theorem 1.4. *Let p, q be projections in a II_1-factor, and let u_t be a free unitary Brownian motion freely independent from $\{p, q\}$; set $p_t = u_t p u_t^*$. Let $\tau(p) = \alpha$ and $\tau(q) = \beta$, and let μ_t denote the spectral measure of the operator-valued angle qp_tq. For $t > 0$ and z in the upper half-plane $\mathbb{C}_+$, let*

$$G(t, z) = \int_0^1 \frac{\mu_t(dx)}{z - x} - \frac{1 - \min\{\alpha, \beta\}}{z}. \tag{1.8}$$

Then the function G is analytic in both $z \in \mathbb{C}_+$ and $t > 0$*, and satisfies the complex PDE*

$$\frac{\partial}{\partial t} G = \frac{\partial}{\partial z}\left[z(z-1)G^2 - (az + b)G\right] \tag{1.9}$$

where $a = 2\min\{\alpha, \beta\} - 1$ and $b = |\alpha - \beta|$.

Remark 1.5. PDE 1.9 is qualitatively similar to the complex (inviscid) Burger's equation, which is a well-known example exhibiting blow-up in finite time with bounded initial data, as well as shock behavior. There is no reason to expect the stated analyticity result to follow from the PDE; in fact, our analysis will use tools from PDE, complex analysis, and from free probability to prove the a priori analyticity (in both variables) of the function G.

A corollary to Theorem 1.4 is the following discontinuity result for the flow of the joint law of (p_t, q).

Corollary 1.6. *Let p, q be projections and let u_t be a free unitary Brownian motion and u_∞ a Haar unitary, both freely independent from $\{p, q\}$. Set $p_t = u_t p u_t^*$ and $p_\infty = u_\infty p u_\infty^*$. Assume that the support of the spectral measure of qpq is not equal to $[r_-, r_+]$ (cf. Equation 1.6). Then, for all sufficiently small $t > 0$, the law of (p_∞, q) is* not absolutely continuous *with respect to the law of (p_t, q).*

Proof. Since the flow $G(t, z)$ of the Cauchy transform of the spectral measure μ_t of qp_tq is analytic in $t > 0$ (cf. Theorem 1.4), the support of μ_t flows continuously, and hence cannot equal the support $[r_-, r_+]$ of the limit free product measure $\mu_p \boxtimes \mu_q$; this proves the result. □

Remark 1.7. This actually shows a somewhat stronger claim: let $(E^t_{ij})_{0\le i,j\le 1}$ be the four projections corresponding to p_t, q as in Section 4.2 below ($E_{11} = p_t \wedge q$, $E_{10} = p_t \wedge q^\perp$, etc.). Then, if the initial support of the spectral measure of qpq is not equal to $[r_-, r_+]$, it follows that the C^**-algebras*

$$C^*(p_t, q, (E^t_{ij})_{0\le i,j\le 1}) \ncong C^*(p_\infty, q, (E^\infty_{ij})_{0\le i,j\le 1})$$

for all sufficiently small $t > 0$.

Section 2 is concerned with the proof of Theorem 1.4. We use free stochastic calculus (Section 1.2) to find a system of ODEs satisfied by the time-dependent moments $\tau((qp_tq)^n)$ for $n \in \mathbb{N}$. These moments are the coefficients of the Laurent-series expansion of G in a neighbourhood of ∞, and on this domain the ODEs combine to give PDE 1.9. Analyticity, and continuation to all of $\mathbb{C}_+$, is then proved with careful estimates on the growth of derivatives given by the original ODEs and iteration of the PDE.

Remark 1.8. The papers [1] and [9] consider similar situations, developing PDEs governing the flow of analytic function transforms of spectral measures associated to the free liberation of two operators; in both cases, the PDEs are compatible with Theorem 1.4 above. In [1], the two initial operators are assumed to be classically independent, hence *commutative*; in this case, the authors were able to solve the PDE explicitly: their solution is a dilation of the law of the free unitary Brownian motion u_t, studied in [3]. The measure analogous to our μ_t is given meaning as a t*-free convolution*, yielding a continuous interpolation from classical independence to free independence. The paper [9] considers a situation similar to ours, with the assumption that one projection dominates the other; hence, the scope is similarly less extensive than the general situation we presently treat. The benefit of these specializations is that they obtain explicit solutions in particular cases, as well as unexpected algebraic identities.

In principle, the measure μ_t can be recovered from the function $G(t,z)$ via the Stieltjes inversion formula (cf. Section 1.1). In practice, understanding the flow of the *boundary values* of the function from a PDE in the interior is a very difficult problem in partial differential equations. We are, as yet, unable to complete the analysis of the measure μ_t in general; however, in the special case $\alpha = \beta = \frac{1}{2}$ (corresponding to $a = b = 0$ in PDE 1.9), we have the following complete analysis.

Theorem 1.9. *Let $\mu_t = \frac{1}{2}\delta_0 + \nu_t$ be the spectral measure in Theorem 1.4 in the special case $\alpha = \beta = \frac{1}{2}$. Then for any $t > 0$, the measure ν_t possesses a continuous density ρ_t on $(0,1)$. For any $t_0 > 0$, there is a constant $C(t_0)$ so that, for all $t \ge t_0$,*

$$\rho_t(x) \le \frac{C(t_0)}{\sqrt{x(1-x)}}. \tag{1.10}$$

Finally, the function ρ_t is real analytic on the set $\{x \in (0,1)\colon \rho_t(x) > 0\}$.

Note that the bound on ρ_t precisely reflects the asymptotic form the measure takes as $t \to \infty$: in the case $\alpha = \beta = \frac{1}{2}$, the Jacobi density of Equation 1.6 reduces to the (shifted) arcsine law $(2\pi)^{-1}(x(1-x))^{-1/2}$ on $[0,1]$; thus, we cannot expect any better behavior at the boundary of the interval. Theorem 1.9 shows that the measure μ_t does not possess any mass at the endpoints: although it may blow up at $x = 1$, the singularity is milder than would reflect the Cauchy transform of a point mass.

The smoothing results of Theorem 1.9 mirror similar properties of the so-called *free heat flow*: if μ is a compactly-supported measure on $\mathbb{R}$ and $\sigma_t(dx) = \frac{1}{2\pi t}\sqrt{(4t-x^2)_+}\,dx$ is the semicircular law of variance t, then the free convolution $\mu \boxplus \sigma_t$ possesses a continuous density which is real analytic on the set where it is > 0, cf. [5]. The techniques used to prove this theorem do not involve any PDEs, but are based on Biane's theory of *subordination* for the Cauchy transform. Motivated by those ideas, our approach to Theorem 1.9 is to develop an analogous theory of subordination for the liberation process: we show that, changing variables $H(t,z) = \sqrt{z}\sqrt{z-1}G(t,z)$, the flow of H may be encoded by a deformation of the identity in the initial condition: $H(t,z) = H(0, f_t(z))$ for a subordinate function f_t which extends to a homeomorphism from the closed upper half-plane $\overline{\mathbb{C}_+}$ onto a domain $\overline{\Omega_t} \subseteq \overline{\mathbb{C}_+}$.

Remark 1.10. In the recent preprint [37], Zhong has studied the multiplicative version of Biane's free heat flow, using related subordination technology to prove similar smoothness properties of the law of the free unitary Brownian motion multiplied by a free unitary. These results bear on the very recent related work of Izumi and Ueda [20], as described below in Remark 1.12 below.

Aside from the motivating question of general position for projections, the main application of the present results is to Voiculescu's theory of free entropy. In a (still ongoing) attempt to resolve the free group factor isomorphism problem, Voiculescu invented free probability as a way to import tools from classical probability, notably entropy and information theory, to produce invariants to distinguish von Neumann algebras. Motivated both by Shannon's original entropy constructions involving spacial microstates, and more sophisticated constructions in information theory (using conjugate variables, for example), he introduced free analogues of entropy, Fisher's information, and mutual information of collections of non-commuting random variables, in a series of six papers from 1993–1996; the most relevant for our purposes is [31]. Classically, there are many relationships that hold between these information measures; the general question of proving the analogous relationships for their free counterparts is known as the *Unification Conjecture* in free probability. In Section 4, we briefly describe the precise form of the unification conjecture, and prove it in the special case of von Neumann algebras generated by two projections (of trace $\frac{1}{2}$). Our main result in this direction is as follows.

Theorem 1.11. *Let $p, q \in (\mathscr{A}, \tau)$ be two projections in a II_1-factor, with trace $\tau(p) = \tau(q) = \frac{1}{2}$. Then the free mutual information of p, q and the free relative entropy of p, q are equal:*

$$i^*(p, q) = -\chi_{\mathrm{proj}}(p, q) + \chi_{\mathrm{proj}}(p) + \chi_{\mathrm{proj}}(q) = -\chi_{\mathrm{proj}}(p, q). \tag{1.11}$$

The assumption on the traces is in place since it is required for Theorem 1.9; we fully expect techniques similar to those presently developed will solve the general problem for all traces.

Remark 1.12. Theorem 1.11 was also recently proved in the preprint [20], which was posted to the arXiv seven months after the first public version of the present manuscript. Indeed, using the Jakowski transform as in [9], Izumi and Ueda show that μ_t can be identified with the spectral measure of the product of a free unitary Brownian motion with a free unitary operator whose distribution is determined by the initial projections (p, q). Thus, the appropriate smoothing analogues of our Theorem 1.9 follow from [37]. They use this to prove Theorem 1.11 much the same way we do below. They go further, and prove subordination results akin to our Section 3.3 below, without the restriction that $\tau(p) = \tau(q) = \frac{1}{2}$, and they use this to give some partial results generalizing Theorem 1.11 beyond the trace $\frac{1}{2}$ regime. The present authors find this to be a promising avenue of research.

This paper is organized as follows. The remainder of Section 1 is devoted to the relevant background for the rest of the work presented: Section 1.1 fixes the basic ideas and notation of free probability; Section 1.2 discusses the free Itô calculus; Section 1.3 is devoted to the free unitary Brownian motion that is central to this paper; and Section 1.4 describes the flaw in the proof of the general position result [2, Thm. 8.2] that partly motivated the present work. The main results of this paper are in Sections 2–4: Section 2 is the proof of Theorem 1.4; Section 3 uses this result to develop local properties of the measure μ_t, including the proof of Theorem 1.9; and Section 4 presents our main application, proving Theorem 1.11.

1.1 Free Probability

Here we briefly record the basic ideas and notation used in the sequel; the reader is directed to the books [22] and [33] for a full treatment. The setting of free probability is a non-commutative probability space; we will work in the richer framework of a W^*-**probability space** $(\mathscr{A}, \tau)$ where $\mathscr{A}$ is a von Neumann algebra, and τ is a normal, faithful tracial state on $\mathscr{A}$. The elements in $\mathscr{A}$ are called (non-commutative) random variables. The motivating example is the space $\mathscr{A} = \mathrm{End}_{\mathbb{C}}(\mathbb{C}^d) \otimes L^\infty(\Omega, \mathscr{F}, \mathbb{P})$ of all random matrices with bounded entries (over a probability space $(\Omega, \mathscr{F}, \mathbb{P})$); here the tracial state is $\tau = \frac{1}{d}\mathrm{Tr}_d \otimes \mathbb{E}$, the expected normalized trace. Voiculescu's key observation was that this W^*-probability space can be used to approximate (in the sense of moments) non-commutative random variables in infinite-dimensional von Neumann algebras, and classical independence combined with random rotation of eigenvectors for random matrices gives rise to a more general independence notion modeled on free groups.

Definition 1.13. *Let $(\mathscr{A}, \tau)$ be a W^*-probability space. The $*$-subalgebras $A_1, \ldots, A_n \subseteq \mathscr{A}$ are called* **free** *or* **freely independent** *if, given any centered elements $a_i \in A_i$, $\tau(a_i) = 0$, and any sequence $i_1, i_2, \ldots, i_m \in \{1, \ldots, n\}$ with $i_k \neq i_{k+1}$ for $1 \leq j < m$, $\tau(a_{i_1} a_{i_2} \cdots a_{i_m}) = 0$. Two random variables $a, b \in \mathscr{A}$ are freely independent if the $*$-subalgebras they generate are freely independent.*

Free independence is a moment-factorization condition. For example, if a and b are freely independent, then $\tau(a^n b^m) = \tau(a^n)\tau(b^m)$ for any $n, m \in \mathbb{N}$, coinciding the classical independence of bounded random variables; for non-commutating variables, freeness includes more complicated factorizations such as $\tau(abab) = \tau(a^2)\tau(b)^2 + \tau(a)^2\tau(b^2) - \tau(a)^2\tau(b)^2$. Freeness is modeled on freeness in group theory. Let $\mathscr{A} = L(\mathbb{F}_k)$ denote the **free group factor** with k generators $u_1, \ldots, u_k$. ($L(\mathbb{F}_k)$ is the von Neumann algebra generated by the left-regular representation of the free group $\mathbb{F}_k$ on $\ell^2(\mathbb{F}_k)$; i.e. the von Neumann algebra generated by convolution on the free group.) Then subalgebras generated by disjoint subsets of the generators $\{u_1, \ldots, u_k\}$ are freely independent.

If $\{(\mathscr{A}_n, \tau_n) : 1 \leq n \leq \infty\}$ are W^*-probability spaces, and $k \in \mathbb{N}$, a sequence $\mathbf{a}_n = (a_n^1, \ldots, a_n^k) \in (\mathscr{A}_n)^k$ is said to **converge in distribution** to $\mathbf{a} = (a^1, \ldots, a^k) \in \mathscr{A}_\infty^k$ if, for any polynomial P in k non-commuting indeterminates, $\tau_n[P(\mathbf{a}_n)] \to \tau_\infty[P(\mathbf{a})]$ as $n \to \infty$. That is: each mixed moment in $\mathbf{a}_n$ converges to that moment in $\mathbf{a}$. Such a sequence is said to be **asymptotically free** if its limit consists of random variables $a^1, \ldots, a^k$ that are freely independent.

Proposition 1.14 ([22, 33])**.** *Let X_n and Y_n be $n \times n$ random matrices with all moments finite, and suppose that each has a limit in distribution separately. Let U_n be a unitary matrix sampled from the Haar measure on $U(n)$, independent from X_n and Y_n. Then $(X_n, U_n Y_n U_n^*)$ are asymptotically free.*

Proposition 1.14 asserts that freeness is an asymptotic statement about the *eigenvectors* of random matrices: freeness means that their eigenspaces are independently, uniformly randomly rotated against each other. This result allows the realization of free random variables and stochastic processes as limits of random matrix ensembles; more on that in Section 1.2.

If $a \in (\mathscr{A}, \tau)$ is a single self-adjoint random variable, it possesses a spectral resolution E^a taking values in the projections of $\mathscr{A}$. The composition $\tau \circ E^a$ produces a probability measure μ_a on the spectrum of a, known as the **spectral measure** of a; it is determined by the moments of a:

$$\tau(a^n) = \int_{\mathbb{R}} x^n \, \mu_a(dx).$$

In general the non-commutative distribution of a random vector $\mathbf{a} \in \mathscr{A}^k$ is the set of all traces of non-commutative polynomials in $\mathbf{a}$; in the case of a single self-adjoint element ($k = 1$), these moments are encoded by the single measure μ_a, coinciding with the law of a classical random variable. In the case of a self-adjoint random matrix a, μ_a is the average empirical eigenvalue distribution (the average of the random probability measure placing a point-mass at each eigenvalue of the matrix).

Definition 1.15. *Let μ be a compactly-supported finite positive measure on $\mathbb{R}$. The* **Cauchy transform** *G_μ is the analytic function on the upper half-plane $\mathbb{C}_+$ defined by*

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z - x} \, \mu(dx). \tag{1.12}$$

The function G_μ is analytic on $\mathbb{C} - \operatorname{supp}\mu$, but does not have a continuous extension across $\operatorname{supp}\mu$. A generally useful uniform estimate for the Cauchy transform in the upper half-plane is

$$|G_\mu(z)| \le \frac{\mu(\mathbb{R})}{|\Im z|}. \tag{1.13}$$

Another reason it is customary to restrict it to the upper half-plane is that the measure can be recovered from its action there, via the **Stieltjes inversion formula**:

$$\mu(dx) = -\frac{1}{\pi}\lim_{\epsilon\downarrow 0}\Im G_\mu(x+i\epsilon). \tag{1.14}$$

The limit in Equation 1.14 is a weak limit: if f is a continuous test function, the integral of f against the ϵ-dependent measure on the right converges to $\int f\,d\mu$; if μ possesses a sufficiently integrable continuous density, the limit is also true pointwise for the density. If $\mu(dx) = \rho(x)\,dx$ where $\rho \in L^p(\mathbb{R})$ for some $p \in (0,\infty)$, then the **Hilbert transform**

$$H\rho(x) \equiv \frac{1}{\pi}\,p.v.\!\int \frac{\rho(y)}{x-y}\,dy \tag{1.15}$$

is also in L^p, and gives the boundary values of the real part of G_μ; that is

$$-\frac{1}{\pi}\lim_{\epsilon\downarrow 0}\Im G_\mu(x+i\epsilon) = \rho(x), \quad \frac{1}{\pi}\lim_{\epsilon\downarrow 0}\Re G_\mu(x+i\epsilon) = H\rho(x). \tag{1.16}$$

The Stieltjes inversion formula is also robust under vague limits of measures.

Theorem 1.16 (Stieltjes continuity theorem)**.** *Let μ_n be a sequence of probability measures on $\mathbb{R}$.*

(1) If $\mu_n \to \mu$ weakly, then $G_{\mu_n} \to G_\mu$ uniformly on compact subsets of $\mathbb{C}_+$.

(2) Conversely, if $G_{\mu_n} \to G$ pointwise on $\mathbb{C}_+$, then G is the Cauchy transform of a finite positive measure μ, and $\mu_n \to \mu$ vaguely. If it is known a priori that $G = G_\mu$ for a probability measure μ, then $\mu_n \to \mu$ weakly.

It is possible for mass to escape at ∞ in a vague limit; for example, it is possible for $G_{\mu_n} \to 0$ pointwise. In our applications, the limit will be directly identified as the Cauchy transform of a probability measure.

Freeness, which is a property of moments (hence of distributions), can be encoded in terms of the Cauchy transform. Given a compactly-supported probability measure μ, the *$\mathscr{R}$-transform $\mathscr{R}_\mu$* is the analytic function on a neighborhood of the identity in the upper half-plane determined by the functional equation

$$G_\mu(\mathscr{R}_\mu(z) + 1/z) = z, \quad z \in \mathbb{C}_+, \quad |z| \text{ small}. \tag{1.17}$$

The $\mathscr{R}$-transform in fact determines the Cauchy transform, by analytic continuation, modulo the constraint $\lim_{|z|\to\infty} zG_\mu(z) = \mu(\mathbb{R}) = 1$. It is the free analogue of the (log-)Fourier transform: it linearizes freeness.

Proposition 1.17 ([33])**.** *Let a, b be self-adjoint random variables. Then a, b are freely independent if and only if*

$$\mathscr{R}_{\mu_{a+b}}(z) = \mathscr{R}_{\mu_a}(z) + \mathscr{R}_{\mu_b}(z)$$

for all sufficiently small $z \in \mathbb{C}_+$.

This highlights the fact that free independence yields a free convolution operation on measures. Given two (compactly-supported) probability measures μ, ν, their **(additive) free convolution** $\mu \boxplus \nu$ is the measure determined by $\mathscr{R}_{\mu\boxplus\nu} = \mathscr{R}_\mu + \mathscr{R}_\nu$. That is: realize μ and ν as the laws of two free random variables a, b; then $\mu \boxplus \nu$ is the law of $a + b$.

There is a similar notion of **(multiplicative) free convolution** for positive operators. If a and b are free, then the law of $a^{1/2}ba^{1/2}$ is denoted $\mu_a \boxtimes \mu_b$. It is determined through another analytic function transform known as the *$\mathscr{S}$-transform.* For a compactly-supported probability measure μ, let $\chi_\mu(z) = \frac{1}{z}G_\mu(\frac{1}{z}) - 1$ be its shifted moment-generating function, defined in a neighborhood of 0. The $\mathscr{S}$-transform $\mathscr{S}_\mu$ is the analytic function defined in a neighborhood of 0 by the functional equation

$$\chi_\mu\left(\tfrac{z}{z+1}\mathscr{S}_\mu(z)\right) = z, \quad |z| \text{ small.} \tag{1.18}$$

Then $\mathscr{S}_{\mu\boxtimes\nu}(z) = \mathscr{S}_\mu(z)\mathscr{S}_\nu(z)$, as shown in [33]. It is through this operation that the free Jacobi law of Equation 1.6 was determined.

1.2 Free Brownian Motion and Free Stochastic Calculus

Let $X_n(t)$ denote an $n \times n$ Hermitian matrix-valued Brownian motion: the upper-triangular entries $[X_n(t)]_{ij}, i < j$ are independent complex Brownian motions of total variance t/n; the diagonal does not matter to the limit. Then there is a W^*-probability space $(\mathscr{A}, \tau)$ in which the limit in distribution of any collection of instances of X_n at different times $(X_n(t_1), X_n(t_2), \ldots, X_n(t_k))$ for $0 \le t_1 < t_2 < \cdots < t_k$ may be realized. For fixed t, $X_n(t)$ is a Gaussian unitary ensemble, whose limit empirical eigenvalue distribution is Wigner's semicircle law $\sigma_t(dx) = \frac{1}{2\pi t}\sqrt{(4t - x^2)_+}\, dx$ [35, 36]; hence the limits $x_{t_1}, \ldots, x_{t_k}$ are semicircular random variables. Mirroring the isonormal Gaussian process construction of Brownian motion, the limit may be realized in any W^*-probability space rich enough to contain an infinite sequence of freely independent identically distributed semicircular random variables (e.g. any free group factor). The limit x_t is a non-commutative stochastic process with properties analogous to Brownian motion.

Definition 1.18. *A(n additive)* **free Brownian motion** *in a W^*-probability space $(\mathscr{A}, \tau)$ is a non-commutative stochastic process $(x_t)_{t\ge 0}$ with the following properties:*

(1) The increments of x_t are freely independent: for $0 \le t_1 < t_2 < \cdots < t_k$,

$$x_{t_2} - x_{t_1}, x_{t_3} - x_{t_2}, \ldots, x_{t_k} - x_{t_{k-1}}$$

are freely independent.

(2) The process is stationary, with semicircular increments: for $0 \le s < t$, the law of $x_t - x_s$ is σ_{t-s}.

(3) The $\mathscr{A}$-valued function $t \mapsto x_t$ is weakly continuous.

Free Brownian motion is the limit of matrix-valued Brownian motion; it can also be constructed as an isonormal process, or through a Fock space construction, cf. [6]. With the properties of Definition 1.18 in hand, the standard construction of the Itô integral may be mirrored. If θ_t is a process adapted to x_t (meaning that θ_t is in the von Neumann subalgebra generated by $\{x_s\}_{s\le t}$ for each $t \ge 0$), then one can define the stochastic integral

$$\int \theta_t\, dx_t$$

as an $L^2(\mathscr{A}, \tau)$-limit of step functions of the form $\sum_k \theta_{t_k}(x_{t_k} - x_{t_{k-1}})$. The relationship $\phi_t = \int \theta_t\, dx_t$ is abbreviated as $d\phi_t = \theta_t\, dx_t$. A (left) **free Itô process** y_t is a stochastic process of the form

$$y_t = \int_0^t \theta_s\, dx_s + \int_0^t \phi_s\, ds \tag{1.19}$$

where θ_t and ϕ_t are adapted processes; here $\int_0^t \theta_s\, dx_s$ is short-hand for $\int \theta_s \mathbb{1}_{[0,t]}(s)\, dx_s$ and so forth. Evidently, a free Itô process is adapted. Equation 1.19 is usually written in the form

$$dy_t = \theta_t\, dx_t + \phi_t\, dt. \tag{1.20}$$

This stochastic differential notation is useful, and allows for succinct description of the rules of free Itô calculus. The most important is the **free Itô formula** which we will use in product form:

$$d(y_t z_t) = (dy_t)z_t + y_t(dz_t) + (dy_t)(dz_t). \tag{1.21}$$

Here y_t and z_t are free Itô processes. For example, if $dy_t = \theta_t\, dx_t + \phi_t\, dt$ and $dz_t = \theta'_t\, dx_t + \phi'_t\, dt$, then

$$y_t(dz_t) = y_t(\theta'_t\, dx_t + \phi'_t\, dt) = (y_t\theta'_t)\, dx_t + (y_t\phi'_t)\, dt$$

while

$$(dy_t)z_t = (\theta_t\, dx_t + \phi_t\, dt)z_t = \theta_t\, dx_t\, z_t + (\theta_t z_t)\, dt.$$

The processes are non-commutative, so we must be able to make sense of such terms; moreover, the product $(dy_t)(dz_t)$ will contain mixed terms $dx_t\, dt$ and so forth. The rules, akin to standard Itô calculus, for these terms are as follows:

$$dx_t\, \theta_t\, dx_t = \tau(\theta_t)\, dt \tag{1.22}$$

$$dx_t\, dt = dt\, dx_t = (dt)^2 = 0 \tag{1.23}$$

Equation 1.22 may seem counter-intuitive from classical Itô calculus: for a standard 1-dimensional Brownian motion B_t, $(dB_t)^2 = dt$. It is easy to calculate, however, that for $n \times n$ matrix-valued Brownian motion $X_n(t)$ and an adapted matrix-valued process $Y_n(t)$, the Itô product-rule takes the form $dX_n(t)\, Y(t) dX_n(t) = \frac{1}{n}\mathrm{Tr}\,(Y_n(t))\, dt$. Therefore, Equation 1.22 follows from standard trace-concentration results, cf. [6]. One final useful result is that free Itô integrals of adapted processes, like their classical cousins, are centered; i.e.

$$\tau(\theta_t\, dx_t) = 0. \tag{1.24}$$

The standard approach from classical stochastic calculus (using the Picard iteration technique, for example) can then be used to solve **free stochastic differential equations** of the form

$$dy_t = a(t, y_t)\, dx_t + b(t, y_t)\, dt \tag{1.25}$$

for sufficiently smooth and slowly-growing functions $a, b\colon \mathbb{R}_+ \times \mathscr{A} \to \mathscr{A}$. The reader is referred to [6, 7] for details.

Remark 1.19. As the calculations following Equation 1.21 demonstrate, the product of two *left* free Itô processes is not, in general, a *left* free Itô process: it may contain terms like $dx_t\, \theta_t$ in differential form, which is not equal to $\theta_t\, dx_t$. A proper treatment of free Itô calculus should be formulated in terms of biprocesses $t \mapsto \omega_t \in \mathscr{A} \otimes \mathscr{A}$ that can act on the left and the right simultaneously. Biane and Speicher develop this theory in [6]; the corresponding stochastic integral is denoted

$$\int \omega_t \sharp dx_t,$$

defined as an L^2-limit of sums of the form $\sum_k \theta_{t_k}(x_{t_k} - x_{t_{k-1}})\phi_{t_k}$ where ω_t is approximated by $\sum_k \theta_{t_k} \otimes \phi_{t_k}$. We have described free Itô calculus in slightly imprecise terms to avoid the new notational complexity; for our purposes, the more general theory is essentially unnecessary. The interested reader can find a careful overview in [21].

1.3 Free Unitary Brownian Motion

Introduced in [3], the **free unitary Brownian motion** is the solution to the free Itô stochastic differential equation

$$du_t = iu_t\, dx_t - \frac{1}{2}u_t\, dt \tag{1.26}$$

with initial condition $u_0 = 1$; here, as usual, x_t is a(n additive) free Brownian motion. It would, perhaps, be most accurate to call the solution of this free SDE a *left* free unitary Brownian motion. To be sure, note that the adjoint u_t^* satisfies

$$du_t^* = (du_t)^* = -idx_t\, u_t^* - \frac{1}{2}u_t^*\, dt \tag{1.27}$$

since x_t is self-adjoint. Equation 1.26 is the exact free analogue of Equation 1.3 for the finite-dimensional unitary Brownian motion. Indeed, this comes from the fact that u_t is the limit in distribution of the Brownian motion on the unitary group, in the same sense that x_t is the limit in distribution of the Hermitian-matrix valued Brownian motion. It is a unitary-valued stochastic process, whose distribution is a continuous function on the unit circle for each $t > 0$. Biane calculated the moments of this measure:

$$\tau((u_t)^k) = e^{-kt/2}\sum_{j=0}^{k-1}\frac{(-t)^j}{j!}\binom{k}{j+1}k^{j-1}, \quad k \geq 0. \tag{1.28}$$

Using these moments, it is possible to compute an implicit description of the density of μ_{u_t} (with respect to the uniform probability measure on the unit circle); for $t < 2$, the measure is supported in a strict symmetric subset, and then achieves full support for $t \geq 2$.

Analogous to Definition 1.18, and in line with the properties of the Brownian motion on the unitary groups, the free unitary Brownian motion has the following properties, which can be derived directly from the stochastic differential equation 1.26.

Proposition 1.20 ([3]). *Let ν_t be the measure on the unit circle possessing the moments on the right-hand-side of Equation 1.28. Then the free unitary Brownian motion satisfies the following properties.*

(1) The multiplicative increments of u_t are freely independent: for $0 \leq t_1 < t_2 < \cdots < t_k$,

$$u_{t_1}^* u_{t_2},\ u_{t_2}^* u_{t_3},\ \ldots,\ u_{t_{k-1}}^* u_{t_k}$$

are freely independent.

(2) The process is stationary: for $0 \leq s < t$, the law of the unitary random variable $u_t u_s^*$ is $\mu_{u_{t-s}}$.

(3) The $\mathscr{A}$-valued function $t \mapsto u_t$ is weakly continuous.

Comparable statements can be made about the right free unitary Brownian motion u_t^*.

1.4 The Flaw in [2, Thm. 8.2]

The final theorem in the first author's paper [2] claimed to prove the motivating theorem of the present paper: if p, q are projections in a W^*-probability space $(\mathscr{A}, \tau)$, and if and u_t is a free unitary Brownian motion, freely independent from $\{p, q\}$, then $p_t = u_t p u_t^*$ and q are in general position for all $t > 0$. The idea of the given proof was as follows. First, by making replacements $p \leftrightarrow 1 - p$ and $q \leftrightarrow 1 - q$ if necessary, it suffices to prove the theorem in the case $\tau(p), \tau(q) \leq \frac{1}{2}$, in which case the general position statement is that $\tau(p_t \wedge q) = 0$. Let $r_t = p_t \wedge q$, and define a function $F_t \colon \mathbb{R}_+ \to \mathbb{R}$ by

$$F_t(s) = \tau[(r_t p_s r_t - r_t)^2],\ s \geq 0.$$

The function F_t is non-negative. Since r_t is a projection onto a subspace of the image of p_t, $r_t p_t r_t = p_t$ and so $F_t(t) = 0$; hence $F_t(s)$ has a minimum at $s = t$. The claim is then made that the function F_t is differentiable, and so $F_t'(t) = 0$. Free Itô calculus is then applied to calculate the derivative; it is then claimed that $F_t'(t) \le -\tau(r_t)^2\tau(p)$. Thus, $F_t'(t) < 0$ unless $\tau(r_t) = 0$ as required.

There are two flaws with this argument. The first, fairly subtle, is the assumption of differentiability. The stochastic process $s \mapsto (r_t p_s r_t - r_t)^2$ is manifestly *not adapted* to the filtration generated by $(u_s)_{s\ge 0}$: for $s < t$, it depends on the value of u_t. It is only adapted for $s \ge t$, in which case the tools of Itô calculus indeed apply as stated. Thus, it was correctly demonstrated that the function $F_t(s)$ has a minimum at $s = t$, and is differentiable for $s \in [t, \infty)$ (meaning *right-differentiable* at $s = t$). If it were indeed true that $\alpha_t'(t) \le -\tau(r_t)^2\tau(p)$, the argument would remain valid: a right-differentiable function cannot have a minimum at a point where the right-derivative is strictly negative. Unfortunately, there was also a calculation error in the determination of this derivative. In fact:

Proposition 1.21. *The right-derivative of F_t satisfies $F_t'(t) = 2\tau(r_t)(1 - \tau(p)) \ge 0$.*

This produces no contradiction to the possibility that $\tau(r_t) > 0$, however: since the function $F_t(s)$ is not known to be differentiable in a neighborhood of $s = t$, its right-derivative may well be strictly positive at the minimum (with behavior akin to the function $s \mapsto |s - t|$, for example). The proof of Proposition 1.21 is delayed until Section 2.2.

It is possible this proof could be mended using a free version of stochastic calculus for non-adapted processes, which has yet to be developed. In the classical case, such techniques are based on the Malliavin calculus [23], which does have an analogue in the world of free probability, developed in [6, 7] and further developed in the second author's recent paper [21]. Nevertheless, even if such a non-adapted calculus were applicable and mirrored the classical behavior, it is likely the function $F_t(s)$ could still not be proved differentiable at the point $s = t$, but only on the complement of this point in $\mathbb{R}_+$. Thus, to prove the general position claim, fundamentally different techniques are required; this is part of the impetus for the present paper.

2 The Flow of the Spectral Measure μ_t

2.1 The Flow of Moments

Let $p, q \in \mathscr{A}$ be projections with $\tau(p) = \alpha$ and $\tau(q) = \beta$. Let u_t be a free unitary Brownian motion, free from p, q, and (as usual) define $p_t = u_t p u_t^*$ for $t \ge 0$. Our present goal is to understand the moments

$$g_n(t) = \tau[(q p_t q)^n] \quad n \ge 1, t \ge 0. \tag{2.1}$$

We will use stochastic calculus to derive a system of ODEs satisfied by the functions g_n on $(0, \infty)$. To that end, we need a generalization of the Itô product rule of Equation 1.21 to products of many adapted processes. An easy induction argument shows that if $a_1(t), \ldots, a_n(t)$ are adapted processes then

$$d(a_1 \cdots a_n) = \sum_{j=1}^n a_1 \cdots a_{j-1}\, da_j\, a_{j+1} \cdots a_n + \sum_{1\le i<j\le n} a_1 \cdots a_{i-1}\, da_i\, a_{i+1} \cdots a_{j-1}\, da_j\, a_{j+1} \cdots a_n.$$

(The induction follows from the fact that $da_i\, da_j\, da_k = 0$ for any i, j, k; this follows from repeated applications of the rules of Equation 1.23.) Specializing to the case $a_1 = \cdots = a_n = a$, we have

$$d(a^n) = \sum_{j=1}^n a^{j-1}\, da\, a^{n-j} + \sum_{1\le i<j\le n} a^{i-1}\, da\, a^{j-i-1}\, da\, a^{n-j} \tag{2.2}$$

(Note: Equation 2.2 only makes sense for $n \geq 2$.) We will apply this to the adapted process $a_t = qp_tq$, and then take the trace to find $dg_n(t)$. This affords immediate simplifications:

$$\begin{aligned}\tau[d(a^n)] &= \sum_{j=1}^{n} \tau[a^{j-1}\,da\,a^{n-j}] + \sum_{1\leq i<j\leq n} \tau[a^{i-1}\,da\,a^{j-i-1}\,da\,a^{n-j}] \\ &= \sum_{j=1}^{n} \tau[a^{n-1}\,da] + \sum_{1\leq i<j\leq n} \tau[a^{n-(j-i)-1}\,da\,a^{(j-i)-1}\,da]\end{aligned}$$

where we have used the trace property and combined terms on the left. The terms in the first sum do not depend on the summation variable j, and so this simply becomes $n\tau[a^{n-1}\,da]$. In the second summation, the summands depend on the summation variables i, j only through their difference $k = j - i$ which ranges from 1 up to $n-1$. We therefore reindex by this variable, and that sum becomes

$$\sum_{1\leq i<j\leq n} \tau[a^{n-(j-i)-1}\,da\,a^{(j-i)-1}\,da] = \sum_{k=1}^{n-1} \sum_{\substack{1\leq i<j\leq n \\ j-i=k}} \tau[a^{n-k-1}\,da\,a^{k-1}\,da].$$

For fixed $k \in \{1, \ldots, n-1\}$, the number of pairs (i, j) with $j - i = k$ is equal to $n - k$, and so this summation S becomes

$$S = \sum_{k=1}^{n-1} (n-k)\tau[a^{n-k-1}\,da\,a^{k-1}\,da]. \tag{2.3}$$

Reindexing $j = n - k$ shows that this sum is also given by

$$S = \sum_{j=1}^{n-1} j\,\tau[a^{j-1}\,da\,a^{n-j-1}\,da]. \tag{2.4}$$

Using the trace property, adding Equations 2.3 and 2.4 gives the simplification

$$2S = \sum_{j=1}^{n-1} (n-j+j)\,\tau[a^{n-j-1}\,da\,a^{j-1}\,da] = n\sum_{j=1}^{n-1} \tau[a^{n-j-1}\,da\,a^{j-1}\,da].$$

Thus, we have

$$\tau[d(a^n)] = n\,\tau[a^{n-1}\,da] + \frac{1}{2}n \sum_{j=1}^{n-1} \tau[a^{n-j-1}\,da\,a^{j-1}\,da]. \tag{2.5}$$

Now, $a_t = qp_tq$ and so (applying the Itô product rule 1.21 twice) $da_t = qdp_tq$. To evaluate the differential dp_t, the product rule again gives

$$dp_t = d(u_tpu_t^*) = (du_t)pu_t^* + u_t[d(pu_t^*)] + (du_t)[d(pu_t^*)],$$

and since p is constant with respect to time,

$$dp_t = (du_t)pu_t^* + u_tpdu_t^* + (du_t)p(du_t^*).$$

We now substitute the stochastic differential equations 1.26 and 1.27 to express dp_t as

$$dp_t = (iu_t\,dx_t - \frac{1}{2}u_t\,dt)pu_t^* + u_tp(-idx_t\,u_t^* - \frac{1}{2}u_t^*\,dt) + (iu_t\,dx_t - \frac{1}{2}u_t\,dt)p(-idx_t\,u_t^* - \frac{1}{2}u_t^*dt).$$

The first two terms simplify to give

$$iu_t\,dx_t\,pu_t^* - iu_tp\,dx_t\,u_t^* - u_tpu_t^*\,dt.$$

The final term has only one surviving factor: by Equation 1.22

$$(iu_t\,dx_t)p(-idx_t\,u_t^*) = u_tdx_t\,p\,dx_tu_t^* = u_t\tau(p)u_t^*\,dt = \tau(p)\,dt.$$

Altogether, then, we have

$$dp_t = -u_tpu_t^*\,dt + iu_t\,dx_t\,pu_t^* - iu_tp\,dx_t\,u_t^* + \tau(p)\,dt. \tag{2.6}$$

Recalling that $\tau(p) = \alpha$ and that $u_tpu_t^* = p_t$, the first and last term combine to $(\alpha - p_t)\,dt$. For the middle terms, it is useful to introduce a new process. Let

$$dy_t = iu_t\,dx_t\,u_t^*. \tag{2.7}$$

While it is easy to see that this SDE has a unique solution y_t (satisfying $y_0 = 0$), we need not concern ourselves with this fact. Indeed, all that is important is that the rules (Equations 1.22 and 1.23) of Itô calculus applied with y_t instead of x_t take the following form. If θ_t is an adapted process, then

$$dy_t\,\theta_t\,dy_t = (iu_t\,dx_t\,u_t^*)\theta_t(iu_t\,dx_t\,u_t^*) = -u_t\tau(u_t^*\theta_tu_t)u_t^*\,dt = -\tau(\theta_t)\,dt \tag{2.8}$$

and

$$dy_t\,dt = dt\,dy_t = (dt)^2 = 0. \tag{2.9}$$

Now, Equation 2.6 can be rewritten as

$$dp_t = (\alpha - p_t)\,dt + dy_t\,p_t - p_t\,dy_t. \tag{2.10}$$

Thus, with $a_t = qp_tq$, it follows that

$$da_t = q\,dp_t\,q = (\alpha q - a_t)\,dt + q\,dy_t\,p_tq - qp_t\,dy_t\,q. \tag{2.11}$$

We now simplify the first term in Equation 2.5.

$$a_t^{n-1}\,da_t = a_t^{n-1}(\alpha q - a_t)\,dt + a_t^{n-1}\,q\,dy_t\,p_tq - a_t^{n-1}qp_t\,dy_t\,q.$$

By Equation 1.24, the last two terms have trace 0, and so we simply have

$$\tau(a_t^{n-1}\,da_t) = \tau[a_t^{n-1}(\alpha q - a_t)]\,dt = [\alpha\,\tau(a_t^{n-1}) - \tau(a_t^n)]\,dt \tag{2.12}$$

where we have simplified $a_t^{n-1}q = (qp_tq)^{n-1}q = (qp_tq)^{n-1} = a_t^{n-1}$ since $q^2 = q$.

For the second term (the summation) in Equation 2.5, it is convenient to make yet another transformation. Define z_t by

$$dz_t = q\,dy_t\,p_tq - qp_t\,dy_t\,q. \tag{2.13}$$

Again, one can use standard theory to show that there is a unique adapted process with $z_0 = 0$ satisfying this SDE, but this is not important for present considerations. The following lemma expresses the form of the Itô calculus in terms of the process z_t.

Lemma 2.1. *Let z_t be defined by Equation 2.13. Then $dz_t\,dt = dt\,dz_t = (dt)^2 = 0$, and if θ_t is an adapted process, then*

$$dz_t\,\theta_t\,dz_t = [-2\tau(a_t\theta_t)a_t + \tau(a_t\theta_t)q + \tau(q\theta_t)a_t]\;dt. \tag{2.14}$$

Proof. Since z_t is a stochastic integral, the Itô rules regarding product with dt apply as usual. For Equation 2.14, we simply expand

$$\begin{aligned} dz_t\,\theta_t\,dz_t &= (q\,dy_t\,p_tq - qp_t\,dy_t\,q)\theta_t(q\,dy_t\,p_tq - qp_t\,dy_t\,q) \\ &= q\,dy_t\,(p_tq\theta_tq)\,dy_t\,p_tq - q\,dy_t\,(p_tq\theta_tqp_t)\,dy_t\,q - qp_t\,dy_t\,(q\theta_tq)\,dy_t\,p_tq + qp_t\,dy_t\,(q\theta_tqp_t)\,dy_t\,q. \end{aligned}$$

Applying Equation 2.8 to each of these four terms yields

$$-q\cdot\tau(p_tq\theta_tq)\cdot p_tq\,dt + q\cdot\tau(p_tq\theta_tqp_t)\cdot q\,dt + qp_t\cdot\tau(q\theta_tq)\cdot p_tq\,dt - qp_t\cdot\tau(q\theta_tqp_t)\cdot q\,dt.$$

Using the trace property, and simplifying with the relations $q^2 = q$, $p_t^2 = p_t$, and $qp_tq = a_t$, yields

$$dz_t\,\theta_t\,dz_t = -\tau(a_t\theta_t)a_t\,dt + \tau(a_t\theta_t)q\,dt + \tau(q\theta_t)a_t\,dt - \tau(a_t\theta_t)a_t\,dt,$$

which simplifies to give Equation 2.14. □

Refer now to the summands in the second term in Equation 2.5. From Equations 2.11 and 2.13, we have

$$da_t = (\alpha q - a_t)\,dt + dz_t.$$

Hence, for $j \in \{1, \ldots, n-1\}$,

$$a_t^{n-j-1}\,da_t\,a_t^{j-1}\,da_t = a_t^{n-j-1}\,[(\alpha q - a_t)\,dt + dz_t]\,a_t^{j-1}\,[(\alpha q - a_t)\,dt + dz_t].$$

We may expand this into four terms. However, since dt commutes with everything and, by Lemma 2.1 $dt\,dz_t = 0$, and as always $(dt)^2 = 0$, the only surviving term is

$$a_t^{n-j-1}\,da_t\,a_t^{j-1}\,da_t = a_t^{n-j-1}\,dz_t\,a_t^{j-1}\,dz_t.$$

Employing Equation 2.14 of Lemma 2.1, we therefore have

$$a_t^{n-j-1}\,da_t\,a_t^{j-1}\,da_t = a_t^{n-j-1}[-2\tau(a_t^j)a_t + \tau(a_t^j)q + \tau(qa_t^{j-1})a_t]\,dt.$$

Taking the trace, we have

$$\tau(a_t^{n-j-1}\,da_t\,a_t^{j-1}\,da_t) = \Big[-2\tau(a_t^j)\tau(a_t^{n-j}) + \tau(a_t^j)\tau(a_t^{n-j-1}q) + \tau(qa_t^{j-1})\tau(a_t^{n-j})\Big]\,dt. \tag{2.15}$$

Provided $j-1 \neq 0$ and $n-j-1 \neq 0$ (i.e. $n \geq 4$ and $j \in \{2, \ldots, n-2\}$), $qa_t^{j-1} = a_t^{j-1}$ and $qa_t^{n-j-1} = a_t^{n-j-1}$, as per the discussion following Equation 2.5. So, in this regime, we have

$$\tau(a_t^{n-j-1}\,da_t\,a_t^{j-1}\,da_t) = \Big[-2\tau(a_t^j)\tau(a_t^{n-j}) + \tau(a_t^j)\tau(a_t^{n-j-1}) + \tau(a_t^{j-1})\tau(a_t^{n-j})\Big]\,dt, \quad 2 \leq j \leq n-2. \tag{2.16}$$

The case $j = 1$ corresponds to $\tau(a_t^{n-2}\,da_t\,da_t)$, while $j = n-1$ corresponds to $\tau(da_t\,a^{n-2}\,da_t)$. By the trace property, these are equal. In each case, one of the q-terms is $\tau(a_t^{n-2}q) = \tau(a_t^{n-2})$, while the other is $\tau(a_t^0q) = \tau(q) = \beta$. So, we have

$$\tau(a_t^{n-j-1}\,da_t\,a_t^{j-1}\,da_t) = \big[-2\tau(a_t)\tau(a_t^{n-1}) + \tau(a_t)\tau(a_t^{n-2}) + \beta\tau(a_t^{n-1})\big]\,dt, \quad j \in \{1, n-1\}. \tag{2.17}$$

Hence, the second term in Equation 2.5 is equal to

$$\begin{aligned} &n[-2\tau(a_t)\tau(a_t^{n-1}) + \tau(a_t)\tau(a_t^{n-2}) + \beta\tau(a_t^{n-1})]\,dt \\ &+\frac{1}{2}n\sum_{j=2}^{n-2}[-2\tau(a_t^j)\tau(a_t^{n-j}) + \tau(a_t^j)\tau(a_t^{n-j-1}) + \tau(a_t^{j-1})\tau(a_t^{n-j})]\,dt \end{aligned} \tag{2.18}$$

where the summation is taken to be 0 in the case $n \le 3$. Combining Equations 2.12 and 2.18 with Equation 2.5, and using the notation $g_n(t) = \tau(a_t^n)$, we therefore have

$$\begin{aligned} dg_n = \tau[d(a^n)] = [n\alpha\, g_{n-1} - n\, g_n - 2n\, g_1 g_{n-1} + n\, g_1 g_{n-2} + n\beta\, g_{n-1}]\, dt \\ + \frac{1}{2} n \sum_{j=2}^{n-2} [-2 g_j g_{n-j} + g_j g_{n-j-1} + g_{j-1} g_{n-j}]\, dt. \end{aligned} \tag{2.19}$$

From here we see that g_n is actually differentiable, and that Equation 2.19 is a differential equation. One more simplification is in order. By making the change of index $k = n - j$, note that

$$\sum_{j=2}^{n-2} g_j g_{n-j-1} = \sum_{k=2}^{n-2} g_{n-k} g_{k-1}.$$

Hence, the second and third summands in the second line of Equation 2.19 have the same sum, and we have

$$g_n' = -n\, g_n + n(\alpha+\beta) g_{n-1} - 2n\, g_1 g_{n-1} + n\, g_1 g_{n-2} + \frac{1}{2} n \sum_{j=2}^{n-2} [-2 g_j g_{n-j} + 2 g_{j-1} g_{n-j}]. \tag{2.20}$$

We can then recombine the third and fourth terms (just before the summation) as follows:

$$\begin{aligned} & - 2n\, g_1 g_{n-1} + n\, g_1 g_{n-2} + \frac{1}{2} n \sum_{j=2}^{n-2} [-2 g_j g_{n-j} + 2 g_{j-1} g_{n-j}] \\ = & \left(-n g_1 g_{n-1} - n \sum_{j=2}^{n-2} g_j g_{n-j} - n g_{n-1} g_1 \right) + \left(n g_1 g_{n-2} + n \sum_{j=2}^{n-2} g_{j-1} g_{n-j} \right) \\ = & - n \sum_{j=1}^{n-1} g_j g_{n-j} + n \sum_{j=2}^{n-1} g_{j-1} g_{n-j}. \end{aligned}$$

Combining this with Equation 2.20, we are led to the following result.

Proposition 2.2. *Let $\{g_n : n \ge 1\}$ be defined as in Equation 2.1. Then g_n is continuous on $[0,\infty)$ and differentiable on $(0,\infty)$ for each n. Furthermore, the functions g_n satisfy the following infinite system of ordinary differential equations.*

$$g_1' = -g_1 + \alpha\beta \tag{2.21}$$

$$g_2' = -2g_2 + 2(\alpha+\beta) g_1 - 2g_1^2 \tag{2.22}$$

$$g_n' = -n g_n + n(\alpha+\beta) g_{n-1} - n \sum_{j=1}^{n-1} g_j g_{n-j} + n \sum_{j=2}^{n-1} g_{j-1} g_{n-j}, \quad n \ge 3. \tag{2.23}$$

Note that similar equations appear in [9] (see also [16, 31]). It will be convenient to define $g_0 \equiv \alpha + \beta$. With this convention, Equations 2.22 and 2.23 can be written in the more compact form

$$g_n' = -n \left[g_n - \sum_{j=1}^{n-1} (g_j - g_{j-1}) g_{n-j} \right], \quad n \ge 2. \tag{2.24}$$

2.2 The Proof of Proposition 1.21

Our goal here is to calculate the right-derivative of $F_t(s) = \tau[(r_t - r_t p_s r_t)^2]$ at $s = t$. The process $s \mapsto (r_t - r_t p_s r_t)^2$ is adapted for $s \in [t, \infty)$, and so we may use the tools of Itô calculus to compute this derivative. To begin, we expand

$$(r_t - r_t p_s r_t)^2 = r_t^2 - r_t^2 p_s r_t - r_t p_s r_t^2 + (r_t p_s r_t)^2 = r_t - 2r_t p_s r_t + r_t p_s r_t p_s r_t.$$

Recalling that t is constant, the stochastic differential is

$$-2r_t dp_s r_t + d(r_t p_s r_t p_s r_t).$$

Using the Itô product rule 1.21, the second term expands to

$$\begin{aligned} d[(r_t p_s r_t)(p_s r_t)] &= (r_t dp_s r_t) p_s r_t + r_t p_s r_t (dp_s r_t) + d(r_t p_s r_t) \cdot d(p_s r_t) \\ &= r_t dp_s r_t p_s r_t + r_t p_s r_t dp_s r_t + r_t dp_s r_t dp_s r_t. \end{aligned}$$

Combining and taking the trace, this yields

$$dF_t(s) = \tau[d(r_t - r_t p_s r_t)^2] = \tau[-2r_t dp_s r_t + r_t dp_s r_t p_s r_t + r_t p_s r_t dp_s r_t + r_t dp_s r_t dp_s r_t].$$

Using the trace property (and the fact that $r_t = r_t^2$), this simplifies to three terms:

$$dF_t(s) = -2\tau(r_t dp_s) + 2\tau(r_t p_s r_t dp_s) + \tau(r_t dp_s r_t dp_s). \tag{2.25}$$

From Equation 2.10 above, we have $dp_s = (\tau(p) - p_s)\, dt + dy_s\, p_s - p_s\, dy_s$ where the process y_s is determined by fSDE 2.7, and obeys the Itô calculus of Equations 2.8 and 2.9.

We consider now the three terms in Equation 2.25 separately. First:

$$\tau(r_t dp_s) = [-\tau(r_t p_s) + \tau(p) r_t]\, ds + i\tau(r_t\, dy_s\, p_s) - i\tau(r_t\, p_s\, dy_s).$$

The last two terms are 0: each can be expressed in the form $\tau(\theta_s\, dx_s)$ where θ_s is adapted (since $s \geq t$), and Itô integrals are centered. Thus

$$\tau(r_t dp_s) = [-\tau(r_t p_s) + \tau(p)\tau(r_t)]\, ds. \tag{2.26}$$

Now for the second term in Equation 2.25. Combining with Equation 2.10, we have

$$\tau(r_t p_s r_t dp_s) = [-\tau(r_t p_s r_t p_s) + \tau(r_t p_s r_t)\tau(p)]\, ds + i\tau(r_t p_s r_t\, dy_s\, p_s) - i\tau(r_t p_s r_t p_s\, dy_s).$$

Again, the last two terms are of the form $\tau(\theta_s\, dx_s)$ for adapted θ_s, and so we have

$$\tau(r_t p_s r_t dp_s) = [-\tau((r_t p_s)^2) + \tau(r_t p_s)]\, ds. \tag{2.27}$$

Finally, we come to the third term in Equation 2.25. We must calculate the trace of

$$(r_t dp_s)^2 = ([-r_t p_s + \tau(p_s) r_t]\, ds + r_t\, dy_s\, p_s - r_t p_s\, dy_s)^2.$$

All products involving the ds term vanish, since $ds^2 = ds dy_s = 0$ (cf. Equation 2.9). Thus, we simply have

$$\begin{aligned} (r_t dp_s)^2 &= (r_t\, dy_s\, p_s - r_t p_s\, dy_s)^2 \\ &= (r_t\, dy_s\, p_s - r_t p_s\, dy_s)^2 = r_t dy_s p_s r_t dy_s p_s - r_t dy_s p_s r_t p_s dy_s - r_t p_s dy_s r_t dy_s p_s + r_t p_s dy_s r_t p_s dy_s. \end{aligned}$$

Now using Equation 2.8, these terms simplify as

- $r_t dy_s p_s r_t dy_s p_s = -\tau(p_s r_t) r_t p_s \, ds$
- $r_t dy_s p_s r_t p_s dy_s = -\tau(p_s r_t p_s) r_t \, ds = -\tau(p_s r_t) r_t \, ds$
- $r_t p_s dy_s r_t dy_s p_s = -r_t p_s \tau(r_t) p_s \, ds = -\tau(r_t) r_t p_s \, ds$
- $r_t p_s dy_s r_t p_s y_s = -\tau(r_t p_s) r_t p_s \, ds$

Summing and taking the trace we have

$$\tau[(r_t dp_s)^2] = [-2(\tau(p_s r_t))^2 + 2\tau(r_t)\tau(p_s r_t)] \, ds. \tag{2.28}$$

Combining Equations 2.26, 2.27, and 2.28 with Equation 2.25, we have

$$\begin{aligned} \frac{dF_t}{ds}(s) &= -2[-\tau(r_t p_s) + \tau(p)\tau(r_t)] + 2[-\tau((r_t p_s)^2) + \tau(r_t p_s)] + [-2(\tau(r_t p_s))^2 + 2\tau(r_t)\tau(r_t p_s)] \\ &= 2\left(2\tau(r_t p_s) + \tau(r_t)[\tau(r_t p_s) - \tau(p)] - [\tau((rp_s)^2) + (\tau(rp_s))^2]\right). \end{aligned}$$

Evaluating at $s = t$, we get (using $r_t = r_t^2 = r_t p_t$)

$$\begin{aligned} \frac{1}{2}\frac{dF_t}{ds}(t) &= 2\tau(r_t p_t) + \tau(r_t)[\tau(r_t p_t) - \tau(p)] - [\tau((rp_t)^2) + (\tau(rp_t))^2] \\ &= 2\tau(r_t) + \tau(r_t)[\tau(r_t) - \tau(p)] - [\tau(r_t) + \tau(r_t)^2] \\ &= \tau(r_t) - \tau(r_t)\tau(p) = \tau(r_t)(1 - \tau(p)) \end{aligned}$$

as claimed in the corollary. □

2.3 Smoothness of the Moment Generating Function in t

Our next goal is to prove that the (centered) moment-generating function of μ_t

$$\psi(t, w) = \sum_{n \geq 1} g_n(t) w^n \tag{2.29}$$

is C^∞ jointly in (t, w) for $t > 0$ and $|w| < 1$. The moments $g_n(t)$ are solutions to Equations 2.21–2.23, which can, in principle, be solved explicitly. Presently, we only use the fact that the solution has a simple analytic form.

Lemma 2.3. *For $n \geq 1$, the function $g_n(t)$ of Proposition 2.2 is a polynomial in t and e^{-t}.*

Proof. To begin, we may solve Equation 2.21 explicitly: $g_1(t) = g_1(0)e^{-t} + \alpha\beta(1 - e^{-t})$, having the desired form. We proceed by induction on n. Equations 2.22 and 2.23 give, for $n \geq 2$, $g_n' + ng_n = h_n$ where h_n is a polynomial in $g_1, \ldots, g_{n-1}$, and is therefore a polynomial in t and e^{-t} by the induction hypothesis. The ODE can then be written in the form $\frac{d}{dt}[e^{nt} g_n(t)] = e^{nt} h_n(t)$, with solution

$$g_n(t) = e^{-nt} \left[\int_0^t e^{ns} h_n(s) \, ds + g_n(0) \right].$$

Since $h_n(s)$ is a polynomial in s and e^{-s}, $e^{ns} h_n(s)$ is a polynomial in s and $e^{\pm s}$ whose positive degree in e^s is $\leq n$. A separate induction argument and elementary calculus show that the antiderivative of $e^{ns} h_n(s)$ is therefore also a polynomial in s and $e^{\pm s}$ whose positive degree in e^s is $\leq n$. Thus $g_n(t)$ has the desired form, proving the corollary. □

We can also iterate the ODEs 2.21–2.23 to find a general recurrence form for the kth derivatives.

Lemma 2.4. *Let $\{g_n\}_{n\geq 1}$ be defined as in Equation 2.1, with $g_0 \equiv \alpha+\beta$ as in Equation 2.24. Then for all $t \geq 0$ and $n, k \geq 1$, the kth derivative $g_n^{(k)}$ satisfies*

$$g_n^{(k)} = \sum_{s=0}^{k+1} \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^{n,k}(j_1,\ldots,j_s) g_{j_1}\cdots g_{j_s} \tag{2.30}$$

for some constants $c_s^{n,k}(j_1,\ldots,j_s)$ satisfying $|c_s^{n,k}(j_1,\ldots,j_s)| \leq (4sn^2)^k$.

Proof. When $n=0$, $g_0^{(k)}=0$ for $k\geq 1$. When $n=1$, iterating Equation 2.21 shows that $g_1^{(k)} = (-1)^k[g_1-\alpha\beta]$, which has the desired form of Equation 2.30 with $c_0^{1,k} = (-1)^{k+1}\alpha\beta$, $c_1^{1,k}(1) = (-1)^k$, and $c_s^{1,k}(j_1,\ldots,j_s)=0$ for $s\geq 2$. For $n\geq 2$, we proceed by induction on k. For the base case $k=1$, we note that, by Equation 2.24,

$$g_n' = -n\left[g_n - \sum_{j=1}^{n-1}(g_j - g_{j-1})g_{n-j}\right] = \sum_{s=0}^{2} \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^1(j_1,\ldots,j_s) g_{j_1}\cdots g_{j_s} \tag{2.31}$$

with $c_0^{n,1}=0$, $c_1^{n,1}(j) = -n\delta_{j,n}$, and

$$c_2^{n,1}(j_1,j_2) = \begin{cases} -n, & j_1+j_2=n \\ n & j_1+j_2 = n-1 \\ 0 & \text{otherwise.}\end{cases}$$

Now for the inductive step. Assume that Equation 2.30 holds up to level $k-1$. Then

$$\begin{aligned} g_n^{(k)} = \frac{d}{dt} g_n^{(k-1)} &= \sum_{s=1}^{k} \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^{n,k-1}(j_1,\ldots,j_s)\frac{d}{dt}\left(g_{j_1}\cdots g_{j_s}\right) \\ &= \sum_{s=1}^{k}\sum_{\ell=1}^{s} \sum_{1\leq j_1',\ldots,j_s'\leq n} c_s^{n,k-1}(\ell;j_1',\ldots,j_s') g_{j_1'}'\cdot g_{j_2'}\cdots g_{j_s'}, \end{aligned} \tag{2.32}$$

where we have reindexed $(j_1',\ldots,j_s') = (j_\ell,\ldots,j_s,j_1,\ldots,j_{\ell-1})$, and the new constants $c_s^{n,k-1}(\ell;j_1',\ldots,j_s')$ are reordered accordingly: $c_s^{n,k-1}(\ell;j_1',\ldots,j_s') = c_s^{n,k-1}(j_{s-\ell+2}',\ldots,j_s',j_1',\ldots,j_{s-\ell+1}')$ (with j_1' in the ℓth slot). We now relabel $j_r' \mapsto j_r$, and do the internal sum over j_1 first:

$$g_{j_1}' = c_0^{j_1,1} + \sum_{i=1}^{j_1} c_1^{j_1,1}(i) g_i + \sum_{1\leq i_1,i_2\leq j_1} c_2^{j_1,1}(i_1,i_2) g_{i_1}g_{i_2}$$

which yields terms of orders $s-1$, s, and $s+1$ in the internal sum in Equation 2.32.

$$\begin{aligned} \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^{n,k-1}(\ell;j_1',\ldots,j_s') = & \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^{n,k-1}(\ell;j_1,\ldots,j_s) c_0^{j_1,1}\cdot g_{j_2}\cdots g_{j_s} && (2.33) \\ & + \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^{k-1}(\ell;j_1,\ldots,j_s)\sum_{i=1}^{j_1} c_1^{j_1,1}(i) g_i g_{j_2}\cdots g_{j_s} && (2.34) \\ & + \sum_{1\leq j_1,\ldots,j_s\leq n} c_s^{k-1}(\ell;j_1,\ldots,j_s) \sum_{1\leq i_1,i_2\leq j_1} c_2^{j_1,1}(i_1,i_2) g_{i_1}g_{i_2}g_{j_2}\cdots g_{j_s}. && (2.35) \end{aligned}$$

Reindexing (2.33) and summing over ℓ gives

$$\sum_{1\le j_1,\dots,j_{s-1}\le n}\left(\sum_{\ell=1}^{s}\sum_{i=1}^{n} c_s^{n,k-1}(\ell;i,j_1,\dots,j_{s-1})c_0^{i,1}\right) g_{j_1}\cdots g_{j_{s-1}} \equiv \sum_{1\le j_1,\dots,j_{s-1}\le n} d_{s,-}^{n,k}(j_1,\dots,j_{s-1})g_{j_1}\cdots g_{j_{s-1}}, \tag{2.36}$$

where in the case $s=1$ this is just a constant. Changing the order of summation in (2.34) and summing over ℓ yields

$$\sum_{1\le j_2,\dots,j_s,i\le n}\ \sum_{\ell=1}^{s}\sum_{j_1=i}^{n} c_s^{k-1}(\ell;j_1,\dots,j_s)c_1^{j_1,1}(i)g_i g_{j_2}\cdots g_{j_s},$$

or, for convenience exchanging $i\leftrightarrow j_1$,

$$\sum_{1\le j_1,\dots,j_s\le n}\left(\sum_{\ell=1}^{s}\sum_{i=j_1}^{n} c_s^{k-1}(\ell;i,j_2,\dots,j_s)c_1^{i,1}(j_1)\right) g_{j_1}\cdots g_{j_s} \equiv \sum_{1\le j_1,\dots,j_s\le n} d_{s,0}^{n,k}(j_1,\dots,j_s)g_{j_1}\cdots g_{j_s}. \tag{2.37}$$

Finally, in (2.35), we change the order of summation between j_1 and $\{i_1,i_2\}$,

$$\sum_{j_1=1}^{n}\sum_{1\le i_1,i_2\le n} = \sum_{1\le i_1,i_2\le n}\left(\mathbb{1}_{i_2\le i_1}\sum_{j_1=i_1}^{n} + \mathbb{1}_{i_2>i_1}\sum_{j_1=i_2}^{n}\right)$$

which yields (2.35) in the form

$$\sum_{1\le j_2,\dots,j_s,i_1,i_2\le n}\ \sum_{\ell=1}^{s}\left(\mathbb{1}_{i_2\le i_1}\sum_{j_1=i_1}^{n} + \mathbb{1}_{i_2>i_1}\sum_{j_1=i_2}^{n}\right) c_s^{k-1}(\ell;j_1,\dots,j_s)c_2^{j_1,1}(i_1,i_2)g_{i_1}g_{i_2}g_{j_2}\cdots g_{j_s},$$

which, after reindexing, gives

$$\sum_{1\le j_1,\dots,j_{s+1}\le n} d_{s,+}^{n,k}(j_1,\dots,j_{s+1})g_{j_1}\cdots g_{j_{s+1}} \tag{2.38}$$

where

$$d_{s,+}^{n,k}(j_1,\dots,j_{s+1}) = \sum_{\ell=1}^{s}\left(\mathbb{1}_{j_2\le j_1}\sum_{i=j_1}^{n} + \mathbb{1}_{j_2>j_1}\sum_{i=j_2}^{n}\right) c_s^{k-1}(\ell;i,j_3,\dots,j_s)c_2^{i,1}(j_1,j_2). \tag{2.39}$$

Now, for any s-tuple $\mathbf{j}_s=(j_1,\dots,j_s)$ in $[n]^s$ where $[n]=\{1,\dots,n\}$, let $g_{\mathbf{j}_s}=g_{j_1}\cdots g_{j_s}$. With this notation, combining (2.32) with (2.33–2.39), we have

$$\begin{aligned} g_n^{(k)} &= \sum_{s=1}^{k}\left(\sum_{\mathbf{j}_{s-1}\in[n]^{s-1}} d_{s,-}^{n,k}(\mathbf{j}_{s-1})g_{\mathbf{j}_{s-1}} + \sum_{\mathbf{j}_s\in[n]^s} d_{s,0}^{n,k}(\mathbf{j}_s)g_{\mathbf{j}_s} + \sum_{\mathbf{j}_{s+1}\in[n]^{s+1}} d_{s,+}^{n,k}(\mathbf{j}_{s+1})g_{\mathbf{j}_{s+1}}\right)\\ &= \sum_{s=1}^{k+1}\sum_{\mathbf{j}_s\in[n]^s}\left(d_{s-1,+}^{n,k}(\mathbf{j}_s) + d_{s,0}^{n,k}(\mathbf{j}_s) + d_{s+1,-}^{n,k}(\mathbf{j}_s)\right)g_{\mathbf{j}_s}, \end{aligned}$$

with the convention that $d_{s,\varepsilon}^{n,k}=0$ for $s\notin[k]$ and $\varepsilon\in\{-,0,+\}$. Hence, if we define

$$c_s^{n,k}(\mathbf{j}_s) \equiv d_{s-1,+}^{n,k}(\mathbf{j}_s) + d_{s,0}^{n,k}(\mathbf{j}_s) + d_{s+1,-}^{n,k}(\mathbf{j}_s), \qquad 1\le s\le k+1,$$

we see that $g_n^{(k)}$ has the desired form (2.30), and it remains only to show that the bound $|c_s^{n,k}(\mathbf{j}_s)| \le (4sn^2)^k$ is satisfied.

Let $C_s^{n,k-1} = \sup_{\mathbf{j}_s \in [n]^s} |c_s^{n,k-1}(\mathbf{j}_s)|$. From (2.33) we have

$$\left| d_{s,-}^{n,k}(\mathbf{j}_{s-1}) \right| = \left| \sum_{\ell=1}^{s} \sum_{i=1}^{n} c_s^{n,k-1}(\ell; i, \mathbf{j}_{s-1}) c_0^{i,1} \right| \le s C_s^{n,k-1} \sum_{i=1}^{n} C_0^{i,1} \le s C_s^{n,k-1},$$

since $C_0^{i,1} = \alpha\beta \le 1$ if $i = 1$ and is 0 if $i > 1$. From (2.34) we have

$$|d_{s,0}^{n,k}(\mathbf{j}_s)| = \left| \sum_{\ell=1}^{s} \sum_{i=j_1}^{n} c_s^{k-1}(\ell; i, j_2, \ldots, j_s) c_1^{i,1}(j_1) \right| \le s C_s^{n,k-1} \sum_{i=1}^{n} C_1^{i,1} \le sn^2 C_s^{n,k-1},$$

since $C_1^{i,1} \le i \le n$ for all i. Finally, from (2.39), we have

$$|d_{s,+}^{n,k}(\mathbf{j}_{s+1})| = \left| \sum_{\ell=1}^{s} \left(\mathbb{1}_{j_2 \le j_1} \sum_{i=j_1}^{n} + \mathbb{1}_{j_2 > j_1} \sum_{i=j_2}^{n} \right) c_s^{k-1}(\ell; i, j_3, \ldots, j_s) c_2^{i,1}(j_1, j_2) \right| \le 2s C_s^{n,k-1} \sum_{i=1}^{n} C_2^{i,1} \le 2sn^2 C_s^{n,k-1}.$$

Thus,

$$C_s^{n,k} = \sup_{\mathbf{j}_s \in [n]^s} |c_s^{n,k}(\mathbf{j}_s)| \le s(3n^2+1) C_s^{n,k-1} \le 4sn^2 C_s^{n,k-1}.$$

The result now follows from the inductive hypothesis that $C_s^{n,k-1} \le (4sn^2)^{k-1}$. $\square$

Next, we use the equations to prove the following family of (blunt) growth estimates for the derivatives of the moments $g_n(t)$.

Lemma 2.5. *Let $\{g_n\}_{n\ge 1}$ be defined as in Equation 2.1. Then for each $k \in \mathbb{N}$, the kth derivative $g_n^{(k)}$ is uniformly bounded by*

$$|g_n^{(k)}| \le (k+1)^k (2n)^{3k+1}.$$

Proof. For all $n \ge 1$, $g_n(t)$ is the nth moment of a probability measure μ_t supported in $[0,1]$; hence $|g_n(t)| \le 1$. By the convention used in Lemma, 2.4 $g_0 \equiv \alpha + \beta$, which is in $[0,2]$. Hence, in general $|g_n(t)| \le 2$ for all $n \ge 0$. Thus, referring to the recursive form of the derivative given in Equation 2.30 in the previous lemma, we have

$$|g_n^{(k)}| \le \sum_{s=0}^{k+1} \sum_{1 \le j_1, \ldots, j_s \le n} |c_s^{n,k}(j_1, \ldots, j_s)| |g_{j_1}| \cdots |g_{j_s}| \le \sum_{s=0}^{k+1} \sum_{1 \le j_1, \ldots, j_s \le n} (4sn^2)^k \cdot 2^s = \sum_{s=0}^{k+1} (2n)^s (4sn^2)^k,$$

and therefore

$$|g_n^{(k)}| \le \sum_{s=0}^{k+1} (2n)^{k+1} (4(k+1)n^2)^k = (k+1)^k (2n)^{3k+1},$$

concluding the proof. $\square$

Remark 2.6. A more careful estimate is possible, showing that $|g_n^{(k)}| = O(n^{2k})$ for each k; for example, direct estimation of Equation 2.23 gives

$$\begin{aligned} |g_n'| &\le n \left(|g_n| + (\alpha+\beta)|g_{n-1}| + \sum_{j=1}^{n-1} |g_j||g_{n-j}| + \sum_{j=2}^{n-1} |g_{j-1}||g_{n-j}| \right) \\ &= n\left(1 + \alpha + \beta + (n-1) + (n-2)\right) \le 2n^2, \end{aligned}$$

much smaller than the $64n^4$ bound proved in above. The improvement this bound represents over the bound proven in Lemma 2.5 is of no consequence to our application, however.

We now prove the main theorem of this section: that the (centered) moment-generating function $\psi(t,w) = \sum_{n\geq 1} g_n(t)w^n$ is C^∞ in both variables.

Proposition 2.7. *The function $\psi(t,w)$ of Equation 2.29 is C^∞ jointly in both variables, for $t > 0$ and $|w| < 1$.*

Proof. Since the coefficients $g_n(t)$ are uniformly bounded in modulus by 1, the power series $\psi(t,\cdot)$ converges uniformly on compact subsets of $\mathbb{D}$, and analyticity is an elementary result from complex variables. Now, for $k \geq 1$ and $m \geq 0$ define

$$\varphi_{k,m}(t,w) = \sum_{n=m}^{\infty} g_n^{(k)}(t) \cdot n(n-1)\cdots(n-m+1)w^{n-m}.$$

Throughout this proof (only), we use the non-standard convention that $0^0 = 0$; thus $\varphi_{0,0} = \psi$. For fixed $t \geq 0$, by Lemma 2.5, the coefficients of this power series in w are bounded by

$$|g_n^{(k)}(t) \cdot n(n-1)\cdots(n-m+1)| \leq (k+1)^k(2n)^{3k+1} \cdot n^m$$

and so by the root test, the power-series converges uniformly on compact subset of the unit disk $\mathbb{D}$, defining a function analytic in w. Similarly, for fixed $|w| < 1$, $\varphi_{k,m}(t,w)$ is absolutely summable; it is a series of functions that are continuous (by Lemma 2.3) and hence, by the Weierstraß M-test, $\varphi_{k,m}(t,w)$ is continuous in t. This shows that $\varphi_{k,m} \in C^0(\mathbb{R}_+ \times \mathbb{D})$.

The uniform convergence of $\varphi_{k,m}$ and pointwise convergence of $\varphi_{k-1,m}$ (at any single point in $\mathbb{D}$) imply (for example by [26, Thm. 7.17]) that $\varphi_{k,m}(\cdot,w)$ is differentiable for each w and that

$$\frac{\partial}{\partial t}\varphi_{k-1,m}(t,w) = \varphi_{k,m}(t,w).$$

By induction on k, this shows that $\varphi_{0,m}(t,w)$ is at least C^k in t for each k, and has time derivatives $\frac{\partial^k}{\partial t^k}\varphi_{0,m}(t,w) = \varphi_{k,m}(t,w)$. Of course, by elementary complex variables,

$$\varphi_{k,m}(t,w) = \frac{\partial^m}{\partial w^m}\sum_{n=1}^{\infty} g_n^{(k)}(t)w^n = \frac{\partial^m}{\partial w^m}\varphi_{k,0}(t,w);$$

thus, we have shown that

$$\varphi_{k,m}(t,w) = \frac{\partial^m}{\partial w^m}\frac{\partial^k}{\partial t^k}\varphi_{0,0}(t,w) = \frac{\partial^m}{\partial w^m}\frac{\partial^k}{\partial t^k}\psi(t,w)$$

for $m,k \geq 0$. As we have shown that $\varphi_{k,m} \in C^0(\mathbb{R}_+ \times \mathbb{D})$, this proves that ψ is $C^\infty(\mathbb{R}_+ \times \mathbb{D})$ as required. □

Our next goal is to extend this result to analyticity of ψ. Analyticity in the spacial variable $w \in \mathbb{D}$ follows immediately from the proof of Proposition 2.7. Analyticity in t is much more involved. It will pay to first translate the ODEs of Equations 2.21–2.23 into a PDE for the Cauchy transform of μ_t (which is simply related to $\psi(t,\cdot)$), and then use analytic function techniques in the spacial variable; this is the purpose of the next two sections.

Remark 2.8. The bound of Lemma 2.5 are woefully inadequate to prove analyticity in t using Taylor's theorem, since the coefficients grow superfactorially in k. In Section 2.5, we will use more sophisticated techniques to prove that these Taylor series coefficients are, in fact, exponentially bounded.

2.4 The Flow of the Cauchy Transform

The moment function $\psi(t,w)$ of Equation 2.29 is closely related to the Cauchy transform of the measure μ_t, which concerns us in this paper. Indeed, the coefficients $g_n(t)$ are the moments of μ_t. Since $\psi(t,w)$ converges uniformly for $w\in\mathbb{D}$, taking $z=1/w$ for $w\neq 0$, we see that $\psi(t,1/z)$ converges uniformly for $|z|>1$, and hence by the Fubini-Tonelli theorem,

$$\int_0^1 \sum_{n\geq 1}\left(\frac{x}{z}\right)^n \mu_t(dx) = \sum_{n\geq 1}\frac{1}{z^n}\int_0^1 x^n\,\mu_t(dx) = \sum_{n\geq 1} g_n(t)\frac{1}{z^n} = \psi(t,1/z), \quad |z|>1.$$

On the other hand, since the support of μ_t is contained in $[0,1]$, for $|z|>1$ the series $\sum_{n\geq 1}(x/z)^n$ converges, and we have

$$\frac{1}{z}\psi(t,1/z) = \frac{1}{z}\int_0^1\left(\frac{1}{1-x/z}-1\right)\mu_t(dx) = \int_0^1 \frac{\mu_t(dx)}{z-x} - \frac{1}{z} = G_{\mu_t}(z) - \frac{1}{z}, \quad |z|>1.$$

Finally, this means that the function G of Equation 1.8 is related to ψ by

$$G(t,z) = G_{\mu_t}(z) - \frac{1}{z} + \frac{\min\{\alpha,\beta\}}{z} = \frac{1}{z}\left(\psi(t,1/z)+\min\{\alpha,\beta\}\right). \tag{2.40}$$

From Equation 2.40 and Proposition 2.7 it follows immediately that

Corollary 2.9. *The function $G(t,z)$ of Equation 1.8 is $C^\infty(\mathbb{R}_+\times(\mathbb{C}-\overline{\mathbb{D}}))$, and for each $t\geq 0$ $G(t,z)$ is analytic in $|z|>1$. Moreover, on this domain,*

$$\frac{\partial}{\partial t}G(t,z) = \frac{\partial}{\partial t}\frac{1}{z}\psi(t,1/z) = \sum_{n\geq 1}\frac{g_n'(t)}{z^{n+1}}. \tag{2.41}$$

We now combine Equation 2.41 with Equations 2.21–2.23 to deduce a partial differential equation satisfied by G. To begin,

$$\begin{aligned}\frac{\partial}{\partial t}G = \sum_{n\geq 1}\frac{1}{z^{n+1}}g_n' &= \frac{1}{z^2}(-g_1+\alpha\beta) + \frac{1}{z^3}(-2g_2+2(\alpha+\beta)g_1-2g_1^2)\\ &\quad + \sum_{n\geq 3}\frac{1}{z^{n+1}}\left(-ng_n + n(\alpha+\beta)g_{n-1} - n\sum_{j=1}^{n-1}g_jg_{n-j} + n\sum_{j=2}^{n-1}g_{j-1}g_{n-j}\right).\end{aligned}$$

By the uniform convergence on the domain $|z|>1$, the order of all summations may be interchanged. It is convenient to recombine the expression into two parts: $\frac{\partial}{\partial t}G = S_1+S_2$ where

$$S_1 = \frac{\alpha\beta}{z^2} - \sum_{n\geq 1}\frac{ng_n}{z^{n+1}} + (\alpha+\beta)\sum_{n\geq 2}\frac{ng_{n-1}}{z^{n+1}}$$

and

$$S_2 = -\frac{2g_1^2}{z^3} - \sum_{n\geq 3}\frac{n}{z^{n+1}}\sum_{j=1}^{n-1}g_jg_{n-j} + \sum_{n\geq 3}\frac{n}{z^{n+1}}\sum_{j=2}^{n-1}g_{j-1}g_{n-j}.$$

To deal with the first sum S_1, note (from Equation 2.40) that

$$G(t,z) = \frac{1}{z}(\varphi(t,1/z)+\min\{\alpha,\beta\}) = \frac{\min\{\alpha,\beta\}}{z} + \sum_{n\geq 1}\frac{g_n(t)}{z^{n+1}}.$$

For convenience, let us define

$$G_1(t,z) = \sum_{n\geq 1} \frac{g_n(t)}{z^{n+1}} = G(t,z) - \frac{\min\{\alpha,\beta\}}{z} = G_{\mu_t}(z) - \frac{1}{z}. \tag{2.42}$$

Then $\frac{\partial}{\partial t}G_1 = \frac{\partial}{\partial t}G = S_1 + S_2$. Now,

$$-\sum_{n\geq 1} \frac{n}{z^{n+1}} g_n = \frac{\partial}{\partial z}\sum_{n\geq 1}\frac{g_n}{z^n} = \frac{\partial}{\partial z}(zG_1) \tag{2.43}$$

which holds true for $|z| > 1$. Similarly,

$$\sum_{n\geq 2}\frac{ng_{n-1}}{z^{n+1}} = \sum_{n\geq 1}\frac{(n+1)g_n}{z^{n+2}} = \frac{1}{z}\left(\sum_{n\geq 1}\frac{g_n}{z^{n+1}} + \sum_{n\geq 1}\frac{n}{z^{n+1}}g_n\right),$$

and employing Equations 2.42 and 2.43 this becomes

$$\sum_{n\geq 2}\frac{ng_{n-1}}{z^{n+1}} = \frac{1}{z}\left(G_1 - \frac{\partial}{\partial z}(zG_1)\right) = -\frac{\partial}{\partial z}G_1. \tag{2.44}$$

Combining Equations 2.43 and 2.44 and simplifying, this shows that

$$\begin{aligned} S_1 &= \frac{\alpha\beta}{z^2} + \frac{\partial}{\partial z}(zG_1) - (\alpha+\beta)\frac{\partial}{\partial z}G_1 \\ &= \frac{\alpha\beta}{z^2} + G_1 + (z-\alpha-\beta)\frac{\partial}{\partial z}G_1. \end{aligned} \tag{2.45}$$

The terms in S_2 can similarly be expressed in terms of G_1: in fact, in terms of G_1^2. For $|z| > 1$,

$$(G_1)^2 = \left(\sum_{n\geq 1}\frac{g_n}{z^{n+1}}\right)^2 = \frac{1}{z^2}\sum_{n\geq 2}\frac{1}{z^n}\sum_{j=1}^{n-1} g_j g_{n-j}, \quad |z| > 1.$$

Whence, for $|z| > 1$,

$$\begin{aligned} \frac{\partial}{\partial z}(zG_1)^2 = \frac{\partial}{\partial z}\sum_{n\geq 2}\frac{1}{z^n}\sum_{j=1}^{n-1} g_j g_{n-j} &= -\sum_{n\geq 2}\frac{n}{z^{n+1}}\sum_{j=1}^{n-1} g_j g_{n-j} \\ &= -\frac{2}{z^3}g_1^2 - \sum_{n\geq 3}\frac{n}{z^{n+1}}\sum_{j=1}^{n-1} g_j g_{n-j}. \end{aligned} \tag{2.46}$$

Regarding the second term in S_2, note that for each $n \geq 3$

$$\sum_{j=2}^{n-1} g_{j-1}g_{n-j} = \sum_{\substack{j+k=n-1\\ j,k\geq 1}} g_j g_k.$$

Whence, we can manipulate the sum as

$$(zG_1)^2 = \sum_{n\geq 2}\frac{1}{z^n}\sum_{\substack{j,k\geq 1\\ j+k=n}} g_j g_k = \sum_{n\geq 3}\frac{1}{z^{n-1}}\sum_{\substack{j+k=n-1\\ j,k\geq 1}} g_j g_k,$$

and so

$$\frac{\partial}{\partial z}\left(z(G_1)^2\right) = \frac{\partial}{\partial z}\sum_{n\geq 3}\frac{1}{z^n}\sum_{\substack{j+k=n-2\\ j,k\geq 1}} g_j g_k = -\sum_{n\geq 3}\frac{1}{z^{n+1}}\sum_{j=1}^{n-2} g_{j-1}g_{n-j} \tag{2.47}$$

holds true for $|z| > 1$. Combining Equations 2.46 and 2.47, we have

$$S_2 = \frac{\partial}{\partial z}(zG_1)^2 - \frac{\partial}{\partial z}\left(z(G_1)^2\right). \tag{2.48}$$

We have thus proved the following result.

Proposition 2.10. *Let $G_1(t,z)$ be defined (for $t > 0$ and $|z| > 1$) as in Equation 2.42. Then G_1 satisfies the partial differential equation*

$$\frac{\partial}{\partial t}G_1 = \frac{\partial}{\partial z}\left(z(z-1)(G_1)^2 + (z-\alpha-\beta)G_1 - \frac{\alpha\beta}{z}\right). \tag{2.49}$$

Proof. As noted following Equation 2.42, $\frac{\partial}{\partial t}G_1 = S_1 + S_2$. Simplifying Equation 2.45 reversing the product rule,

$$S_1 = \frac{\alpha\beta}{z^2} + G_1 + (z-\alpha-\beta)\frac{\partial}{\partial z}G_1 = \frac{\partial}{\partial z}\left(-\frac{\alpha\beta}{z} + (z-\alpha-\beta)G_1\right).$$

On the other hand, Equation 2.48 immediately yields

$$S_2 = \frac{\partial}{\partial z}(zG_1)^2 - \frac{\partial}{\partial z}\left(z(G_1)^2\right) = \frac{\partial}{\partial z}\left((z^2-z)G_1^2\right).$$

Combining these equations, which are valid for $|z| > 1$, yields Equation 2.49. □

Remark 2.11. Since $\frac{\partial}{\partial t}G_1 = \frac{\partial}{\partial t}G_{\mu_t}$, we can rewrite Equation 2.49 in terms of the Cauchy transform directly; it is easy to check that the result is

$$\frac{\partial}{\partial t}G_{\mu_t} = \frac{\partial}{\partial z}\left(z(z-1)(G_{\mu_t})^2 + \left((1-\alpha)+(1-\beta)-z\right)G_{\mu_t} + \frac{(1-\alpha)(1-\beta)}{z}\right). \tag{2.50}$$

While we have shown the equation holds only in the regime $|z| > 1$, we will show below that it actually holds true on all of $\mathbb{C}_+$. Hence, the poor behaviour near $z = 0$ becomes a technical issue. It is partly for this reason that the shifted transform $G(t,z) = G_{\mu_t}(z) - \frac{1-\min\{\alpha,\beta\}}{z}$ is useful: it encapsulates the singularity in a static form, leaving a smooth flow in the vicinity of 0, as demonstrated by the form of Equation 2.51 below.

Corollary 2.12. *Let G be the shifted Cauchy transform of Equation 1.8. Then for $t > 0$ and $|z| > 1$,*

$$\frac{\partial}{\partial t}G = \frac{\partial}{\partial z}\left[z(z-1)G^2 - (az+b)G\right] \tag{2.51}$$

where $a = 2\min\{\alpha,\beta\} - 1$ and $b = |\alpha-\beta|$.

Proof. From Equation 2.42, $G_1 = G - \frac{\min\{\alpha,\beta\}}{z}$. Consider the case $\alpha \leq \beta$, so $G_1 = G - \frac{\alpha}{z}$. We simply change variables in Equation 2.49, which holds for $|z| > 1$. As $\frac{\partial}{\partial t}G_1 = \frac{\partial}{\partial t}G$, we need only transform the quantity differentiated with respect to z in Equation 2.49.

$$
\begin{aligned}
& z(z-1)(G_1)^2 + (z-\alpha-\beta)G_1 - \frac{\alpha\beta}{z} \\
=& z(z-1)\left(G - \frac{\alpha}{z}\right)^2 + (z-\alpha-\beta)\left(G - \frac{\alpha}{z}\right) - \frac{\alpha\beta}{z} \\
=& z(z-1)\left(G^2 - \frac{2\alpha}{z}G + \frac{\alpha^2}{z^2}\right) + (z-\alpha-\beta)G - \alpha + \frac{\alpha(\alpha+\beta)}{z} - \frac{\alpha\beta}{z} \\
=& z(z-1)G^2 + (z-\alpha-\beta-2\alpha(z-1))\,G + \frac{\alpha^2}{z}(z-1) + \frac{\alpha(\alpha+\beta)}{z} - \frac{\alpha\beta}{z} - \alpha \\
=& z(z-1)G^2 + ((1-2\alpha)z + \alpha - \beta)\,G + \alpha^2 - \alpha.
\end{aligned}
$$

Differentiating yields Equation 2.51 in the case $\alpha \leq \beta$; the reader may readily verify the formula also holds true in the case $\alpha \geq \beta$. □

Remark 2.13. Without prior knowledge of the structural singularity in the measure μ_t as $t \to \infty$ (cf. Equation 1.6), one might simply try to change variables by removing a pole of unknown mass at 0: $G = G_1 + \frac{m}{z}$. Easy calculations show that the only masses m that transform Equation 2.49 to a form without an explicit singularity at 0 are $m = \alpha$ and $m = \beta$; hence, the choice here is natural.

2.5 Analyticity of the Cauchy Transform in t

Our goal in this section is to extend Corollary 2.9 to show that $G(t, z)$ is not only C^∞ in t but, in fact, analytic in t for $|z| > 1$. To do so, we will actually use the PDE 2.51 proved to hold in Corollary 2.12, together with our a priori knowledge of analyticity in z.

Remark 2.14. PDE 2.51 is a semilinear complex PDE similar in form to the complex inviscid Burger's equation. In general, it exhibits all the hallmark pathological behaviour of nonlinear PDEs: blow-up in finite time, even with uniformly bounded initial data, and shock formation. Since the equation is non-linear, Holmgren's uniqueness theorem does not apply. Thus, although the Cauchy-Kowalewski theorem proves the existence of an analytic solution with given analytic initial condition, it is only unique amongst potential *analytic* solutions, and there may well be non-analytic solutions with the same initial data $G(0, z)$, so the C^∞ result of Corollary 2.9 does not immediately prove analyticity. In this section, we use the form of the PDE, together with the a priori knowledge of analyticity in z, to show that our particular solution $G(t, z)$ is, indeed, analytic in t as well.

We begin with the following recursion for the t-derivatives of G.

Lemma 2.15. *For $t > 0$ and $|z| > 1$, and for $k \geq 0$, define*

$$G_k = \frac{1}{k!}\frac{\partial^k}{\partial t^k}G(t, z).$$

Let $p = p(z) = z(z-1)$ and $q = q(z) = -(az + b)$, cf. Equation 2.51, so that the $C^\infty(\mathbb{R}_+ \times (\mathbb{C} - \overline{\mathbb{D}}))$ function G satisfies the PDE $\partial_t G = \partial_z[pG^2 + qG]$ on its domain. Then

$$(k+1)G_{k+1} = \frac{\partial}{\partial z}\left[p\sum_{j=0}^{k} G_j G_{k-j} + qG_k\right], \quad k \geq 0. \tag{2.52}$$

Proof. The case $k = 0$ is the statement of PDE 2.51. Proceeding by induction, since G is C^∞ we may commute t and z derivatives. We have

$$(k+2)G_{k+2} = \frac{k+2}{(k+2)!}\frac{\partial^{k+2}}{\partial t^{k+2}}G = \frac{\partial}{\partial t}\frac{1}{(k+1)!}\frac{\partial^{k+1}}{\partial t^{k+1}}G = \frac{\partial}{\partial t}G_{k+1}, \tag{2.53}$$

and so by the inductive hypothesis

$$(k+2)G_{k+2} = \frac{1}{k+1}\frac{\partial}{\partial t}\frac{\partial}{\partial z}\left[p\sum_{j=0}^{k} G_j G_{k-j} + qG_k\right] = \frac{\partial}{\partial z}\left[\frac{p}{k+1}\sum_{j=0}^{k}\frac{\partial}{\partial t}(G_j G_{k-j}) + \frac{q}{k+1}\frac{\partial}{\partial t}G_k\right].$$

Shifting the index down one in Equation 2.53 shows that $\frac{1}{k+1}\frac{\partial}{\partial t}G_k = G_{k+1}$, as desired. For the quadratic term, we use the product rule. Again utilizing Equation 2.53,

$$\frac{\partial}{\partial t}(G_j G_{k-j}) = \frac{\partial G_j}{\partial t}G_{k-j} + G_j\frac{\partial G_{k-j}}{\partial t} = (j+1)G_{j+1}G_{k-j} + (k-j+1)G_j G_{k-j+1}.$$

Thus

$$\begin{aligned}\sum_{j=0}^{k}\frac{\partial}{\partial t}(G_j G_{k-j}) &= \sum_{j=0}^{k}(j+1)G_{j+1}G_{k-j} + \sum_{j=0}^{k}(k-j+1)G_j G_{k-j+1} \\ &= \sum_{i=1}^{k+1} iG_i G_{k-i+1} + \sum_{j=0}^{k}(k-j+1)G_j G_{k-j+1}\end{aligned}$$

where we have made the substitution $i = j+1$ in the first sum. Separating off the $i = k+1$ term in the first sum and the $j = 0$ term in the second sum, and relabeling $j \to i$ in the second sum, this yields

$$\sum_{j=0}^{k}\frac{\partial}{\partial t}(G_j G_{k-j}) = (k+1)G_{k+1}G_0 + \sum_{i=1}^{k} iG_i G_{k-i+1} + \sum_{i=1}^{k}(k-i+1)G_i G_{k-i+1} + (k+1)G_0 G_{k+1}$$

which simplifies to

$$\sum_{j=0}^{k}\frac{\partial}{\partial t}(G_j G_{k-j}) = (k+1)\left[G_{k+1}G_0 + \sum_{i=1}^{k} G_i G_{k+1-i} + G_0 G_{k+1}\right] = (k+1)\sum_{i=0}^{k+1} G_i G_{k+1-i}.$$

Combining with the equation following 2.53 and the following comment yields the result. $\square$

We will use the recursion Equation 2.52, in conjunction with analyticity of G, to prove much tighter bounds on the derivatives $G_k(t,z)$ than those discussed in Remark 2.8, allowing us to prove convergence of the Taylor series of $G(t,z)$ centered at any $t > 0$. First we show that the t-derivatives G_k are also analytic.

Lemma 2.16. *For each $k \geq 0$, and each $t \geq 0$, the function $G_k(t,\cdot)$ from Lemma 2.15 is analytic on $\mathbb{C} - \overline{\mathbb{D}}$.*

Proof. First, $G_0(t,\cdot) = G(t,\cdot)$ is analytic on $\mathbb{C} - [0,1]$ (cf. Equation 1.8 and Definition 1.15), verifying the claim in this base case. We proceed by induction. Let K be any compact subset of $\mathbb{C} - \overline{\mathbb{D}}$, let $t_0 > 0$ and let $0 < \epsilon < t_0$. By Corollary 2.9, $G_k(t,z)$ is C^∞ for $t \geq 0$ and $|z| > 1$, and thus $|G_{k+1}(t,z)| = \frac{1}{(k+1)!}|\frac{\partial^{k+1}}{\partial t^{k+1}}G(t,z)|$ is uniformly bounded on $[t_0-\epsilon, t_0+\epsilon] \times K$. Now, for $|h| < \epsilon$, the fundamental theorem of calculus asserts that

$$\frac{G_k(t_0+h,z) - G_k(t_0,z)}{h} = \int_0^1 \frac{\partial}{\partial t}G_k(t_0+sh,z)\,ds = (k+1)\int_0^1 G_{k+1}(t_0+sh,z)\,ds$$

with the second equality following from Equation 2.53. Thus we have

$$\left|\frac{G_k(t_0+h,z) - G_k(t_0,z)}{h}\right| \leq (k+1)\sup_{\substack{z\in K\\ |t-t_0|<\epsilon}}|G_{k+1}(t,z)|. \tag{2.54}$$

Let $|h_n| < \epsilon$ be any sequence tending to 0, and let $\upsilon_n(z) = \frac{1}{h_n}[G_k(t_0 + h_n, z) - G_k(t_0, z)]$. By the inductive hypothesis, υ_n is analytic on a neighborhood of K, and Inequality 2.54 shows that the family $\{\upsilon_n\}$ is uniformly bounded on K. By Montel's theorem, there is a subsequence that converges normally to an analytic function on K. But the sequence υ_n converges to $\partial_t G_k(t_0, \cdot) = (k+1)G_{k+1}(t_0, \cdot)$. This proves that G_{k+1} is analytic on K, and hence on the domain $\mathbb{C} - \overline{\mathbb{D}}$ as claimed. □

Before proceeding to the main estimates, we state and prove a lemma which is a version of the Cauchy estimates from complex analysis.

Lemma 2.17. *Let $z_0 \in \mathbb{C}$ and $\eta > 0$. If h is analytic on a neighborhood of the closed disk $\overline{\mathbb{D}(z_0, \eta)}$, and $0 < \eta' < \eta$, then*

$$|h'(z)| \leq \frac{1}{\eta - \eta'} \max_{\zeta \in \overline{\mathbb{D}(z_0, \eta)}} |h(\zeta)|, \quad \textit{for all} \quad z \in \mathbb{D}(z_0, \eta'). \tag{2.55}$$

Proof. From the Cauchy integral formula, if $r > 0$ is such that h is analytic on a neighborhood of $\overline{\mathbb{D}(z, r)}$, then

$$h'(z) = \frac{1}{2\pi i} \oint_{\partial \mathbb{D}(z,r)} \frac{h(\zeta)}{(\zeta - z)^2} \, d\zeta.$$

Thus

$$|h'(z)| \leq \frac{1}{2\pi} \oint_{\partial \mathbb{D}(z,r)} \frac{|h(\zeta)|}{|\zeta - z|^2} d\zeta = \frac{1}{2\pi r^2} \oint_{\partial \mathbb{D}(z,r)} |h(\zeta)| d\zeta \leq \frac{1}{2\pi r^2} \cdot 2\pi r \cdot \max_{\zeta \in \partial \mathbb{D}(z,r)} |h(\zeta)|. \tag{2.56}$$

If $z \in \mathbb{D}(z_0, \eta')$, then the closed disk $\overline{\mathbb{D}(z, \eta - \eta')}$ is contained in $\mathbb{D}(z_0, \eta)$ where h is holomorphic; so, taking $r = \eta - \eta'$ in Inequality 2.56 yields

$$|h'(z)| \leq \frac{1}{\eta - \eta'} \max_{\zeta \in \partial \mathbb{D}(z, \eta - \eta')} |h(\zeta)| \leq \frac{1}{\eta - \eta'} \max_{\zeta \in \overline{\mathbb{D}(z_0, \eta)}} |h(\zeta)|$$

as desired. □

Let us introduce the following maximal functions. For $k \geq 0$, $t_0 > 0$, $|z_0| > 1$, and $0 < \eta < |z_0| - 1$,

$$M_k(t_0, z_0; \eta) = \max_{\zeta \in \overline{\mathbb{D}(z_0, \eta)}} |G_k(t_0, \zeta)|, \qquad M'_k(t_0, z_0; \eta) = \max_{\zeta \in \overline{\mathbb{D}(z_0, \eta)}} |G'_k(t_0, \zeta)|. \tag{2.57}$$

When the point (t_0, z_0) is understood from context, we will shorten the notation to $M_k(\eta) \equiv M_k(t_0, z_0; \eta)$ and $M'_k(\eta) \equiv M'_k(t_0, z_0; \eta)$. By the analyticity result of Lemma 2.16, the Cauchy estimates of Lemma 2.17 yield the following *maximal Cauchy estimate*:

$$M'_k(\eta') \leq \frac{M_k(\eta)}{\eta - \eta'}. \tag{2.58}$$

We now use Inequality 2.58, together with the recursion of Equation 2.52, to prove exponential-growth bounds for G_k inductively. The key idea is to apply the maximal Cauchy estimate repeatedly with an appropriately chosen $\eta'' \in (\eta', \eta)$. It is important that η'' be chosen to minimize the bound for each individual term, or else the resulting estimates blow up super-exponentially.

Proposition 2.18. *Let $t_0 > 0$ and $|z_0| > 1$, and let $0 < \eta < \min\{|z_0| - 1, 1\}$, so that the disc $\mathbb{D}(z_0, \eta)$ of radius η centered at z_0 is contained in $\mathbb{C} - \overline{\mathbb{D}}$. There is a constant $c = c(z_0, \eta)$ so that, for all $0 < \eta' < \eta$ and all $k \geq 0$,*

$$M_k(\eta') \leq \frac{c^k (\max\{M_0(\eta), \frac{1}{2}\})^{k+1}}{(\eta - \eta')^k}. \tag{2.59}$$

Remark 2.19. It will be important that the constant c does not depend on t_0. Indeed, we will see that c does not depend on the value of G_0 at all; it can be taken to equal 52 times the maximum modulus of the polynomials p, q, p', q' on $\mathbb{D}(z_0, \eta)$, with $p(z) = z(z-1)$ and $q(z) = -(az+b)$.

Proof. The function $G_0(t_0, \cdot)$ is analytic on a neighborhood of $\overline{\mathbb{D}(z_0, \eta)}$, so it is uniformly bounded by $M_0(\eta)$ on the disk, which proves the inequality in the base case $k = 0$. We proceed by induction. Fix $k \geq 0$ and suppose that we have shown that, for $0 \leq \ell \leq k$, there are constants $c_\ell = c_\ell(\eta, z_0)$ so that

$$M_\ell(\eta') \leq \frac{c_\ell}{(\eta - \eta')^\ell}, \quad 0 \leq \ell \leq k. \tag{2.60}$$

Proceeding to $k+1$, we use the recursion of Equation 2.52, which we now expand:

$$\begin{aligned} (k+1)G_{k+1} &= \frac{\partial}{\partial z}\left[p\sum_{j=0}^{k} G_j G_{k-j} + qG_k\right] \\ &= p'\sum_{j=0}^{k} G_j G_{k-j} + 2p\sum_{j=0}^{k} G_j' G_{k-j} + q'G_k + qG_k'. \end{aligned} \tag{2.61}$$

We will bound each term in this recursion separately. Recall that $p(z) = z(z-1)$ and $q(z) = -(az+b)$ are polynomials. Hence there is a constant $\lambda \geq 1$ so that $\max\{|p(z)|, |p'(z)|, |q(z)|, |q'(z)|\} \leq \lambda$ for all $z \in \overline{\mathbb{D}(z_0, \eta)}$. Thus, on $\mathbb{D}(z_0, \eta')$,

$$\left|p'\sum_{j=0}^{k} G_j G_{k-j}\right| \leq \lambda\sum_{j=0}^{k} |G_j||G_{k-j}| \leq \lambda\sum_{j=0}^{k} M_j(\eta')M_{k-j}(\eta') \leq \lambda\sum_{j=0}^{k} \frac{c_j}{(\eta-\eta')^j}\frac{c_{k-j}}{(\eta-\eta')^{k-j}}.$$

Hence, the first term is bounded by

$$\left|p'\sum_{j=0}^{k} G_j G_{k-j}\right| \leq \frac{1}{(\lambda - \lambda')^k}\cdot\lambda\sum_{j=0}^{k} c_j c_{k-j} \leq \frac{1}{(\eta-\eta')^{k+1}}\cdot\lambda\sum_{j=0}^{k} c_j c_{k-j}, \tag{2.62}$$

where the final inequality is justified by the assumption that $\eta \leq 1$ so that $\eta - \eta' < 1$.

For the second term in 2.61, we can make the initial estimate

$$\left|2p\sum_{j=0}^{k} G_j' G_{k-j}\right| \leq 2\lambda\sum_{j=0}^{k} M_j'(\eta')M_{k-j}(\eta') \leq 2\lambda\sum_{j=0}^{k} \frac{c_{k-j}}{(\eta-\eta')^{k-j}}M_j'(\eta').$$

Now, for any $\eta'' \in (\eta', \eta)$, Inequality 2.58 with η'' in the role of η yields

$$M_j'(\eta') \leq \frac{M_j(\eta'')}{\eta'' - \eta'}. \tag{2.63}$$

Now applying the inductive hypothesis Inequality 2.60, this time with η'' in the role of η', yields

$$M_j(\eta'') \leq \frac{c_j}{(\eta - \eta'')^j}. \tag{2.64}$$

Thus, we have the estimate

$$\left|2p\sum_{j=0}^{k} G_j' G_{k-j}\right| \leq 2\lambda\sum_{j=0}^{k} \frac{c_j c_{k-j}}{(\eta''-\eta')(\eta-\eta'')^j(\eta-\eta')^{k-j}} \tag{2.65}$$

which holds for any $\eta'' \in (\eta', \eta)$. We now optimize the inequality over η'' separately in each term in the sum. By elementary calculus, we find that the minimum occurs at $\eta'' = \frac{j}{j+1}\eta' + \frac{1}{j+1}\eta$. At this point,

$$\eta - \eta'' = \tfrac{j}{j+1}(\eta - \eta'), \qquad \eta'' - \eta' = \tfrac{1}{j+1}(\eta - \eta').$$

Thus,

$$\inf_{\eta' < \eta'' < \eta} \frac{1}{(\eta'' - \eta')(\eta - \eta'')^j} = \left(\tfrac{j+1}{j}\right)^j (j+1) \frac{1}{(\eta - \eta')^{j+1}} \leq \frac{e \cdot (j+1)}{(\eta - \eta')^{j+1}}. \tag{2.66}$$

Inserting these estimates into the terms in Inequality 2.65, we have

$$\left| 2p \sum_{j=0}^{k} G'_j G_{k-j} \right| \leq 2\lambda \sum_{j=0}^{k} \frac{e \cdot (j+1) \cdot c_j c_{k-j}}{(\eta - \eta')^{j+1} (\eta - \eta')^{k-j}} = \frac{1}{(\eta - \eta')^{k+1}} \cdot 2e\lambda(k+1) \sum_{j=0}^{k} c_j c_{k-j}. \tag{2.67}$$

The third term in 2.61 is straightforward to estimate:

$$|q' G_k| \leq \lambda M_k(\eta') \leq \lambda \frac{c_k}{(\eta - \eta')^k} \leq \frac{1}{(\eta - \eta')^{k+1}} \cdot \lambda c_k \tag{2.68}$$

again using the assumption $\eta - \eta' < 1$.

For the fourth term, we use the same approach as the second term. First we have $|qG'_k| \leq \lambda M'_k(\lambda')$. Let $\lambda'' = \frac{k}{k+1}\eta' + \frac{1}{k+1}\eta$. Then Inequalities 2.63, 2.64, and 2.66 give

$$|qG'_k| \leq \frac{c_k}{(\eta'' - \eta')(\eta - \eta'')^k} \leq \frac{1}{(\eta - \eta')^{k+1}} \cdot e\lambda(k+1)c_k. \tag{2.69}$$

Finally, combining Inequalities 2.62, 2.67, 2.68, and 2.69 with Inequality 2.61 shows that

$$(k+1)|G_{k+1}| \leq \frac{1}{(\eta - \eta')^{k+1}} \cdot \left[\lambda \sum_{j=0}^{k} c_j c_{k-j} + 2e\lambda(k+1) \sum_{j=0}^{k} c_j c_{k-j} + \lambda c_k + e\lambda(k+1)c_k \right]$$

and thus

$$M_{k+1}(\eta') \leq \frac{1}{(\eta - \eta')^{k+1}} \cdot (1 + 2e)\lambda \left[\sum_{j=0}^{k} c_j c_{k-j} + c_k \right].$$

This completes the induction to show that Inequality 2.60 holds additionally for $\ell = k+1$, provided that

$$c_{k+1} \geq (1 + 2e)\lambda \left[\sum_{j=0}^{k} c_j c_{k-j} + c_k \right]. \tag{2.70}$$

To this end, we now recursively define

$$c_{k+1} = 2(1 + 2e)\lambda \sum_{j=0}^{k} c_j c_{k-j} \ \text{ for } k \geq 0, \quad c_0 = \max\{M_0(\eta), \tfrac{1}{2}\}. \tag{2.71}$$

Note that $2c_0 \geq 1$ and (by induction) $c_j \geq 0$ for all j, and so

$$\begin{aligned} c_{k+1} = 2(1+2e)\lambda \left[\sum_{j=0}^{k} c_j c_{k-j} + \sum_{j=0}^{k} c_j c_{k-j} \right] &\geq (1+2e)\lambda \left[\sum_{j=0}^{k} c_j c_{k-j} + 2c_0 c_k \right] \\ &\geq (1+2e)\lambda \left[\sum_{j=0}^{k} c_j c_{k-j} + c_k \right] \end{aligned}$$

as desired. Thus, to conclude the proof, it only remains to show that the constants defined by Equation 2.71 are bounded by the given exponential form. In fact, it is straightforward to verify that the solution to the recursion Equation 2.71 is given by the scaled Catalan numbers

$$c_k = (2(1+2e)\lambda)^k (\max\{M_0(\eta), \tfrac{1}{2}\})^{k+1} C_k,$$

where $C_k = \frac{1}{k+1}\binom{2k}{k} \le 4^k$. Therefore, taking $c = 8(1+2e)\lambda$ proves the result.

□

This brings us to the main result of this section: that $G(t,z)$ is analytic in t for $|z| > 1$.

Proposition 2.20. *The function $G(t,z)$ of Equation 1.8 is analytic in both variables (t,z) for $t > 0$ and $z \in \mathbb{C}_+$ such that $|z| > 1$.*

Proof. Since $G(t,\cdot)$ is analytic on $\mathbb{C} - [0,1]$ (cf. Equation 1.8 and Definition 1.15), we need only concern ourselves with (real) analyticity in t. Let $z_0 \in \mathbb{C}_+$ with $|z_0| > 1$, and let $t_0 > 0$. Corollary 2.9 shows that $t \mapsto G(t,z_0)$ is C^∞, and so it suffices to show that $G(t,z_0)$ is the limit of its Taylor series in a neighborhood of $t = t_0$; i.e. it suffices to show that the remainder term in Taylor's theorem tends to 0. From Lemma 2.15 and Taylor's theorem, we have for any $n \in \mathbb{N}$

$$\begin{aligned} G(t,z_0) &= \sum_{k=0}^{n} \frac{1}{k!}\frac{\partial^k}{\partial t^k} G(t_0,z_0)(t-t_0)^k + \frac{1}{(n+1)!}\frac{\partial^{n+1}}{\partial t^{n+1}} G(t_1,z_0)(t-t_0)^{n+1} \\ &= \sum_{k=0}^{n} G_k(t_0,z_0)(t-t_0)^k + G_{n+1}(t_1,z_0)(t-t_0)^{n+1}, \end{aligned} \tag{2.72}$$

for some time t_1 between t_0 and t.

Let $0 < \eta < \min\{|z_0| - 1, 1\}$, and let $0 < \eta' < \eta$; then Inequality 2.59 gives exponential bounds

$$|G_k(t_1,z_0)| \le M_k(t_1,z_0;\eta') \le \frac{c^k (\max\{M_0(t_1,z_0;\eta), \frac{1}{2}\})^{k+1}}{(\eta-\eta')^k}. \tag{2.73}$$

From the end of the proof of Proposition 2.18 and Remark 2.19, we can take

$$c \le 52 \max_{z \in \mathbb{D}(z_0,\eta)} \max\{|z(z-1)|, |2z-1|, |az+b|, |a|\} \le 52(|z_0| + \eta + 1)^2.$$

On the other hand, $M_0(t_1,z_0;\eta) = \max\{|G_0(t_1,z)| \colon |z - z_0| \le \eta\}$ can be estimated from the a priori bound on $G_0 = G$: it is the Cauchy transform of a sub-probability measure, and hence $|G(t_1,z)| \le \frac{1}{\Im z}$ for $z \in \mathbb{C}_+$. We therefore restrict $\eta < \frac{1}{2}\Im z_0$, so that the assumption $z_0 \in \mathbb{C}_+$ implies that $|M_0(t_1,z_0;\eta)| \le 2/\Im z_0$.

Thence, the t_1-independent bound in (2.73) shows that the remainder term in (2.72) tends to 0 for $|t - t_0| < (\eta - \eta')/(52(|z_0| + \eta + 1)^2)\max\{2/\Im z_0, \frac{1}{2}\}$ provided $0 < \eta' < \eta < \min\{|z_0| - 1, \frac{1}{2}\Im z_0, 1\}$. This proves the proposition. □

2.6 Analytic Continuation to $\mathbb{C}_+$ and the Proof of Theorem 1.4

We have now proven the statement of Theorem 1.4 *restricted to* $\mathbb{C}_+ - \overline{\mathbb{D}}$: $G(t,z)$ is analytic in both $t > 0$ and $z \in \mathbb{C}_+$ subject to the constraint $|z| > 1$, and it satisfies the PDE 1.9 on this domain (cf. Corollary 2.12). We knew a priori that $G(t,z)$ is analytic in z for all $z \in \mathbb{C}_+$ (cf. Equation 1.8 and Definition 1.15); it thus remains to extend the analyticity in t into the larger z-domain. This is actually quite simple, once we reinterpret t as a complex variable.

Lemma 2.21. *For each $z_0 \in \mathbb{C}_+ - \overline{\mathbb{D}}$, there is $\epsilon = \epsilon(z_0) > 0$ so that $t \mapsto G(t, z_0)$ has an analytic continuation to the strip $\{t \in \mathbb{C}\colon \Re t > 0, |\Im t| < \epsilon(z_0)\}$. The tolerance $\epsilon(z_0)$ is independent of t.*

Proof. This follows immediately from the proof of Proposition 2.20: for $t > 0$, $G(t, z_0)$ is equal to its Taylor series series centered at t_0, with radius of convergence at least $(\eta - \eta')/(52(|z_0| + \eta + 1)^2) \max\{2/\Im z_0, \frac{1}{2}\}$ provided $0 < \eta' < \eta < \min\{|z_0|-1, \frac{1}{2}\Im z_0\}$. We can therefore take $\eta' \downarrow 0$ and $\eta \uparrow \eta_0 \equiv \min\{|z_0|-1, \frac{1}{2}\Im z_0, 1\}$, and therefore define $\epsilon(z_0) = \eta_0/(52(|z_0|+\eta_0+1)^2) \max\{2/\Im z_0, \frac{1}{2}\} > 0$, independent of t_0; these power series, for all base points $t_0 > 0$, define the analytic continuation. □

Hence $G(t, z)$ can be viewed as a complex analytic function of two variables. This brings to bear all the tools of many complex variables. We can now complete the proof of Theorem 1.4, with the help of a lemma of Hartog.

Proof of Theorem 1.4. Fix a point $(t_0, z_0) \in \mathbb{R}_+ \times (\mathbb{C}_+ - \overline{\mathbb{D}})$. Let D be the largest disk centered at z_0 that does not intersect $[0, 1]$, and let D' be the largest disk centered at z_0 that does not intersect $\mathbb{D}$ (so $D' \subseteq D$). Then $G(t, z)$ is complex analytic in $t > 0$ for $z \in D'$ by Lemma 2.21. Since $G(t, z)$ is the Cauchy transform of a positive measure $\nu_t = \mu_t - (1 - \min\{\alpha, \beta\})\delta_0$ of total mass ≤ 1, supported in $[0, 1]$ (cf. Equation 1.8 and Definition 1.15), it is complex analytic in z for all $z \in D$, and is also uniformly bounded on compact subsets of $\mathbb{C}_+$ (cf. Inequality 1.13). It therefore follows from a lemma of Hartog [19, Lemma 2.2.11] that $G(t, z)$ is jointly analytic in (t, z) for $t > 0$ and z in the larger disk D. Applying this at each point of $z_0 \in \mathbb{C}_+ - \overline{\mathbb{D}}$ shows that $G(t, z)$ is analytic on $\mathbb{R}_+ \times \mathbb{C}_+$, as desired. Thus, the functions on both sides of PDE 1.9 are analytic on this domain, and by Corollary 2.12 the are equal on the open set $\mathbb{R}_+ \times (\mathbb{C}_+ - \overline{\mathbb{D}})$, it follows that they are equal on their larger analytic domain $\mathbb{R}_+ \times \mathbb{C}_+$, concluding the proof. □

3 Local Properties of the Flow μ_t

In this section, we develop properties of the measure μ_t directly from the PDE of Theorem 1.4 that determines its Cauchy transform. Let us define a new positive finite measure ν_t, supported in $[0, 1]$, by

$$\mu_t = \nu_t + (1 - \min\{\alpha, \beta\})\delta_0 \tag{3.1}$$

so that $G(t, z) = G_{\nu_t}(z)$ is the Cauchy transform of ν_t, cf. Equation 1.8. Since μ_t is a probability measure, the total mass of ν_t is

$$\nu([0, 1]) = \min\{\alpha, \beta\} \ge 0. \tag{3.2}$$

3.1 Steady-State Solution

To begin, as a sanity check, note that the steady-state equation (determined by $\frac{\partial}{\partial t} G(t, z) = 0$) takes the form

$$\partial_z[z(z - 1)G^2 - (az + b)G] = 0.$$

Due to the analyticity of G on the connected domain $\mathbb{C}_+$, this forces $z(z-1)G^2 - (az+b)G$ to be constant. The constant can be determined from the known limit behaviour of the Cauchy transform:

$$\lim_{|z|\to\infty} zG(z) = \nu_t([0, 1]) = \min\{\alpha, \beta\}.$$

Thus

$$\lim_{|z|\to\infty} [z(z - 1)G(z)^2 - (az + b)G(z)] = \min\{\alpha, \beta\}^2 - a \min\{\alpha, \beta\}.$$

Using $a = 2 \min\{\alpha, \beta\} - 1$, it follows that the steady state solution G_∞ satisfies

$$z(z - 1)G_\infty(z)^2 - (az + b)G_\infty(z) = \min\{\alpha, \beta\}(1 - \min\{\alpha, \beta\})$$

This quadratic equation has (simplified) solutions

$$G_\infty(z) = \frac{(az+b) \pm \sqrt{z^2 - 2(\alpha+\beta-2\alpha\beta)z + (\alpha-\beta)^2}}{2z(z-1)}.$$

The discriminant matches that in Equation 1.5, but the terms outside the radical do not match; of course, that equation describes the Cauchy transform of the full limit measure μ, while our measures ν_t have the point mass at $z = 0$ removed. Adding it back in,

$$\frac{1-\min\{\alpha,\beta\}}{z} + G_\infty(z) = \frac{z+\alpha+\beta-2 \pm \sqrt{z^2 - 2(\alpha+\beta-2\alpha\beta)z + (\alpha-\beta)^2}}{2z(z-1)}$$

which precisely matches the Jacobi measure's Cauchy transform from Equation 1.5, as expected. In other words, using the Stieltjes continuity Theorem 1.16, we have:

Proposition 3.1. *The spectral measure μ_t of qp_tq converges weakly to the free Jacobi measure μ of Equation 1.6 as $t \to \infty$.*

This was already known, as a consequence of the fact that p_t, q are asymptotically free as $t \to \infty$; it is interesting that this can be seen directly from PDE 1.9.

3.2 Conservation of Mass and Propagation of Singularities

The support of the measure μ_t in the limit as $t \to \infty$ is the full interval $[0,1]$; however, the initial condition μ_0 is only constrained to have point masses of appropriate magnitudes at the endpoints $\{0,1\}$, cf. Equation 1.6, and thus $\operatorname{supp}\mu_0$ may be any closed subset of $[0,1]$. It is therefore possible that, for some $t > 0$, $\operatorname{supp}\mu_t$ is disconnected. The following results deal with the flow of such support "bumps" under PDE 1.9.

Lemma 3.2. *Let $t_0 > 0$. Let U_1, U_2 be two disjoint open subintervals of $[0,1]$ (with the relative topology), and let $K_1 \subset U_1$ and $K_2 \subset U_2$ be closed subsets. Suppose that $\operatorname{supp}\mu_{t_0} \subseteq K_1 \sqcup K_2$. Then, for some $\epsilon > 0$, $\operatorname{supp}\mu_t \subset U_1 \sqcup U_2$ for all $t \in (t_0-\epsilon, t_0+\epsilon)$.*

Proof. Applying the analytic continuation argument in the proof of Theorem 1.4 (on page 32) to the larger domain $\mathbb{C} - \overline{U_1 \sqcup U_2}$ shows that $G(t,z)$ is analytic in both variables for z in this domain, and PDE 1.9 also holds there in a neighborhood of time t_0. The result follows immediately. ☐

In particular, this shows that, if for some $t_0 > 0$ the support of μ_{t_0} consists of a finite (or countable) collection of disjoint closed intervals in $[0,1]$, the same holds true for all $t \geq t_0$. Moreover, we can quantify the motion of the endpoints of these intervals. For the following, we assume that G_{ν_t} has a continuous extension to $(0,1)$ (i.e. μ_t has a continuous density); we will prove this assumption holds true for all $t > 0$ in Section 3.3, at least in the special case $\alpha = \beta = \frac{1}{2}$.

Lemma 3.3. *Suppose $G(t,z)$ has a continuous extension to $z \in (0,1)$ for all $t > 0$. Let x_t be a point in the boundary of the support of μ_t. Then $t \mapsto x_t$ satisfies the ODE*

$$\dot{x}_t = 2G(t,x_t)x_t(1-x_t) + ax_t + b, \quad t > 0. \tag{3.3}$$

Proof. The Cauchy transform of a compactly-supported measure is one-to-one on a neighborhood of ∞ in $\mathbb{C}_+$, so at least for small w, there is an analytic function $K(t,w)$ so that $w = G(t, K(t,w))$. To simplify notation, denote $G(t,z) = G_t(z)$ and $K(t,w) = K_t(w)$; let $G_t'(z) = \frac{\partial}{\partial z}G(t,z)$ and $K_t'(w) = \frac{\partial}{\partial w}K(t,w)$. Noting that $w = G_t(K_t(w))$ for small w and differentiating with respect to t, we have

$$0 = \frac{\partial}{\partial t}G_t(K_t(w)) = \frac{\partial G_t}{\partial t}(K_t(w)) + G_t'(K_t(w))\frac{\partial K_t}{\partial t}(w). \tag{3.4}$$

Now, PDE 1.9 can be written in the form

$$\frac{\partial G_t}{\partial t}(z) = (2z-1)G_t(z)^2 + 2z(z-1)G_t(z)G_t'(z) - aG_t(z) - (az+b)G_t'(z),$$

and so, using $G_t(K_t(w)) = w$, Equation 3.4 yields

$$\begin{aligned} &- G_t'(K_t(w))\frac{\partial K_t}{\partial t}(w) \\ &\quad = (2K_t(w)-1)w^2 + 2K_t(w)(K_t(w)-1)wG_t'(K_t(w)) - aw - (aK_t(w)+b)G_k'(K_t(w)). \end{aligned} \tag{3.5}$$

The inverse function theorem shows that $G_t'(K_t(w)) = 1/K_t'(w)$. Combining this with Equation 3.5 and simplifying, we have

$$-\frac{\partial K_t}{\partial t}(w) = (2K_t(w)-1)K_t'(w)w^2 + 2K_t(w)(K_t(w)-1)w - aK_t'(w)w - (aK_t(w)+b) \tag{3.6}$$

which can be written in the form

$$\frac{\partial K_t}{\partial t}(w) = \frac{\partial}{\partial w}\left[w^2K_t(w)(1-K_t(w)) + awK_t(w)\right] + b. \tag{3.7}$$

The existence of a continuous extension of G_t to a neighborhood of $(0,1)$, coupled with standard analytic continuation arguments, then implies that the PDE 3.6 holds up to the boundary $G_t((0,1))$.

Now, let x_t be a boundary point of a support interval of ν_t; then G_t is singular at x_t. Let $y_t = G_t(x_t)$ be the corresponding singular value; then $K_t'(y_t) = 0$. Hence, with $w = y_t$, equation 3.6 asserts that

$$-\frac{\partial K_t}{\partial t}(y_t) = 2K_t(y_t)(K_t(y_t)-1)y_t - (aK_t(y_t)+b) = 2x_t(x_t-1)y_t - (ax_t+b).$$

On the other hand,

$$\frac{dx_t}{dt} = \frac{\partial}{\partial t}K_t(y_t) = \frac{\partial K_t}{\partial t}(y_t) + K_t'(y_t)\frac{dy_t}{dt} = \frac{\partial K_t}{\partial t}(y_t)$$

since y_t is (by assumption) a critical point for K_t. Hence, we have the ODE

$$\dot{x}_t = 2y_tx_t(1-x_t) + (ax_t+b)$$

which yields the result, since $y_t = G_t(x_t)$. □

Equation 3.3 gives a precise (implicit) formula for the speed of propagation of singularities on the boundary: the edges of the support move with finite speed $\dot{x}_t$ so long as they stay away from the endpoints $\{0,1\}$, since $G(t,x)x(1-x)$ is continuous on $(0,1)$. In the sequel, we will see that, although $G(t,x)$ may blow up at $x=0,1$, the function $G(t,x)\sqrt{x(1-x)}$ remains bounded; thus, Equation 3.3 yields further information: as any support "bump" approaches the endpoints, its speed decreases to 0.

Remark 3.4. Equation 3.3 at first seems to suggest the boundary point x_t, which is of course in $[0,1]$, evolves into the lower half-plane: indeed, the function $G(t,z)$ takes values in the closed lower half-plane even for $z \in [0,1]$. However, x_t is assumed to be at the boundary of the support of μ_t which, under the assumption of the lemma, possesses a continuous density. The Stieltjes inversion formula of Equation 1.14 then shows that $\Im G(t,x_t) = 0$, since the density of μ_t is 0 at a boundary point of the support; hence, $G(t,x_t)$ is, in fact, real.

Remark 3.5. If the support of μ_t is not the full interval, or more generally if G_t possesses a singular point in $[0,1]$, this singular point cannot dissipate in finite time; this follows from Hartog's second theorem and the continuity principle, cf. [19]. Lemmas 3.2 and 3.3 are in line with this observation: singular points propagate with finite speed and slow down as they approach the endpoints, therefore never dissipating.

One final result that follows from this framework is conservation of mass of support "bumps".

Lemma 3.6. *Let t_0, U_1, U_2 be as in Lemma 3.2. The total mass of $\mu_t|_{U_1}$ is preserved for $t \in (t_0 - \epsilon, t_0 + \epsilon)$.*

Proof. Since, by Lemma 3.2, $\operatorname{supp} \nu_t \subseteq \operatorname{supp} \mu_t$ is contained in $U_1 \sqcup U_2$ for $t \in (t_0 - \epsilon, t_0 + \epsilon)$, the Cauchy transform $G(t,z) = G_{\nu_t}(z)$ is analytic for $z \in \mathbb{C} - \overline{U_1 \sqcup U_2}$. Let α_1 be a closed C^1 curve in $\mathbb{C} - \overline{U_1 \sqcup U_2}$ which encloses $\overline{U_1}$ and has winding number 0 around each point in $\overline{U_2}$. Then

$$\mu_t(\overline{U_1}) = \frac{1}{2\pi i} \oint_{\alpha_1} G(t,z)\, dz. \tag{3.8}$$

Equation 3.8 holds by the standard Residue Theorem if μ_t is a discrete measure – a convex combination of point-masses – supported in $\overline{U_1}$; any measure may be weakly approximated by discrete measures, and by the Stieltjes continuity Theorem 1.16 weak convergence of measures implies uniform convergence of the Cauchy transforms on compact subsets of $\mathbb{C} - \overline{U_1 \sqcup U_2}$, justifying Equation 3.8.

Lemma 3.2 justifies that PDE 1.9 holds in $\mathbb{C} - \overline{U_1 \sqcup U_2}$, and the solution is analytic in z and t. In particular, it follows that the integral on the right-hand-side of Equation 3.8 can be differentiated with respect to t under the integral, so $\mu_t(\overline{U_1})$ is differentiable, and PDE 1.9 then gives

$$\frac{\partial}{\partial t}\mu_t(\overline{U_1}) = \frac{1}{2\pi i} \oint_{\alpha_1} \frac{\partial}{\partial t} G(t,z)\, dz = \frac{1}{2\pi i} \oint_{\alpha_1} \frac{\partial}{\partial z}\left[z(z-1)G(t,z)^2 + (az+b)G(t,z)\right]\, dz \tag{3.9}$$

for $t \in (t_0 - \epsilon, t_0 + \epsilon)$. The analyticity of $G(t,z)$ in a neighborhood of the image of α_1 now guarantees the integral on the right-hand-side of Equation 3.9 is 0, by the fundamental theorem of calculus. $\square$

Remark 3.7. Of course, if support "bumps" merge in finite time, the total mass combines additively.

3.3 Subordination for the Liberation Process

This section is devoted to the proof of Theorem 1.9. We assume, from this point forward, that $\alpha = \beta = \frac{1}{2}$. It is plausible that similar techniques may apply more generally (indeed [20] gives promising progress in this direction), but that discussion is left to a future publication. The proof is fairly involved: we essentially develop an analogue in the present context of Biane's theory of subordination for the additive free convolution. To clarify the proof, we begin with an outline. The case $\alpha = \beta = \frac{1}{2}$ corresponds to $a = b = 0$ in Equation 1.9. So ν_t is a positive measure of mass $\frac{1}{2}$ supported in $[0,1]$, and its Cauchy transform $G(t,z) = G_{\nu_t}(z)$ satisfies the PDE

$$\frac{\partial}{\partial t} G(t,z) = \frac{\partial}{\partial z}\left[z(z-1)G(t,z)^2\right], \quad t > 0,\ z \in \mathbb{C}_+. \tag{3.10}$$

We begin by making a change of variables. The function $z \mapsto \sqrt{z}\sqrt{z-1}$ (where we use the standard branch of the square root function) is analytic on $\mathbb{C} - [0,1]$. Let

$$H(t,z) = H_t(z) = \sqrt{z}\sqrt{z-1}G(t,z), \quad t > 0,\ z \in \mathbb{C}_+. \tag{3.11}$$

Then $H(t,z)$ is analytic in both variables for $z \in \mathbb{C}_+$ and $t > 0$, and satisfies the PDE

$$\frac{\partial}{\partial t} H(t,z) = \sqrt{z}\sqrt{z-1}\frac{\partial}{\partial z}\left[H(t,z)^2\right] \tag{3.12}$$

on its analytic domain. We introduce the auxiliary function M, and the domain S:

$$M(w) = \tfrac{1}{2}e^{-2w}(e^{2w} + \tfrac{1}{2})^2, \qquad S = \{w = u + iv \in \mathbb{C} \colon u > \tfrac{1}{2}\ln\tfrac{1}{2}, 0 < v < \tfrac{\pi}{2}\}. \tag{3.13}$$

The function M is entire; when restricted to the strip S, $M|_S$ is injective, and its image $M(S) = \mathbb{C}_+$. Let $L = (M|_S)^{-1}$; explicitly

$$L(z) = \tfrac{1}{2}\ln M(z) = \tfrac{1}{2}\ln\left(z - \tfrac{1}{2} + \sqrt{z}\sqrt{z-1}\right), \quad z \in \mathbb{C}. \tag{3.14}$$

In fact L has an analytic continuation to $\mathbb{C} - (-\infty, 1]$ (given by Equation 3.14), and this extension is injective: it is the inverse of M on the larger domain $\overline{S} \cup \bar{S}$ (closure union conjugate). On this domain, $L'(z) = \frac{1}{2\sqrt{z}\sqrt{z-1}}$. More generally, L extends to a multivalued holomorphic function (on the infinite-sheet Riemann surface covering the slit plane $\mathbb{C} - [0, 1]$).

Given the solution H_t to PDE 3.12, define the **subordinator** f_t as

$$f_t(z) = M[L(z) + tH_t(z)], \qquad z \in \mathbb{C}_+. \tag{3.15}$$

This subordinator is a deformation of the identity: $f_0(z) = M(L(z)) = z$. (One should be wary, however: when $w \in \mathbb{C}_+$, $L(M(w))$ is only equal to w for $w \in S$.) Using the method of characteristics, we will show that $H(t, z)$ satisfies the fixed-point equation

$$H_t(z) = H_0(f_t(z)). \tag{3.16}$$

Equation 3.16 transfers the dynamics of PDE 3.12 to a deformation f_t of the identity, changing the role of H_t from active to passive. Immediately from this equation, we see that smoothness is propagated: if $H_0|_{\mathbb{C}_+}$ happens to have an analytic continuation to a complex neighborhood of $(0, 1)$, then the same must be true for H_t for all $t > 0$.

In fact, we will prove (Theorem 3.19) that, for any initial measure ν_0 and any $t > 0$, f_t extends to a homeomorphism from the closed upper half-plane $\overline{\mathbb{C}_+}$ to a region in $\overline{\mathbb{C}_+}$ bounded by a Jordan curve; it then follows from Carathéodory's Theorem (cf. [24, Thm 2.6]) that H_t has a continuous extension to $\overline{\mathbb{C}_+}$. Since the measure ν_t is given by the boundary values of G_t, it follows that ν_t has a density ρ_t which is continuous (with potential blow up at $\{0, 1\}$ due to the factor $\sqrt{x(1-x)}$ relating $H_t(x)$ and $G_t(x)$ for $x \in [0, 1]$). Finally, we show that the extension of H_t actually has an analytic continuation to a complex neighborhood of any point in the interior of $\operatorname{supp}\nu_t$, concluding the proof of Theorem 1.9.

To begin, we record the relevant properties of the auxiliary function L; the following lemma is a straightforward exercise.

Lemma 3.8. *Let $L \colon \mathbb{C}_+ \to \mathbb{C}$ be defined as in Equation 3.14. Then L is holomorphic and injective, its range is $L(\mathbb{C}_+) = S$, and L satisfies*

$$\frac{\partial L}{\partial z} = \frac{1}{2\sqrt{z}\sqrt{z-1}}, \qquad z \in \mathbb{C}_+. \tag{3.17}$$

The extension of L to a multivalued function satisfies

$$L[1, \infty) = [\tfrac{1}{2}\ln\tfrac{1}{2}, \infty), \quad L(-\infty, 0] = [\tfrac{1}{2}\ln\tfrac{1}{2}, \infty) + i\tfrac{\pi}{2}, \quad L(0, 1) = \tfrac{1}{2}\ln\tfrac{1}{2} + i(0, \tfrac{\pi}{2}). \tag{3.18}$$

In particular, L has an extension to a holomorphic map on a complex neighborhood of $(0, 1)$.

The first task is to prove the fixed-point Equation 3.16.

Lemma 3.9. *Let H_t be the solution in the upper half-plane to PDE 3.12 with initial condition H_0 analytic on $\mathbb{C}_+$. Define f_t in terms of H_t by Equation 3.15. Then $H_t(z) = H_0(f_t(z))$ for $z \in \mathbb{C}_+$.*

Proof. PDE 3.12 is well set-up for applying the method of characteristics, cf. [13]. Writing it in the form

$$\frac{\partial H}{\partial t}(t, z) = 2\sqrt{z}\sqrt{z-1}H(t, z)\frac{\partial H}{\partial z}(t, z)$$

we see that it is a homogeneous semilinear equation of the form

$$\mathbf{b}(t, z, H(t, z)) \cdot \nabla H(t, z) = 0 \tag{3.19}$$

where $\mathbf{b}(t, z, w) = \left[1, -2\sqrt{z}\sqrt{z-1}w\right]$. Fix $(t_0, z_0) \in \mathbb{R}_+ \times \mathbb{C}_+$, and define $\mathbf{z}(t) = (t, z(t))$ by the ODE

$$\frac{d\mathbf{z}}{dt} = \mathbf{b}\left[\mathbf{z}(t), H(t, z(t))\right] \tag{3.20}$$

passing through the point $\mathbf{z}(t_0) = (t_0, z_0)$. Then taking $w(t) = H(\mathbf{z}(t)) = H(t, z(t))$ and applying the chain rule, Equation 3.19 shows that the *characteristic curve* $(\mathbf{z}(t), w(t))$ is contained in the graph $\{(t, z, w) \in \mathbb{R}_+ \times \mathbb{C}_+ \times \mathbb{C} \colon w = H(t, z)\}$; moreover, the set of all such curves (for all choices (t_0, z_0)) traces out the entire graph. (This follows from the a priori knowledge that $H(t, z)$ is analytic in both variables.) It is customary to write Equation 3.20 as a system for the separate variables z, w. Because of the homogeneity in Equation 3.19, the characteristic equations in our case are

$$\frac{dz}{dt} = -2\sqrt{z}\sqrt{z-1}w, \tag{3.21}$$

$$\frac{dw}{dt} = 0. \tag{3.22}$$

Thus, $w(t) = H(t, z(t))$ is constant; so, in particular, we have

$$H(0, z(0)) = w(0) = w(t_0) = H(t_0, z(t_0)) = H(t_0, z_0). \tag{3.23}$$

We can determine the position $z(0)$ of the characteristic at time 0 (given its position $z(t_0) = z_0$) from Equation 3.21. Since $w = w(t_0)$ is constant, this can be explicitly solved in the upper half-plane, where the function $\frac{1}{2\sqrt{z}\sqrt{z-1}}$ has antiderivative L (cf. Equation 3.17). The solution is

$$\tfrac{dz}{2\sqrt{z}\sqrt{z-1}} = -w(t_0)\,dt \qquad \Longrightarrow \; L(z(t)) = -w(t_0)t + C$$

for a constant C, which is determined by the constraint $z(t_0) = z_0$; that is, $C = L(z_0) + w(t_0)t_0$. Thus, at time 0 the curve's value $z(0)$ is determined by

$$L(z(0)) = -w(t_0) \cdot 0 + L(z_0) + w(t_0)t_0 = L(z_0) + t_0 H(t_0, z_0) \tag{3.24}$$

from Equation 3.23. It follows that $L(z_0) + t_0 H(t_0, z_0) \in S$, and so $z(0) = M[L(z_0) + t_0 H(t_0, z_0)] = f_{t_0}(z_0)$ by definition Equation 3.15. Thus, Equation 3.23 asserts that

$$H(0, f_{t_0}(z_0)) = H(t_0, z_0).$$

Since $(t_0, z_0) \in \mathbb{R}_+ \times \mathbb{C}_+$ was arbitrary, this concludes the proof. □

Remark 3.10. The function $f_t = M \circ (L + tH_t)$ need not, a priori, satisfy $L \circ f_t = L + tH_t$. Nevertheless, Equation 3.24 actually clarifies that this further restriction does hold; in other words, since that equation holds true for all $(t_0, z_0) \in \mathbb{R}_+ \times \mathbb{C}_+$, we have $f_t(z) \in L(\mathbb{C}_+) = S$ for each $z \in \mathbb{C}_+$, and

$$L(f_t(z)) = L(z) + tH_t(z) = L(z) + tH_0(f_t(z)), \quad \forall\, z \in \mathbb{C}_+. \tag{3.25}$$

The second equality follows from the statement of Lemma 3.9. It is also worth noting here that the domain of definition for the function H_t in Lemma 3.9 is $\mathbb{C}_+$; as such, the characteristic $z(t)$ passing through z_0 at time t_0 is necessarily contained in $\mathbb{C}_+$, and so $f_{t_0}(z_0) = z(0) \in \mathbb{C}_+$. In other words, it follows from the proof that

$$f_t(\mathbb{C}_+) \subseteq \mathbb{C}_+, \qquad t > 0. \tag{3.26}$$

This can also be seen from the definition of f_t, Equation 3.15, at least for small $t > 0$, using the Taylor expansion of M about $L(z)$ together with a small-angle argument in a neighborhood of $\mathbb{R}$ in $\mathbb{C}_+$; the details are left to the interested reader.

Remark 3.11. It is possible to derive the relation $H(t,z) = H(0, f_t(z))$ in the following alternative fashion: define $J(t,z) = H_0(f_t(z))$. Note that $J(0,z) = H_0(f_0(z)) = H_0(z)$. Elementary but laborious calculations show that J also satisfies PDE 3.12, and is analytic. It therefore follows from the Cauchy-Kowalewski theorem that $J = H$, as claimed. We prefer the method of characteristics approach above for two reasons: the calculations are much shorter, and they avoid some technical difficulties that arise differentiating $J(t,z)$ at points z where $f_t(z)$ happens to be in $[0,1]$, outside the known analytic domain of H_0. These difficulties can be overcome with restriction and then analytic continuation techniques once continuity is proven; we prefer this more direct approach, which aids in the proof of continuity. It is worth noting that points z where $f_t(z) \in [0,1]$ will play an important role in what follows.

Remark 3.12. One might hope to use the method of characteristics to similarly deduce the existence of a subordinator for the solution of the general case, for $(\alpha,\beta) \neq (\frac{1}{2},\frac{1}{2})$, i.e. $(a,b) \neq (0,0)$. The same change of coordinates $H = \sqrt{z}\sqrt{z-1}G$ transforms the general PDE 1.9 into the *in*homogeneous semilinear equation

$$\frac{\partial H}{\partial t} - \left[2\sqrt{z}\sqrt{z-1}H - (az+b)\right]\frac{\partial H}{\partial z} + \frac{(a+2b)z-b}{2z(z-1)}H = 0.$$

The corresponding characteristic equations are

$$\begin{aligned} \frac{dz}{dt} &= -2\sqrt{z}\sqrt{z-1}w - (az+b), \\ \frac{dw}{dt} &= \frac{(a+2b)z-b}{2z(z-1)}w. \end{aligned}$$

The inhomogeneity generates a fully intertwined system of ODEs that is not explicitly solvable except in the special case $a = b = 0$; this is one demonstration of how the behavior in the case we presently consider is better than the general case.

We next use the fixed-point equation to show that the solution H_t is bounded for $t > 0$, uniformly in t away from 0.

Lemma 3.13. *For any compact subset $K \subset \mathbb{C}$, there is a constant C_K so that*

$$|H(t,z)| \le \max\{C_K/t, 1\}, \quad z \in K \cap \mathbb{C}_+.$$

In particular, it follows that, for any $\delta > 0$,

$$\sup_{t\ge\delta}\sup_{z\in\mathbb{C}_+}|H(t,z)| < \infty.$$

Proof. Consider the function $H_0 \circ M$. For fixed $\epsilon > 0$, let $B_\epsilon = B_{1/2+\epsilon}(1/2)$ be the open ball of radius $1/2+\epsilon$ centered at $1/2$ (i.e. a complex open ball containing $[0,1]$). Since $H_0(z) = \sqrt{z}\sqrt{z-1}G_0(z)$ and $G_0(z) = G_{\nu_0}(z)$ is the Cauchy transform of a measure supported in $[0,1]$, H_0 is analytic on $\mathbb{C} - [0,1]$ and hence is bounded on the closed set $\overline{\mathbb{C} - B_\epsilon}$. Thus, $H_0 \circ M$ is bounded on $L(\overline{\mathbb{C} - B_\epsilon})$. Since the function L is bounded on compact sets, there is an ϵ-dependent constant R so that

$$\overline{\mathbb{C} - B_R(0)} \subseteq L(\overline{\mathbb{C} - B_\epsilon}). \tag{3.27}$$

Now, let K be any compact subset of $\mathbb{C}$. Since $\sup_K |L| < \infty$, we can find a constant C_K such that

$$\forall w \in \mathbb{C} \quad |w| + \sup_K |L| \ge C_K \quad \Longrightarrow \quad |H_0 \circ M(w)| \le 1. \tag{3.28}$$

To see why, note that we may first choose C_K larger than $R - \sup_K |L|$; then the set of w in question is contained in the set $\overline{\mathbb{C} - B_\epsilon}$ where $H_0 \circ M$ is bounded. Since we also know that the limit as $|z| \to \infty$ of $H_0(z)$ is $1/2$, and that $|M(w)| \to \infty$ as $|w| \to \infty$, the estimate follows from the continuity of $H_0 \circ M$ on $\overline{\mathbb{C} - B_R(0)}$.
Now, for any compact $K \subset \mathbb{C}$, suppose there is a $z \in K - [0,1]$ such that $|H_t(z)| \geq C_K/t$. Set $w = L(z) + tH_t(z)$. Then

$$|w| = |tH_t(z) + L(z)| \geq t|H_t(z)| - |L(z)| \geq C_K - |L(z)|.$$

Therefore

$$|w| + \sup_K |L(z)| \geq |w| + |L(z)| \geq C_K.$$

From Equation 3.28, it follows that $|H_0 \circ M(w)| \leq 1$. But the fixed-point Equation 3.16 for H_t then says that, for $z \in K \cap \mathbb{C}_+$,

$$|H_t(z)| = |(H_0 \circ M)[L(z) + tH_t(z)]| = |H_0 \circ M(w)| \leq 1.$$

Thus, for any $z \in K \cap \mathbb{C}_+$, if $|H_t(z)| \geq C_K/t$, then $|H_t(z)| \leq 1$. This proves the first statement of the lemma. For the second, note (as used above) that $\lim_{|z|\to\infty} zG(t,z) = \nu_t[0,1] = \frac{1}{2}$ for any $t \geq 0$, and so $\lim_{|z|\to\infty} H_t(z) = \frac{1}{2}$. Moreover, the analyticity in t and the convergence to the steady state (Section 3.1) shows that this convergence is uniform in $t \geq \delta$. Thus, there is a compact set K and a fixed constant C so that $\sup_{t\geq\delta} \sup_{z\notin K} |H(t,z)| \leq C$. Combining this with the first statement of the lemma proves the second. □

With Lemma 3.13 in hand, we now make the following assumption, without loss of generality.

Assumption 1. H_0 is bounded. Thus $\sup_{t\geq 0} \sup_{z\in\mathbb{C}_+} |H(t,z)| < \infty$.

This assumption is justified by the semigroup property for the solution H_t of PDE 3.12: given $t_0 > 0$, the solution $H_t^{t_0}$ of the PDE with initial condition $H_0^{t_0} = H_{t_0}$ is equal to $H_t^{t_0} = H_{t+t_0}$. So, in each of the following statements that hold for all $t > 0$, we may simply begin the proof by selecting some $t_0 \in (0,t)$ and proving the theorem instead for $t - t_0 > 0$, then use the semigroup property; in so doing, we translate to initial condition H_{t_0} which satisfies Assumption 1 by Lemma 3.13. So we may make this assumption without loss of generality, and freely do so for the remainder of this section.

Remark 3.14. It is important to note that, while H_t satisfies the semigroup property, its subordinator f_t does not. Indeed, if we denote $f_t^{t_0}$ as the subordinator corresponding to the solution with initial condition H_{t_0}, then Equation 3.15 yields

$$f_t^{t_0}(z) = M[L(z) + tH_t^{t_0}(z)] \neq M[L(z) + (t+t_0)H_t(z)] = f_{t_0+t}(z).$$

We now proceed to demonstrate that the subordinator f_t is a homeomorphism. First, we identify its range.

Definition 3.15. *For $t > 0$, define the region $\Omega_t \in \mathbb{C}_+$ as follows:*

$$\Omega_t = \{w \in \mathbb{C}_+ \colon L(w) - tH_0(w) \in L(\mathbb{C}_+) = S\}. \tag{3.29}$$

Lemma 3.16. *The map f_t is a conformal (one-to-one) bijection from $\mathbb{C}_+$ onto Ω_t.*

Proof. Let $z_1, z_2 \in \mathbb{C}_+$. If $f_t(z_1) = f_t(z_2)$, then by Equation 3.25, it follows that

$$L(z_1) + tH_0(f_t(z_1)) = L(f_t(z_1)) = L(f_t(z_2)) = L(z_2) + tH_0(f_t(z_2)).$$

But the assumption $f_t(z_1) = f_t(z_2)$ then implies that $L(z_1) + tH_0(f_t(z_1)) = L(z_2) + tH_0(f_t(z_1))$ and so $L(z_1) = L(z_2)$; so $z_1 = z_2$ by Lemma 3.8, and thus f_t is one-to-one on $\mathbb{C}_+$. By Equation 3.25, for any $z \in \mathbb{C}_+$,

$L(f_t(z)) = L(z) + tH_0(f_t(z))$ which means that $(L - tH_0)(f_t(z)) = L(z) \in L(\mathbb{C}_+)$. This, coupled with Equation 3.26, shows that $f_t(z) \in \Omega_t$ for all $z \in \mathbb{C}_+$; and so $f_t(\mathbb{C}_+) \subseteq \Omega_t$. Conversely, if $w \in \Omega_t$, then there exists $z \in \mathbb{C}_+$ such that $L(w) - tH_0(w) = L(z)$. Equation 3.25 shows that $w = f_t(z)$ is a solution to this fixed-point equation, and moreover the injectivity of L proves that it is the unique solution; hence $w = f_t(z)$, so $w \in f_t(\mathbb{C}_+)$; and so $\Omega_t \subseteq f_t(\mathbb{C}_+)$. □

We will next show that f_t extends continuously to a homeomorphism $\overline{\mathbb{C}_+} \to \overline{\Omega_t}$. Were we to follow the approach in [5], here we would use the inverse function theorem applied to the putative inverse of f_t. Indeed, Equation 3.25 states that $(L - tH_0) \circ f_t = L$, meaning that the inverse h_t of f_t, should it exist, must be $h_t = M \circ (L - tH_0)$. A strictly-positive lower-bound on the Lipschitz constant of h_t would imply h_t extends to a continuous one-to-one map from $\overline{\Omega_t} \to \overline{\mathbb{C}_+}$, yielding the corresponding result for f_t. Unfortunately, this approach is not possible in the present context, as the following example illustrates.

Example 3.17. Suppose μ_0 is the Bernoulli measure $\mu_0 = \frac{1}{2}(\delta_0 + \delta_1)$. Hence $\nu_0 = \frac{1}{2}\delta_1$, and so $G_0(z) = 1/2(z-1)$, and $H_0(z) = \sqrt{z}/2\sqrt{z-1}$. Simple calculation shows that $\frac{\partial}{\partial w}[L(w) - tH_0(w)] = 0$ at the point $w = w_t = 1 - t/2$, which is in the unit interval for $0 \le t \le 2$; so $h_t'(w_t) = 0$. In fact, this point is also in $\partial\Omega_t$ during this time interval. Restricting to $t \in (0,1)$ and letting $s = 1 - t$, we have

$$L(w_t) = \tfrac{1}{2}\ln\left[\tfrac{1}{2}\left(1 - t + i\sqrt{(2-t)t}\right)\right] = \tfrac{1}{2}\ln\tfrac{1}{2} + \tfrac{1}{2}\ln\left[s + i\sqrt{1-s^2}\right] = \tfrac{1}{2}\ln\tfrac{1}{2} + \tfrac{1}{2}i\tan^{-1}\left(\tfrac{\sqrt{1-s^2}}{s}\right).$$

Also

$$tH_0(w_t) = t\frac{\sqrt{1-t/2}}{\sqrt{-t/2}} = -i\sqrt{1-s^2}.$$

Hence

$$L(w_t) - tH_0(w_t) = \tfrac{1}{2}\ln\tfrac{1}{2} + i\cdot\tfrac{1}{2}\left[\sqrt{1-s^2} + \tan^{-1}\left(\tfrac{\sqrt{1-s^2}}{s}\right)\right].$$

Elementary calculus shows that, for $s \in (0,1)$, this is in $\frac{1}{2}\ln\frac{1}{2} + i(0,\pi/2) \subset \overline{L(\mathbb{C}_+)}$. Thus $w_t \in \overline{\Omega_t}$, and $h_t'(w_t) = 0$. So there is no strictly positive lower-bound on $\|h_t\|_{\mathrm{Lip}(\Omega_t)}$.

Nevertheless, f_t does indeed possess a continuous, one-to-one extension to the boundary. They key issue is identifying the nature of $\partial\Omega_t$.

Lemma 3.18. *The boundary $\partial\Omega_t$ is a Jordan curve in $\overline{\mathbb{C}_+}$.*

Proof. To begin, let us note that it suffices to prove the claim for all sufficiently small $t > 0$, by the semigroup property of the solution H_t. Consider the fibration of $\mathbb{C}_+$ provided by the vertical line-segments $y \mapsto x + iy$ for $y > 0$ and fixed $x \in \mathbb{R}$. We will prove the following claim.
Claim. For fixed $x \in \mathbb{R}$, if $x + iy_0 \in \Omega_t$ then $x + iy \in \Omega_t$ for any $y > y_0$.
This suffices to prove the lemma: it demonstrates that Ω_t is the region above the graph of real-valued function of $x \in \mathbb{R}$. Since Ω_t is open, this function is automatically upper semi-continuous, and therefore does not have any oscillatory discontinuities; hence, by appending vertical line segments to any jump discontinuities, we see that Ω_t is bounded by a Jordan curve.

To prove the claim, we use the convexity of the strip $S = L(\mathbb{C}_+)$. Indeed, fix two $\frac{1}{2}$-probability measures ν_0, ν_1, and let Ω_t^0 and Ω_t^1 be the regions corresponding to these measures. Fix $s \in (0,1)$, and set $\nu_s = (1-s)\nu_0 + s\nu_1$, their convex combination, with corresponding region Ω_t^s. Let $H_0^s(z) = \sqrt{z}\sqrt{z-1}G_{\nu_s}(z)$; then

$$\begin{aligned} L(z) - tH_0^s(z) &= L(z) - t\sqrt{z}\sqrt{z-1}\int_0^1 \frac{1}{z-u}\left[(1-s)\nu_0(du) + s\nu_1(du)\right] \\ &= L(z) - t[(1-s)H_0^0(z) + sH_0^1(z)] \\ &= (1-s)[L(z) - tH_0^0(z)] + s[L(z) - tH_0^1(z)]. \end{aligned}$$

Now, suppose $z \in \Omega_t^j$ for $j = 0, 1$. This means (cf. 3.29) that $L(z) - tH_0^j(z) \in L(\mathbb{C}_+) = S$. Since S is convex, any convex combination of the two points $L(z) - tH_0^j(z)$, $i = 0, 1$, is also in S; thus, we have shown that $L(z) - tH_0^s(z) \in L(\mathbb{C}_+)$, and so $z \in \Omega_t^s$ for $0 < s < 1$. It follows immediately that, if the Claim holds for Ω_t^0 and for Ω_t^1, then it holds for Ω_t^s for $0 < s < 1$. Since every $\frac{1}{2}$-probability measure is a weak limit of convex combinations of $\frac{1}{2}$-point masses in $[0, 1]$, it therefore suffices to prove the claim only for the special case that the initial measure of the form $\nu_0 = \frac{1}{2}\delta_a$ for some point $a \in [0, 1]$.

The Cauchy transform of $\frac{1}{2}\delta_a$ is $z \mapsto \frac{1}{2(z-a)}$; hence, we must study the function

$$F_{t,a}(z) = L(z) - t\frac{\sqrt{z}\sqrt{z-1}}{2(z-a)} = \frac{1}{2}\ln\left(z - \frac{1}{2} + \sqrt{z}\sqrt{z-1}\right) - t\frac{\sqrt{z}\sqrt{z-1}}{2(z-a)}$$

To prove the claim, it suffices to show that, for each $x \in \mathbb{R}$, the image of the line segment $y \mapsto x + iy$ under $F_{t,a}$ intersects the boundary of S at most once for $y > 0$. The following facts may be verified by elementary calculus.

(1) For $x < a$, $\Im F_{t,a}(x + iy) > 0$ for all $y > 0$.

(2) For $x \geq a$, $\frac{\partial}{\partial y}\Im F_{t,a}(x + iy) > 0$ for all $y > 0$.

(3) For $x \in [0, 1]$, $\frac{\partial}{\partial y}\Re F_{t,a}(x + iy)$ possesses at most one 0, and is > 0 for large $y > 0$.

(4) For $x \notin [0, 1]$, $\frac{\partial}{\partial y}\Re F_{t,a}(x + iy) > 0$ for all $y > 0$.

Item (1) shows that the image of $\Im F_{t,a}$ never intersects the lower boundary $y = 0$ when $x < a$, and item (2) shows that it intersects the lower boundary at most once when $x \geq a$. In both cases, since $t\sqrt{z}\sqrt{z-1}/2(z-a) \to t/2$ as $|z| \to \infty$, its imaginary part tends to 0; thus for sufficiently small t (independent of a), $\Im F_{t,a}$ never intersects the upper boundary $y = \frac{\pi}{2}$. As for $\Re F_{t,a}$, when $x \in [0, 1]$ note that $\Re F_t(x + i0) = \Re L(x + i0) = \frac{1}{2}\ln\frac{1}{2}$ constantly, and so item (3) shows that the image curve $y \mapsto \Re F_{t,a}(x + iy)$ may initially dip below this level and return to intersect the line $y = \frac{1}{2}\ln\frac{1}{2}$ once, or it may stay above this line for all $y > 0$; in either case, it intersects the line at most once. Similarly, for $x \notin [0, 1]$, $\Re F_{t,a}(x + i0) > \frac{1}{2}\ln\frac{1}{2}$ for sufficiently small $t > 0$ (independent of a), and so item (4) shows that in this case $\Re F_{t,a}(x + iy) > \frac{1}{2}\ln\frac{1}{2}$ for $y > 0$.

We have thus shown that, for all sufficiently small $t > 0$, independent of $a \in [0, 1]$, for any $x \in \mathbb{R}$, if $F_{t,a}(x + iy_0) \in L(\mathbb{C}_+)$ then $F_{t,a}(x + iy) \in L(\mathbb{C}_+)$ for all $y > y_0$. This proves the claim in the case of initial measure $\frac{1}{2}\delta_a$, and thence by the above convexity argument, proves the lemma. □

This actually suffices to prove the main result: that f_t extends continuously (and, in fact, injectively) to the boundary. This illustrates the general principle that pathological boundary behavior of conformal maps is observable in the topology of the image of the boundary.

Theorem 3.19. *For $t > 0$, the subordinator $f_t\colon \mathbb{C}_+ \to \Omega_t$ extends to a homeomorphism $\overline{\mathbb{C}_+} \to \overline{\Omega_t}$.*

Proof. Since f_t is a conformal map defined on all of $\mathbb{C}_+$, and by Lemma 3.18 the boundary of $f_t(\mathbb{C}_+)$ is a Jordan curve, the result follows immediately from Carathéodory's Theorem, cf. [24, Thm 2.6]. □

Remark 3.20. Example 3.17 demonstrates that the continuous extension of the conformal map f_t to $\mathbb{R}$ need not be smooth: it can certainly have singularities along the line (at the critical values of $f_t^{-1} = h_t$). We will quantify exactly where such singularities may occur in Lemma 3.26 below.

Corollary 3.21. *For $t > 0$, the solution H_t to PDE 3.12 possesses a continuous extension to $(0, 1)$.*

Proof. Referring to Equation 3.25, we have $L(f_t(z)) = L(z) + tH_t(z)$ for $z \in \mathbb{C}_+$. Thus

$$H_t(z) = \frac{1}{t}\left[L(f_t(z)) - L(z)\right]. \tag{3.30}$$

Theorem 3.19 shows that f_t has a continuous extension to $\overline{\mathbb{C}_+}$, whose range is $\overline{\Omega_t}$, contained in $\overline{\mathbb{C}_+}$. By Lemma 3.8, L possesses an analytic continuation to a complex neighborhood of $(0, 1)$, concluding the proof. □

Remark 3.22. From here on, we will use H_t and f_t to refer to the extended continuous maps defined on $(0, 1)$. The continuity of all involved functions and the fact that Equation 3.30 holds on $\mathbb{C}_+$ shows that it also holds for the extensions to $(0, 1)$.

Corollary 3.23. *For each $t > 0$, the measure ν_t possesses a density ρ_t, which is continuous on $(0, 1)$, and satisfies the bound*

$$\rho_t(x) \le \frac{C}{\sqrt{x(1-x)}}, \quad x \in (0, 1)$$

for some constant $C > 0$ independent of t.

Proof. By Corollary 3.21, the map H_t has a continuous extension to $(0, 1)$ (which we also refer to as H_t). Thus

$$G_t(z) = \frac{1}{\sqrt{z}\sqrt{z-1}} H_t(z)$$

possesses a continuous extension to $(0, 1)$. The Stieltjes inversion Formula 1.14 in pointwise form then shows that, for $x \in (0, 1)$,

$$\rho_t(x) = -\frac{1}{\pi}\Im G_t(x) = \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}}\Re H_t(x) \tag{3.31}$$

is the density of ν_t, which is therefore continuous. Finally, Assumption 1 (i.e. Lemma 3.13, together with the semigroup property) shows that $\|H_t\|_{L^\infty(\mathbb{R})} < \infty$, and so the stated result holds true with $C = \|H_t\|_{L^\infty(\mathbb{R})}/\pi$. □

This proves most of our main Theorem 1.9. Let us also note one more immediate Corollary of Theorem 3.19, together with Lemma 3.9, that will be useful in Section 4.

Corollary 3.24. *Suppose that ν_0 has a strictly positive density ρ_0. Then the density ρ_t of ν_t is strictly positive for any $t > 0$.*

Proof. From Equation 3.31, $\rho_t(x)$ is positive at any $x \in (0, 1)$ for which $\Re H_t(x) > 0$. Equation 3.16 asserts that $H_t(x) = H_0(f_t(x))$. The harmonic function $\Re H_0$ is equal to $\pi\sqrt{x(1-x)}\rho_0(x) \ge 0$ on $\mathbb{R}$ and $\lim_{|z|\to\infty} H_0(z) = \frac{1}{2}$ so it is strictly positive for large $z \in \mathbb{C}_+$; by the minimum principle, $\Re H_0 > 0$ in $\mathbb{C}_+$. Thus, since $f_t(x)$ is either in $[0, 1]$ or in $\mathbb{C}_+$, the assumption that $\Re H_0(y) > 0$ for $y \in [0, 1]$ implies that $H_0(f_t(x)) > 0$ for all $x \in (0, 1)$. Hence $\rho_t(x) > 0$. □

Having Corollaries 3.21 and 3.23 in hand, we may apply the semigroup property as in Assumption 1, and so we freely make the following assumption from here forward, without loss of generality.

Assumption 2. H_0 has a continuous, bounded extension to $(0, 1)$, and the initial measure ν_0 possesses a density ρ_0 which is continuous on $(0, 1)$, for which $\sup_{x\in\mathbb{R}} \sqrt{x(1-x)}\rho_0(x) < \infty$.

It remains to prove the smoothness claim of Theorem 1.9 for the measure ν_t. First, we need the following results on the behavior of the continuous extension of the subordinator f_t on the boundary set $[0, 1]$.

Lemma 3.25. *Let $t > 0$ and $x \in [0,1]$. If $f_t(x) \in [0,1]$, then $\rho_t(x) = 0$, and $\rho_0(f_t(x)) = 0$.*

Proof. If $f_t(x) \in [0,1]$, then $L(f_t(x)) - L(x)$ is purely imaginary, since $\Re L \equiv \frac{1}{2}\ln\frac{1}{2}$ is constant on $[0,1]$, cf. Lemma 3.8. Hence, by Equation 3.30,

$$\Re H_t(x) = \frac{1}{t}\Re[L(f_t(x)) - L(x)] = 0.$$

It follows immediately from Equation 3.31 that $\rho_t(x) = 0$, as claimed. Also, from Equation 3.16 (extended to the boundary by continuity) $H_0(f_t(x)) = H_t(x)$, so the assumption that $f_t(x) \in [0,1]$ shows again by Equation 3.31 that $\rho_0(f_t(x)) = 0$ in this case. □

Lemma 3.26. *Let $t > 0$ and let $x \in (0,1)$. If $\rho_t(x) > 0$, then f_t is analytic in a neighborhood of x, and $f_t(x) \in \mathbb{C} - [0,1]$.*

Proof. First, from Lemma 3.25, since $\rho_t(x) \neq 0$ it follows that $f_t(x) \notin [0,1]$, as claimed. Now, from Theorem 3.19, $f_t\colon \overline{\mathbb{C}_+} \to \overline{\Omega_t}$ is a homeomorphism with inverse $h_t = M \circ (L - tH_0)$ on $\overline{\Omega_t}$. In particular, $L - tH_0$ is one-to-one on this domain, and so $(L - tH_0)^{\langle -1\rangle}$ is well-defined on $L(\overline{\mathbb{C}_+})$. What's more, since L has an analytic continuation to a neighborhood of $(0,1)$, and H_0 is analytic on $\mathbb{C} - [0,1]$, the fact that $L - tH_0$ is one-to-one on all of $L(\overline{\mathbb{C}_+})$ implies that $(L - tH_0)^{\langle -1\rangle}$ is analytic on the complement of $(L - tH_0)([0,1])$. For $x \in [0,1]$, if $L(x) \in (L - tH_0)([0,1])$ then there is some $y \in [0,1]$ so that $L(x) = L(y) - tH_0(y)$, which is precisely to say that $x = M(L(y) - tH_0(y)) = h_t(y)$, and so $y = f_t(x)$. Since $y \in [0,1]$, this implies by the first statement of the lemma that $\rho_t(x) = 0$. Hence, if $\rho_t(x) > 0$ then $(L - tH_0)^{-1}$ is analytic in a neighborhood of $L(x)$, and so f_t is analytic in a neighborhood of x. □

This completes all the elements needed for the proof of Theorem 1.9.

Proof of Theorem 1.9. By Corollary 3.23, the ν_t possesses a continuous density ρ_t with the appropriate behavior at the boundary points $\{0,1\}$. Lemma 3.9 asserts that $H_t(z) = H_0(f_t(z))$ for $z \in \mathbb{C}_+$; the continuity of H_t and f_t on $(0,1)$ means that this holds for $z = x \in (0,1)$ as well, and so $H_t(x) = H_0(f_t(x))$ where $f_t(x) \in \mathbb{C}_+$. Let $x \in [0,1]$ be a point where $\rho_t(x) > 0$. Lemma 3.26 proves that f_t is analytic in a neighborhood of x, and that $f_t(x) \notin [0,1]$. Since H_0 is analytic on $\mathbb{C} - [0,1]$, the composition $H_t(x) = H_0 \circ f_t(x)$ is therefore analytic in a neighborhood of x. □

Thus, ν_t has continuous density which is analytic except at the boundary of its support. We conclude this section with a corollary regarding the nature of the zero set of ρ_t.

Lemma 3.27. *For $t > 0$, let $Z_t = \{x \in \mathbb{R}\colon \rho_t(x) = 0\}$ be the 0-set of the measure ν_t. Then $Z_t = f_t^{-1}(Z_0)$.*

Note: we are explicitly under Assumption 2 here. The semigroup property has been applied and time has been shifted some small positive amount so that ν_0 possesses a continuous density ρ_0. The result does not say that the topology of $\operatorname{supp}\nu_t$ in any way resembles the topology of the support of the original (unliberated) measure μ_{qpq}, which can be any closed set in $[0,1]$.

Proof. Suppose $x \in Z_t$, so $\rho_t(x) = 0$. Equation 3.31 shows that $\Re H_t(x) = 0$, and Equation 3.16 then yields $\Re H_0(f_t(x)) = 0$. Now, $\Re H_0(y) = \pi\sqrt{y(1-y)}\rho_t(y) \geq 0$ for $y \in \mathbb{R}$, and $\lim_{|z|\to\infty} H_0(z) = \frac{1}{2} > 0$, so $\Re H_0 > 0$ for $|z|$ sufficiently large in $\mathbb{C}_+$. H_0 is harmonic in $\mathbb{C}_+$, so it follows from the minimum principle that $\Re H_0(z) > 0$ for $z \in \mathbb{C}_+$. Thus, $\Re H_0(y) = 0$ implies that $y \in \mathbb{R}$. Hence, $\Re H_0(f_t(x)) = 0$ implies that $f_t(x) = 0$, and so by definition $f_t(x) \in Z_0$. Thus $Z_t \subseteq f_t^{-1}(Z_0)$.

Conversely, if $f_t(x) \in Z_0$, then $y = f_t(x) \in \mathbb{R}$ and

$$0 = \rho_0(y) = \frac{1}{\pi\sqrt{y(1-x)}}\Re H_0(y) \implies 0 = \Re H_0(y) = \Re H_0(f_t(x)) = \Re H_t(x)$$

which gives $\rho_t(x) = 0$ by Equation 3.31 once more. Hence $x \in Z_t$, so $f_t^{-1}(Z_0) \subseteq Z_t$. □

Remark 3.28. It is tempting to conclude from Corollary 3.27 that $\operatorname{supp}\nu_t = f_t^{-1}(\operatorname{supp}\nu_0)$. This is not generally the case, since f_t maps the interior of the support of ν_t into the upper half-plane $\mathbb{C}_+$. The subordinator f_t is continuous and one-to-one, but it is not a map from $[0,1] \to [0,1]$. In particular, it does not follow that $\operatorname{supp}\nu_t$ is homeomorphic to $\operatorname{supp}\nu_0$ – it is perfectly possible for components of the support to merge in finite time.

4 The Unification Conjecture for Projections

This section is devoted to the Unification Conjecture for free entropy and information of projections. We briefly describe the setting of free entropy and information, and then formulate and prove a special case of the conjecture in the present context. The reader is directed to the excellent introduction [32], and the papers [31] and [17, 18], for a detailed treatment of the background.

4.1 Free Entropy, Fisher Information, and Mutual Information

Let $x_1, \dots, x_n$ be self-adjoint operators in a II_1-factor $(\mathscr{A}, \tau)$. For parameters $N, m \in \mathbb{N}$ and $R, \epsilon > 0$, the set of **matricial microstates** of these operators, denoted $\Gamma_R(x_1, \dots, x_n; N, m, \epsilon)$, is the set of self-adjoint $N \times N$ matrices $X_1, \dots, X_n$, with norm $\le R$, all of whose mixed (normalized) trace moments of order $\le m$ are within tolerance $\epsilon > 0$ of the corresponding mixed τ moments of $x_1, \dots, x_n$.

$$\begin{aligned}&\Gamma_R(x_1, \dots, x_n; N, m, \epsilon)\\ &\quad= \left\{X_1, \dots, X_n \in M_N^{sa}(\mathbb{C})\colon |\mathrm{tr}_N(X_{i_1}\cdots X_{i_r}) - \tau(x_{i_1}\cdots x_{i_r})| < \epsilon \;\forall r \in [m], (i_1, \dots, i_r) \in [n]^r\right\}\end{aligned}$$

where $[n] = \{1, \dots, n\}$. The volume of this set (in the metric given by the trace norms) grows or decays exponentially in the square of the dimension N. The (microstates) **free entropy** $\chi(x_1, \dots, x_n)$ is defined to be

$$\chi(x_1, \dots, x_n) = \sup_{R>0} \inf_{m\in\mathbb{N}, \epsilon>0} \limsup_{N\to\infty} \frac{1}{N^2} \log \mathrm{Vol}\left(\Gamma_R(x_1, \dots, x_n; N, m, \epsilon)\right). \tag{4.1}$$

In the $n = 1$ case of a single self-adjoint operator x, Voiculescu calculated that the free entropy is equal to the *logarithmic energy* of the spectral measure μ_x (up to an additive constant): if μ_x has a density $\mu_x(du) = \rho_x(u)\,du$, then

$$\chi(x) - \frac{3}{4} - \frac{1}{2}\ln 2\pi = \int \ln|u - v|\rho_x(u)\rho_x(v)\,dudv \equiv \Sigma(\mu_x). \tag{4.2}$$

(If μ_x is singular, $\chi(x) \equiv -\infty$.) For several non-commuting operators $x_1, \dots, x_n$, there is no analytic formula for the free entropy; although we will see below that, changing the definition as appropriate, the case of two *projections* does afford a closed-form analysis, cf. Equation 4.9.

Let $A, B \subset \mathscr{A}$ be algebraically free unital $*$-subalgebras; $A \vee B$ denotes the unital $*$-subalgebra generated by $A \cup B$. There is a unique derivation $\delta_{A:B}\colon A \vee B \to (A \vee B) \otimes (A \vee B)$ determined by

$$\begin{cases}\delta_{A:B}(a) = a \otimes 1 - 1 \otimes a & a \in A\\ \delta_{A:B}(b) = 0 & b \in B\end{cases}.$$

(Uniqueness is guaranteed by the algebraic freeness of A, B; without this assumption, $\delta_{A:B}$ is not well-defined. It is important to note that algebraic freeness is not related to free independence in general.) The **liberation gradient** of A relative to B, $j(A : B)$, is defined (if it exists) through integration by parts: $j(A : B) \in L^1(W^*(A \vee B), \tau)$ satisfies

$$\tau\left[j(A : B)x\right] = \tau \otimes \tau\left(\delta_{A:B}(x)\right), \quad x \in A \vee B. \tag{4.3}$$

The (liberation) **free Fisher information** of A relative to B, $\varphi^*(A:B)$, is the square-L^2-norm of the liberation gradient:

$$\varphi^*(A:B) \equiv \|j(A:B)\|_2^2 = \tau[j(A:B)^* j(A:B)]. \tag{4.4}$$

(If $j(A:B)$ does not exist, or exists in L^1 but is not in L^2, $\varphi^*(A:B) \equiv \infty$.) This definition precisely mirrors the conjugate variables approach to classical Fisher information, cf. [32]. As with free entropy, free Fisher information can rarely be computed explicitly. One important exception is in the case of two projections: if $\mathscr{A} = W^*(p,q)$, then $\varphi^*(W^*(p) : W^*(q))$ can be computed directly, cf. Equation 4.7 below.

The **mutual free information** of subalgebras A and B, $i^*(A:B)$, is defined in terms of the mutual free Fisher information, via the liberation process:

$$i^*(A:B) = \frac{1}{2}\int_0^\infty \varphi^*(u_t A u_t^* : B)\, dt \tag{4.5}$$

where u_t is a free unitary Brownian motion, freely independent from $A \vee B$. This definition is arrived at in [31, Section 4] from the classical relation between Information and Entropy. Indeed, if $S(X_1, \ldots, X_n)$ denotes the Shannon entropy of random variables $X_1, \ldots, X_n$, the mutual information $I(X:Y)$ of a pair is defined to be $I(X:Y) = -S(X,Y) + S(X) + S(Y)$. Starting here, and using a microstates-free *infinitesimal* version of free entropy χ^*, Voiculescu gave a convincing heuristic (modulo continuity issues at $t = 0, \infty$) that if $i^*(x:y) \equiv -\chi^*(x,y) + \chi^*(x) + \chi^*(y)$ then Equation 4.5 should hold for $A = W^*(x)$ and $B = W^*(y)$.

At present, it is unknown whether $\chi = \chi^*$ in general, and it is similarly unknown whether the heuristic used in [31] to give the definition of i^* can be made rigorous. This question (along with the claim that $\chi = \chi^*$) is known as the Unification Conjecture.

Conjecture 4.1 (Unification Conjecture)**.** *Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be self-adjoint operators in a II_1-factor $(\mathscr{A}, \tau)$, such that $\chi(x_1, \ldots, x_n)$, $\chi(y_1, \ldots, y_n)$, and $\chi(x_1, \ldots, x_n, y_1, \ldots, y_n)$ are all finite. Then*

$$i^*(W^*(x_1, \ldots, x_n) : W^*(y_1, \ldots, y_n)) = -\chi(x_1, \ldots, x_n, y_1, \ldots, y_n) + \chi(x_1, \ldots, x_n) + \chi(y_1, \ldots, y_n).$$

Remark 4.2. One benefit of free mutual information i^*, defined in Equation 4.5, is that it is finite in many situations where free entropy is not; for this reason, the finiteness assumption on χ is needed for the statement of Conjecture 4.1 to be plausible.

As noted in [32] and still true today, the result of Conjecture 4.1 is a long way from where the theory is at present.

4.2 Free Entropy for Projections

In the special case $n = 2$ and generators p, q that are projections, the abstract quantities of Section 4.1 can be described more directly. Following [18] and [31, Section 12], denote

$$E_{11} = p \wedge q, \quad E_{10} = p \wedge q^\perp, \quad E_{01} = p^\perp \wedge q, \quad E_{00} = p^\perp \wedge q^\perp$$

where $p^\perp = 1 - p$ and $q^\perp = 1 - q$. Define $E = 1 - (E_{00} + E_{01} + E_{10} + E_{11})$, and let $\alpha_{ij} = \tau(E_{ij})$ for $i, j \in \{0, 1\}$. Then E_{ij} (and hence E) are in the center of $W^*(p,q)$, and so the compression $\Big(EW^*(p,q)E,\, \tau|_{EW^*(p,q)E}\Big)$ is isomorphic to the 2×2-matrix-valued L^∞ space of a measure ν: the representations $EpE \leftrightarrow M_p$, $EqE \leftrightarrow M_q$ given by

$$M_p(x) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad M_q(x) = \begin{bmatrix} x & \sqrt{x(1-x)} \\ \sqrt{x(1-x)} & 1-x \end{bmatrix}$$

identify the compression as $W^*(M_p, M_q) \cong L^\infty((0,1), \nu; M_2(\mathbb{C}))$ for a uniquely-defined positive measure on $(0,1)$ with mass $\nu((0,1)) = 1 - (\alpha_{00} + \alpha_{01} + \alpha_{10} + \alpha_{11})$. The trace restricted to $EW^*(p,q)E$ is then given by

$$\tau(a) = \int_0^1 \mathrm{tr}\,[M_a(x)]\,\nu(dx).$$

In this context, the measure ν encodes all the structure of the algebra; we will soon see that ν is indeed related to the spectral measure of the operator valued projection qpq, our central object of study. In [31, Prop 12.7], it was shown that the (liberation) free Fisher information of $W^*(p)$ relative to $W^*(q)$ can be explicitly computed in terms of the measure ν.

Proposition 4.3 ([31])**.** *Suppose $\alpha_{00}\alpha_{11} = \alpha_{10}\alpha_{01} = 0$. Suppose that $\nu(dx) = \rho(x)\,dx$ has a density $\rho \in L^3((0,1), x(1-x)\,dx)$. Assume that*

$$\int_0^1 \left(\frac{\alpha_{01}+\alpha_{10}}{x} + \frac{\alpha_{00}+\alpha_{11}}{1-x}\right)\rho(x)\,dx < \infty. \tag{4.6}$$

Let ϕ denote the Hilbert transform of ρ, modified by the α_{ij} as follows:

$$\phi(x) = H\rho(x) + \frac{\alpha_{01}+\alpha_{10}}{x} + \frac{\alpha_{00}+\alpha_{11}}{1-x}.$$

Then the (liberation) free Fisher information of $W^*(p)$ relative to $W^*(q)$ is given by

$$\varphi^*(W^*(p) : W^*(q)) = \int_0^1 \phi(x)^2\rho(x)x(1-x)\,dx \tag{4.7}$$

and is finite.

Remark 4.4. If $\tau(p) = \alpha$ and $\tau(q) = \beta$ as in the foregoing, then the assumption $\alpha_{00}\alpha_{11} = \alpha_{10}\alpha_{01} = 0$ is equivalent to the identifications $\alpha_{11} = \max\{\alpha+\beta-1, 0\}$, $\alpha_{00} = \max\{1-\alpha-\beta, 0\}$, $\alpha_{10} = \max\{\alpha-\beta, 0\}$, and $\alpha_{01} = \max\{\beta-\alpha, 0\}$. In particular, in our special case $\alpha = \beta = \frac{1}{2}$, this assumption is tantamount to $\alpha_{ij} = 0$, meaning that $E_{ij} = 0$ for $i, j = \{0, 1\}$; thus, the assumption is that p **and** q **are in general position**. If they are not, we cannot expect a simple relationship between the Fisher information and the density. Note that, in this case, Equation 4.6 holds trivially.

With Proposition 4.3 in hand, we can hope to explicitly verify the Unification Conjecture 4.1 for the case of two projections. There is a twist, however. Going back to the definition of free entropy in Equation 4.1, dimension considerations show that, unless $p = q = 1$, $\chi(p) = \chi(q) = \chi(p,q) = -\infty$; hence, these operators do not fit the statement of the Unification Conjecture per se. Instead, a modified version of free entropy for projections, χ_{proj}, is required.

Let $\mathcal{G}(N,k)$ denote the Grassmannian manifold of rank-k projections on $\mathbb{C}^N$. Any projection in $\mathcal{G}(N,k)$ is conjugate to the projection $\mathrm{diag}(1,1,\ldots,1,0,0,\ldots,0)$ with k 1s; the conjugating unitary $U \in U(N)$ is only determined up to the block structure of this diagonal projection, and so U is invariant under the action of $U(k)\times U(N-k)$. This gives an identification of $\mathcal{G}(N,k) \cong U(N)/(U(k)\times U(N-k))$ as a symmetric space. Let $\pi_{N,k}\colon U(N) \to \mathcal{G}(N,k)$ be the quotient map, and define

$$\gamma_{\mathcal{G}(N,k)} = \mathrm{Haar}_{U(N)} \circ \pi_{N,k}^{-1}.$$

Thus $\gamma_{\mathcal{G}(N,k)}$ is the unique unitarily invariant probability measure on the Grassmannian of appropriate dimension/rank. They key to defining microstates free entropy for projections is to use this measure in place of Euclidean volume: if the rank of the limit projections is not full, then they cannot be properly approximated by microstates of full rank.

Fix projections $p_1, \ldots, p_n$. For $1 \le i \le n$, let $(k_i(N))_{N=1}^\infty$ be sequences of positive integers with the property that $k_i(N)/N \to \tau(p_i)$ as $N \to \infty$. Define $\Gamma_{\text{proj}}(p_1, \ldots, p_n; k_1(N), \ldots, k_n(N); N, m, \epsilon)$, the **projection microstates**, to be the set of projection matrices $P_1, \ldots, P_n$ with $P_i \in \mathcal{G}(N, k_i(N))$, all of whose mixed (normalized) trace moments of order $\le m$ are within tolerance $\epsilon > 0$ of the corresponding mixed τ moments of $p_1, \ldots, p_n$. That is, Γ_{proj} is

$$\Big\{(P_1, \ldots, P_n) \in \prod_{i=1}^n \mathcal{G}(N, k_i(N)) \colon |\text{tr}_N(P_{i_1} \cdots P_{i_r}) - \tau(p_{i_1} \cdots p_{i_r})| < \epsilon \; \forall r \in [m], (i_1, \ldots, i_r) \in [n]^r\Big\}.$$

Following Voiculescu's remarks in [31, Sect 14], in [17, 18] Hiai and Petz defined the **projection free entropy** as follows:

$$\begin{aligned}&\chi_{\text{proj}}(p_1, \ldots, p_n) \\ &\qquad = \inf_{m \in \mathbb{N}, \epsilon > 0} \limsup_{N \to \infty} \frac{1}{N^2} \log \bigotimes_{i=1}^n \gamma_{\mathcal{G}(N, k_i(N))} \left(\Gamma_{\text{proj}}(p_1, \ldots, p_n; k_1(N), \ldots, k_n(N); N, m, \epsilon)\right).\end{aligned} \tag{4.8}$$

The value does not depend on the specific sequences $k_i(N)$. Note that, when $n = 1$, all moments $\text{tr}_N(P^m) = \text{tr}_N(P) = k_i(N)$ and $\tau(p^m) = \tau(p)$, so the microstate space in this case is the full Grassmannian, which means that $\chi_{\text{proj}}(p) = 0$ for a single projection. For two projections, χ_{proj} can be explicitly calculated using large deviations results for projections, cf. [16, Thm 3.2, Prop 3.3]. The result is as follows.

Proposition 4.5 ([16])**.** *Suppose $\alpha_{00}\alpha_{11} = \alpha_{10}\alpha_{01} = 0$. There is an explicit constant $C(\alpha, \beta)$ (where $\alpha = \tau(p)$ and $\beta = \tau(q)$ as usual) such that*

$$\chi_{\text{proj}}(p, q) = \frac{1}{4}\Sigma(\nu) + \frac{\alpha_{10} + \alpha_{01}}{2} \int_0^1 \log x \, \nu(dx) + \frac{\alpha_{11} + \alpha_{00}}{2} \int_0^1 \log(1 - x) \, \nu(dx) - C(\alpha, \beta). \tag{4.9}$$

Remark 4.6. Following Remark 4.4, in the special case $\alpha = \beta = \frac{1}{2}$ the assumption $\alpha_{00}\alpha_{11} = \alpha_{10}\alpha_{01} = 0$ is equivalent to the assumption that p, q are in general position: $\tau(p \wedge q) = 0$. In this case $\alpha_{ij} = 0$ for $i, j \in \{0, 1\}$; moreover, the complicated constant $C(\alpha, \beta)$ satisfies $C(\frac{1}{2}, \frac{1}{2}) = 0$ (cf. [18, Equation 1.2]). Thus, $\chi_{\text{proj}}(p, q) = \frac{1}{4}\Sigma(\nu)$ in our case.

Thus, a natural extension of Conjecture 4.1 is to ask that it hold for projections, where the free entropy terms on right-hand-side are replaced with *projection* free entropy terms, $\chi \to \chi_{\text{proj}}$. We state this formally in the case $n = 1$ as follows.

Conjecture 4.7. *Let p, q be projections in a II_1-factor. Then*

$$i^*(W^*(p) : W^*(q)) = -\chi_{\text{proj}}(p, q) + \chi_{\text{proj}}(p) + \chi_{\text{proj}}(q) = -\chi_{\text{proj}}(p, q). \tag{4.10}$$

The remainder of this paper is devoted to the proof of Conjecture 4.7, in the special case $\tau(p) = \tau(q) = \frac{1}{2}$; i.e. Theorem 1.11.

4.3 The Proof of Theorem 1.11

The theorem is stated on page 6; it asserts that Conjecture 4.7 holds true in the special case $\alpha = \beta = \frac{1}{2}$. Our method is to employ Propositions 4.3 and 4.5 to precisely calculate the two sides of Equation 4.10. Note, as explained in Remarks 4.4 and 4.6, these Propositions require the assumption that p, q are in general position; otherwise, both sides of Equation 4.10 are $-\infty$.

The idea for the present proof is essentially due to Hiai and Ueda, and is very briefly outlined in [18, Sect 3.2]. Using the explicit formulas of Equations 4.7 and 4.9, direct calculation demonstrates that $\frac{d}{dt}\chi_{\mathrm{proj}}(p_t, q) = \frac{1}{2}\varphi^*(p_t : q)$. (This is in line with Voiculescu's original heuristics in defining free Fisher information: a version of this equality – differentiating entropy along a Brownian perturbation yields Fisher information – does hold for classical entropy and Fisher information.) Integrating this using Equation 4.5 defining mutual free information, the fundamental theorem of calculus shows that $i^*(W^*(p) : W^*(q)) = -\lim_{t\downarrow 0}\chi_{\mathrm{proj}}(p_t, q)$; passing the limit inside χ_{proj} is then justified using the work of [15, 18]. The technical difficulties arise in the derivative calculations; it is here that our main smoothing Theorem 1.9 will play a central role in the analysis.

To begin, we identify the measure ν in Propositions 4.3 and 4.5 with the (continuous part of the) operator-valued angle studied in Sections 2–3.

Lemma 4.8. *Let p, q be projections with trace $\tau(p) = \tau(q) = \frac{1}{2}$. Let $(p_t, q)_{t>0}$ be their free liberation. Denote by ν^t the measure characterizing $W^*(p_t, q)$, cf. the discussion at the beginning of Section 4.2. Then $\nu^t = 2\nu_t$, the measure of Theorem 1.4 and 1.9.*

Proof. Let ν_t be as in Theorems 1.4 and 1.9. By Corollary 3.24 and Theorem 1.9, the density ρ_t of ν_t is continuous and piecewise real analytic on $(0, 1)$. Let u_t, v_t be the real and imaginary parts of the Cauchy transform of ν_t: $G(t, z) = G_{\nu_t}(x+iy) = u_t(x, y)+iv_t(x, y)$. Equation 1.16 identifies the density ρ_t as $v_t(x, 0) = -\pi\rho_t(x)$, and its (scaled) Hilbert transform $\phi_t = -\pi H\rho_t$ is $\phi_t(x) = Hv_t(x, 0) = -u_t(x, 0)$. (We scale the Hilbert transform by $-\pi$ here to match up to the notation in [18].) In each interval on which $\rho_t > 0$, ρ_t is analytic, and hence $G_t|_{\overline{\mathbb{C}_+}}$ has an analytic continuation to a neighborhood of this interval. Hence, PDE 3.10 extends to $(0, 1)$, and we have

$$\frac{\partial}{\partial t}v_t(x, 0) = \frac{\partial}{\partial t}\Im G_t(x) = \frac{\partial}{\partial x}\left[x(x-1)\Im(G_t(x)^2)\right] = 2\frac{\partial}{\partial x}\left[x(x-1)u_t(x, 0)v_t(x, 0)\right].$$

Thus

$$-\frac{\partial}{\partial t}\pi\rho_t(x) = 2\frac{\partial}{\partial x}[x(x-1)(-\phi_t(x)\pi\rho_t(x))].$$

Dividing by π and and writing this in terms of the scaled functions $2\rho_t$ and $2\phi_t$ yields

$$\frac{\partial}{\partial t}(2\rho_t(x)) = \frac{\partial}{\partial x}[x(x-1)2\phi_t(x)\cdot 2\rho_t(x)]. \tag{4.11}$$

This is precisely the differential equation for the density of the measure ν^t, Equation 3.1 in [18], in the case $\alpha = \beta = \frac{1}{2}$ (so $\alpha_{ij} = 0$ for $i, j \in \{0, 1\}$) under the assumption that measure has a smooth (enough) density. Thus, Theorem 1.9 and the Stieltjes inversion formula 1.16 show that $2\rho_t$ is indeed the density of the measure ν^t; in other words, $\nu^t = 2\nu_t$. $\square$

Remark 4.9. This technique may be carried out for general α, β to show that the measure ν is proportional to the continuous part of the spectral measure of qpq, provided smoothness may be proved first. It should be possible to see directly that the two measures are (up to a proportionality constant) equal, since they are both defined naturally in terms of the operator-valued angle qp_tq and the intersection $p_t \wedge q$; however, such a direct identification is not obvious to the authors.

Henceforth, write $\rho^t = 2\rho_t$ and $\phi^t = 2\phi^t$. Thus, Equation 4.11, which is now proved rigorously as a consequence of Theorems 1.4 and 1.9 and Corollary 3.24, takes the form

$$\frac{\partial}{\partial t}\rho^t(x) = \frac{\partial}{\partial x}[x(x-1)\phi^t(x)\rho^t(x)]. \tag{4.12}$$

We now wish to use Propositions 4.3 and 4.5 to calculate $i^*(p_t, q)$ and $\chi_{\mathrm{proj}}(p_t, q)$; to do so, we must verify the integrability condition of Proposition 4.3.

Lemma 4.10. *For $t \geq 0$, $\rho^t \in L^3(x(1-x)\,dx)$.*

Proof. This follows immediately from the bound of Equation 1.10:

$$\begin{aligned}
\|\rho_t\|^3_{L^3(x(1-x)\,dx)} &= \int_0^1 \rho_t(x)^3 x(1-x)\,dx \\
&\leq \int_0^1 \frac{C(t)^3}{(x(1-x))^{3/2}} x(1-x)\,dx = C(t)^3 \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = C(t)^3\pi < \infty.
\end{aligned}$$

□

We now prove the main comparison that will lead to the proof of Theorem 1.11.

Lemma 4.11. *For $t > 0$,*

$$\frac{d}{dt}\chi_{\mathrm{proj}}(p_t, q) = \frac{1}{2}\varphi^*(W^*(p_t) : W^*(q)). \tag{4.13}$$

In this proof, we make the simplifying assumption that ρ_t is real analytic on $(0,1)$. The same proof justifies the statement in the more general case, that ρ_t is continuous and real analytic on the set where $\rho_t > 0$ (hence Hölder continuous), with additional notation that obscures the idea of the proof.

Proof. By Proposition 1.2, p_t and q are in general position for $t > 0$. In particular, $\alpha_{ij} = 0$ for $i, j \in \{0, 1\}$. Hence, Proposition 4.5 and Remark 4.6 say that

$$\chi_{\mathrm{proj}}(p_t, q) = \frac{1}{4}\Sigma(\nu^t) = \frac{1}{4}\int \ln|x-y|\rho^t(x)\rho^t(y)\,dxdy. \tag{4.14}$$

Now, Equation 4.12 implies that

$$\begin{aligned}
\frac{\partial}{\partial t}(\rho^t(x)\rho^t(y)) &= \rho^t(x)\frac{\partial}{\partial t}\rho^t(y) + \rho_t(y)\frac{\partial}{\partial t}\rho^t(x) \\
&= \rho^t(x)\frac{\partial}{\partial y}[y(y-1)\phi^t(y)\rho^t(y)] + \rho^t(y)\frac{\partial}{\partial x}[x(x-1)\phi^t(x)\rho^t(x)].
\end{aligned}$$

Integrating against the logarithmic energy kernel, by symmetry we have

$$\frac{1}{4}\int \ln|x-y|\frac{\partial}{\partial t}[\rho^t(x)\rho^t(y)]\,dxdy = \frac{1}{2}\int \rho^t(y)\ln|x-y|\frac{\partial}{\partial x}[x(x-1)\phi^t(x)\rho^t(x)]\,dxdy.$$

Let $k^t(x) = x(x-1)\phi^t(x)\rho^t(x)$. The Hilbert transform interchanges harmonic conjugates; hence, the analyticity of ρ^t implies the analyticity of ϕ^t, and so k^t is analytic on $(0,1)$. The dependence of the integrand on t is thus also analytic (cf. Section 2.5), and this easily justifies differentiating under the integral sign. We now integrate by parts, first writing the integral as a principal value:

$$\begin{aligned}
&\frac{d}{dt}\chi_{\mathrm{proj}}(p_t, q) \\
=&\frac{1}{2}\int \rho^t(y)\ln|x-y|\frac{\partial}{\partial x}k^t(x)\,dxdy \\
=&\frac{1}{2}\int_0^1 \rho^t(y)\,dy \lim_{\epsilon\downarrow 0}\left(\int_0^{y-\epsilon}\ln|x-y|\frac{\partial}{\partial x}k^t(x)\,dx + \int_{y+\epsilon}^1 \ln|x-y|\frac{\partial}{\partial x}k^t(x)\,dx\right) \\
=&\frac{1}{2}\int_0^1 \rho^t(y)\,dy \lim_{\epsilon\downarrow 0}\left(\ln\epsilon\cdot k^t(y-\epsilon) - \int_0^{y-\epsilon}\frac{k^t(x)}{x-y}\,dx - \ln\epsilon\cdot k^t(y+\epsilon) - \int_{y+\epsilon}^1 \frac{k_t(x)}{x-y}\,dx\right)
\end{aligned}$$

where the first equality follows from the continuity of $\partial k^t(x)/\partial x$. The boundary terms from the integration by parts combine to

$$\ln\epsilon[k^t(y-\epsilon)-k^t(y+\epsilon)] = -\epsilon\ln\epsilon\frac{k^t(y+\epsilon)-k^t(y-\epsilon)}{\epsilon} \to 0 \text{ uniformly on compact subset of } (0,1)$$

since k^t analytic (so, in particular, C^1). Hence, recombining,

$$\begin{aligned}\frac{d}{dt}\chi_{\mathrm{proj}}(p_t,q) &= -\frac{1}{2}\int_0^1 dy \lim_{\epsilon\downarrow 0}\left(\int_0^{y-\epsilon}\frac{k^t(x)}{x-y}\,dx + \int_{y+\epsilon}^1 \frac{k^t(x)}{x-y}\,dx\right)\\ &= -\frac{1}{2}\int_0^1 \rho^t(y)\,dy\left(p.v.\int_0^1 \frac{k^t(x)}{x-y}\,dx\right).\end{aligned}$$

Since ρ^t is supported in $[0,1]$, we can write this as

$$\frac{d}{dt}\chi_{\mathrm{proj}}(p_t,q) = -\frac{\pi}{2}\int_{\mathbb{R}}\rho^t(y)H(k^t)(y)\,dy.$$

Write the kernel as $k^t(x) = \sqrt{x(1-x)}\phi^t(x)\cdot\sqrt{x(1-x)}\rho^t(x)$; up to factors of π, these are the real and imaginary parts of $H_t(x)$ (cf. Equation 3.11) which, by Lemma 3.13, is uniformly bounded. Thus $k^t \in L^\infty$ and is supported in $[0,1]$; hence $k^t \in L^2(\mathbb{R})$. The Hilbert transform is self-adjoint on $L^2(\mathbb{R})$, cf. [27], and so

$$\frac{d}{dt}\chi_{\mathrm{proj}}(p_t,q) = -\frac{\pi}{2}\int_{\mathbb{R}}\rho^t(y)H(k^t)(y)\,dy = -\frac{\pi}{2}\int_{\mathbb{R}}H(\rho^t)(y)k^t(y)\,dy.$$

In the proof of Lemma 4.8, we saw that $\phi^t = -\pi H\rho^t$; thus, we have shown that

$$\frac{d}{dt}\chi_{\mathrm{proj}}(p_t,q) = \frac{1}{2}\int_{\mathbb{R}}\phi^t(y)k^t(y)\,dy = \frac{1}{2}\int_0^1 \phi^t(y)x(1-x)\phi^t(y)\rho^t(y)\,dy.$$

By Lemma 4.10, the density ρ^t is in $L^3(x(1-x)\,dx)$, and so Proposition 4.3 applies. Hence, Equation 4.7 concludes the proof:

$$\frac{d}{dt}\chi_{\mathrm{proj}}(p_t,q) = \int_0^1 \phi^t(t)^2\rho_t(y)x(1-x)\,dx = \varphi^*(W^*(p_t):W^*(q)).$$

□

Corollary 4.12. *If p,q are projections of trace $\frac{1}{2}$, then*

$$i^*(W^*(p)\colon W^*(q)) = -\lim_{t\downarrow 0}\chi_{\mathrm{proj}}(p_t,q). \tag{4.15}$$

Proof. Lemma 4.11 shows that $t\mapsto\chi_{\mathrm{proj}}(p_t,q)$ is differentiable for $t>0$, with derivative equal to $\frac{1}{2}\varphi^*(W^*(p_t):W^*(q))$. Equation 4.4 defining the (liberation) Fisher information φ^* shows that it is manifestly ≥ 0, and so the function $t\mapsto\chi_{\mathrm{proj}}(p_t,q)$ is non-decreasing. Moreover, Theorem 2.1 in [18] (the main result of that paper) shows that $-\chi_{\mathrm{proj}}(p_t,q)\leq\varphi^*(W^*(p_t):W^*(q))$ under our assumptions; whence

$$-\int_0^\infty \chi_{\mathrm{proj}}(p_t,q)\,dt \leq \int_0^\infty \varphi^*(W^*(p_t):W^*(q))\,dt = 2i^*(W^*(p):W^*(q)) \tag{4.16}$$

by Equation 4.5 defining i^*. Since $\varphi^*(W^*(p)\colon W^*(q))$ is finite (by Proposition 4.3 and Lemma 4.10), [31, Prop 10.11(c)] guarantees that $i^*(W^*(p) : W^*(q))$ is finite. Hence, Inequality 4.16 certainly guarantees that $\lim_{t\to\infty}\chi_{\mathrm{proj}}(p_t,q)=0$. Ergo, by Lemma 4.11,

$$\begin{aligned} i^*(W^*(p) : W^*(q)) = \frac{1}{2}\int_0^\infty \varphi^*(W^*(p_t) : W^*(q))\,dt &= \int_0^\infty \frac{d}{dt}\chi_{\mathrm{proj}}(p_t,q)\,dt \\ &= \lim_{t\uparrow\infty}\chi_{\mathrm{proj}}(p_t,q) - \lim_{t\downarrow 0}\chi_{\mathrm{proj}}(p_t,q) \\ &= 0 - \lim_{t\downarrow 0}\chi_{\mathrm{proj}}(p_t,q). \end{aligned}$$

□

Finally, this brings us to the proof of Theorem 1.11.

Proof of Theorem 1.11. Note that, as $t\downarrow 0$, (p_t,q) converges in non-commutative distribution to (p,q). (Indeed, for any non-commutative polynomial P in two variables, the function $t\mapsto\tau[P(p_t,q)]$ is $C^\infty[0,\infty)$.) Hence, by [17, Proposition 1.2(iii)],

$$-\chi_{\mathrm{proj}}(p,q) \le \liminf_{t\downarrow 0} -\chi_{\mathrm{proj}}(p_t,q). \tag{4.17}$$

On the other hand, by [15, Proposition 4.6], we have the reverse inequality: taking $v_1=u_t$ and $v_2=1$ (which are freely independent), it follows that $-\chi_{\mathrm{proj}}(p,q) \ge -\chi_{\mathrm{prom}}(v_1pv_1^*, v_2qv_2^*) = -\chi_{\mathrm{proj}}(p_t,q)$ for all $t\ge 0$. Thus, we have

$$-\chi_{\mathrm{proj}}(p,q) \ge \limsup_{t\downarrow 0} -\chi_{\mathrm{proj}}(p_t,q). \tag{4.18}$$

Equations 4.17 and 4.18 show that $-\chi_{\mathrm{proj}}(p,q) = \lim_{t\downarrow 0} -\chi_{\mathrm{proj}}(p_t,q)$. Combined with Corollary 4.12, this concludes the proof. □

Acknowledgments. The authors are grateful to Serban Belinschi and Yoann Dabrowski for insightful conversations, particularly regarding the results in Section 3. Thanks are also due to Michael Anshelevich, Nizar Demni, Peter Ebenfelt, and Jacob Sterbenz, for helpful feedback on a preliminary version of this paper. Finally, we wish to thank the anonymous referee for thoroughly reading an earlier version of this manuscript and providing constructive feedback on the results. In particular, thanks are due for making us aware of the results in [15, 17] that strengthened the main result of this work, removing a technical assumption that had been needed in an earlier proof of Theorem 1.11.

References

[1] Benaych-Georges, F.; Lévy, T.: *A continuous semigroup of notions of independence between the classical and the free one.* Ann. Probab. **39**, no. 3, (2011), 904-938

[2] Bercovici, H.; Collins, B.; Dykema, K.; Li, W. S.; Timotin, D.: *Intersections of Schubert varieties and eigenvalue inequalities in an arbitrary finite factor.* J. Funct. Anal. 258 (2010), no. 5, 1579–1627.

[3] Biane, P.: *Free Brownian motion, free stochastic calculus and random matrices.* Fields. Inst. Commun. **12** Amer. Math. Soc. Providence, RI, 1997. 1–19.

[4] Biane, P.: *Processes with free increments.* Math. Z. **227** (1998), no. 1 143–174.

[5] Biane, P.: *On the free convolution with a semi-circular distribution*. Indiana Univ. Math. J. **46** (1997), no. 3, 705–718.

[6] Biane, P.; Speicher, R.: *Stochastic calculus with respect to free Brownian motion and analysis on Wigner space.* Probab. Theory Related Fields **112** (1998), no. 3, 373–409.

[7] Biane, P.; Speicher, R.: *Free diffusions, free entropy and free Fisher information.* Ann. Inst. H. Poincaré Probab. Statist. **37** (2001), no. 5, 581–606

[8] Collins, B.: *Product of random projections, Jacobi ensembles and universality problems arising from free probability.* Probab. Theory Related Fields **133** (2005), no. 3, 315–344.

[9] Demni, N.; Hamdi, T.; Hmidi, T.: *On the spectral distribution of the free Jacobi process.* To appear in the Indiana University Mathematics Journal. arXiv:1204.6227.

[10] Demni, N.; Zani, M.: *Large deviations for statistics of the Jacobi process.* Stochastic Process. Appl. **119** (2009), no. 2, 518533.

[11] Demni, N.: *Free Jacobi process.* J. Theoret. Probab. **21** (2008), no. 1, 118143.

[12] Erdos, L.; Farrell, B.: *Local eigenvalue density for general MANOVA matrices.* Preprint, arXiv:1207.0031.

[13] Evans, L.: *Partial differential equations.* Graduate Studies in Mathematics, **19**. American Mathematical Society, Providence, RI, 1998

[14] Farrell, B.: *Limiting empirical singular value distribution of restrictions of discrete Fourier matrices.* Journal of Fourier Analysis and Applications **17** (2011), 733-753

[15] Hiai, F.; Miyamoto, T.; Ueda, Y.: *Orbital approach to microstate free entropy.* Internat. J. Math. 20 (2009), no. 2, 227–273.

[16] Hiai, F.; Petz, D.: *Large deviations for functions of two random projection matrices.* Acta Sci. Math. (Szeged) **72** (2006), no. 3-4, 581–609

[17] Hiai, F.; Ueda, Y.: *Notes on microstate free entropy of projections.* Publ. Res. Inst. Math. Sci. **44** (2008), no. 1, 49–89

[18] Hiai, F.; Ueda, Y.: *A log-Sobolev type inequality for free entropy of two projections*. Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 1, 239–249

[19] Hörmander, Lars.: *An introduction to complex analysis in several variables.* Third edition. North-Holland Mathematical Library, 7. North-Holland Publishing Co., Amsterdam, 1990.

[20] Izumi, M.; Ueda, Y.: *Remarks on free mutual information and orbital free entropy.* Preprint (2013) arXiv:1306.5372

[21] Kemp, T.; Nourdin, I.; Pecatti, G.; Speicher, R.: *Wigner Chaos and the Fourth Moment.* Ann. Prob. **40** (2012), 1577–1635

[22] Nica, A.; Speicher, R.: *Lectures on the combinatorics of free probability.* London Mathematical Society Lecture Note Series, 335. Cambridge University Press, Cambridge, 2006.

[23] Nualart, D.: *The Malliavin calculus and related topics.* Second edition. Probability and its Applications (New York). Springer-Verlag, Berlin, 2006.

[24] Pommerenke, C.: *Boundary Behaviour of Conformal Maps.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 299. Springer-Verlag, Berlin, 1992

[25] Raeburn, I.; Sinclair, A.: *The C^*-algebra generated by two projections.* Math. Scand. **65** (1989), no. 2, 278–290.

[26] Rudin, W.: *Principles of mathematical analysis.* Second edition McGraw-Hill Book Co., New York 1964

[27] Stein, E.; Weiss, G.: *Introduction to Fourier analysis on Euclidean spaces.* Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.

[28] Tao,T.: *Topics in random matrix theory.* Graduate Studies in Mathematics, 132. American Mathematical Society, Providence, RI, 2012.

[29] Varopoulos, N.: *The heat kernel on Lie groups.* Rev. Mat. Iberoamericana **12** (1996), no. 1, 147–186.

[30] Voiculescu, D.: *The analogues of entropy and of Fisher's information measure in free probability theory. I.* Comm. Math. Phys. **155** (1993), no. 1, 71–92.

[31] Voiculescu, D.: *The analogues of entropy and of Fisher's information measure in free probability theory. VI. Liberation and mutual free information.* Adv. Math. 146 (1999), no. 2, 101–166.

[32] Voiculescu, D.: *Free entropy.* Bull. London Math. Soc. **34** (2002), no. 3, 257–278.

[33] Voiculescu, D.; Dykema, K.; Nica, A.: *Free random variables.* CRM Monograph Series, 1. American Mathematical Society, Providence, RI, 1992.

[34] von Neumann, J. *Functional Operators. II. The Geometry of Orthogonal Spaces.* Annals of Mathematics Studies, no. 22. Princeton University Press, Princeton, N. J., 1950.

[35] Wigner, E.: *Characteristic Vectors of Bordered Matrices with Infinite Dimensions.* Ann. of Math. **62**, 548-564 (1955)

[36] Wigner, E.: *On the Distribution of the Roots of Certain Symmetric Matrices.* Ann. of Math. **67**, 325-328 (1958)

[37] Zhong, P.: *On the free convolution with a free multiplicative analogue of the normal distribution.* Preprint (2013) arXiv:1211.3.160v2