Optimal tail estimates for directed last passage site percolation with geometric random variables

Jinho Baik[*][†], **Percy Deift**[‡], **Ken McLaughlin**[§]
Peter Miller[¶] and **Xin Zhou**[‖]

May 30, 2007

Abstract

In this paper, we obtain optimal uniform lower tail estimates for the probability distribution of the properly scaled length of the longest up/right path of the last passage site percolation model considered by Johansson in [12]. The estimates are used to prove a lower tail moderate deviation result for the model. The estimates also imply the convergence of moments, and also provide a verification of the universal scaling law relating the longitudinal and the transversal fluctuations of the model.

1 Introduction

In [12], Johansson considered directed last passage site percolation on $\mathbb{Z}_+^2 = \{(m,n) : m,n \in \mathbb{N}\}$ with geometric random variables. More precisely, for $(i,j) \in \mathbb{Z}_+^2$, let $w(i,j)$ be independent, identically distributed geometric random variables with

$$\mathbb{P}(w(i,j) = k) = (1-t^2)(t^2)^k, \qquad k = 0,1,2,\cdots, \tag{1.1}$$

and $0 < t < 1$. An up/right path π from $(1,1)$ to (M,N) is, by definition, a collection of sites $\{(i_k,j_k) : k = 1,2,\cdots,n\}$, $n := M+N-1$ such that $(i_1,j_1) = (1,1)$, $(i_n,j_n) = (M,N)$ and $(i_{k+1},j_{k+1})-(i_k,j_k) = (1,0)$ or $(0,1)$. Let $(1,1) \nearrow (M,N)$ be the (finite) set of all such up/right paths from $(1,1)$ to (M,N). Now define the maximal 'length',

$$G(M,N) := \max\Big\{ \sum_{(i,j)\in\pi} w(i,j) : \pi \in (1,1) \nearrow (M,N) \Big\}. \tag{1.2}$$

[*]Deparment of Mathematics, Princeton University, Princeton, NJ 08544, USA, jbaik@math.princeton.edu
[†]Institute for Advanced Study, Princeton, NJ 08540, USA
[‡]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA, deift@cims.nyu.edu
[§]Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599, USA, mcl@amath.unc.edu
[¶]Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA, millerpd@umich.edu
[‖]Department of Mathematics, Duke University, Durham, NC 27708, USA, zhou@math.duke.edu

Fix $\gamma \geq 1$ and set $M = [\gamma N]$, the integer part of γN. The main result in [12] is the following: for any $x \in \mathbb{R}$,

$$\lim_{N\to\infty} \mathbb{P}\Big(\frac{G([\gamma N], N) - \frac{1}{a_0}N}{b_0 N^{1/3}} \leq x\Big) = F(x), \tag{1.3}$$

where

$$a_0 = \frac{1-t^2}{t((\gamma+1)t + 2\sqrt{\gamma})}, \tag{1.4}$$

and

$$b_0 = \frac{t^{1/3}\gamma^{-1/6}}{1-t^2}(t+\sqrt{\gamma})^{2/3}(1+t\sqrt{\gamma})^{2/3}, \tag{1.5}$$

and where $F(x)$ is the Tracy-Widom distribution [23] for the largest eigenvalue of a random matrix chosen from the Gaussian unitary ensemble (GUE). In addition to (1.3), Johansson also proved large deviation results,

$$\lim_{N\to\infty} \frac{1}{N^2} \log \mathbb{P}\big(G([\gamma N], N) \leq N(\frac{1}{a_0} - y)\big) = -\ell(y), \tag{1.6}$$

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(G([\gamma N], N) \geq N(\frac{1}{a_0} + y)\big) = -i(y), \tag{1.7}$$

for some explicit positive functions $\ell(y)$ and $i(y)$, $y > 0$.

The result (1.3) parallels an earlier result in [1] on the length of the longest increasing subsequence $\ell_N(\sigma)$ of a random permutation σ of N letters. The main result in [1] is the following: for any $x \in \mathbb{R}$,

$$\lim_{N\to\infty} \mathbb{P}\Big(\frac{\ell_N - 2\sqrt{N}}{N^{1/6}} \leq x\Big) = F(x), \tag{1.8}$$

where again $F(x)$ is the Tracy-Widom distribution appearing in (1.3). The authors in [1] also proved the convergence of moments: if $\chi_N(\sigma) := \frac{\ell_N - 2\sqrt{N}}{N^{1/6}}$ and χ is a random variable with distribution function $F(x)$, then for $m = 0, 1, 2, \cdots$,

$$\lim_{N\to\infty} \mathbb{E}(\chi_N^m) = \mathbb{E}(\chi^m). \tag{1.9}$$

In earlier work other authors proved large deviation results for ℓ_N,

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}\big(\ell_N \leq \sqrt{N}(2-y)\big) = -H(y), \tag{1.10}$$

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \log \mathbb{P}\big(\ell_N \geq \sqrt{N}(2+y)\big) = -I(y), \tag{1.11}$$

for $y > 0$, where $H(y), I(y)$ are certain explicit positive functions. The result (1.10) is due to Deuschel and Zeitouni [9] and the result (1.11) is essentially due to Seppäläinen [21].

In two recent papers, [18] [17], the authors have considered ℓ_N in the moderate deviation regime. More precisely, for $0 < \alpha < \frac{1}{3}$, they showed [18] that for $y > 0$

$$\lim_{N\to\infty} \frac{\log \mathbb{P}\big(\ell_N \leq (2 - yN^{-\alpha})\sqrt{N}\big)}{y^3 N^{1-3\alpha}} = -\frac{1}{12}, \tag{1.12}$$

and [17]

$$\lim_{N\to\infty} \frac{\log \mathbb{P}\big(\ell_N \geq (2+yN^{-\alpha})\sqrt{N}\big)}{y^{3/2}N^{(1-3\alpha)/2}} = -\frac{4}{3}. \tag{1.13}$$

These moderate deviation results can be motivated by noting that

$$F(x) \sim e^{x^3/12} \qquad \text{as } x \to -\infty, \tag{1.14}$$

and

$$1 - F(x) \sim \frac{e^{-(4/3)x^{3/2}}}{16\pi x^{3/2}} \qquad \text{as } x \to +\infty. \tag{1.15}$$

Thus from (1.8), one anticipates that as $N \to \infty$,

$$\begin{aligned}\mathbb{P}\big(\ell_N \leq (2-yN^{-\alpha})\sqrt{N}\big) &= \mathbb{P}\big(\ell_N \leq 2\sqrt{N} - (yN^{1/3-\alpha})N^{1/6}\big)\\ &\sim \log F(-yN^{1/3-\alpha}) \sim -\frac{1}{12}y^3N^{1-3\alpha},\end{aligned} \tag{1.16}$$

and

$$\begin{aligned}\mathbb{P}\big(\ell_N \geq (2+yN^{-\alpha})\sqrt{N}\big) &= \mathbb{P}\big(\ell_N \geq 2\sqrt{N} + (yN^{1/3-\alpha})N^{1/6}\big)\\ &\sim \log(1-F(yN^{1/3-\alpha})) \sim -\frac{4}{3}y^{3/2}N^{(1-3\alpha)/2}.\end{aligned} \tag{1.17}$$

Of course, when $\alpha = 0$, we are in the large deviation regime, and when $\alpha = \frac{1}{3}$, we are in the GUE central limit theorem regime. The above moderate deviation results can also be motivated by estimating the functions $I(y)$ and $H(y)$ for the large-deviation regime. In [17], the authors proved (1.13) by refining certain estimates in [1] and using a careful summation argument. In [18], the authors utilized an analogous summation argument together with the estimate Lemma 6.3 (ii) in [1].

Calculations similar to (1.16), (1.17), motivate the following moderate deviation results for $G([\gamma N], N)$: for $0 < \alpha < \frac{2}{3}$,

$$\lim_{N\to\infty} \frac{\log \mathbb{P}\big(G([\gamma N], N) \leq (\frac{1}{a_0} - yb_0N^{-\alpha})N\big)}{y^3N^{2-3\alpha}} = -\frac{1}{12}, \tag{1.18}$$

and

$$\lim_{N\to\infty} \frac{\log \mathbb{P}\big(G([\gamma N], N) \geq (\frac{1}{a_0} + yb_0N^{-\alpha})N\big)}{y^{3/2}N^{1-3\alpha/2}} = -\frac{4}{3}. \tag{1.19}$$

One of the principal goals in this paper is to prove (1.18). Relation (1.19) is slightly simpler and can also be approached using the techniques in this paper. We hope to return to this problem in the future.

Relation (1.18) is a consequence of the following result.

Theorem 1.1. *Fix $0 < t < 1$ and $\gamma_0 \geq 1$. Then there exist a (large) constant $L > 0$ and a (small) constant $\delta > 0$, such that for large N,*

$$\log \mathbb{P}\big(G([\gamma N], N) \leq \frac{1}{a_0}N - xb_0N^{1/3}\big) = -\frac{1}{12}x^3 + O(x^4N^{-2/3}) + O(\log x) \tag{1.20}$$

uniformly for all $L \le x \le \delta N^{2/3}$ and $1 \le \gamma \le \gamma_0$. In particular, for the variables x, γ in the same range,

$$\mathbb{P}\big(G([\gamma N], N) \le \frac{1}{a_0} N - x b_0 N^{1/3}\big) \le e^{-c|x|^3} \tag{1.21}$$

for some constant $c > 0$.

Setting $x = yN^{2/3-\alpha}$ in (1.20), we immediately obtain (1.18) together with error estimates.

Corollary 1.2 (Estimate for lower moderate deviation). *For $0 < \alpha < \frac{2}{3}$ and $y > 0$, as $N \to \infty$,*

$$\frac{\log \mathbb{P}\big(G([\gamma N], N) \le (\frac{1}{a_0} - y b_0 N^{-\alpha}) N\big)}{y^3 N^{2-3\alpha}} = -\frac{1}{12} + O(yN^{-\alpha}) + O\big(\frac{\log(yN^{2/3-\alpha})}{y^3 N^{2-3\alpha}}\big), \tag{1.22}$$

Theorem 1.1 can also be used for other applications.

Corollary 1.3 (Convergence of moments). *For $\gamma \ge 1$, set $\theta_N := \frac{G([\gamma N], N) - a_0^{-1} N}{b_0 N^{1/3}}$ and let χ be the random variable with distribution function $F(x)$ as above. Then for $m = 0, 1, 2, \cdots$,*

$$\lim_{N \to \infty} \mathbb{E}(\theta_N^m) = \mathbb{E}(\chi^m). \tag{1.23}$$

Proof. By Remark 2.5 of [12], (1.23) follows from the estimate (1.21). □

In particular, setting $m = 2$, we see that the fluctuation $\sqrt{Var(G([\gamma N], N))}$ of G is of order N^η where $\eta = \frac{1}{3}$. It is believed (see e.g., [16]) that the transversal fluctuations of G have order N^ξ where ξ and η are related by a dimension-independent universal scaling law $2\xi = \eta + 1$. In other words, it is expected that in our case $\xi = \frac{2}{3}$. In [13], Johansson considered transversal fluctuations for the Poissonized version of the longest increasing subsequence problem and showed in that case that the scaling law $2\xi = \eta + 1$ is satisfied. By [1], η is again 1/3 and it follows therefore that the scaling law $2\xi = \eta + 1$ is satisfied for this case. A key role in his analysis was again played by Lemma 6.3 (ii) of [1]. This Poissonized problem can be viewed as a continuum version of the above site percolation problem and in Remark 1.2 of [13] Johansson notes that the scaling law $2\xi = \eta + 1$ for the site percolation problem would follow from an estimate of type (1.21) above. The modifications in the argument in [13] that are needed for the site percolation problem are detailed in [14]. We thus have

Corollary 1.4 (Transversal fluctuations). *For any $\gamma \ge 1$, the above coefficients η and ξ for longitudinal and transversal fluctuations of the site percolation model obey the scaling law*

$$2\xi = \eta + 1. \tag{1.24}$$

In order to prove Corollary 1.3, 1.4, weaker bounds than (1.21) suffice. Indeed, using an observation of Harold Widom [24] (see in particular Lemma 2), it is possible to prove the bound

$$\mathbb{P}\big(G([\gamma N], N) \le \frac{1}{a_0} N - x b_0 N^{1/3}\big) \le e^{-c'|x|^{3/2}} \tag{1.25}$$

for x, γ in the range of Theorem 1.1, for some constant $c' > 0$. As opposed to the proof of (1.21), which requires a steepest-descent Riemann-Hilbert analysis (see below), the proof of (1.25) uses only classical steepest-descent methods. A key role in [24] is played by a beautiful conjecture of Widom for the spectral properties of a class of singular integral operators (see [24]). This conjecture can be verified in our case, as in the case considered considered by Widom in [24], by using an elegant formula of Borodin and Okounkov [4] (see identity (4.9)). The method in [24] is itself motivated by earlier calculations in [2]. The estimate (1.25) is enough to prove Corollary 1.3, 1.4, but does not suffice to prove Corollary 1.2.

Remark 1. The bound in Lemma 6.3 (ii) of [1] was also used by Seppäläinen [22] to control fluctuations for the "stick process" introduced in [20]. In [22], Seppäläinen also mentioned that a similar result could be obtained for a certain continuous-time totally asymmetric simple exclusion process, provided the appropriate analogue of Lemma 6.3 (ii) could be established. The same should be true for a discrete-time version of this process. The above estimate (1.25), and of course also the stronger estimate (1.21), then suffices to control the fluctuations as in the stick process.

Remark 2. Our results are given for $G(M,N)$ where $M = [\gamma N] \geq N$, but as the statistics of $G(M,N)$ are clearly the same as for $G(N,M)$, it is immediate that our results, suitably scaled, also apply to $G([\gamma N], N)$ for $0 < \gamma < 1$.

As indicated above, the proof of Theorem 1.1 is based on the steepest-descent method for Riemann-Hilbert problems (RHP's) introduced by Deift and Zhou [8], and further developed in [7]. The method has been used to solve a wide variety of asymptotic problems in pure and applied mathematics (see, for example, [5, 6] and the references therein). The steepest-descent calculations in this paper are closely related to the calculations in [5, 6] and particularly [1]. Our analysis is based on the algebraic formula (2.7) below, which relates $\mathbb{P}(G(M,N) \leq n)$ to the solution Y to an associated RHP on the unit circle $\Sigma = \{|z| = 1\}$ (see (2.6)). It follows then that our problem reduces to the asymptotic analysis of a RHP with large oscillatory parameters. The steepest-descent method in [8] was introduced precisely for this purpose. The key step in the method is to identify the leading order asymptotics for the solution of the RHP and this is done following [7], [5, 6] and [1] by introducing a so-called g-function with certain specific properties on an appropriate contour $\overline{\Gamma_1 \cup \Gamma_2}$ (see Proposition 4.1 below). Using g, one transforms the RHP for Y as follows: $U := e^{-\frac{1}{2}k\ell\sigma_3} Y e^{-k(g-\frac{1}{2}\ell)\sigma_3}$ where ℓ is a specific constant to be determined and σ_3 is the Pauli matrix $\sigma_3 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. A simple calculation shows that U solves the RHP

$$\begin{cases} U(z) \text{ is analytic in } z \in \mathbb{C} \setminus \Sigma, \text{ and continous up to the boundary,} \\ U_+(z) = U_-(z) \begin{pmatrix} e^{-k(g_+ - g_-)} & e^{k(g_+ + g_- - W - \ell)} \\ 0 & e^{k(g_+ - g_-)} \end{pmatrix}, \qquad z \in \Sigma, \end{cases} \tag{1.26}$$

where W is given (2.17), and $g_\pm$ denote the boundary values of g. In addition, one requires $g(z) = \log z + O(z^{-1})$ as $z \to \infty$, so that the RHP for U is normalized at infinity,

$$U(z) = I + O(1/z), \qquad \text{as } z \to \infty. \tag{1.27}$$

The choice of the properties of g mentioned above is made precisely such that the leading contribution to the RHP (1.26) is immediate. Further information on the steepest-descent method can be found, for example, in [7], [5, 6], [1].

In [7], [5, 6], the RHP's are given on the real line $\mathbb{R}$ and the analogues of Γ_1, Γ_2 are subintervals of $\mathbb{R}$: in [1], the RHP is given on the unit circle $\Sigma = \{|z| = 1\}$ and the analogues of Γ_1, Γ_2 are again subintervals of Σ. The main new technical feature of the RHP in this paper is that Γ_1 and Γ_2 cannot be chosen as subintervals of the original contour Σ, and the central problem is to discover the shape of Γ_1, Γ_2. The situation is similar to the problem confronted by Kamvissis, McLaughlin and Miller in [15], where the authors considered the semi-classical limit of the solution of the Cauchy problem for the focusing nonlinear Schrödinger equation. Motivated by the calculations in [15], we show that the construction of Γ_1, Γ_2 is equivalent to the problem of determining the global structure of the trajectories $Q(z)(dz)^2 > 0$ and orthogonal trajectories $Q(z)(dz)^2 < 0$ of a particular quadratic differential $Q(z)(dz)^2$ (see (3.53) below).

The outline of the paper is as follows. In Section 2, we derive the basic algebraic formula (2.7) relating $\mathbb{P}(G([\gamma N], N) \leq n)$ and the RHP (2.6), and state our basic asymptotic estimate, Proposition 2.2, for $Y_{21}(0; k)$. In Section 3, which is the heart of the paper, we construct Σ_1 and Σ_2 using the theory of quadratic differentials and verify the desired properties of $h = g'$. In Section 4, the constant ℓ mentioned above is defined (see (4.2)) and the desired properties of $g = \int^z h$ are verified (Proposition 4.1). In Section 5, we use the g-function to analyze the RHP and eventually give the proof of Proposition 2.2. Finally, in Section 6, we use the estimate in Proposition 2.2 together with a careful summation argument as in [18] to prove the main result Theorem 1.1.

Acknowledgments. The authors would like to thank Nick Ercolani for useful discussions and Kurt Johansson for making available to us his calculations in [14]. The authors would also like to thank Harold Widom for providing us with his preprint [24]. The first author would like to thank Anne Boutet de Monvel for kindly inviting him to Université Paris 7, where a part of work is done, and also acknowledge that a part of work is conducted while he is visiting Korea Institute for Advanced Study for 2 weeks of August, 2001. The work of the first author was supported in part by NSF Grant # DMS 97-29992. The work of the second author was supported in part by NSF Grant # DMS 00-03268. The work of the third author was supported in part by NSF Grant # DMS-9970328. The work of the fourth author was supported in part by NSF Grant # DMS 01-03909. The work of the fifth author was supported in part by NSF Grant # DMS 0071398.

2 Basic relations and formulae

For $M, N \geq 1$, let

$$Z_{M,N} := (1 - t^2)^{-MN}. \tag{2.1}$$

Set

$$\varphi(z) = (1 + tz)^M (1 + \frac{t}{z})^N, \tag{2.2}$$

and consider the $n \times n$ Toeplitz determinant

$$D_n(\varphi) = D_n := \det(\varphi_{j-k})_{0 \le j,k < n}, \tag{2.3}$$

where φ_j is the j^{th} Fourier coefficient of φ:

$$\varphi_j := \int_{|z|=1} z^{-j} \varphi(z) \frac{dz}{2\pi i z}. \tag{2.4}$$

Here and below the integration contour $|z| = 1$ is assumed to be oriented in the counter-clockwise direction. Let $G(M, N)$ be the maximal length introduced in the Introduction. From earlier result of Gessel [11] an Johansson [12], Baik and Rains [3] extracted the relation

$$\mathbb{P}(G(M, N) \le n) = \frac{1}{Z_{M,N}} D_n(\varphi), \tag{2.5}$$

which plays the basic role in our analysis.

Let Σ be the unit circle $|z| = 1$ in the complex plane, oriented counter-clockwise and let $Y(z) = Y(z; k) = (Y_{ij}(z; k))_{1 \le i,j \le 2}$ be the solution to the following 2×2 matrix RHP:

$$\begin{cases} Y(z) \text{ is analytic in } z \in \mathbb{C} \setminus \Sigma, \text{ and continous up to the boundary,} \\ Y_+(z) = Y_-(z) \begin{pmatrix} 1 & z^{-k}\varphi(z) \\ 0 & 1 \end{pmatrix}, \qquad z \in \Sigma, \\ Y(z) z^{-k\sigma_3} = I + O(1/z), \qquad \text{as } z \to \infty, \end{cases} \tag{2.6}$$

where $\sigma_3 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ is the standard third Pauli matrix, and $Y_+(z)$, (resp., $Y_-(z)$), $z \in \Sigma$, are the boundary values of $Y(z')$ as $z' \to z$ from the inside (resp., outside) of the circle.

Lemma 2.1. *The solution Y to the above RHP* (2.6) *exists and is unique. Moreover,*

$$\mathbb{P}(G(M, N) \le n) = \prod_{k=n}^{\infty} (-Y_{21}(0; k + 1)). \tag{2.7}$$

Proof. We will construct the solution Y explicitly using computations similar to [10]. First note that from the equality (2.5), $D_n(\varphi) \ne 0$ for $n \ge 0$ since the probability $\mathbb{P}(G(M, N) = 0) = \mathbb{P}(w(i, j) = 0, 1 \le \forall i \le M, 1 \le \forall j \le N) = (1 - q)^{MN}$, and hence $\mathbb{P}(G(M, N) \le n) \ge \mathbb{P}(G(M, N) = 0) > 0$ for $n \ge 0$. Consider for $k \ge 0$ the polynomials of degree k

$$\pi_k(z) := \frac{1}{D_k} \det \begin{pmatrix} \varphi_0 & \varphi_{-1} & \cdots & \varphi_{-k} \\ \varphi_1 & \varphi_0 & \cdots & \varphi_{-k+1} \\ \vdots & \vdots & & \vdots \\ \varphi_{k-1} & \varphi_{k-2} & \cdots & \varphi_{-1} \\ 1 & z & \cdots & z^k \end{pmatrix}, \quad \pi_k^*(z) := \frac{1}{D_k} \det \begin{pmatrix} \varphi_0 & \varphi_1 & \cdots & \varphi_k \\ \varphi_{-1} & \varphi_0 & \cdots & \varphi_{k-1} \\ \vdots & \vdots & & \vdots \\ \varphi_{-k+1} & \varphi_{-k+2} & \cdots & \varphi_1 \\ z^k & z^{k-1} & \cdots & 1 \end{pmatrix}. \tag{2.8}$$

A direct check shows that π_k and π_k^* satisfy the following orthogonality conditions:

$$\int_{|z|=1} z^{-j}\pi_k(z)\varphi(z)\frac{dz}{2\pi iz} = N_k\delta_{jk} \qquad 0 \le j \le k, \tag{2.9}$$

$$\int_{|z|=1} z^{-j}\pi_k^*(z)\varphi(z)\frac{dz}{2\pi iz} = N_k\delta_{j0} \qquad 0 \le j \le k. \tag{2.10}$$

where

$$N_k = \frac{D_{k+1}}{D_k}. \tag{2.11}$$

Let $(Ch)(z) = \frac{1}{2\pi i}\int_\Sigma \frac{h(s)}{s-z}ds$, $z \in \mathbb{C}\backslash\Sigma$, denote the Cauchy transform of h. Let $(C_\pm h)(z) = \lim_{z'\to z}(Ch)(z')$ where z' approaches z from the $\pm$ side respectively, denote its boundary values as in the Introduction. We claim that

$$Y(z;k) = \begin{pmatrix} \pi_k(z) & C((\cdot)^{-k}\varphi\pi_k)(z) \\ -N_{k-1}^{-1}\pi_{k-1}^*(z) & -N_{k-1}^{-1}C((\cdot)^{-k}\varphi\pi_{k-1}^*)(z) \end{pmatrix}. \tag{2.12}$$

is a solution to (2.6). The analyticity of Y in $\mathbb{C}\setminus\Sigma$ is clear, while the jump condition follows from the relation $C_+ - C_- = 1$. The asymptotic condition follows from the orthogonality (2.9), (2.10). On the other hand, the uniqueness of the solution to the RHP (2.6) is standard (cf. for example, Lemma 4.1 of [1]). Hence (2.12) is the unique solution to the RHP (2.6).

For the proof of (2.7), first note that by taking $n \to \infty$ in (2.5),

$$\lim_{n\to\infty} D_n(\varphi) = Z_{M,N}. \tag{2.13}$$

(This can also be seen directly from the Szegö strong limit theorem for Toeplitz determinants.) Thus we have, using (2.11),

$$\mathbb{P}(G(M,N) \le n) = \prod_{k=n}^{\infty}\frac{D_k}{D_{k+1}} = \prod_{k=n}^{\infty} N_k^{-1} \tag{2.14}$$

Finally, from (2.8) and (2.12), we observe that $Y_{21}(0;k) = -N_{k-1}^{-1}$, which completes the proof. □

Remark 3. From (2.7),

$$-Y_{21}(0;k+1) = \frac{\mathbb{P}(G(M,N) \le k)}{\mathbb{P}(G(M,N) \le k+1)}, \tag{2.15}$$

and hence we have $-Y_{21}(0;k) > 0$ for $k \ge 1$.

There are three parameters in the RHP (2.6): M, N, k. We regard t, $0 < t < 1$, as a fixed number throughout this paper. For convenience we introduce the following notation:

$$\gamma := \frac{M}{N}, \qquad a := \frac{N}{k}. \tag{2.16}$$

Instead of M, N, k, we now regard γ, a, k as our parameters in the RHP (2.6). Using this notation, the jump condition for Y takes the form

$$Y_+(z) = Y_-(z)\begin{pmatrix} 1 & e^{-kW(z)} \\ 0 & 1 \end{pmatrix}, \qquad W(z) := -\gamma a \log(1+tz) - a\log(1+t/z) + \log z, \tag{2.17}$$

where $\log w$ is defined to be analytic in $\mathbb{C} \setminus (-\infty, 0]$, and $\log w = \log|w|$ for $w > 0$. We are interested in the asymptotics of Y as $k \to \infty$, while γ and a remain in appropriate bounded regions. The main technical result we are going to prove in the rest of this paper is the following:

Proposition 2.2. *Set*

$$a_0 = \frac{1-t^2}{t((\gamma+1)t + 2\sqrt{\gamma})}. \tag{2.18}$$

Fix $0 < t < 1$ and $\gamma_0 \geq 1$. There are L_0, δ_0, $k_0 > 0$ such that for γ, a, k satisfying

$$a_0 + \frac{L_0}{k^{2/3}} \leq a \leq (1+\delta_0)a_0, \tag{2.19}$$

$k \geq k_0$ and $1 \leq \gamma \leq \gamma_0$, we have

$$\log(-Y_{21}(0;k)) = -c_2 k(a-a_0)^2 + O(k|a-a_0|^3) + O(|a-a_0|) + O\Big(\frac{1}{k|a-a_0|}\Big), \tag{2.20}$$

where

$$c_2 = \frac{t^2(t+t\gamma+2\sqrt{\gamma})^3\sqrt{\gamma}}{4(1+t\sqrt{\gamma})^2(t+\sqrt{\gamma})^2}. \tag{2.21}$$

In particular, for the variables in the same range,

$$\log(-Y_{21}(0;k)) \leq -c_0 k(a-a_0)^2 \tag{2.22}$$

for some constant $c_0 > 0$.

3 The h-function

As noted in the Introduction, we seek a g-function and associated contours $\overline{\Gamma_1 \cup \Gamma_2}$. Rather than analyzing g directly, we first seek an *h-function*, $h' = g$. We start from an ansatz, to be verified a posteriori, that Γ_1, Γ_2 has the shape indicated in Figure 1 with the endpoints ξ, $\overline{\xi}$. Recall from (2.17)

$$W(z) = -\gamma a\log(1+tz) - a\log(1+t/z) + \log z, \qquad W'(z) = -\frac{\gamma a}{z+t^{-1}} - \frac{a}{z+t} + \frac{a+1}{z}, \tag{3.1}$$

and recall from (2.18)

$$a_0 = \frac{1-t^2}{t((\gamma+1)t+2\sqrt{\gamma})}. \tag{3.2}$$

Notation from the theory of quadratic differentials (see e.g., [19]): Given a meromorphic function $f(z)$ and a simple oriented curve C lying outside the zeros and poles of f, the notation $f(z)(dz)^2 > 0$ means that $f(z(t))(\frac{dz}{dt})^2$ is real and positive for all $t \in (a,b)$ where $z(t)$, $t \in (a,b)$ is the arc length parameterization of C. Similarly, $f(z)(dz)^2 < 0$ means that $f(z(t))(\frac{dz}{dt})^2$ is real and negative.

Proposition 3.1. *Fix $0 < t < 1$, $\gamma \geq 1$, $a > a_0$. Then there exist (cf. Figure 1 below)*

- *a point ξ with $Im(\xi) > 0$,*
- *a simple open curve Γ_1 connecting $\overline{\xi}$ and ξ, oriented from $\overline{\xi}$ to ξ such that (i) it does not intersect $(-\infty, 0]$ and (ii) Γ_1 is symmetric with reflection about the real line,*
- *a simple open curve Γ_2 connecting ξ and $\overline{\xi}$, oriented from ξ to $\overline{\xi}$ such that (i) it does not intersect $(-\infty, -1/t] \cup [-t, \infty)$ and (ii) Γ_2 is symmetric with respect to reflection about the real line,*
- *a function $h(z)$ analytic in $\mathbb{C} \setminus \overline{\Gamma_1}$ and continuous up to the boundary*

such that the following properties are satisfied:

(a) $h_+(z) + h_-(z) = W'(z)$ *for* $z \in \Gamma_1$.

(b) $h(z) = \frac{1}{z} + O(z^{-2}))$ *as* $z \to \infty$.

(c) $i(h_+(z) - h_-(z))dz > 0$ *for* $z \in \Gamma_1$.

(d) $(2h(z) - W'(z))dz < 0$ *for* $z \in \Gamma_2 \cap \mathbb{C}_+$ *and* $(2h(z) - W'(z))dz > 0$ *for* $z \in \Gamma_2 \cap \mathbb{C}_-$.

In addition, we have the following properties:

(i) The function h has the form

$$\begin{aligned} h(z) &= \frac{R(z)}{2\pi i} \int_{\Gamma_1} \frac{W'(s)}{R_+(s)(s-z)} ds \\ &= \frac{1}{2} W'(z) + \frac{1}{2} R(z) \left(\frac{\gamma a}{(z+t^{-1})R(-t^{-1})} + \frac{a}{(z+t)R(-t)} - \frac{a+1}{zR(0)} \right), \end{aligned} \tag{3.3}$$

where $R(z) = \sqrt{(z-\xi)(z-\overline{\xi})}$ is defined to be analytic in $\mathbb{C} \setminus \Gamma_1$ and $R(z) \sim z$ as $z \to \infty$.

(ii) (Endpoint condition)

$$\int_{\Gamma_1} \frac{W'(s)}{R_+(s)} ds = 0, \qquad \frac{1}{2\pi i} \int_{\Gamma_1} \frac{sW'(s)}{R_+(s)} ds = -1. \tag{3.4}$$

(iii) Set

$$\Phi(z) := (1+\gamma a) \frac{R(z)(z+z_0)}{z(z+t)(z+t^{-1})}, \qquad z \in \mathbb{C} \setminus \Gamma_1, \tag{3.5}$$

where

$$z_0 = \frac{a+1}{-R(0)(1+\gamma a)}. \tag{3.6}$$

Then

$$2h(z) - W'(z) = \Phi(z), \qquad z \in \mathbb{C} \setminus \overline{\Gamma_1}. \tag{3.7}$$

In particular,

$$h_+(z) - h_-(z) = \Phi_+(z), \qquad z \in \Gamma_1, \tag{3.8}$$

$$2h(z) - W'(z) = \Phi(z), \qquad z \in \Gamma_2. \tag{3.9}$$

(iv) The curve Γ_2 intersects the real axis at $-z_0$.

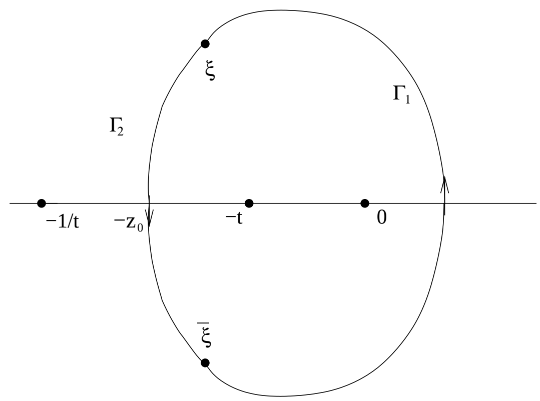

Figure 1: Sketch of the contours Γ_1 and Γ_2.

Idea of proof:

Suppose that the curve Γ_1 is known. Let

$$R(z) = \sqrt{(z-\xi)(z-\overline{\xi})}, \tag{3.10}$$

which is defined to be analytic in $\mathbb{C} \setminus \Gamma_1$, and $R(z) \sim z$ as $z \to \infty$. From the Plemelj formula, it is easy to check that

$$h(z) = \frac{R(z)}{2\pi i} \int_{\Gamma_1} \frac{W'(s)}{R_+(s)(s-z)} ds, \tag{3.11}$$

satisfies condition (a). By a residue calculation, h can be written as

$$h(z) = \frac{1}{2}W'(z) + \frac{1}{2}R(z)\left(\frac{\gamma a}{(z+t^{-1})R(-t^{-1})} + \frac{a}{(z+t)R(-t)} - \frac{a+1}{zR(0)} \right). \tag{3.12}$$

For this h to satisfy (b), the following two conditions are necessary and sufficient,

$$\int_{\Gamma_1} \frac{W'(s)}{R_+(s)} ds = 0, \qquad \frac{1}{2\pi i} \int_{\Gamma_1} \frac{sW'(s)}{R_+(s)} ds = -1, \tag{3.13}$$

or equivalently by residue calculations,

$$\frac{\gamma a}{R(-t^{-1})} + \frac{a}{R(-t)} = \frac{a+1}{R(0)}, \qquad \frac{\gamma a}{tR(-t^{-1})} + \frac{at}{R(-t)} = -(1+\gamma a). \tag{3.14}$$

As we will see, these conditions determine the endpoint ξ.

Now for $z \in \Gamma_1$, from (3.12),

$$i(h_+(z) - h_-(z)) = iR_+(z)\left(\frac{\gamma a}{(z+t^{-1})R(-t^{-1})} + \frac{a}{(z+t)R(-t)} - \frac{a+1}{zR(0)}\right). \tag{3.15}$$

Substituting $R(-t^{-1})$ and $R(-t)$ in terms of $R(0)$ using the endpoint conditions (3.14) we obtain, after some algebra, for $z \in \Gamma_1$,

$$i(h_+(z) - h_-(z)) = i(1+\gamma a)R_+(z)\frac{(z+z_0)}{z(z+t)(z+t^{-1})}, \tag{3.16}$$

where

$$z_0 = \frac{a+1}{-R(0)(1+\gamma a)}. \tag{3.17}$$

Hence we have

$$Q(z) := (i(h_+ - h_-))^2 = -(1+\gamma a)^2 \frac{(z-\xi)(z-\overline{\xi})(z+z_0)^2}{z^2(z+t)^2(z+t^{-1})^2}, \tag{3.18}$$

which is meromorphic in $\mathbb{C}$. We then use the theory of the quadratic differentials to find the trajectories for $Q(z)(dz)^2 > 0$ and this leads us to the determination of the contour Γ_1 for which condition (c) is satisfied.

The contour Γ_2 for which condition (d) is satisfied turns out to be obtained by finding the so-called orthogonal trajectories corresponding to $Q(z)(dz)^2 < 0$. Clearly if a trajectory and an orthogonal trajectory meet at a point z in the plane where $Q(z)$ is analytic and nonzero, they do so at right angles to each other.

The rest of this section consists of a proof of the above Proposition.

3.1 The endpoint ξ

In this subsection, we are going to prove that there is a unique ξ, $Im(\xi) > 0$, for which the following two conditions (cf. (3.13)) are satisfied,

$$\int_\Gamma \frac{W'(s)}{R_+(s)}ds = 0, \qquad \frac{1}{2\pi i}\int_\Gamma \frac{sW'(s)}{R_+(s)}ds = -1, \tag{3.19}$$

where Γ is *any* simple oriented curve connecting ξ and $\overline{\xi}$, oriented from $\overline{\xi}$ to ξ such that (i) it does not intersect $(-\infty, 0]$ (ii) it is symmetric under reflection about the real line, and

$$R(z) = \sqrt{(z-\xi)(z-\overline{\xi})}, \tag{3.20}$$

which is defined to be analytic in $\mathbb{C} \setminus \Gamma$ with $R(z) \sim z$ as $z \to \infty$.

Remark 4. A priori we should look for a pair of unrelated points ξ_1, ξ_2 such that (3.19) is satisfied for any contour Γ connecting them as above. It turns out, however, that it is sufficient to look for ξ_1 and ξ_2 in the form $\xi_1 = \xi$ and $\xi_2 = \overline{\xi}$. The reason for this symmetry lies in the form of the equations (3.19).

Indeed, both $W'(s)$ and $sW'(s)$ are real analytic functions, and in each integral the path of integration can be doubled along the "minus" side of the branch cut for $R(s)$ and then deformed into a closed loop containing ξ_1 and ξ_2 and the branch cut Γ connecting them but no singularities of $W'(s)$; this loop is otherwise arbitrary. If we take the loop to be symmetric with respect to reflection in the real axis, then it is easy to see that the only way for both integrals to be purely imaginary as required by (3.19) is for $R(s)$ itself to be a real analytic function, which forces $\xi_2 = \overline{\xi_1}$.

As in (3.14), these conditions become

$$\frac{\gamma a}{R(-t^{-1})} + \frac{a}{R(-t)} = \frac{a+1}{R(0)}, \qquad \frac{\gamma a}{tR(-t^{-1})} + \frac{at}{R(-t)} = -(1+\gamma a). \tag{3.21}$$

Set

$$r := |\xi| = -R(0), \qquad x := |t^{-1} + \xi| = -R(-t^{-1}), \qquad y := |t + \xi| = -R(-t). \tag{3.22}$$

Then conditions (3.21) have the form

$$\frac{\gamma a}{x} + \frac{a}{y} = \frac{a+1}{r}, \qquad \frac{\gamma a}{tx} + \frac{at}{y} = 1 + \gamma a. \tag{3.23}$$

Set

$$r_1 := \frac{t(a+1)}{1+\gamma a}, \qquad r_2 := \frac{a+1}{t(1+\gamma a)}. \tag{3.24}$$

The conditions (3.23) are now, after simple algebra, equivalent to

$$\frac{a\gamma(1-t^2)}{x} = t(1+\gamma a)\big(1 - \frac{r_1}{r}\big), \tag{3.25}$$

$$\frac{a(1-t^2)}{y} = t(1+\gamma a)\big(\frac{r_2}{r} - 1\big). \tag{3.26}$$

Also from the definitions, r, x, y satisfy the relation

$$r^2 = 1 + \frac{y^2 - t^2x^2}{1-t^2}. \tag{3.27}$$

Inserting x, y of (3.25), (3.26) into (3.27), we obtain an equation for r:

$$\frac{1}{r^2} + \frac{(1-t^2)a^2}{(1+\gamma a)^2}\left(\frac{1}{t^2(r_2-r)^2} - \frac{\gamma^2}{(r-r_1)^2}\right) - 1 = 0. \tag{3.28}$$

Since $x, y > 0$, from (3.25), (3.26), we must have $r_1 < r < r_2$. Thus we seek x, y, r satisfying

$$x > 0, \qquad y > 0, \qquad r_1 < r < r_2. \tag{3.29}$$

Lemma 3.2. *For each fixed* $0 < t < 1$, $\gamma \geq 1$, $a > a_0$, *there is a unique solution* r *to* (3.28) *satisfying* $r_1 < r < r_2$.

Proof. Set

$$H(r) := \frac{1}{r^2} + \frac{(1-t^2)a^2}{(1+\gamma a)^2}\left(\frac{1}{t^2(r_2-r)^2} - \frac{\gamma^2}{(r-r_1)^2}\right) - 1. \tag{3.30}$$

Clearly, $H(r) \to -\infty$ as $r \downarrow r_1$, and $H(r) \to +\infty$ as $r \uparrow r_2$. Thus there is $r_1 < r_c < r_2$ satisfying $H(r_c) = 0$. We want to show that such an r_c is unique. By direct calculation, for $r_1 < r < r_2$,

$$H(r) + \frac{r}{2}H'(r) = \frac{a^2(1+a)(1-t^2)}{(1+\gamma a)^3}\left(\frac{1}{t^3(r_2-r)^3} + \frac{t\gamma^2}{(r-r_1)^3}\right) - 1. \tag{3.31}$$

The minimum of this function on (r_1, r_2) is obtained at

$$r_* = \frac{r_1 + r_2\sqrt{\gamma}t}{1+\sqrt{\gamma}t}, \tag{3.32}$$

and for $r_1 < r < r_2$,

$$H(r) + \frac{r}{2}H'(r) \geq H(r_*) + \frac{r_*}{2}H'(r_*) = \left(\frac{a(1+\sqrt{\gamma}t)^2}{(1+a)(1-t^2)}\right)^2 - 1. \tag{3.33}$$

But since $a > a_0$,

$$H(r) + \frac{r}{2}H'(r) > \left(\frac{a_0(1+\sqrt{\gamma}t)^2}{(1+a_0)(1-t^2)}\right)^2 - 1 = 0. \tag{3.34}$$

Therefore, if $H(r_c) = 0$ for $r_1 < r_c < r_2$, we must have $H'(r_c) > 0$. A simple calculus argument then proves the uniqueness of the solution. □

Thus if we define x, y by (3.25), (3.26), we have obtained the unique solution (r, x, y) to the equations (3.25), (3.26), (3.27) subject to (3.29). Now we need to prove that the (r, x, y) defined in this way determines ξ, $Im(\xi) > 0$, uniquely from (3.22). In order for ξ, $Im(\xi) > 0$, to satisfy (3.22), we must have $\xi = re^{i\theta}$ for some $0 < \theta < \pi$ satisfying

$$\cos\theta = \frac{x^2 - r^2 - t^{-2}}{2rt^{-1}} = \frac{y^2 - r^2 - t^2}{2rt}. \tag{3.35}$$

Conversely, if there exists $\theta \in (0, \pi)$ satisfying (3.35), then $\xi := re^{i\theta}$ and $\overline{\xi} = re^{-i\theta}$ are the desired endpoints. However, the second inequality follows from (3.27) and so it is sufficient to prove that for (x, y, r) satisfying (3.25), (3.26), (3.27), (3.29), we have the relation

$$-1 < \frac{x^2 - r^2 - t^{-2}}{2rt^{-1}} < 1. \tag{3.36}$$

In order to prove (3.36), we first prove the following Lemma.

Lemma 3.3. *For each* $0 < t < 1$, $\gamma \geq 1$, $a > a_0$, *the solution* (r, x, y) *to* (3.25), (3.26), (3.27), *subject to* (3.29) *satisfies*

$$x + y > t^{-1} - t. \tag{3.37}$$

Proof. From (3.25), (3.26), we have

$$\frac{x+y}{t^{-1}-t} = \frac{a}{1+\gamma a}\Big(\gamma - 1 + \frac{r_1\gamma}{r-r_1} + \frac{r_2}{r_2-r}\Big). \tag{3.38}$$

The minimum of the right-hand side, regarded as a function in r, is again obtained at $r = r_*$, where r_* is defined in (3.32), and hence, by evaluating the minimum, we obtain

$$\frac{x+y}{t^{-1}-t} \geq \frac{a(t+\sqrt{\gamma})^2}{(1+\gamma a)(1-t^2)}. \tag{3.39}$$

But since $a > a_0$, we have

$$\frac{x+y}{t^{-1}-t} > \frac{a_0(t+\sqrt{\gamma})^2}{(1+\gamma a_0)(1-t^2)} = 1. \tag{3.40}$$

□

Now we prove (3.36).

Lemma 3.4. *For each* $0 < t < 1$, $\gamma \geq 1$, $a > a_0$*, the solution* (r, x, y) *to* (3.25), (3.26), (3.27), *subject to* (3.29) *satisfies*

$$-1 < \frac{x^2 - r^2 - t^{-2}}{2rt^{-1}} < 1. \tag{3.41}$$

Proof. Suppose that

$$\frac{x^2 - r^2 - t^{-2}}{2rt^{-1}} \geq 1. \tag{3.42}$$

Then $x^2 \geq (r + t^{-1})^2$, and thus from (3.27),

$$y^2 = t^2x^2 + (r^2-1)(1-t^2) \geq t^2(r+t^{-1})^2 + (r^2-1)(1-t^2) = (r+t)^2. \tag{3.43}$$

Hence we have

$$x \geq r + t^{-1}, \qquad y \geq r + t. \tag{3.44}$$

Inserting (3.44) into (3.25), (3.26), we have

$$\frac{a\gamma(1-t^2)}{r+t^{-1}} \geq t(1+\gamma a)\big(1 - \frac{r_1}{r}\big), \qquad \frac{a(1-t^2)}{r+t} \geq t(1+\gamma a)\big(\frac{r_2}{r} - 1\big) \tag{3.45}$$

We multiply the first inequality by $r(r + t^{-1})$, and multiply the second inequality by $r(r + t)$. Then by adding the two inequalities, we obtain, after some algebra, $0 \geq 2$, which is a contradiction.

Now suppose that

$$\frac{x^2 - r^2 - t^{-2}}{2rt^{-1}} \leq -1. \tag{3.46}$$

This implies that $(tx)^2 \leq (1 - rt)^2$. Since $r < r_2 \leq \frac{1}{t}$ for $\gamma \geq 1$, we have

$$tx \leq 1 - rt, \tag{3.47}$$

and from (3.27),

$$y^2 = t^2x^2 + (r^2-1)(1-t^2) \le (1-rt)^2 + (r^2-1)(1-t^2) = (r-t)^2. \tag{3.48}$$

Hence we have $y \le |r-t|$. We distinguish two cases $r \ge t$ and $r < t$. For the first case when $r \ge t$, from (3.47) we have

$$x + y \le (t^{-1} - r) + (r - t) = t^{-1} - t, \tag{3.49}$$

which contradicts Lemma 3.3. For the second case when $r < t$,

$$y \le t - r. \tag{3.50}$$

Inserting (3.47), (3.50) into (3.25), (3.26), we obtain

$$\frac{a\gamma(1-t^2)}{1-rt} \le 1 + \gamma a - \frac{t(a+1)}{r}, \qquad \frac{a(1-t^2)}{t-r} \le \frac{a+1}{r} - t(1+\gamma a). \tag{3.51}$$

Multiply the first inequality by $(1-rt)r$, multiply the second inequality by $(t-r)r$, and add the resulting two inequalities. Then after some algebra, we find $a \le 0$, which is a contradiction. This proves the lemma. □

It now follows from the preceding discussion that there is a unique solution ξ, $Im(\xi) > 0$, to (3.21), or equivalently (3.19).

3.2 The contour Γ_1

As in (3.6), set

$$z_0 := \frac{a+1}{-R(0)(1+\gamma a)}, \tag{3.52}$$

where $R(0)$ is given in (3.22). We emphasize that z_0 is uniquely determined by the endpoint ξ, and is independent of the curve Γ in subsection 3.1, as long as Γ does not intersect $(-\infty, 0]$, Note that $t < z_0 < t^{-1}$ from the condition $r_1 < r := -R(0) < r_2$ of (3.29).

Now define

$$Q(z) := -(1+\gamma a)^2 \frac{(z-\xi)(z-\overline{\xi})(z+z_0)^2}{z^2(z+t)^2(z+t^{-1})^2}, \tag{3.53}$$

and we consider the quadratic differential $Q(z)(dz)^2$. In this subsection, we are interested in the trajectories $Q(z)(dz)^2 > 0$. In the next subsection we consider the orthogonal trajectories $Q(z)(dz)^2 < 0$. A general reference for quadratic differentials is Chapter 8 of [19]. Note, in particular, that if two trajectories (or two orthogonal trajectories) meet at a point z where $Q(z)$ is analytic and nonzero, then they must be identical. The quadratic differential $Q(dz)^2$ has double poles at $0, -t, -t^{-1}$ and ∞, a double zero at $-z_0$ and simple zeros at $\xi, \overline{\xi}$. We first consider the local structure of the trajectories near the poles and the zeros.

3.2.1 Local structure

- Near 0: $Q(dz)^2 \sim \frac{-c^2}{z^2}(dz)^2$, where $c = (1+\gamma a)|\xi|z_0 > 0$. Thus we want trajectories $z = z(t)$ satisfying $\frac{ic}{z}\frac{dz}{dt} \sim c_1$ for some real constant c_1. The solution is given by $z \sim c_2 e^{-i(c_1/c)t}$, and hence the trajectories near $z = 0$ are circular.

- Near $-t$: $Q(dz)^2 \sim \frac{-c^2}{(z+t)^2}(dz)^2$, where $c = \frac{(1+\gamma a)|t+\xi||-t+z_0|}{t(t^{-1}-t)} > 0$. As in the $z = 0$ case, the trajectories are circular.

- Near $-t^{-1}$: $Q(dz)^2 \sim \frac{-c^2}{(z+t^{-1})^2}(dz)^2$, where $c = \frac{(1+\gamma a)|t^{-1}+\xi||-t^{-1}+z_0|}{t^{-1}(t^{-1}-t)} > 0$. Again, the trajectories are circular.

- Near ∞, $Q(dz)^2 \sim \frac{-c^2}{z^2}(dz)^2 = \frac{-c^2}{w^2}(dw)^2$, where $c = 1+\gamma a > 0$, $w = \frac{1}{z}$. Thus once again the trajectories are circular.

- Near $-z_0$: $Q(dz)^2 \sim -c^2(z+z_0)^2(dz)^2$, where $c = \frac{(1+\gamma a)|z+\xi|}{z_0(z_0-t)(t^{-1}-z_0)} > 0$. Thus we seek trajectories $z = z(t)$ satisfying $ic(z+z_0)\frac{dz}{dt} \sim c_1$ for some real constant c_1. The solution with $z(0) = -z_0$ is given by $(z+z_0)^2 \sim -i(2c_1/c)t$, thus we have $arg(z+z_0) = \frac{\pi}{4} + \frac{k\pi}{4}$, $k = 0,1,2,3$. Hence there are 4 trajectories starting from $-z_0$, all making an angle $\frac{\pi}{4}$ with the real line.

- Near ξ: $Q(dz)^2 \sim c(z-\xi)(dz)^2$ for some $c \in \mathbb{C}$. Thus we seek trajectories $z = z(t)$ satisfying $c\sqrt{z-\xi}\frac{dz}{dt} \sim c_1$ for some real constant c_1. The solution with $z(0) = \xi$ is given by $(z-\xi)^{3/2} \sim (c_1/c)t$, and hence $arg(z-\xi) = c_2 + \frac{2k\pi}{3}$, $k = 0,1,2$. Thus there are 3 trajectories starting from ξ, making an angle $\frac{2\pi}{3}$ between themselves.

- Near $\overline{\xi}$: $Q(dz)^2 \sim c(z-\overline{\xi})(dz)^2$ for some $c \in \mathbb{C}$. Again, there are 3 trajectories starting from $\overline{\xi}$, making an angle $\frac{2\pi}{3}$ between themselves.

The local structure of the trajectories of $Q(dz)^2$ near the poles and the zeros are summarized in Figure 2.

3.2.2 Global structure

Note that as $Q(z) = \overline{Q(\overline{z})}$, if $\{z(t) : \alpha < t < \beta\}$ is a trajectory of $Q(dz)^2 > 0$, then $\{\overline{z(t)} : \alpha < t < \beta\}$ is also a trajectory. Also from (3.18), we see that $Q(z) < 0$ for $z \in \mathbb{R} \setminus \{-z_0, 0, -t, -t^{-1}\}$, hence all trajectories that cross the real axis do so at $\pi/2$, and if $\{z(t) : \alpha < t < \beta\}$ is a trajectory that satisfies $Im(z(t)) > 0$ for $\alpha < t < \beta$, and $Im(z(\beta)) = 0$, $Re(z(\beta)) \neq -t^{-1}, -z_0, -t, 0$, then $\{\overline{z(2\beta - t)} : \beta < t < 2\beta - \alpha\}$ gives a smooth continuation of $\{z(t)\}$ into $Im(z) < 0$.

We need the following lemma.

Lemma 3.5. *Let $\xi, \overline{\xi}$ be as in subsection 3.1. Let Γ be a simple curve with endpoints $\xi, \overline{\xi}$ which does not intersect $(-\infty, 0]$ and is symmetric under reflection about the real axis. Choose the branch of $\sqrt{Q(z)}$ to be analytic in $\mathbb{C} \setminus \overline{\Gamma}$, and satisfy $\sqrt{Q(z)} \sim \frac{i(1+\gamma a)}{z}$ as $z \to \infty$. For any real number x satisfying*

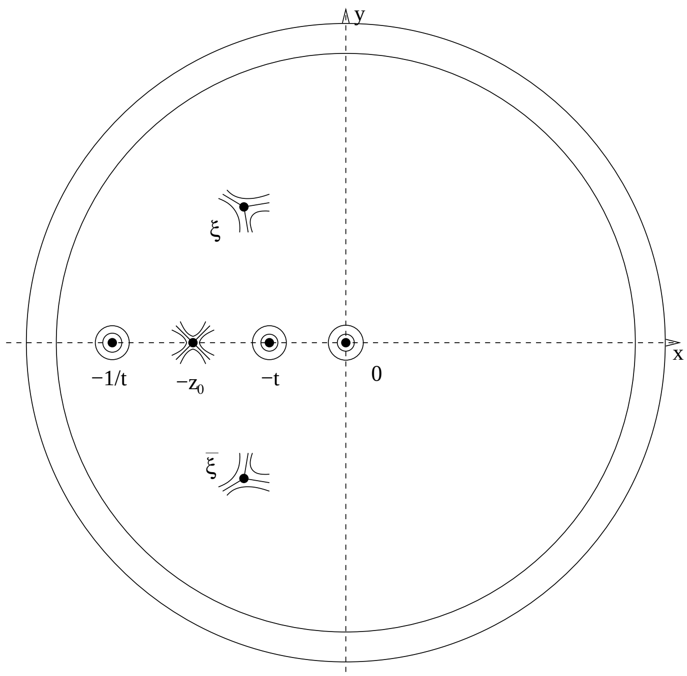

Figure 2: Local structure of the trajectories of $Q(dz)^2$ near the poles and the zeros.

$-t^{-1} < x < -t$, let C be a smooth curve in $\mathbb{C}_+$ with endpoints ξ and x, oriented from ξ to x, which does not intersect $\overline{\Gamma}$ (see Figure 3). Then we have

$$Re \int_C \sqrt{Q} dz = 0. \tag{3.54}$$

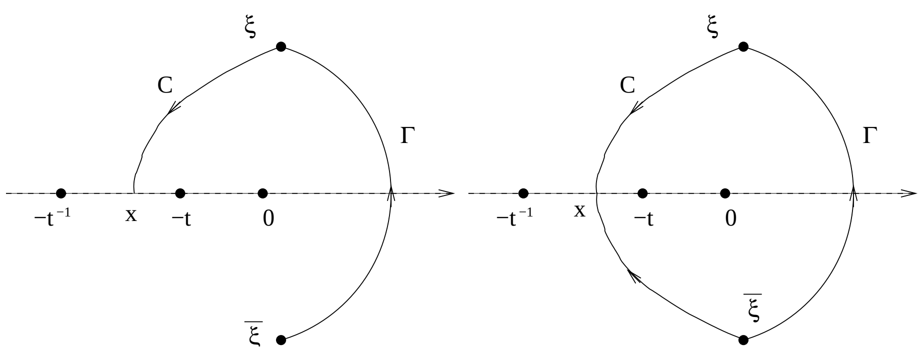

Figure 3: The contours C and C'.

Proof. Let C' be the closure of $C \cup \overline{C}$, oriented from ξ to $\overline{\xi}$. Hence C' is a curve which has endpoints $\xi, \overline{\xi}$, intersects the real axis at $-t^{-1} < x < -t$, and satisfies $C' = \overline{C'}$ (see Figure 3). Let $C^* = \{\overline{z} : z \in C\}$, oriented from $\overline{\xi}$ to x. Using $\sqrt{Q(z)} = -\overline{\sqrt{Q(\overline{z})}}$ and the realness of x, we have

$$
\begin{aligned}
Re \int_C \sqrt{Q(z)}dz &= \frac{1}{2}\Big[\int_C \sqrt{Q(z)}dz + \overline{\int_C \sqrt{Q(z)}dz}\Big] = \frac{1}{2}\Big[\int_C \sqrt{Q(z)}dz + \int_{C^*} \overline{\sqrt{Q(\overline{z})}}dz\Big] \\
&= \frac{1}{2}\Big[\int_C \sqrt{Q(z)}dz + \int_{-C^*} \sqrt{Q(z)}dz\Big] = \frac{1}{2}\int_{C'} \sqrt{Q(z)}dz.
\end{aligned}
\tag{3.55}
$$

Hence we want to prove that the last integral is 0.

Set (cf. (3.11), (3.12))

$$
\begin{aligned}
h(z) &= \frac{R(z)}{2\pi i}\int_\Gamma \frac{W'(s)}{R_+(s)(s-z)}ds \\
&= \frac{1}{2}W'(z) + \frac{1}{2}R(z)\Big(\frac{\gamma a}{(z+t^{-1})R(-t^{-1})} + \frac{a}{(z+t)R(-t)} - \frac{a+1}{zR(0)}\Big) \\
&= \frac{1}{2}W'(z) + \frac{1}{2}(1+\gamma a)\frac{R(z)(z+z_0)}{z(z+t)(z+t^{-1})},
\end{aligned}
\tag{3.56}
$$

where the second equality follows from a residue calculation, while the third equality follows from the endpoint conditions (3.21). Thus from the definition (3.53) of Q and the choice of $\sqrt{Q}$, we have $\sqrt{Q} = i(2h - W')$. Now

$$
\begin{aligned}
\frac{1}{i}\int_{C'} \sqrt{Q(z)}dz &= \int_{C'} (2h - W')dz = \int_{C'} \frac{R(z)}{\pi i}dz \int_\Gamma \frac{W'(s)}{R_+(s)(s-z)}ds - \int_{C'} W'(s)ds \\
&= \frac{1}{\pi i}\int_\Gamma \frac{W'(s)}{R_+(s)}ds \int_{C'} \frac{R(z)}{s-z}dz - \int_{C'} W'(s)ds.
\end{aligned}
\tag{3.57}
$$

By a residue calculation, for $s \in \Gamma$,

$$
\int_{C'} \frac{R(z)}{s-z}dz = \lim_{r\to\infty} \frac{1}{2}\int_{|z|=r} \frac{R(z)}{s-z}dz - \pi i R_+(s) = \pi i\Big(\frac{\xi+\overline{\xi}}{2} - s - R_+(s)\Big). \tag{3.58}
$$

Hence using the endpoints conditions (3.19), we have

$$
\begin{aligned}
\frac{1}{i}\int_{C'} \sqrt{Q(z)}dz &= \int_\Gamma \frac{W'(s)}{R_+(s)}\Big(\frac{\xi+\overline{\xi}}{2} - s - R_+(s)\Big)ds - \int_{C'} W'(s)ds \\
&= 2\pi i - \int_\Gamma W'(s)ds - \int_{C'} W'(s)ds \\
&= 2\pi i - \int_{\Gamma'} W'(s)ds,
\end{aligned}
\tag{3.59}
$$

where Γ', the closure of $\overline{\Gamma \cup C'}$, encloses $-t, 0$, but not $-t^{-1}$, and is oriented counter-clockwise. But a direct calculation shows $\int_{\Gamma'} W'(s)ds = 2\pi i$, and we obtain the lemma. □

Remark 5. The authors are indebted to Nick Ercolani who suggested that a formula such as (3.54) should be true.

The following general results for the trajectories of quadratic differentials are given in Lemmas 8.3, 8.4 of [19], and are used in several places in the arguments that follow.

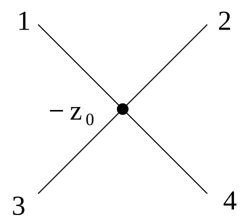

Figure 4: Local structure of the trajectories of $Q(dz)^2$ near $-z_0$.

Lemma 3.6. *Let $Q(z)(dz)^2$ be a quadratic differential in a simply connected domain G.*

(i). If there is at most one pole of $Q(z)$ in G and this pole is simple, then there is no closed Jordan curve in G consisting only of trajectories (or orthogonal trajectories) and their endpoints.

(ii). Suppose that $Q(z)$ has no poles in G and let Γ be a trajectory (or an orthogonal trajectory). Then in both directions, Γ ends at a zero of Q or converges to ∂G.

Remark 6. At various points in the argument that follows, we will use Lemma 3.6 (i) in a slightly extended form. For example, we will want to consider Jordan curves consisting of trajectories of the form in the first picture of Figure 5 for which the hypotheses of the Lemma are not fully satisfied. However, if we, for example, make a change of variables $z \mapsto \zeta = z^{1-\epsilon}$, $0 < \epsilon < 1$, then the figure takes the form as the second picture in Figure 5 and as the change of variable takes trajectories to trajectories, it is easy to verify that the hypotheses of Lemma 3.6 (i) are now satisfied. We will use this extended form of the Lemma without further comment below.

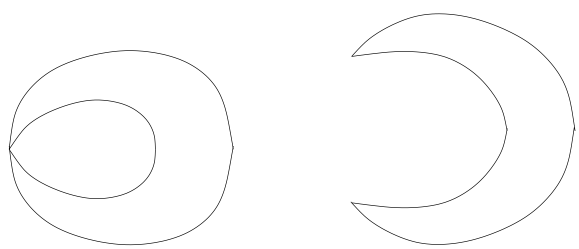

Figure 5: Unfolding of a "closed" trajectory meeting itself at a point.

Denote the four trajectories emerging from $-z_0$ by $1, 2, 3, 4$ as shown in Figure 4. First, consider the trajectory emerging from $-z_0$ along the ray 1. Since there are no poles in $\mathbb{C}_+$, which is a simply connected region, this trajectory must either go to ξ, the zero of Q in $\mathbb{C}_+$, or escape from $\mathbb{C}_+$. Suppose that the ray 1 does not escape from $\mathbb{C}_+$. Then from the definition of a trajectory, $\int_1 \sqrt{Q}dz \in \mathbb{R}\backslash\{0\}$, and hence the ray 1 (and similarly the ray 2) can not go to ξ due to Lemma 3.5. So the ray 1 must exit from $\mathbb{C}_+$. Now it can not exit through $-z_0$ because then by the local structure of the trajectories, the ray 1 comes back to z_0 through ray 2, and the interior of the loop has no poles of Q, contradicting Lemma 3.6 (i). Also, again by the local structure of the trajectories, the ray 1 can not exit through $-t^{-1}, -t, 0$

or ∞. Now there are five possibilities: the ray 1 exits from C_+ through $(-\infty, -t^{-1})$, $(-t^{-1}, -z_0)$, $(-z_0, -t)$, $(-t, 0)$, or $(0, \infty)$. We examine each case.

(i) The ray 1 can not exit through $(-t^{-1}, -z_0)$, for if it does, the trajectory can be continued by complex conjugation as remarked before, and the interior of the loop contains no poles of $Q(dz)^2$, contradicting Lemma 3.6 (i).

(ii) By the same argument, the ray can not exit through $(-z_0, -t)$.

(iii) Suppose the ray 1 exits at $z_1 \in (-t, 0)$. Then extending the ray by conjugation, we obtain the closed loop (see the first Picture in Figure 6). Now the trajectory along ray 2 from $-z_0$ also can not go to ξ and hence must exit $\mathbb{C}_+$. Also the ray 2 can not cross the trajectory emerging from

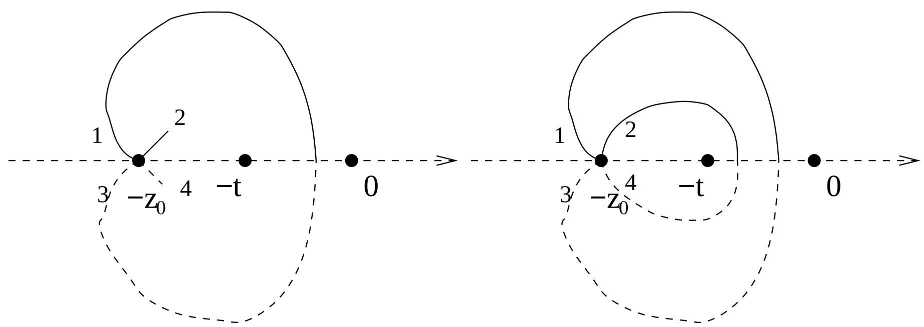

Figure 6: Ray 1, case (iii).

the ray 1. Hence it must cross the real axis between $(-z_0, -t)$ or $(-t, z_1)$. The first case can not hold as the closed loop obtained by extending the ray 2 by conjugation has no poles inside. In the latter case, the simply connected region formed by the two loops has no poles, which is again a contradiction (see the second picture in Figure 6). Thus ray 1 can not exit through $(-t, 0)$.

(iv) Suppose that the ray 1 exits at $z_2 \in (0, \infty)$. Then by arguing as in case (iii), the trajectory along the ray 2 from $-z_0$ must exit $\mathbb{C}_+$ as some point $z_3 \in (-t, 0)$. We denote the open regions in $\mathbb{C}_+$

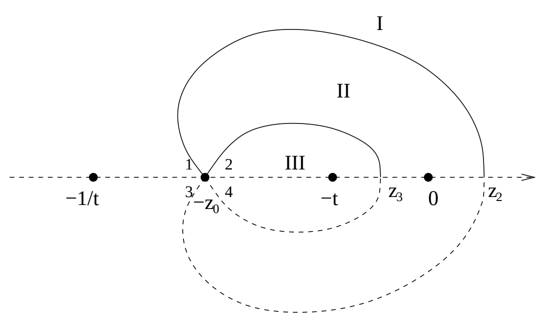

Figure 7: Ray 1, case (iv).

divided by the trajectories emerging from $-z_0$ by I, II, III as shown in Figure 7. Now the zero ξ

of Q lies in one of the regions I, II, III. (Note that ξ can not be on the trajectory emerging from $-z_0$ due to Lemma 3.5.)

(iv-1) If $\xi \in III$, then as before, each of the three trajectories emerging from ξ exit through $(-z_0, -t)$ or $(-t, z_3)$. Hence at least two of the trajectories exit through the same interval. Then the simply connected region bounded by these two trajectories has no pole, which is again a contradiction.

(iv-2) If $\xi \in II$, we reach a contradiction by a similar argument.

(iv-3) If $\xi \in I$, the three trajectories emerging from ξ must exit, one through $(-\infty, -t^{-1})$, one through $z_4 \in (-t^{-1}, -z_0)$, and one through (z_2, ∞). But this in turn gives a contradiction, because $Re \int_{\xi}^{z_4} \sqrt{Q} dz = 0$ by Lemma 3.5.

Hence in each case, we have a contradiction.

Thus we conclude that the trajectory along the ray 1 emerging from $-z_0$ exits $\mathbb{C}_+$ through $z_5 \in (-\infty, -t^{-1})$.

Now we consider the trajectory along the ray 2 emerging from $-z_0$. As before, it must exit $\mathbb{C}_+$ through either of $(-\infty, z_5)$, $(-z_0, -t)$, $(-t, 0)$ or $(0, \infty)$.

(i) If it exits through $(-\infty, z_5)$, the two trajectories from the rays 1 and 2, extended as in the remark above to $\mathbb{C}_+$, form a simply connected region which does not contain poles, this is again a contradiction.

(ii) Suppose that the trajectory along the ray 2 exits through $(-z_0, -t)$. Then the ($\mathbb{C}_-$-extended) loop of the trajectory 2 contains no poles, and again we have a contradiction.

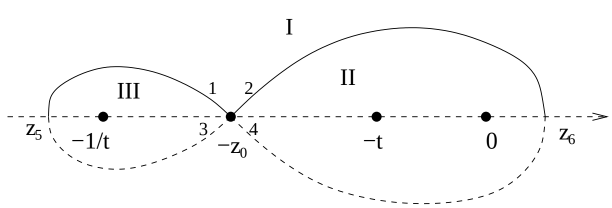

Figure 8: Ray 2, case (iii).

(iii) Suppose that the trajectory along the ray 2 exits through $z_6 \in (0, \infty)$. Then ξ lies either in the region I, II or III as of Figure 8. If $\xi \in I$, at least two of the trajectories from ξ exit together through $(-\infty, z_5)$ or (z_6, ∞). This yields a contradiction as in (iv-1) above. The case $\xi \in III$ leads to a similar contradiction. If $\xi \in II$, then the three trajectories exit, one through $(-z_0, -t)$, one through $(-t, 0)$, and one through $(0, z_6)$. But this gives in turn a contradiction by Lemma 3.5 as in the case (iv-3) above.

Therefore the trajectory along the ray 2 emerging from $-z_0$ must exit $\mathbb{C}_+$ through $z_7 \in (-t, 0)$.

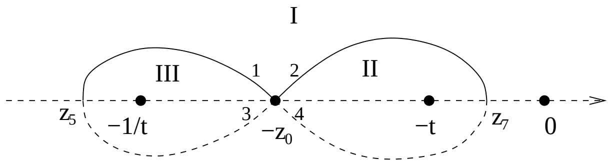

Figure 9: Rays 1 and 2.

Thus ξ lies either in regions I, II or III of Figure 9. But by a now familiar argument as above, ξ can not lie in II or III. Hence $\xi \in I$. Then the three trajectories emerging from ξ exit at some points z_8, z_9, z_{10} where $z_8 \in (-\infty, z_5), z_9 \in (z_7, 0), z_{10} \in (0, \infty)$. This shows that the global structure of the trajectories of $Q(dz)^2$ is given in Figure 10.

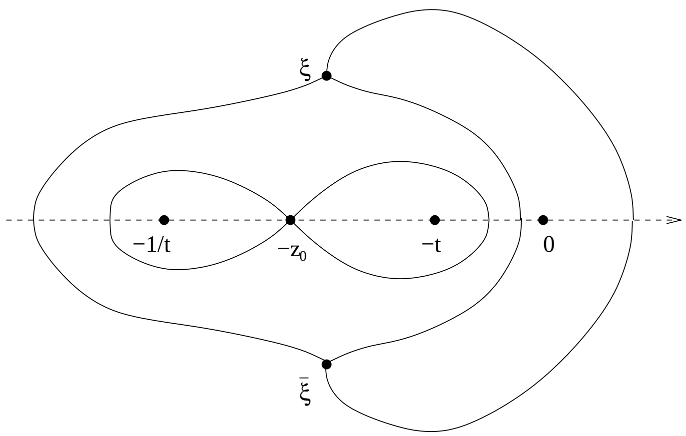

Figure 10: Global structure of the trajectories of $Q(dz)^2$.

The above considerations prove, in particular, that there is a trajectory emerging from ξ and ending at $\overline{\xi}$ which crosses the real axis to the right of 0. We take Γ_1 to be this curve, oriented from $\overline{\xi}$ to ξ. Define h by (3.56) with the choice of the contour $\Gamma = \Gamma_1$: let $R(z) = \sqrt{(z-\xi)(z-\overline{\xi})}$ be analytic in $\mathbb{C} \setminus \Gamma_1$, $R(z) \sim z$ as $z \to \infty$. Thus

$$\begin{aligned} h(z) &= \frac{R(z)}{2\pi i} \int_{\Gamma_1} \frac{W'(s)}{R_+(s)(s-z)} ds \\ &= \frac{1}{2} W'(z) + \frac{1}{2} R(z) \left(\frac{\gamma a}{(z+t^{-1}) R(-t^{-1})} + \frac{a}{(z+t) R(-t)} - \frac{a+1}{z R(0)} \right) \\ &= \frac{1}{2} W'(z) + \frac{1}{2} (1+\gamma a) \frac{R(z)(z+z_0)}{z(z+t)(z+t^{-1})}. \end{aligned} \tag{3.60}$$

Condition (a) of Proposition 3.1, $h_+(z) + h_-(z) = W'(z)$ for $z \in \Gamma_1$, now follows by properties of the Cauchy operator, and the condition (b) by the endpoint condition (3.19).

Now we consider condition (c). Define $\sqrt{Q(z)} = i(1+\gamma a) \frac{R(z)(z+z_0)}{z(z+t)(z+t^{-1})}$, so that it is analytic in

$\mathbb{C} \setminus \overline{\Gamma_1}$, and $\sqrt{Q} \sim \frac{i(1+\gamma a)}{z}$ as $z \to \infty$. Then

$$i(h_+(z) - h_-(z)) = \sqrt{Q(z)}_+, \qquad z \in \Gamma_1. \tag{3.61}$$

Along the trajectory Γ_1, we must have $\sqrt{Q}dz \in \mathbb{R} \setminus \{0\}$. Let $z_{10} \in (0, \infty)$ be the point at which Γ_1 crosses the real axis. As $R(z) = \sqrt{(z-\xi)(z-\overline{\xi})} > 0$ for $z \in (z_{10}, \infty)$, it follows that for all upward-oriented trajectories that cross the real axis at points $x \in (z_{10}, \infty)$, $\sqrt{Q}dz|_{z=x} < 0$. It particular, for Γ_1, $\sqrt{Q}_- dz|_{z_{10}} < 0$, and hence $\sqrt{Q(z)}_- dz < 0$ for all $z \in \Gamma_1$. Now since $\sqrt{Q}_+ = -\sqrt{Q}_-$ on Γ_1, we have

$$\sqrt{Q}_+ dz \in \mathbb{R}_+, \quad \text{on } \Gamma_1, \tag{3.62}$$

and hence from (3.61),

$$i(h_+(z) - h_-(z))dz > 0, \quad z \in \Gamma_1, \tag{3.63}$$

which proves condition (c).

3.3 The contour Γ_2

Again we choose the branch of $\sqrt{Q}$ so that $\sqrt{Q}$ is analytic in $\mathbb{C} \setminus \Gamma_1$ and $\sqrt{Q} \sim \frac{i(1+\gamma a)}{z}$ as $z \to \infty$. Now we consider the orthogonal trajectories $Q(z)(dz)^2 < 0$. As before, the local structure is easy to determine. We summarize the local structure of the orthogonal trajectories near the finite poles and the zeros of $Q(dz)^2$ in Figure 11. Near ∞, any straight rays emerging from ∞ are orthogonal trajectories. We note that the real axis is an orthogonal trajectory. Hence the orthogonal trajectories can cross the

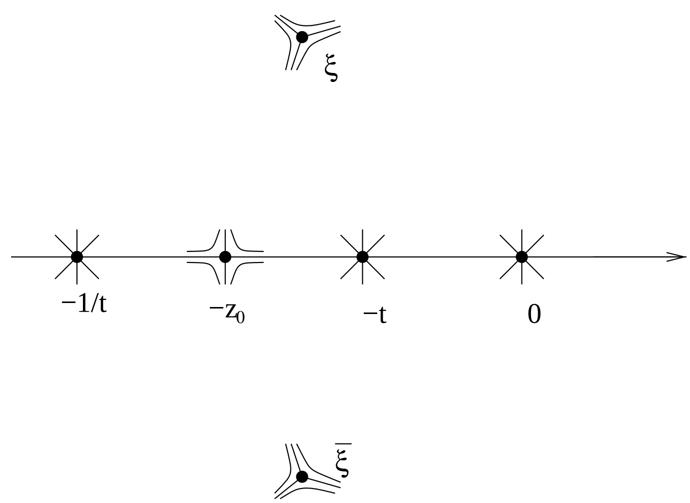

Figure 11: Local structure of the orthogonal trajectories near the finite poles and zeros of $Q(dz)^2$.

real axis only at $0, -t, z_0, -t^{-1}$ or ∞.

Now we consider the global structure. Again by the symmetry $Q(z) = \overline{Q(\overline{z})}$, the orthogonal trajectories are symmetric under reflection about the real axis. There are three orthogonal trajectories, denoted by $1, 2, 3$, emerging from ξ, each of which bisects the angle between two trajectories emerging from ξ (see Figure 12). Now we show that the orthogonal trajectory 1 can not cross either of the two

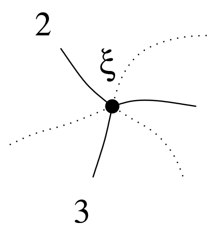

Figure 12: Orthogonal trajectories and trajectories emerging from ξ.

adjacent trajectories emerging from ξ. If so, there is a closed loop of the form shown the Figure 13 consisting of a part of a trajectory and a part of an orthogonal trajectory, and by analyticity of $\sqrt{Q}$, the integral of $\sqrt{Q}dz$ is zero around the loop. But the integral of $\sqrt{Q}dz$ along the trajectory is in

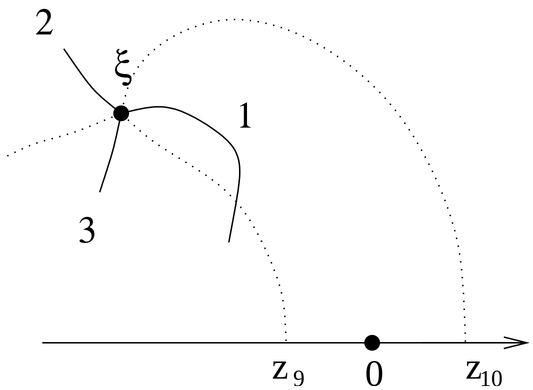

Figure 13: Orthogonal trajectories emerging from ξ.

$\mathbb{R} \setminus \{0\}$, and along the orthogonal trajectory, the integral of $\sqrt{Q}dz$ along the orthogonal trajectory is in $i\mathbb{R} \setminus \{0\}$. Hence the sum can not be zero, which is a contradiction. Therefore the orthogonal trajectory 1 must exit $\mathbb{C}_+$ between z_9 and z_{10}. But then from the local structure, it must exit at 0. Similarly, the orthogonal trajectory 2 must go to ∞. The orthogonal trajectory 3, by a similar argument, must exit $\mathbb{C}_+$ either through $-t^{-1}, z_0$ or $-t$. Suppose it exits through $-t$. By the local structure, the orthogonal trajectory approaches along an angle; indeed it is easy to show that $z(s) \sim -t + e^{i\phi}e^{-cs}$ as $s \to \infty$ for some $0 < \phi < \pi$, $c > 0$. Also by the definition of an orthogonal trajectory, we have

$$\lim_{s\to\infty} Re \int_{\xi}^{z(s)} \sqrt{Q(z)}dz = 0 \tag{3.64}$$

along the orthogonal trajectory 3. For s large, but fixed, write $z(s) := -t + \epsilon e^{i\theta}$ for $\epsilon > 0$, $0 < \theta < \pi$, and consider the curve C_0 from ξ to $z(s)$ along the orthogonal trajectory 3. Let C_s be the curve $\{-t + \epsilon e^{i\beta} : \theta \le \beta \le \pi\}$, oriented from $z(s)$ to $-t - \epsilon$. Thus $C_0 \cup C_s$ is a curve from ξ to a point in

$(-t^{-1}, -t)$. Then by Lemma 3.5, we have $\int_{C_0 \cup C_s} Re\sqrt{Q}dz = 0$. Hence we have

$$\begin{aligned}\lim_{s\to\infty} Re \int_{C_0} \sqrt{Q}dz &= \lim_{s\to\infty} Re \int_{C_0\cup C_s} \sqrt{Q}dz - Re\int_{C_s} \sqrt{Q}dz \\ &= -\lim_{s\to\infty} Re \int_{C_s} \sqrt{Q}dz \\ &= -\lim_{s\to\infty} Re \int_{C_s} i(1+\gamma a)\frac{R(z)(z+z_0)}{z(z+t^{-1})}\frac{dz}{z+t} \\ &= (1+\gamma a)\frac{R(-t)(-t+z_0)}{-t(-t+t^{-1})}(\pi-\theta) > 0.\end{aligned} \tag{3.65}$$

This contradicts (3.64), and hence the trajectory 3 cannot exit $\mathbb{C}_+$ through $-t$. Similar argument shows that it cannot exit $\mathbb{C}_+$ through $-t^{-1}$. Hence the trajectory 3 exits $\mathbb{C}_+$ through $-z_0$. Thus we have obtained the global structure for the orthogonal trajectories of $Q(dz)^2$ emerging from the zeros ξ, $\overline{\xi}$, $-z_0$ as shown by the solid curves in the Figure 14. The dotted curves in the Figure 14 denote the trajectories already displayed in Figure 10.

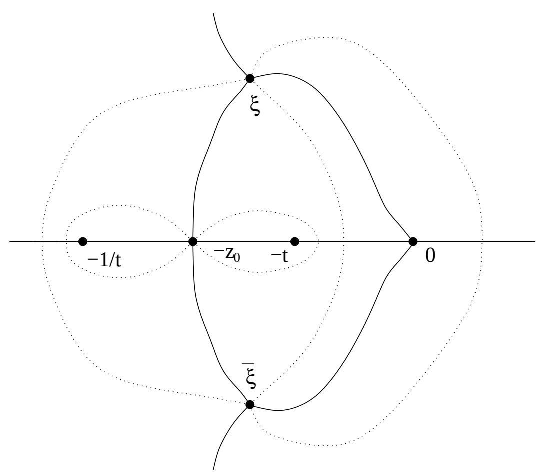

Figure 14: Global structure for the orthogonal trajectories and the trajectories emerging from the zeros ξ, $\overline{\xi}$ and $-z_0$.

In particular, we have shown that there is an orthogonal trajectory (more precisely, a union of two orthogonal trajectories, one in $\mathbb{C}_+$ and the other, its conjugate, in $\mathbb{C}_-$, meeting at the point $-z_0$ which is a zero of $Q(dz)^2$) emerging from ξ and ending at $\overline{\xi}$ and which crosses the real axis at $-z_0$. We denote the curve by Γ_2, and take the orientation from ξ to $\overline{\xi}$. Since Γ_2 is a union of two orthogonal trajectories, we have $\sqrt{Q}dz \in i\mathbb{R}\setminus\{0\}$ for $z \in \Gamma_2\setminus\{0\}$ and an explicit computation using $\sqrt{Q(z)} = i(1+\gamma a)\frac{R(z)(z+z_0)}{z(z+t)(z+t^{-1})}$ shows that

$$\sqrt{Q}dz \in i\mathbb{R}_-, \qquad \text{on } \Gamma_2 \cap \mathbb{C}_+ \tag{3.66}$$

$$\sqrt{Q}dz \in i\mathbb{R}_+, \qquad \text{on } \Gamma_2 \cap \mathbb{C}_- \tag{3.67}$$

with the orientation from ξ to $\overline{\xi}$. From (3.60), $i(2h - W') = \sqrt{Q}$. Thus we have

$$(2h - W')dz < 0, \quad z \in \Gamma_2 \cap \mathbb{C}_+, \tag{3.68}$$

$$(2h - W')dz > 0, \quad z \in \Gamma_2 \cap \mathbb{C}_-. \tag{3.69}$$

This proves condition (d). The reader will observe that the remaining conditions and formulae in Proposition 3.1 have been proved en route in this section, and this completes the proof of the Proposition.

These facts can be illustrated by numerical computations of trajectories and orthogonal trajectories associated with the quadratic differential $Q(z)\,(dz)^2$. The (orthogonal) trajectories can be obtained simply with a Runge-Kutta scheme. For (orthogonal) trajectories that emerge from zeros or poles of $Q(z)$, which amount to singularities of the vector field in the complex plane, the only additional difficulty is to determine the initial direction, which is not unique. But the possible initial directions are easily determined by the sort of local analysis that has already been presented above. An example of the results of such a calculation is presented in Figure 15.

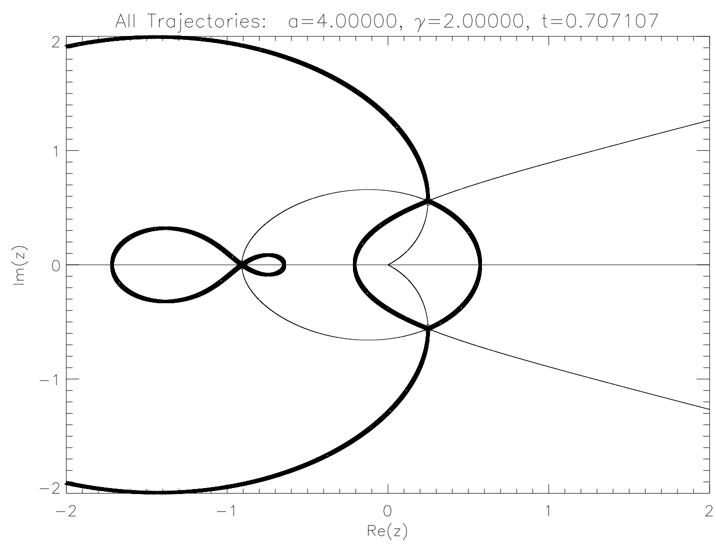

Figure 15: The trajectories and orthogonal trajectories of the quadratic differential $Q(z)\,(dz)^2$ emerging from the points $z = \xi$, $z = \overline{\xi}$, and $z = -z_0$. The parameter values are $a = 4$, $\gamma = 2$, and $t = 1/\sqrt{2}$. The trajectories where $Q(z)\,(dz)^2 > 0$ are shown with thick curves and the orthogonal trajectories where $Q(z)\,(dz)^2 < 0$ are shown with thin curves.

A computer program for generating numerical approximations to the contours Γ_1 and Γ_2 is of course useful because it allows one to explore/illustrate the dependence of the contours on the parameters a, γ, and t. As an example, Figure 16 illustrates the deformation of the contours as a is varied while γ and t are held fixed. Here, we see clearly what happens as a is decreased to a_0, its minimum value which depends on γ and t. Namely, the two endpoints converge to a common point on the negative real axis, and at the same time, the contour Γ_1 closes without collapsing, while Γ_2 disappears. Similarly,

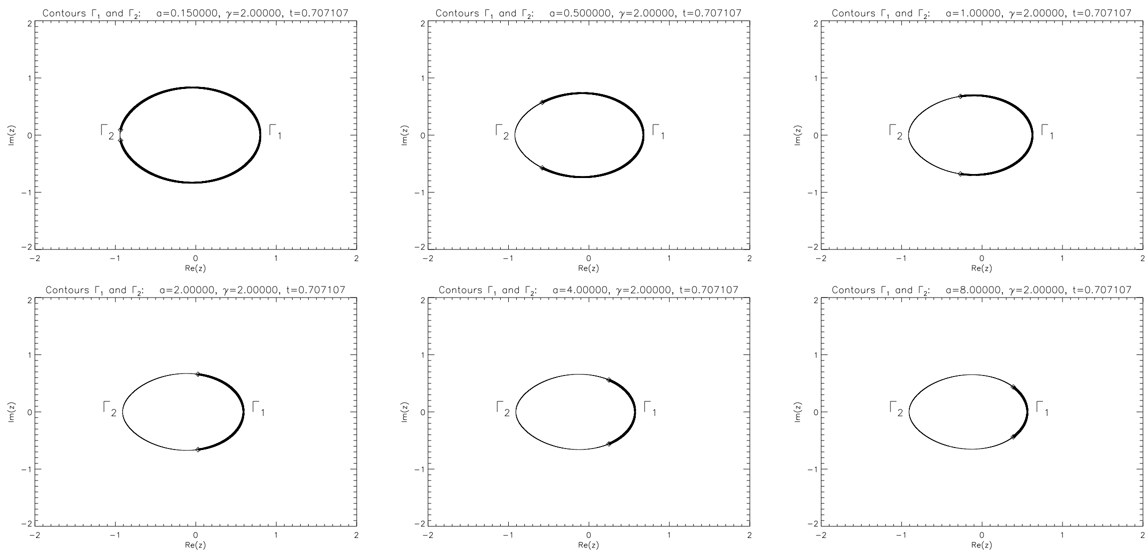

Figure 16: The dependence of the contours Γ_1 and Γ_2 on the parameter a. As a tends to a_0 from above, the endpoints ξ and $\overline{\xi}$ coalesce on the negative real axis.

as a increases without bound, the opposite situation prevails, with the endpoints ξ and $\overline{\xi}$ coalescing on the positive real axis, while Γ_2 closes without collapsing and Γ_1 disappears. It is also possible to see in these pictures that the closed curve $\Gamma_1 \cup \Gamma_2$ deforms somewhat throughout this process; the endpoints ξ and $\overline{\xi}$ do not simply slide along a fixed closed curve as a is varied.

4 The g-function

Proposition 4.1. *Fix $0 < t < 1$, $\gamma \geq 1$, $a > a_0$. Let ξ, Γ_1, Γ_2 be as in the Proposition 3.1. Let $\ln z$ denote the principal branch of logarithm, $\ln z \in \mathbb{R}$ for $z > 0$, and set*

$$g(z) := \int_\infty^z \Big(h(s) - \frac{1}{s}\Big) ds + \ln z, \qquad z \in \mathbb{C} \setminus (\overline{\Gamma_1} \cup (-\infty, p_i]), \tag{4.1}$$

where $p_i > 0$ is the intersection point of Γ_1 and $\mathbb{R}$, and the integral is taken over a curve from ∞ to z which does not intersect $\overline{\Gamma_1} \cup (-\infty, p_i]$. Let $\Phi(z)$ be as in (3.5) of Proposition 3.1. Also set

$$\ell := 2g(\xi) - W(\xi). \tag{4.2}$$

Then g and ℓ satisfy the following properties:

(1) $g(z)$ is well-defined and analytic in $\mathbb{C} \setminus (\overline{\Gamma_1} \cup (-\infty, p_i])$, and $e^{g(z)}$ is analytic in $\mathbb{C} \setminus \overline{\Gamma_1}$.

(2) $e^{g(z)} = z(1 + O(1/z))$ as $z \to \infty$.

(3) $e^{g_+(z)+g_-(z)-W(z)-\ell} = 1$ for $z \in \Gamma_1$.

(4) $e^{g_+(z)-g_-(z)} = e^{\Psi_{1+}(z)}$ *for* $z \in \Gamma_1$*, where* $e^{\Psi_1(z)} = \exp\{\int_\xi^z \Phi(s)ds\}$*,* $z \in \mathbb{C} \setminus \overline{\Gamma_1 \cup \mathbb{R}_-}$ *with the integral taken along any curve not intersecting* $\overline{\Gamma_1 \cup \mathbb{R}_-}$*.*

(5) $e^{2g(z)-W(z)-\ell} = e^{\Psi_2(z)}$ *for* $z \in \Gamma_2$*, where* $\Psi_2(z) = \int_\xi^z \Phi(s)ds$ *with the integral taken along* Γ_2*.*

Remark 7. It follows from (3.3) that $h(z) = \overline{h(\overline{z})}$ for $z \in \mathbb{C} \setminus \overline{\Gamma_1}$, and hence from (4.1), we see that $g(z) = \overline{g(\overline{z})}$ for $z \in \mathbb{C} \setminus (\overline{\Gamma_1} \cup (-\infty, p_i))$.

Proof. Since h is analytic in $\mathbb{C} \setminus \overline{\Gamma_1}$ and continuous up to the boundary, we have g given by (4.1) is well-defined and analytic in $\mathbb{C} \setminus \overline{\Gamma_1} \cup (-\infty, p_i])$, and continuous up to the boundary. Now let C be a simple closed curve, oriented counter-clockwise enclosing Γ_1. Using the formula (3.3) for h,

$$\int_C h(z)dz = \int_C \frac{R(z)}{2\pi i} dz \int_{\Gamma_1} \frac{W'(s)}{R_+(s)(s-z)} ds = \frac{1}{2\pi i} \int_{\Gamma_1} \frac{W'(s)}{R_+(s)} ds \int_C \frac{R(z)}{s-z} dz. \tag{4.3}$$

Using the residue at infinity, we then have

$$\frac{1}{2\pi i} \int_C \frac{R(z)}{s-z} dz = s - \frac{\xi + \overline{\xi}}{2}, \tag{4.4}$$

and

$$\int_C h(z)dz = \int_{\Gamma_1} \frac{W'(s)}{R_+(s)} \big(s - \frac{\xi + \overline{\xi}}{2}\big) ds = -2\pi i \tag{4.5}$$

from the endpoint conditions (3.4). Now from the definition of g, for $z \in (-\infty, p_i)$, $g_+(z) - g_-(z) = \int_C h(s)ds = -2\pi i$, where $g_\pm(z) = \lim_{\epsilon \downarrow 0} g(z \pm \epsilon i)$. Therefore $e^{g(z)}$ is analytic in $\mathbb{C} \setminus \overline{\Gamma_1}$ and continuous up to the boundary. This proves property (1).

Since $h(z) = \frac{1}{z} + O(z^{-2})$ as $z \to \infty$, we have $g(z) = \ln z + O(z^{-1})$ as $z \to \infty$, which proves property (2).

For $z \in \Gamma_1 \cap \mathbb{C}_+$,

$$g_\pm(z) = g(\xi) + \int_\xi^z \big(h_\pm(s) - \frac{1}{s}\big) ds + \ln z - \ln \xi = g(\xi) + \int_\xi^z h_\pm(s) ds, \tag{4.6}$$

where the integral from ξ to z is taken along Γ_1. Hence from $h_+(z) + h_-(z) = W'(z)$, $z \in \Gamma_1$, we have, using (4.2),

$$g_+(z) + g_-(z) = 2g(\xi) + \int_\xi^z W'(s) ds = W(z) + \ell, \tag{4.7}$$

for $z \in \Gamma_1 \cap \mathbb{C}_+$. For $z \in \Gamma_1 \cap \mathbb{C}_-$,

$$g_\pm(z) = g(\overline{\xi}) + \int_{\overline{\xi}}^z h_\pm(s) ds = g(\overline{\xi}) + \int_{\overline{\xi}}^{\xi} h_\pm(s) ds + \int_\xi^z h_\pm(s) ds, \tag{4.8}$$

where the integrals are again taken along Γ_1. But from (4.5) (recall that h is analytic in $\mathbb{C} \setminus \overline{\Gamma_1}$), $\int_{\overline{\xi}}^{\xi} (h_+(s) - h_-(s)) ds = 2\pi i$ and $g(\overline{\xi}) - g(\xi) = \int_\xi^{\overline{\xi}} h_-(s) ds$. Hence for $z \in \Gamma_1 \cap \mathbb{C}_-$,

$$g_\pm(z) = g(\xi) - \int_{\overline{\xi}}^{\xi} h_-(s) ds + \int_{\overline{\xi}}^{\xi} h_\pm(s) ds + \int_\xi^z h_\pm(s) ds, \tag{4.9}$$

and hence

$$g_+(z) + g_-(z) = 2g(\xi) + \int_{\overline{\xi}}^{\xi} (h_+(s) - h_-(s))ds + \int_{\xi}^{z} W'(s)ds = W(z) + \ell + 2\pi i. \tag{4.10}$$

Therefore, since e^{g_+}, e^{g_-} and W are continuous for $z \in \Gamma_1$, we have $e^{g_+(z)+g_-(z)-W(z)-\ell} = 1$ for all $z \in \Gamma_1$, which verifies property (3).

From (3.8), (4.6), (4.9), and the above relation $\int_{\overline{\xi}}^{\xi}(h_+(s) - h_-(s))ds = 2\pi i$, we have for $z \in \Gamma_1$,

$$e^{g_+(z)-g_-(z)} = \exp\Big\{\int_{\xi}^{z} (h_+(s) - h_-(s))ds\Big\} = e^{\int_\xi^z \Phi(s)_+ ds}, \tag{4.11}$$

and hence the property (4) follows if we prove that $\exp\{\int_\xi^z \Phi(s)ds\}$, $z \in \mathbb{C} \setminus \overline{\Gamma_1 \cup \mathbb{R}_-}$ does not depend on the choice of the integration path. For this purpose, it is enough to prove that

$$\int_C \Phi(z)dz \in 2\pi i\mathbb{Z} \tag{4.12}$$

for any simple closed contour C which encloses Γ_1 and does not intersect $(-\infty, 0]$. From (3.7) and the fact that $W'(z)$ is analytic away from $-t^{-1}, -t, 0$, we have

$$\int_C \Phi(z)dz = \int_C (2h(z) - W'(z))dz = 2\int_C h(z)dz. \tag{4.13}$$

Using (3.3) for h, the above integral is equal to

$$\int_C \frac{R(z)}{\pi i}dz \int_{\Gamma_1} \frac{W'(s)}{R_+(s)(s-z)}ds = \frac{1}{\pi i}\int_{\Gamma_1} \frac{W'(s)}{R_+(s)}ds \int_C \frac{R(z)}{s-z}dz = 2\int_{\Gamma_1} \frac{W'(s)}{R_+(s)}\Big(\frac{\xi + \overline{\xi}}{2} - s\Big)ds. \tag{4.14}$$

But this is equal to $4\pi i$ from the endpoint condition (3.4), and so (4.12) is established.

For $z \in \Gamma_2 \cap \mathbb{C}_+$, as in (4.6)

$$g(z) = g(\xi) + \int_\xi^z h(s)ds \tag{4.15}$$

and

$$2g(z) - W(z) - \ell = 2\int_\xi^z h(s)ds - W(z) + W(\xi) = \int_\xi^z (2h(s) - W'(s))ds. \tag{4.16}$$

For $z \in \Gamma_2 \cap \mathbb{C}_-$, as in (4.9),

$$g(z) = g(\overline{\xi}) + \int_{\overline{\xi}}^z h(s)ds = g(\xi) - \int_{\overline{\xi}}^{\xi} h_-(s)ds + \int_{\overline{\xi}}^{\xi} h_+(s) + \int_\xi^z h(s)ds = g(\xi) + \int_\xi^z h(s)ds + 2\pi i, \tag{4.17}$$

and hence for $z \in \Gamma_2$,

$$e^{2g(z)-W(z)-\ell} = \exp\Big\{\int_\xi^z (2h(s) - W'(s))ds\Big\}, \tag{4.18}$$

and property (5) follows from (3.9). This completes the proof of Proposition 4.1. ∎

5 RHP analysis

Set

$$\Gamma = \overline{\Gamma_1 \cup \Gamma_2}, \tag{5.1}$$

oriented counter-clockwise. It is a simple closed curve which has 0 and $-t$ inside and $-t^{-1}$ outside. Since the jump matrix $V_Y = \left(\begin{smallmatrix} 1 & z^{-k}\varphi(z) \\ 0 & 1 \end{smallmatrix}\right)$ for Y in (2.6) is analytic in $\mathbb{C}\setminus\{0\}$, we can deform the contour Σ for Y to Γ, as follows. Set

$$\widetilde{Y}(z) = \begin{cases} Y(z) & \text{for } z \text{ inside both } \Gamma \text{ and } \Sigma, \text{ and for } z \text{ outside both } \Gamma \text{ and } \Sigma, \\ Y(z)V_Y(z) & \text{for } z \text{ inside } \Gamma \text{ and outside } \Sigma, \\ Y(z)V_Y^{-1}(z) & \text{for } z \text{ outside } \Gamma \text{ and inside } \Sigma. \end{cases} \tag{5.2}$$

Then $\widetilde{Y}$ is analytic in $\mathbb{C}\setminus\Gamma$ and continuous up to the boundaries, satisfies $\widetilde{Y}_+(z) = \widetilde{Y}_- V_Y(z)$ for $z \in \Gamma$, and $\widetilde{Y}(z)z^{-k\sigma_3} = I + O(z^{-1})$ as $z \to \infty$.

Now (see the Introduction) we define

$$M(z) = e^{-\frac{1}{2}\ell k\sigma_3}\widetilde{Y}(z)e^{-(g(z)-\frac{1}{2}\ell)k\sigma_3}. \tag{5.3}$$

Then from Proposition 4.1 (1), (2), M satisfies

$$\begin{cases} M(z) \text{ is analytic in } z \in \mathbb{C}\setminus\Gamma, \\ M_+(z) = M_-(z)\begin{pmatrix} e^{-k(g_+(z)-g_-(z))} & e^{k(g_+(z)+g_-(z)-W(z)-\ell)} \\ 0 & e^{k(g_+(z)-g_-(z))} \end{pmatrix}, \qquad z \in \Gamma, \\ M(z) = I + O(1/z), \qquad \text{as } z \to \infty. \end{cases} \tag{5.4}$$

From the Proposition 4.1 (3), (4), the jump matrix V for M is now

$$V(z) = \begin{pmatrix} e^{-k\Psi_{1+}(z)} & 1 \\ 0 & e^{k\Psi_{1+}(z)} \end{pmatrix}, \qquad z \in \Gamma_1, \tag{5.5}$$

and from the Proposition 4.1 (1), (5), we have

$$V(z) = \begin{pmatrix} 1 & e^{k\Psi_2(z)} \\ 0 & 1 \end{pmatrix}, \qquad z \in \Gamma_2. \tag{5.6}$$

For the jump matrix on $z \in \Gamma_1$, note that $\Psi_+(z) = -\Psi(z)_-$ and

$$V(z) = \begin{pmatrix} e^{-k\Psi_{1+}(z)} & 1 \\ 0 & e^{-k\Psi_{1-}(z)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ e^{-k\Psi_{1-}(z)} & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ e^{-k\Psi_{1+}(z)} & 1 \end{pmatrix}. \tag{5.7}$$

Clearly Ψ_{1+} has an analytic continuation to the (+)-side of the contour Γ_1. Now for $z \in \Gamma_1$, it is easy to see that $Re(\Psi_{1_+}(z)) = 0$, and hence from the Proposition 3.1 (c) and (3.8), the derivative

$\frac{d}{dz}\Psi_{1+}(z) = \Phi_+(z)$ along the contour Γ_1 satisfies $Im(\frac{d}{dz}\Psi_{1+}(z)) < 0$, $z \in \Gamma_1$. Thus the Cauchy-Riemann condition for the analyticity implies that $Re(\Psi_1(z)) > 0$ for z on the $(+)$-side of Γ_1 and close to the contour. Therefore we can take a contour $\Gamma_1^{(1)}$ with endpoints ξ, $\overline{\xi}$ for which $Re(\Psi(z)) > 0$ for $z \in int(\Gamma^{(1)})$. Similarly, Ψ_{1-} has an analytic continuation to the $(-)$-side of Γ_1 and its real part is positive for z on the $(+)$-side of Γ_1 close to the contour Γ_1 and we take a contour $\Gamma_1^{(2)}$ for which $Re(\Psi(z)) > 0$ on its interior. We take the orientation of $\Gamma_1^{(j)}$, $j = 1, 2$ to be from $\overline{\xi}$ to ξ. See Figure 17 for the general shape of the contours $\Gamma_1^{(1)}$, $\Gamma_1^{(2)}$.

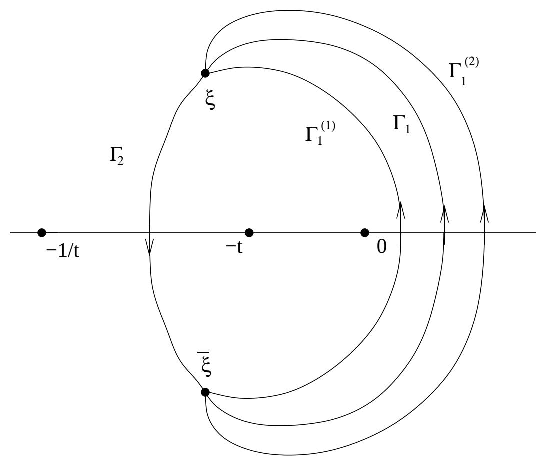

Figure 17: The contours $\Gamma_1^{(1)}$ and $\Gamma_1^{(2)}$.

Define $\widetilde{M}(z)$ to be $M(z)$ for z in the region bounded by Γ_2 and $\Gamma_1^{(1)}$ and also in the unbounded region. For the region bounded by Γ_1 and $\Gamma_1^{(1)}$, define

$$\widetilde{M} = M \begin{pmatrix} 1 & 0 \\ -e^{-k\Psi_1(z)} & 1 \end{pmatrix}, \tag{5.8}$$

and for the region bounded by Γ_1 and $\Gamma_1^{(2)}$, define

$$\widetilde{M} = M \begin{pmatrix} 1 & 0 \\ e^{-k\Psi_1(z)} & 1 \end{pmatrix}. \tag{5.9}$$

Set $\Gamma' = \overline{\Gamma_1 \cup \Gamma_1^{(1)} \cup \Gamma_1^{(2)} \cup \Gamma_2}$. Then $\widetilde{M}$ satisfies the new RHP

$$\begin{cases} \widetilde{M}(z) \text{ is analytic in } z \in \mathbb{C} \setminus \Gamma', \\ \widetilde{M}_+(z) = \widetilde{M}_-(z)\widetilde{V}(z) \qquad z \in \Gamma', \\ \widetilde{M}(z) = I + O(1/z), \qquad \text{as } z \to \infty, \end{cases} \tag{5.10}$$

where the jump matrix $\widetilde{V}$ is

$$\widetilde{V}(z) = \begin{cases} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & z \in \Gamma_1, \\ \begin{pmatrix} 1 & 0 \\ e^{-k\Psi_1(z)} & 1 \end{pmatrix}, & z \in \Gamma_1^{(1)} \cup \Gamma_1^{(2)}, \\ \begin{pmatrix} 1 & e^{k\Psi_2(z)} \\ 0 & 1 \end{pmatrix}, & z \in \Gamma_2. \end{cases} \tag{5.11}$$

Now we take the limit $k \to \infty$ with $\gamma \geq 1$ in a compact set and $a_0 < a \leq a^*$, for some a^*. From the signature of $Re(\Psi_1(z))$ on $\Gamma_1^{(j)}$, $j = 1, 2$, we see that the jump matrix $\widetilde{V} \to I$. For $z \in \Gamma_2$, we have $\Psi_2(z) < 0$ from the Proposition 4.1 and 3.1. Indeed $\Psi_2(z)$ is decreasing as z follows from ξ to $-z_0$ along Γ_2, and then increasing as z follows from $-z_0$ to $\overline{\xi}$ along Γ_2. On the other hand, $\Psi_2(\xi) = 0$ and $\Psi_2(\overline{\xi}) = \int_{\xi}^{\overline{\xi}} \Phi(s)ds = 0$, as $\Phi(s)ds = \overline{\Phi(s)ds} = -\Phi(\overline{s})d\overline{s}$, and the negativity of $\Psi_2(z)$ on Γ_2 follows. Hence as $k \to \infty$, $\widetilde{V} \to I$ on Γ_2. Therefore we have $\widetilde{V} \to V^\infty$ where

$$V^\infty = \begin{cases} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & z \in \Gamma_1, \\ I & z \in \Gamma_1^{(1)} \cup \Gamma_1^{(2)} \cup \Gamma_2. \end{cases} \tag{5.12}$$

Let M^∞ be the solution to the RHP with the jump matrix V^∞ and normalized at infinity. The solution is given by

$$M^\infty(z) = \begin{pmatrix} \frac{\beta+\beta^{-1}}{2} & \frac{\beta-\beta^{-1}}{2i} \\ -\frac{\beta-\beta^{-1}}{2i} & \frac{\beta+\beta^{-1}}{2} \end{pmatrix}, \tag{5.13}$$

where

$$\beta(z) = \left(\frac{z-\xi}{z-\overline{\xi}} \right)^{1/4} \tag{5.14}$$

which is defined to be analytic $\mathbb{C} \setminus \Gamma_1$ and satisfies $\beta(z) \sim 1$ as $z \to \infty$. We expect that $\widetilde{M} \sim M^\infty$ as $k \to \infty$, and hence by tracking the algebraic transformations $Y \to M \to \widetilde{M}$, we expect that

$$Y_{21}(0)e^{-k(g(0)-\ell)} \sim M_{21}^\infty(0), \qquad k \to \infty, \tag{5.15}$$

for $\gamma \geq 1$ in a compact set and $a_0 < a \leq a^*$.

Indeed we have:

Proposition 5.1. *Let $1 \leq \gamma \leq \gamma_1$ for any fixed $\gamma_1 \geq 1$. There are $L_1, \delta_1 > 0$ such that for*

$$a_0 + \frac{L_1}{k^{2/3}} \leq a \leq (1+\delta_1)a_0, \tag{5.16}$$

we have

$$Y_{21}(0)e^{-k(g(0)-\ell)} = M_{21}^\infty(0)\big(1 + O(\frac{1}{k|a-a_0|})\big), \tag{5.17}$$

for sufficiently large k.

The convergence $\widetilde{V} \to V^\infty$ is not uniform on Γ' and this considerably complicates the analysis. As in [6] in order to obtain the above error bound, we need to introduce local parametrices for the solution of the RHP near each of the endpoints ξ and $\overline{\xi}$. As in [1], a suitable local parametrix near each endpoint can be obtained in terms of Airy functions. Also since a is not fixed, but is allowed to approach a_0, we need to vary the magnitude of the parametrix according to the size of $a - a_0$. A similar situation arises in Lemma 6.2 (ii) of [1]. The proof of the above Proposition is parallel to that of Lemma 6.2 (ii), [1] (γ, q in [1] play the same role as a, k in this paper, respectively), and we do not repeat the argument here.

Lemma 5.2. *We have* $\Delta := g(0 + i0) - \ell \in \mathbb{R}$. *In particular,* $e^{g(0)-\ell} = e^{g(0+i0)-\ell} > 0$.

Proof. For $x \in \mathbb{R} \setminus \{p_i\}$, by Remark 7, $g(x + i0) = \overline{g(x - i0)}$, and hence by Proposition 4.1 (1), $e^{g(x+i0)} = e^{g(x-i0)} = e^{\overline{g(x+i0)}}$, and so $e^{g(x)} = e^{g(x+i0)} = e^{g(x-i0)}$ is real. In particular, it follows that $Im(g(x + i0)) \in \mathbb{Z}\pi$, and hence by continuity, $Im(g(x + i0))$ is constant for $x < p_i$. From (4.2) and the proof of Proposition 4.1 (3), $\ell = g_+(\xi) + g_-(\xi) - W(\xi) = g_+(p_i + i0) + g_-(p_i) - W(p_i)$. Hence $e^{g(0)-\ell} = e^{g(0+i0)-g_+(p_i+i0)}e^{-g_-(p_i)+W(p_i)}$. But by the above, $Im(g(0 + i0)) = Im(g_+(p_i + i0))$. Hence $g(0 + i0) - g_+(p_i + 0i) \in \mathbb{R}$. Clearly $g_-(p_i)$ and $W(p_i)$ are also real, and this proves the lemma. ☐

If we set $\xi = |\xi|e^{i\theta_c}$, $0 < \theta_c < \pi$, then we can check $\beta(0) = e^{i\theta_c/2}$ and $M^\infty_{21}(0) = -\frac{1}{2i}(\beta(0)-\beta(0)^{-1}) = -\sin\frac{\theta_c}{2}$. Hence the above proposition yields that

$$-Y_{21}(0) = e^{k\Delta} \sin\frac{\theta_c}{2}\big(1 + O(\frac{1}{k|a - a_0|})\big). \tag{5.18}$$

Note that $-Y_{21}(0)$ is indeed real and positive from Lemma 2.1. This is consistent with Lemma 5.2.

Now we compute $e^{g(0)-\ell}$. Let

$$\alpha = \frac{\xi + \overline{\xi}}{2} = |\xi| \cos\theta_c. \tag{5.19}$$

From the formula (3.3) for h, one can check directly that an anti-derivative of $2(h(z) - \frac{1}{z})$ is

$$\begin{aligned}
&-\gamma a \log(z + t^{-1}) - a \log(z + t) + (a - 1) \log(z)\\
&+ \Big(-\frac{\gamma a}{x} - \frac{a}{y} + \frac{a + 1}{r}\Big)R(z) - \Big(-\frac{\gamma a}{x}(\alpha + t^{-1}) - \frac{a}{y}(\alpha + t) + \frac{a + 1}{r}\alpha\Big) \log(z - \alpha + R(z))\\
&-\gamma a \log\Big(\frac{z + R(z) + t^{-1} - x}{z + R(z) + t^{-1} + x}\Big) - a \log\Big(\frac{z + R(z) + t - y}{z + R(z) + t + y}\Big) + (a + 1) \log\Big(\frac{z + R(z) - r}{z + R(z) + r}\Big)
\end{aligned} \tag{5.20}$$

where the logarithms are taken to be the analytic in $\mathbb{C} \setminus (-\infty, 0]$ and $\log z = \log|z|$ for $z > 0$, and $r = -R(0)$, $x = -R(-t^{-1})$, $y = -R(-t)$ as in (3.22). It is straightforward, but tedious, to check that (5.20) is analytic in $\mathbb{C} \setminus (\overline{\Gamma_1} \cup (-\infty, p_i])$ as in the definition of g in (4.1). Using the endpoint conditions (3.4), or(3.23), (5.20) is equal to

$$\begin{aligned}
&-\gamma a \log(z + t^{-1}) - a \log(z + t) + (a - 1) \log(z) + (1 + \gamma a) \log(z - \alpha + R(z))\\
&-\gamma a \log\Big(\frac{z + R(z) + t^{-1} - x}{z + R(z) + t^{-1} + x}\Big) - a \log\Big(\frac{z + R(z) + t - y}{z + R(z) + t + y}\Big) + (a + 1) \log\Big(\frac{z + R(z) - r}{z + R(z) + r}\Big).
\end{aligned} \tag{5.21}$$

Evaluating the asymptotics as $z \to \infty$, we see that

$$\begin{aligned} 2g(z) =& -\gamma a \log(z+t^{-1}) - a\log(z+t) + (a+1)\log(z) \\ &+ (1+\gamma a)\log\big((z-\alpha+R(z))/2\big) - \gamma a \log\left(\frac{z+R(z)+t^{-1}-x}{z+R(z)+t^{-1}+x}\right) \\ &- a\log\left(\frac{z+R(z)+t-y}{z+R(z)+t+y}\right) + (a+1)\log\left(\frac{z+R(z)-r}{z+R(z)+r}\right), \end{aligned} \tag{5.22}$$

and

$$\begin{aligned} 2Re(g(0+i0)) =& (\gamma-1)a\log t + (1+\gamma a)\log((\alpha+r)/2) \\ &- \gamma a \log\left|\frac{-r+t^{-1}-x}{-r+t^{-1}+x}\right| - a\log\left|\frac{-r+t-y}{-r+t+y}\right| + (a+1)\log\left(\frac{2r}{1+\alpha/r}\right). \end{aligned} \tag{5.23}$$

Also from (5.22) and (4.2), we have

$$\begin{aligned} \ell =& \gamma a \log t + (1+\gamma a)\log(\xi-\alpha)/2 \\ &- \gamma a\log\left(\frac{\xi+t^{-1}-x}{\xi+t^{-1}+x}\right) - a\log\left(\frac{\xi+t-y}{\xi+t+y}\right) + (a+1)\log\left(\frac{\xi-r}{\xi+r}\right). \end{aligned} \tag{5.24}$$

By (3.36),

$$\cos\theta_c = \frac{x^2-r^2-t^{-2}}{2rt^{-1}} = \frac{y^2-r^2-t^2}{2rt}. \tag{5.25}$$

Thus using (5.23) and (5.24), we can express Δ in terms of r, x, y. After some algebra, we find

$$\begin{aligned} \Delta =& -\gamma a \log t + (2+a+\gamma a)\log 2 + (1+a)\log r - \frac{1}{2}\log(r+t^{-1}-x) \\ &- \frac{1}{2}(1+2\gamma a)\log(r+t^{-1}+x) - \frac{1}{2}\log(r+t-y) - \frac{1}{2}(1+2a)\log(r+t+y). \end{aligned} \tag{5.26}$$

To emphasize the dependence on a, we write $\Delta = \Delta(a)$, etc.

Lemma 5.3. *Fix $1 \le \gamma_2 < \infty$. Then there exists $\delta_2 > 0$ such that for $a_0 \le a \le (1+\delta_2)a_0$ and $1 \le \gamma \le \gamma_2$, we have*

$$\Delta(a) = -c_2(a-a_0)^2\big(1+O(|a-a_0|)\big) \tag{5.27}$$

where

$$c_2 = \frac{t^2(t+t\gamma+2\sqrt{\gamma})^3\sqrt{\gamma}}{4(1+t\sqrt{\gamma})^2(t+\sqrt{\gamma})^2}, \tag{5.28}$$

and the order term $O(|a-a_0|)$ is uniform for a and γ as above.

For the proof, we need the following lemma, which considers the case when $a = a_0$. This case is specifically excluded from Lemma 3.2.

Lemma 5.4. *For fixed $0 < t < 1$, $\gamma \ge 1$, and $a = a_0 = \frac{1-t^2}{t((\gamma+1)t+2\sqrt{\gamma})}$, there is a unique solution r to (3.28) satisfying $r_1 < r < r_2$, given by*

$$r = r(a_0) = r_0 := \frac{1+t\sqrt{\gamma}}{t+\sqrt{\gamma}}. \tag{5.29}$$

The function $a \mapsto r(a)$ is smooth for all $a \ge a_0$ and

$$r'(a_0) = -\frac{3(\gamma-1)(t\gamma+t+2\sqrt{\gamma})^2 t^2}{4(1+t\sqrt{\gamma})(t+\sqrt{\gamma})^3}. \tag{5.30}$$

Proof. When $a = a_0$, we have $r_1 = tr_0^2$ and $r_2 = \frac{1}{t}r_0^2$. Since $t < r_0 < \frac{1}{t}$, r_0 satisfies $r_1 < r_0 < r_2$. It is then a direct calculation to check that r_0 satisfies the equation (3.28). Now we want to show the uniqueness of the solution. Let H be as in (3.30) in the proof of Lemma 3.2. Now when $a = a_0$, the value r_* of (3.32) is r_0. Thus we have

$$H(r) + \frac{r}{2}H'(r) \ge H(r_0) + \frac{r_0}{2}H'(r_0) = 0, \tag{5.31}$$

for $r_1 < r < r_2$, and the inequality is strict for $r \ne r_0$. Hence if there is a zero $r_c \ne r_0$, it should satisfy $H'(r_c) > 0$. On the other hand, by direct calculation, we have

$$H'(r_0) = H''(r_0) = 0, \qquad H^{(3)}(r_0) = \frac{24t(t+\sqrt{\gamma})^5}{\sqrt{\gamma}(1-t^2)^2(1+t\sqrt{\gamma})^3} > 0, \tag{5.32}$$

and hence H is also increasing at the zero $r = r_0$. As in Lemma 3.2, there is no other zero $r_c \ne r_0$ in (r_1, r_2).

Now consider $H = H(a, r)$. By direct calculations, we find

$$H(a_0, r_0) = H_r(a_0, r_0) = H_a(a_0, r_0) = H_{rr}(a_0, r_0) = 0 \tag{5.33}$$

and

$$H_{ra}(a_0, r_0) = \frac{4(t+\sqrt{\gamma})(t\gamma+t+2\sqrt{\gamma})^2 t^2}{(1+t\sqrt{\gamma})^3(1-t^2)} \ne 0, \tag{5.34}$$

$$H_{aa}(a_0, r_0) = \frac{6(\gamma-1)(t\gamma+t+2\sqrt{\gamma})^4 t^4}{(1+t\sqrt{\gamma})^4(t+\sqrt{\gamma})^2(1-t^2)}. \tag{5.35}$$

Hence near (a_0, r_0), the power series of $H(a, r)$ has the form

$$H(a, r) = H_{ra}(a_0, r_0)(r-r_0)(a-a_0) + \frac{1}{2}H_{aa}(a_0, r_0)(a-a_a)^2 + \frac{1}{6}H_{rrr}(a_0, r_0)(r-r_0)^3 + \cdots. \tag{5.36}$$

Motivated by this expansion, we set

$$\eta := \frac{r-r_0}{a-a_0}, \tag{5.37}$$

and substitute $r = r_0 + \eta(a-a_0)$ in H, and define

$$F(a, \eta) := \frac{H(a, r_0 + \eta(a-a_0))}{(a-a_0)^2}. \tag{5.38}$$

Setting

$$\eta_0 := -\frac{H_{aa}(a_0, r_0)}{2H_{ra}(a_0, r_0)} = -\frac{3(\gamma-1)(t\gamma+t+2\sqrt{\gamma})^2t^2}{4(1+t\sqrt{\gamma})(t+\sqrt{\gamma})^3}, \tag{5.39}$$

a direct calculation shows that $F(a, \eta)$ is a smooth function near (a_0, η_0) and

$$F(a_0, \eta_0) = 0, \qquad F_\eta(a_0, \eta_0) = H_{ra}(a_0, r_0) \neq 0. \tag{5.40}$$

Therefore by the implicit function theorem, there is the smooth function $\eta = \eta(a)$, $a_0 \le a < a_0(1+\delta)$ for some $\delta > 0$ such that $F(a, \eta(a)) = 0$. Then $H(a, r_0 + \eta(a)(a - a_0)) = 0$, and by the uniqueness of the solution, $r = r_0 + \eta(a)(a - a_0)$ is smooth in a for $a_0 \le a < a_0(1+\delta)$. But for $a > a_0$, $H_a(a, r(a)) > 0$, and the smoothness of $r = r(a)$ is elementary. Hence $r(a)$ is a smooth function for $a \ge a_0$. □

Let

$$x_0 := x(a_0) = \frac{\sqrt{\gamma}(1-t^2)}{t(t+\sqrt{\gamma})}, \qquad y_0 := y(a_0) = \frac{1-t^2}{t+\sqrt{\gamma}}, \tag{5.41}$$

which are obtained by setting $a = a_0$, $r = r(a_0) = r_0$ in (3.25), (3.26). Then

$$\frac{x_0^2 - r_0^2 - t^{-2}}{2r_0t^{-1}} = -1, \tag{5.42}$$

and hence from (3.35), the point ξ is on the negative real line. Thus when $a = a_0$, the two endpoints ξ and $\overline{\xi}$ collapse to the point $-r_0$ on the real line. This is an extreme case of the deformation illustrated in Figure 16.

Proof of Lemma 5.3. When $a = a_0$, we have $r = r_0$, $x = x_0$, $y = y_0$, and we can direct check from (5.26) that $\Delta(a_0) = 0$. We have

$$\begin{aligned}\Delta'(a) = &-\gamma\log t + (1+\gamma)\log 2 + \log r - \gamma\log(r + t^{-1} + x) - \log(r + t + y)\\ &+ (1+a)\frac{r'}{r} - \frac{1}{2}\frac{r'-x'}{r+t^{-1}-x} - \frac{1}{2}(1+2\gamma a)\frac{r'+x'}{r+t^{-1}+x} - \frac{1}{2}\frac{r'-y'}{r+t-y} - \frac{1}{2}(1+2a)\frac{r'+y'}{r+t+y}.\end{aligned} \tag{5.43}$$

At $a = a_0$, from from (5.29), (5.41), we have

$$-\gamma\log t + (1+\gamma)\log 2 + \log r_0 - \gamma\log(r_0 + t^{-1} + x_0) - \log(r_0 + t + y_0) = 0 \tag{5.44}$$

and hence again from (5.29), (5.41), after some algebra,

$$\Delta'(a_0) = -\frac{(\gamma-1)(1-t^2)}{(t+t\gamma+2\sqrt{\gamma})(1+t\sqrt{\gamma})}\big[(1+t\sqrt{\gamma})r'(a_0) + t\sqrt{\gamma}x'(a_0) - y'(a_0)\big]. \tag{5.45}$$

Now from the relation (3.27) between r, x, y, we have

$$rr' = \frac{1}{1-t^2}(yy' - t^2xx'). \tag{5.46}$$

This implies by (5.29), (5.41),

$$(1+t\sqrt{\gamma})r'(a_0) = y'(a_0) - t\sqrt{\gamma}x'(a_0), \tag{5.47}$$

and hence

$$\Delta'(a_0) = 0. \tag{5.48}$$

Now we compute $\Delta''(a_0)$. We have

$$\begin{aligned}\Delta''(a) =& 2\frac{r'}{r} + (1+a)\Big(\frac{r''}{r} - (\frac{r'}{r})^2\Big) - \frac{1}{2}\Big(\frac{r''-x''}{r+t^{-1}-x} - (\frac{r'-x'}{r+t^{-1}-x})^2\Big) \\ & - 2\gamma\frac{r'+x'}{r+t^{-1}+x} - \frac{1}{2}(1+2\gamma a)\Big(\frac{r''+x''}{r+t^{-1}+x} - (\frac{r'+x'}{r+t^{-1}+x})^2\Big) \\ & - \frac{1}{2}\Big(\frac{r''-y''}{r+t-y} - (\frac{r'-y'}{r+t-y})^2\Big) - 2\frac{r'+y'}{r+t+y} - \frac{1}{2}(1+2a)\Big(\frac{r''+y''}{r+t+y} - (\frac{r'+y'}{r+t+y})^2\Big).\end{aligned} \tag{5.49}$$

First consider the terms with double derivatives. At $a = a_0$,

$$\begin{aligned}&\Big((1+a)\frac{r''}{r} - \frac{1}{2}\frac{r''-x''}{r+t^{-1}-x} - \frac{1}{2}(1+2\gamma a)\frac{r''+x''}{r+t^{-1}+x} - \frac{1}{2}\frac{r''-y''}{r+t-y} - \frac{1}{2}(1+2a)\frac{r''+y''}{r+t+y}\Big)(a_0) \\ &= -\frac{(\gamma-1)(1-t^2)}{4(t+t\gamma+2\sqrt{\gamma})(1+t\sqrt{\gamma})}\big[(1+t\sqrt{\gamma})r''(a_0) + t\sqrt{\gamma}x''(a_0) - y''(a_0)\big].\end{aligned} \tag{5.50}$$

From (5.46),

$$rr'' - \frac{1}{1-t^2}(yy'' - t^2xx'') = -(r')^2 + \frac{1}{1-t^2}\big((y')^2 - t^2(x')^2\big), \tag{5.51}$$

and hence the right-hand side of (5.50) is equal to

$$-\frac{(\gamma-1)(1-t^2)(t+\sqrt{\gamma})}{4(t+t\gamma+2\sqrt{\gamma})(1+t\sqrt{\gamma})}\big[-(r'(a_0))^2 + \frac{1}{1-t^2}\big((y'(a_0))^2 - t^2(x'(a_0))^2\big)\big]. \tag{5.52}$$

Thus $\Delta''(a_0)$ is given by (5.49) at $a = a_0$ where the terms with double derivatives are replaced by (5.52), which involves only the first derivatives of r, x, y at a_0. From (5.30) and (3.25), (3.26), we have

$$r'(a_0) = -\frac{3(\gamma-1)(t\gamma+t+2\sqrt{\gamma})^2t^2}{4(1+t\sqrt{\gamma})(t+\sqrt{\gamma})^3} \tag{5.53}$$

$$x'(a_0) = \frac{(t+4\sqrt{\gamma}+3t\gamma)(t+t\gamma+2\sqrt{\gamma})^2t}{4(1+t\sqrt{\gamma})(t+\sqrt{\gamma})^3} \tag{5.54}$$

$$y'(a_0) = \frac{(4t\sqrt{\gamma}+\gamma+3)(t+t\gamma+2\sqrt{\gamma})^2t^2}{4(1+t\sqrt{\gamma})(t+\sqrt{\gamma})^3}, \tag{5.55}$$

and we obtain, after some calculation,

$$\Delta''(a_0) = -\frac{t^2(t+t\gamma+2\sqrt{\gamma})^3\sqrt{\gamma}}{2(1+t\sqrt{\gamma})^2(t+\sqrt{\gamma})^2}. \tag{5.56}$$

By Taylor's formula, for $a \geq a_0$,

$$\Delta(a) = \frac{1}{2}\Delta''(a_0)(a-a_0)^2 + \frac{1}{6}\Delta'''(\widetilde{a})(a-a_0)^3, \tag{5.57}$$

for some $a_0 \leq \widetilde{a} \leq a$. For $\delta > 0$ and $1 \leq \gamma_2 < \infty$, set

$$C := \sup\{|\Delta'''(\widetilde{a},\gamma)| : 1 \leq \gamma \leq \gamma_2, a_0(\gamma) \leq \widetilde{a} \leq a_0(\gamma)(1+\delta)\}, \tag{5.58}$$

where we have made the dependence on γ of Δ''' explicit. It follows from the smooth dependence of $\Delta(a,\gamma)$ on γ, as well as on a, that given γ_2, we can choose $\delta = \delta_2$ such that $C < \infty$. Therefore

$$\Delta(a) = \frac{1}{2}\Delta''(a_0)(a-a_0)^2\big(1 + O(|a-a_0|\big), \tag{5.59}$$

where $O(|a-a_0|)$ is uniform for $1 \leq \gamma \leq \gamma_2$ and $a_0 \leq a \leq a_0(1+\delta_2)$. Indeed $O(|a-a_0|) \leq C|a-a_0|$. $\square$

In order to prove Proposition 2.2, we use (5.18). As

$$\cos\theta_c = \frac{x(a)^2 - r(a)^2 - t^{-2}}{2r(a)t^{-1}}, \tag{5.60}$$

and $\cos\theta_c \to -1$ as $a \to a_0$, we see that $\sin\frac{\theta_c}{2} = \sqrt{\frac{1-\cos\theta_c}{2}}$ is a smooth function of a in $[a_0,\infty)$. Thus $\sin\frac{\theta_c}{2} = 1 + O(|a-a_0|)$ for a near a_0, $a \geq a_0$. Inserting this information into (5.18) and using (5.27), we obtain (2.20).

6 Proof of Theorem 1.1

Take

$$L > \frac{2L_0}{a_0^{4/3}b_0}, \qquad 0 < \delta < \frac{\delta_0}{a_0b_0(1+\delta_0)}, \tag{6.1}$$

where L_0, δ_0 are given in Proposition 2.2. Let

$$n = \frac{1}{a_0}N - xb_0N^{1/3}. \tag{6.2}$$

Set

$$b := \frac{1}{a_0}N - \frac{L_0}{a_0^{4/3}}N^{1/3}. \tag{6.3}$$

Then for $L \leq x \leq \delta N^{2/3}$, the condition (2.19) for $a = \frac{N}{k}$ in Proposition 2.2 is satisfied for any k in the range $n \leq k \leq b$, Following [18], we consider

$$\log\frac{\mathbb{P}(G([\gamma N],N) \leq n)}{\mathbb{P}(G([\gamma N],N) \leq b)} \tag{6.4}$$

which equals

$$\sum_{k=n+1}^{b} \log(-Y_{21}(0;k)) \tag{6.5}$$

by (2.7). Inserting (2.20) into (6.5), we obtain (cf. [18])

$$\begin{aligned} \log \frac{\mathbb{P}(G([\gamma N], N) \le n)}{\mathbb{P}(G([\gamma N], N) \le b)} &= -\frac{1}{3} c_2 a_0^3 b_0^3 x^3 + O(x^4 N^{-2/3}) + O(\log x) \\ &= -\frac{1}{12} x^3 + O(x^4 N^{-2/3}) + O(logx) \end{aligned} \tag{6.6}$$

as $c_2 a_0^3 b_0^3 = \frac{1}{4}$. But by the result of Johansson [12], as $N \to \infty$,

$$\mathbb{P}(G([\gamma N], N) \le b) = F(-L_0 a_0^{-4/3} b_0^{-1}) + o(1), \tag{6.7}$$

where $F(x)$ is the Tracy-Widom distribution. This proves Theorem 1.1.

References

[1] J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[2] J. Baik, P. Deift, and E. Rains. A Fredholm determinant identy and the convergence of moments for random young tableaux. *Comm. Math. Phys.*, 223(3):627–672, 2001.

[3] J. Baik and E. M. Rains. Algebraic aspects of increasing subsequences. *Duke Math. J.*, 109(1):1–65, 2001.

[4] A. Borodin and A. Okounkov. A Fredholm determinant formula for Toeplitz determinants. *Integral Equations Operator Theory*, 37(4):386–396, 2000.

[5] P. Deift, T. Kriecherbauer, K. McLaughlin, S. Venakides, and X. Zhou. Strong asymptotics of orthogonal polynomials with respect to exponential weights. *Comm. Pure Appl. Math.*, 52(12):1491–1552, 1999.

[6] P. Deift, T. Kriecherbauer, K. McLaughlin, S. Venakides, and X. Zhou. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.*, 52(11):1335–1425, 1999.

[7] P. Deift, S. Venakides, and X. Zhou. New results in small dispersion KdV by an extension of the steepest descent method for Riemann-Hilbert problems. *Internat. Math. Res. Notices*, 6:285–299, 1997.

[8] P. Deift and X. Zhou. A steepest descent method for oscillatory Riemman-Hilbert problems; asymptotics for the MKdV equation. *Ann. of Math.*, 137:295–368, 1993.

[9] J.-D. Deuschel and O. Zeitouni. On increasing subsequences of i.i.d. samples. *Combin. Probab. Comput.*, 8(3):247–263, 1999.

[10] A. Fokas, A. Its, and V. Kitaev. Discrete Painlevé equations and their appearance in quantum gravity. *Comm. Math. Phys.*, 142:313–344, 1991.

[11] I. Gessel. Symmetric functions and P-recursiveness. *J. Combin. Theory Ser. A*, 53:257–285, 1990.

[12] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.

[13] K. Johansson. Transversal fluctuations for increasing subsequences on the plane. *Probab. Theory Related Fields*, 116(4):445–456, 2000.

[14] Kurt Johansson. personal communications.

[15] S. Kamvissis, K. McLaughlin, and P. Miller. Semiclassical soliton ensembles for the focusing nonlinear schroedinger equation. nlin.SI/0012034 ; http://xxx.lanl.gov/abs/.

[16] J. Krug and H. Spohn. Kinetic roughening of growth interfaces. In C. Godr‘eche, editor, *Solids far from Equilibrium: Growth, Morphology and Defect*, pages 479–582. Cambridge University Press, 1992.

[17] M. Löwe and F. Merkl. Moderate deviations for longest increasing subsequences: the upper tail. preprint, 2001.

[18] M. Löwe, F. Merkl, and S. Rolles. Moderate deviations for longest increasing subsequences: the lower tail. preprint, 2001.

[19] C. Pommerenke. *Univalent functions*. Vandenhoeck & Ruprecht, Göttingen, 1975.

[20] T. Seppäläinen. A microscopic model for the burgers equation and longest increasing subsequences. *Electron. J. Prob.*, 1(5), 1995.

[21] T. Seppäläinen. Large deviations for increasing sequences on the plane. *Probab. Theory Related Fields*, 112(2):221–244, 1998.

[22] T. Seppäläinen. Perturbation of the equilibrium for a totally asymmetric stick process in one dimension. *Ann. Probab.*, 29(1):176–204, 2001.

[23] C. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159:151–174, 1994.

[24] H. Widom. On convergence of moments for random Young tableaux and a random growth models. math.CO/0108008 ; http://xxx.lanl.gov/abs/.