



Bothner, T. (2021). On the origins of Riemann-Hilbert problems in mathematics. *Nonlinearity*, 34(4), R1-R73.
<https://doi.org/10.1088/1361-6544/abb543>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1088/1361-6544/abb543](https://doi.org/10.1088/1361-6544/abb543)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Institute of Physics at <https://doi.org/10.1088/1361-6544/abb543> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

INVITED ARTICLE • **OPEN ACCESS**

On the origins of Riemann–Hilbert problems in mathematics^{*}

To cite this article: Thomas Bothner 2021 *Nonlinearity* **34** R1

View the [article online](#) for updates and enhancements.

Invited Article

On the origins of Riemann–Hilbert problems in mathematics*

Thomas Bothner 

School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol, BS8 1UG, United Kingdom

E-mail: thomas.bothner@bristol.ac.uk

Received 16 April 2020, revised 30 July 2020

Accepted for publication 4 September 2020

Published 22 February 2021



CrossMark

Abstract

This article is firstly a historic review of the theory of Riemann–Hilbert problems with particular emphasis placed on their original appearance in the context of Hilbert’s 21st problem and Plemelj’s work associated with it. The secondary purpose of this note is to invite a new generation of mathematicians to the fascinating world of Riemann–Hilbert techniques and their modern appearances in nonlinear mathematical physics. We set out to achieve this goal with six examples, including a new proof of the integro-differential Painlevé-II formula of Amir *et al* (2011 *Commun. Pure Appl. Math.* **64** 466–537) that enters in the description of the Kardar–Parisi–Zhang crossover distribution. Parts of this text are based on the author’s Szegő prize lecture at the 15th International Symposium on Orthogonal Polynomials, Special Functions and Applications (OPSFA) in Hagenberg, Austria.

Keywords: Riemann–Hilbert problem, Hilbert’s 21st problem, Fuchsian systems, monodromy group, singular integral equations, defocusing nonlinear Schrödinger equation, Painlevé-II equation

Mathematics Subject Classification numbers: primary 30E25, secondary 45M05, 60B20.

*The author is greatly indebted to A Its, P Bleher, P Deift, A Kuijlaars, M Bertola, J Baik and P Miller for countless discussions on Riemann–Hilbert problems over the past 10 years. This article is dedicated to Boris Dubrovin (1950–2019) and his outstanding legacy in the field of mathematical physics. The author would also like to thank the referees for their constructive suggestions which improved the paper in a variety of ways.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

(Some figures may appear in colour only in the online journal)

1. Introduction

Riemann–Hilbert problems, in short RHPs, are useful in the analysis of orthogonal polynomials (OPs), special functions (SFs) and thus several applications (A) in mathematics or physics. Still, these modern appearances of RHPs are quite detached from the original meaning of the term RHP and it is our primary objective to pin down the occurrence of the first problem in mathematics that was coined an RHP. Once done, we will follow Plemelj's traces whose work on the original RHP gave rise to an analytic apparatus, the *Riemann–Hilbert techniques* or *Riemann–Hilbert methods*, which found various applications in the OPSFA realm. We will showcase and advertise the latter modern aspects of the theory with examples from mathematical physics, in particular examples from nonlinear wave and integrable systems theory, statistical mechanics, random matrix theory and integrable probability.

2. The original Riemann–Hilbert problem

In order to answer the question ‘What is . . . a Riemann–Hilbert problem’ one should go back to 1900, the year of the second international congress of mathematics. Held at the turn of the century in Paris, the ICM 1900 proved to be highly influential for the development of 20th century mathematics, in particular since Hilbert delivered his famous address in the section *Bibliographie et Histoire. Enseignement et methodes*. During his lecture he presented (because of time constraints) ten mathematical problems ‘from the discussion of which an advancement of science may be expected’, see [70, p 445]. The full, twenty-three long, list of Hilbert's problems appeared in German print even before the French proceedings of the congress. Scrolling through this list we eventually encounter problem #21, the main focus point of this section, see [69, pp 289–290]:

21. *Beweis der Existenz linearer Differentialgleichungen mit vorgeschriebener Monodromiegruppe*

Aus der Theorie der linearen Differentialgleichungen mit einer unabhängigen Veränderlichen z möchte ich auf ein wichtiges Problem hinweisen, welches wohl bereits Riemann im Sinne gehabt hat, und welches darin besteht, zu zeigen, daß es stets eine lineare Differentialgleichung der Fuchsschen Klasse mit gegebenen singulären Stellen und einer gegebenen Monodromiegruppe giebt. Die Aufgabe verlangt also die Auffindung von n Functionen der Variablen z , die sich überall in der complexen z -Ebene regulär verhalten, außer etwa in den gegebenen singulären Stellen: in diesen dürfen sie nur von endlich hoher Ordnung unendlich werden und beim Umlauf der Variablen z um dieselben erfahren sie die gegebenen linearen Substitutionen. Die Existenz solcher Differentialgleichungen ist durch Constantenzählung wahrscheinlich gemacht worden, doch gelang der strenge Beweis bisher nur in dem besonderen Falle, wo die Wurzeln der Fundamentalgleichungen der gegebenen Substitutionen sämtlich vom absoluten Betrage 1 sind. Diesen Beweis hat L. Schlesinger auf Grund der Poincaréschen Theorie der Fuchsschen ζ -Functionen erbracht. Es würde offenbar die Theorie der linearen Differentialgleichungen ein wesentlich abgeschlosseneres Bild zeigen, wenn die allgemeine Erledigung des bezeichneten Problems gelänge.

In this problem Hilbert asks the reader to *show that there always exists a linear differential equation of the Fuchsian class, with given singular points and monodromic group*. Before reviewing some of the necessary background material in section 3 below it will be important to point out the following peculiar wording in Hilbert's original formulation: *the problem requires the production of n functions of the variable z , regular throughout the complex z plane except at the given singular points; at these points the functions may become infinite of only finite order*. Thus, Hilbert does not use the word pole singularity—he refers to what is nowadays called a regular singularity, see [83, definition 16.2]—and he also never speaks of Fuchsian linear systems. These subtleties in the formulation of the problem played an important role as we will soon learn.

Right now we follow Anosov and Bolibrukh [5, p 7] and coin Hilbert's 21st problem the original RHP: Hilbert formulated it and he explicitly mentions Riemann in his wording as this problem is extremely close to Riemann's ideas¹ (namely the global construction of a function from given local analytic data, i.e. the singularity locations and associated monodromy). In short,

Riemann–Hilbert Problem 2.1 (The original RHP ([69], 1900)). *Proof of the existence of linear differential equations having a prescribed monodromic group.*

This problem seems quite different from the modern RHPs discussed in section 6 below, yet it is this problem on which the modern theory is based upon. Thus, being so influential, it is only natural to ask if RHP 2.1 has been solved 120 years after Hilbert's address? Well, according to the Encyclopedia of Mathematics [51], the 'solution is negative or positive depending on how the problem is understood'. This answer points at a curious mathematical misunderstanding which persisted from 1908 till 1983 and which relates to the aforementioned atypicalities in Hilbert's wording. Indeed, we highlight two possible sources of confusion:

- *equation vs. system*: The monodromy of a p th order scalar linear ODE with solution y matches the monodromy of the $p \times p$ linear ODE system which the vector (y, y', \dots, y^{p-1}) solves. Still, at the time of the congress, see [123], it was known that Hilbert's question for scalar Fuchsian equations has a negative solution unless additional singularities are included. So by a p th order scalar Fuchsian *equation*, Hilbert likely meant a generic $p \times p$ Fuchsian *system*.
- *regular singularities vs Fuchsian singularities*: Both notions exist for p th order scalar linear ODEs in the complex plane. Although up front different in appearance, these notions are in fact equivalent (this is a result by Fuchs from 1868, see [83, theorem 19.20]). However, this is no longer the case when one considers the same singularities for $p \times p$ linear ODE systems, see section 4.

In turn we could think of at least three different interpretations of RHP 2.1: given a monodromy group with encoded singularity locations, are we asked to realize it by

- a Fuchsian linear p th order differential equation?
- a linear $p \times p$ system having only regular singularities?
- a Fuchsian system on the whole Riemann sphere \mathbb{CP}^1 ?

The answer to (a) is *negative* as shown by Poincaré [123]: the number of parameters in a p th order linear Fuchsian equation with singularities a_1, \dots, a_n is less than the dimension of the

¹ Although [5, p 7] also argues that 'Riemann never spoke exactly of something like it.'

space \mathcal{M} of monodromy representations². Interpretation (b) was *positively* solved by Plemelj in 1908 [121] and believed to settle (c) as well. This belief persisted for 75 years until Kohn [94] and Arnold and Il'yashenko [7] discovered a serious gap in Plemelj's argument. As it turned out, interpretation (c) is much more subtle and a *negative* answer to Hilbert's 21st problem in said context was eventually given by Bolibrukh in 1989, see [17, 18]. We will now discuss in detail interpretations (b), (c) and highlight the pioneering work of Plemelj in this context. His contributions laid the foundation for several aspects of modern Riemann–Hilbert theory which we showcase through a series of OPSFA-type examples in sections 6, 8 and 9—however only section 9 contains original content. For an in-depth discussion of interpretation (a) we refer the interested reader to [5, chapter 7] or more recently [66, section 3].

3. Background terminology

We are considering $p \times p$ linear ODE systems of the form

$$\frac{d\Psi}{dz} = A(z)\Psi, \quad \Psi = \Psi(z) \in \mathbb{C}^{p \times p}, \tag{3.1}$$

where $A(z) \in \mathbb{C}^{p \times p}$ is analytic in a disk $\mathbb{D}_r(z_0) \subset \mathbb{C}$ of radius $r > 0$ centred at some $z_0 \in \mathbb{C}$. Then, by the Picard–Lindelöf theorem [83, theorem 1.1], any fundamental solution of (3.1) is analytic in $\mathbb{D}_r(z_0)$. Now suppose $A(z)$ is analytic on the punctured Riemann sphere $S := \mathbb{CP}^1 \setminus \{a_1, \dots, a_n\}$ where a_1, \dots, a_n are n distinct points on \mathbb{CP}^1 . Through the use of overlapping disks and the previous local existence theorem, any fundamental solution of (3.1) can be continued along any path in S . This statement is part of the monodromy theorem [10, theorem 67] which furthermore asserts that the continuation only depends on the homotopy class of the path. With this in mind, we now define the *monodromy* of (3.1) as follows: fix $a_0 \in S$ distinct from a_1, \dots, a_n and continue any fundamental solution $\Psi(z)$ of (3.1) along some $\gamma \in \pi_1(S, a_0)$ (the fundamental group of the surface S with base point a_0). The continuation yields $\tilde{\Psi}(z)$, another fundamental solution of (3.1), which links to $\Psi(z)$ via

$$\tilde{\Psi}(z) = \Psi(z)G_\gamma,$$

for some invertible $p \times p$ matrix G_γ , i.e. $G_\gamma \in \text{GL}(p, \mathbb{C})$. Strictly speaking, G_γ only depends on the homotopy class of the loop γ . In turn, the map $[\gamma] \mapsto G_\gamma$ defines a representation

$$\chi : \pi_1(S, a_0) \rightarrow \text{GL}(p, \mathbb{C}), \tag{3.2}$$

of $\pi_1(S, a_0)$, called the *monodromy* or *monodromy representation* of system (3.1). Next, fix n fundamental loops $\gamma_1, \dots, \gamma_n$ as indicated in figure 1 below. Each loop γ_i wraps once around its corresponding blue dashed cut, i.e. γ_i wraps once around the singularity a_i , and all loops compose to the identity, $\gamma_1 \circ \dots \circ \gamma_n = e \in \pi_1(S, a_0)$.

The *monodromy matrix* at the singular point a_i is defined as

$$G_i := \chi([\gamma_i]), \quad i = 1, \dots, n,$$

and we record the cyclic constraint $G_1 \cdot \dots \cdot G_n = \mathbb{I} \in \text{GL}(p, \mathbb{C})$. Observe that the image of $\pi_1(S, a_0)$ under (3.2), the *monodromy group* of system (3.1), is generated by the matrices $\{G_1, \dots, G_n\}$.

²Thus the need for additional, so called *apparent*, singular points in the construction of Fuchsian scalar equations with prescribed monodromy. At these (from a_1, \dots, a_n different) singular points the coefficients in the equation are singular but the solutions single-valued analytic.

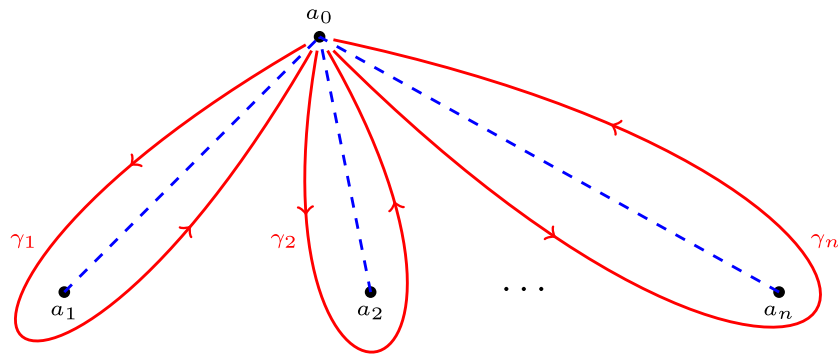


Figure 1. A possible visualization of the red fundamental loops γ_i .

We are left with two obvious ambiguities in our definitions: first, if we were to start from a different fundamental solution $\Psi^*(z)$ instead of $\Psi(z)$, then $\Psi^*(z) = \Psi(z)H$ for some $H \in GL(p, \mathbb{C})$ and the associated monodromy matrices change accordingly,

$$G_i^* = H^{-1}G_iH, \quad i = 1, \dots, n. \tag{3.3}$$

Second, the dependence of $\pi_1(S, a_0)$ and thus $G_i = \chi([\gamma_i])$ on the base point a_0 : given that S is path connected we know that $\pi_1(S, a_0)$ and $\pi_1(S, a_0^*)$ are isomorphic for any two base points a_0, a_0^* different from a_1, \dots, a_n , thus under the representation (3.2) the associated monodromy matrices change also in the style of (3.3). Hence, summarizing the two ambiguities, the monodromy of (3.1) is defined up to conjugation equivalence and so an element of the space

$$\mathcal{M} := \text{Hom}(\pi_1(S, a_0), GL(p, \mathbb{C})) / GL(p, \mathbb{C}),$$

of conjugacy classes of representations of $\pi_1(S, a_0)$. Alternatively, in the language of monodromy matrices, we have

$$\mathcal{M} \cong \{(G_1, \dots, G_n) : G_1 \cdot \dots \cdot G_n = \mathbb{I}\} / GL(p, \mathbb{C}). \tag{3.4}$$

One notion remains: we say system (3.1) is *Fuchsian* if all its singular points $\{a_1, \dots, a_n\}$ are first order poles of $A(z)$, i.e. without loss of generality (thanks to a fractional linear map) we have

$$A(z) = \sum_{i=1}^n \frac{B_i}{z - a_i}, \quad \text{res}_{z=\infty} A(z) = -\sum_{i=1}^n B_i = 0.$$

Using the above terminology we are now able to rephrase RHP 2.1 in interpretation (c) as follows:

Riemann–Hilbert Problem 3.1 (The original RHP, case (3)). *Is the monodromy map $\mu : \mathcal{M}^* \rightarrow \mathcal{M}$ from the space*

$$\mathcal{M}^* := \left\{ (B_1, \dots, B_n) : B_i \in \mathbb{C}^{p \times p}, \sum_{i=1}^n B_i = 0 \right\} / GL(p, \mathbb{C})$$

of Fuchsian systems with fixed singularities a_1, \dots, a_n into \mathcal{M} surjective?

This problem concludes our content on the necessary background terminology of Hilbert’s 21st problem. We will now review some of the mathematical works devoted to its solution and which were published between 1908 and 1989. These works introduced several key ideas and techniques that are nowadays used in the Riemann–Hilbert analysis of OPSFA-type problems.

4. Plemelj’s contributions

In 1908, Plemelj [121] (upgraded to book form 55 years later in [122]) published a solution of RHP 3.1 that was widely accepted until the early 1980s. In his work, he reduced RHP 3.1 to a Hilbert boundary value problem³ in the theory of singular integral equations and the idea goes as follows: join all singularities $a_1, \dots, a_n \in \mathbb{C}$ by a simple closed oriented contour Γ as indicated in figure 2 below.

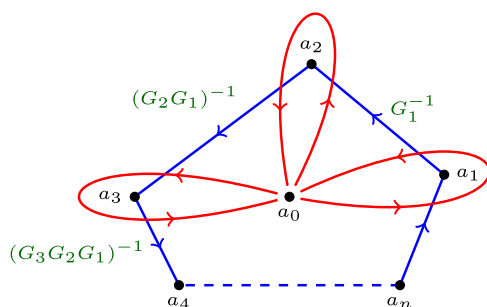


Figure 2. The oriented contour Γ in blue together with the aforementioned red fundamental loops γ_i . Some values of $G(z)$ are indicated in green.

Now define the piecewise constant, matrix-valued and invertible function

$$G(z) := (G_i G_{i-1} \cdots G_1)^{-1}, \quad z \in [a_i, a_{i+1}), \quad i = 1, \dots, n; \quad a_{n+1} := a_1,$$

which, thanks to the cyclic constraint (3.4), satisfies $G(z) = \mathbb{I}$ for $z \in [a_n, a_1)$. With Ω^+ denoting the region in \mathbb{C} bounded by Γ and Ω^- the complement of the closure $\overline{\Omega^+}$ in $\mathbb{C}\mathbb{P}^1$, we then consider the following boundary value problem:

Hilbert Boundary Value Problem 4.1 ([121, pp 213, 214, 229]). Find all vector-valued functions $X = X(z) \in \mathbb{C}^{1 \times P}$ such that

- (a) $X(z)$ is analytic in $\Omega^\pm \setminus \{\infty\}$ and extends continuously from either side to the punctured contour $\Gamma \setminus \{a_1, \dots, a_n\}$.
- (b) On the open segments $(a_i, a_{i+1}), i = 1, \dots, n$ the pointwise limits

$$X_\pm(z) := \lim_{\substack{w \rightarrow z \\ w \in \Omega^\pm}} X(w)$$

satisfy the boundary condition $X_+(z) = X_-(z)G(z)$, compare figure 3 below.

³ In standard textbooks on singular integral equations, see [112, section 39], a *Riemann–Hilbert problem*, named after the original works [71, 72, 125], generally refers to the problem of constructing a function which is analytic in a domain $\Omega \subset \mathbb{C}$, continuous on the closure $\overline{\Omega}$ and with prescribed boundary values on $\partial\Omega$. We will not follow this tradition but instead use the term *Hilbert boundary value problem* like in [138, section 4] as to not confuse ourselves with the original RHPs 2.1 and 3.1.

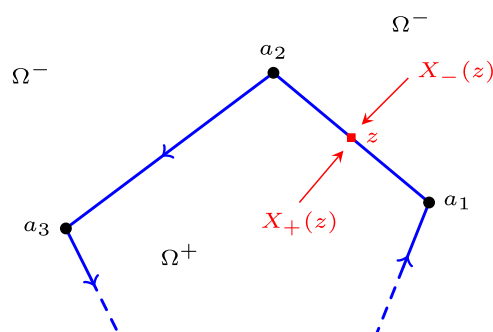


Figure 3. The pointwise limits $X_{\pm}(z)$ at some $z \in (a_i, a_{i+1}), i = 1, \dots, n$ in red.

(c) $X(z)$ is of finite degree at $z = \infty$, that is

$$X(z) = \gamma(z) + \mathcal{O}(z^{-1}), \quad |z| \rightarrow \infty, \tag{4.1}$$

with some vector-valued polynomial $\gamma \in \mathbb{C}^{1 \times p}[z]$. Moreover, in a small neighbourhood of a_i ,

$$\|X(z)\| \leq \frac{C}{|z - a_i|^\alpha}, \quad 0 \leq \alpha < 1, \quad C > 0.$$

With problem 4.1 as starting point, Plemelj first proved its solvability through an application of Fredholm’s theory of integral equations, see [61, 119], at the time a novel analytic tool. We will lay out some of his steps below subject to the following two temporary assumptions:

Assumption 4.2. We seek solutions of problem 4.1 which extend Hölder continuously from either side up to the punctured contour $\Gamma \setminus \{a_1, \dots, a_n\}$.

Assumption 4.3. The invertible jump matrix $G(z), z \in \Gamma$ in condition (b) of problem 4.1 is Hölder continuous on all of Γ and not just on each open segment (a_i, a_{i+1}) separately. In turn, the blow up constraint near $z = a_i$ in problem 4.1(c) becomes unnecessary.

Contingent on these assumptions, problem 4.1 is equivalent to the problem of finding all Hölder continuous functions $X_{\pm}(z) \in \mathbb{C}^{1 \times p}$ defined for $z \in \Gamma$ such that

$$\forall z \in \Omega^- : 0 = \frac{1}{2\pi i} \int_{\Gamma} \frac{X_{-}(\lambda)G(\lambda)}{\lambda - z} d\lambda;$$

and

$$\forall z \in \Omega^+ : 0 = -\frac{1}{2\pi i} \int_{\Gamma} \frac{X_{-}(\lambda)}{\lambda - z} d\lambda + \gamma(z). \tag{4.2}$$

Indeed, by Cauchy’s theorem⁴ and Plemelj’s own 1908 formulæ [120] (the Plemelj–Sokhotski formulæ which Plemelj rediscovered) the first constraint in (4.2) holds precisely when $X_{-}(z)G(z), z \in \Gamma$ is the boundary value of a function $X(z) \in \mathbb{C}^{1 \times p}$ which is analytic in Ω^+ and Hölder continuous on $\Omega^+ \cup \Gamma$. Likewise the second constraint in (4.2) holds if and only if

⁴ In the slightly more general form $\oint_{\partial\Omega} f(z)dz = 0$ where $f : \overline{\Omega} \rightarrow \mathbb{C}$ is analytic in the region $\Omega \subset \mathbb{C}$, continuous on the closure $\overline{\Omega}$ and the boundary $\partial\Omega$ of Ω is a rectifiable Jordan curve, see [139].

$X_-(z), z \in \Gamma$ is the boundary value of some $X(z) \in \mathbb{C}^{1 \times p}$, analytic in $\Omega^- \setminus \{\infty\}$, Hölder continuous on $(\Omega^- \setminus \{\infty\}) \cup \Gamma$ and of finite degree at $z = \infty$, in the sense of (4.1). Next, Plemelj noticed that both constraints in (4.2) are equivalent (again by Cauchy’s theorem and the Plemelj–Sokhotski formulæ) to the principal value integral equations,

$$\begin{aligned} X_-(z) &= -\frac{1}{\pi i} \text{pv} \int_{\Gamma} \frac{X_-(\lambda)}{\lambda - z} d\lambda + 2\gamma(z), \\ X_-(z)G(z) &= \frac{1}{\pi i} \text{pv} \int_{\Gamma} \frac{X_-(\lambda)G(\lambda)}{\lambda - z} d\lambda, \quad z \in \Gamma, \end{aligned} \tag{4.3}$$

and from these one obtains in turn the quasi-regular integral equation

$$X_-(z) - \frac{1}{\pi i} \text{pv} \int_{\Gamma} X_-(\lambda) \frac{K(z, \lambda)}{\lambda - z} d\lambda = \gamma(z), \quad z \in \Gamma, \tag{4.4}$$

with the Hölder continuous matrix-valued kernel function

$$K(z, \lambda) := \frac{1}{2} (G(\lambda) - G(z))(G(z))^{-1}, \quad (z, \lambda) \in \Gamma \times \Gamma.$$

Remark 4.4. In more abstract and general terms, the left-hand side in (4.4) defines a singular Fredholm operator

$$(K\phi)(z) := \phi(z) - \frac{1}{\pi i} \text{pv} \int_{\Gamma} \phi(\lambda) \frac{K(z, \lambda)}{\lambda - z} d\lambda, \quad z \in \Gamma,$$

of index zero which acts on Hölder continuous functions defined on Γ , see [112, section 45], and we assume that $K(z, \lambda)$ satisfies a Hölder condition on Γ in both variables. Plemelj did not work with such singular Fredholm equations in [121] since he managed to transform the piecewise constant jumps in problem 4.1 to differentiable ones. In turn, his [121, (4)] is an ordinary Fredholm equation of the second kind which can be studied by Fredholm’s 1903 theory. We will discuss a more generally applicable reduction of piecewise Hölder continuous jumps to Hölder continuous ones and en route lift assumption 4.3 on $G(z)$. In fact, the general theory of singular Fredholm integral equations of the type (4.4) was systematically developed some 30 years after Plemelj’s paper with prominent contributions by Giraud [64], Muskhelishvili–Vekua [113] and Gakhov [62].

We now make (4.4) the starting point for the solvability analysis of problem 4.1. Precisely, given that any continuous solution $X_-(z), z \in \Gamma$ of (4.4) will be automatically Hölder continuous as a consequence of assumption 4.3, we answer the following two questions:

- Q1 Under which conditions is (4.4) solvable in the space of continuous functions on Γ ?
- Q2 Does each continuous solution of (4.4) yield a solution of problem 4.1?

Regarding Q2, we recall that every continuous solution $X_-(z), z \in \Gamma$ of (4.4) will produce a solution of problem 4.1 with assumption 4.3 in place if and only if $X_-(z), z \in \Gamma$ solves both equations in (4.3). Now rephrase (4.3) in terms of the function

$$\widehat{X}(z) := \begin{cases} \frac{1}{2\pi i} \int_{\Gamma} \frac{X_-(\lambda)}{\lambda - z} d\lambda - \gamma(z), & z \in \Omega^+ \\ \frac{1}{2\pi i} \int_{\Gamma} \frac{X_-(\lambda)G(\lambda)}{\lambda - z} d\lambda, & z \in \Omega^- \end{cases}, \tag{4.5}$$

which is analytic on Ω^\pm , Hölder continuous on $\Omega^\pm \cup \Gamma$ and vanishes at $z = \infty$. In fact,

$$(4.3) \text{ holds } \Leftrightarrow \widehat{X}_+(z) = 0 = \widehat{X}_-(z), \quad z \in \Gamma \Leftrightarrow \widehat{X}(z) \equiv 0 \text{ in } \mathbb{C}. \quad (4.6)$$

Furthermore, our central integral equation (4.4) is equivalent to the jump condition

$$\widehat{X}_+(z) = \widehat{X}_-(z)(G(z))^{-1}, \quad z \in \Gamma,$$

which motivates the introduction of the below, after Plemelj the *accompanying*, Hilbert boundary value problem:

Hilbert Boundary Value Problem 4.5 ([121, p 215]). Find all vector-valued functions $\widehat{X} = \widehat{X}(z) \in \mathbb{C}^{1 \times p}$ such that

- (a) $\widehat{X}(z)$ is analytic in Ω^\pm and extends Hölder continuously from either side up to Γ .
- (b) The pointwise limits

$$\widehat{X}_\pm(z) := \lim_{\substack{w \rightarrow z \\ w \in \Omega^\pm}} \widehat{X}(w), \quad z \in \Gamma,$$

satisfy the boundary condition $\widehat{X}_+(z) = \widehat{X}_-(z)(G(z))^{-1}$, $z \in \Gamma$.

- (c) $\widehat{X}(z)$ vanishes at $z = \infty$, that is

$$\widehat{X}(z) = \mathcal{O}(z^{-1}), \quad |z| \rightarrow \infty.$$

The special homogeneous problem 4.5 is useful since it allows us to summarize our previous chain (4.6) as the following exclusive alternative: let $X_-(\lambda)$, $\lambda \in \Gamma$ be a continuous solution of (4.4),

- If $\widehat{X}(z)$ defined in terms of said $X_-(\lambda)$, $\lambda \in \Gamma$ in (4.5) is identically zero, then $X_-(\lambda)$ solves (4.3) and thus produces a solution of the initial problem 4.1.
- If $\widehat{X}(z)$ defined in terms of said $X_-(\lambda)$, $\lambda \in \Gamma$ in (4.5) is not identically zero, then $\widehat{X}(z)$ is a non-trivial solution of the accompanying problem 4.5.

Most importantly we can now formulate Plemelj’s first criterion as response to Q2 above:

Lemma 4.6 ([121, p 215]). If the accompanying problem 4.5 has no non-trivial solutions, then every continuous solution $X_-(z)$, $z \in \Gamma$ of (4.4) yields a solution of problem 4.1.

Evidently, lemma 4.6 does not guarantee solvability of (4.4). Indeed, in order to answer Q1, Plemelj used Fredholm’s theory of integral equations⁵. First, consider two additional homogeneous Hilbert boundary value problems: the *associated* Hilbert boundary value problem which consists in finding $Y = Y(z) \in \mathbb{C}^{1 \times p}$ that is analytic in Ω^\pm , extends Hölder continuously up to Γ , vanishes at infinity and instead of the jump condition in problem 4.5 satisfies

$$Y_+(z) = Y_-(z)(G^\top(z))^{-1}, \quad z \in \Gamma, \quad (4.7)$$

with the matrix transpose $G^\top(z)$ of $G(z)$. Moreover, the problem which accompanies the associated problem (4.7) and which seeks $\widehat{Y} = \widehat{Y}(z) \in \mathbb{C}^{1 \times p}$ as in problem 4.5 but with jump constraint

$$\widehat{Y}_+(z) = \widehat{Y}_-(z)G^\top(z), \quad z \in \Gamma. \quad (4.8)$$

⁵Fredholm’s theorems for singular integral equations of the form (4.4) were proven in [115] and are called Noether’s theorems. If however the index of the underlying Fredholm operator is zero, compare remark 4.4, then those theorems are exactly the same as for standard Fredholm integral equations of the second kind, see [61, 119].

Repeating the logic that took us from problem 4.1 (with assumptions 4.2 and 4.3) to equation (4.4) we easily see that (4.8) leads to the integral equation

$$\hat{Y}_+(z) + \frac{1}{\pi i} \text{pv} \int_{\Gamma} \hat{Y}_+(\lambda) \frac{K^T(\lambda, z)}{\lambda - z} d\lambda = 0, \quad z \in \Gamma, \tag{4.9}$$

and which is the adjoint equation of (4.4). We can now derive Plemelj’s second criterion:

Lemma 4.7 ([121, pp 217, 218]). *If the associated problem (4.7) has no non-trivial solutions, then (4.4) is solvable in the space of continuous functions on Γ for any given $\gamma \in \mathbb{C}^{1 \times p}[z]$.*

Proof. Problem (4.8) accompanies the associated problem (4.7), or equivalently, (4.7) is the accompanying problem of (4.8) (as we invert the jump matrix and homogenize at infinity in the accompanying problems). But by assumption, (4.7) has no non-trivial solutions, so by lemma 4.6 every continuous solution $\hat{Y}_+(z), z \in \Gamma$ of the adjoint equation (4.9) yields a solution of the Hilbert boundary value problem (4.8). Let $\hat{Y}_+(z), z \in \Gamma$ denote such a non-trivial continuous solution of (4.9),⁶ so $\hat{Y}_+(z), z \in \Gamma$ is the boundary value of some analytic function $\hat{Y}(z), z \in \Omega^+$. However, by [115, (7)], the inhomogeneous equation (4.4) is continuously solvable if and only if

$$\int_{\Gamma} \gamma(z) \hat{Y}_+^T(z) dz = 0,$$

and which is guaranteed by our last conclusion and the fact that $\gamma \in \mathbb{C}^{1 \times p}[z]$ (any entire $\gamma(z)$ would do). This concludes the proof of the lemma. □

To summarize, lemmas 4.6 and 4.7 guarantee solvability of our initial problem 4.1 (with assumption 4.3 in place, but no longer assumption 4.2), provided the accompanying problem 4.5 and the associated problem (4.7) are only trivially solvable. Although these two conditions seem peculiar they are in fact easily verifiable from our assumption that $\gamma(z)$ is polynomial and thus our solutions of problem 4.1 meromorphic at $z = \infty$. Indeed, we first recall, see [115, section 2], that the homogeneous version of (4.4) (or any of the other homogeneous equations in this section) has finitely many, say $s \in \mathbb{Z}_{\geq 1}$, linearly independent continuous solutions⁷. Then

Proposition 4.8 ([121, p 219]). *Any solution $X(z)$ of problem 4.1 (with assumption 4.3) has a zero of order at most $s \in \mathbb{Z}_{\geq 1}$ at $z = \infty$.*

Proof. If $X(z)$ has a zero of order $k \in \mathbb{Z}_{\geq 1}$ at $z = \infty$, then

$$X(z), zX(z), z^2X(z), \dots, z^{k-1}X(z)$$

all solve the homogeneous version of problem 4.5 (with assumption 4.3) and thus $X_-(z), zX_-(z), z^2X_-(z), \dots, z^{k-1}X_-(z), z \in \Gamma$ all solve (4.4) with $\gamma \equiv 0$. Since these solutions are linearly independent we must have $k \leq s$ by our previous discussion. □

⁶ If none exists, i.e. the adjoint equation (4.9) has only the trivial solution, then by [115, (3''), (5')], equation (4.4) is continuously solvable for any right-hand side $\gamma(z)$. This is the Fredholm alternative.

⁷ In the abstract setting $K\phi = 0$ of remark 4.4, this statement is part of the Riesz-Schauder theorem, asserting that all eigenspaces of the associated Fredholm integral operator are finite-dimensional.

So we can always find an integer $r \in \mathbb{Z}_{\geq 0}$ such that neither the accompanying problem 4.1 nor the associated problem (4.7) admit solutions with a zero of order greater than r at $z = \infty$. Hence, if $X(z)$ solves problem 4.1 where $\gamma \in \mathbb{C}^{1 \times p}[z]$ satisfies $\deg(\gamma) \leq r$, then

$$X^\circ(z) := \begin{cases} X(z), & z \in \Omega^+ \\ \frac{X(z)}{(z - z_0)^r}, & z \in \Omega^- \end{cases} \text{ for some arbitrary } z_0 \in \Omega^+, \quad (4.10)$$

solves problem 4.1 with jump matrix $G^\circ(z) := (z - z_0)^r G(z)$ and is bounded at $z = \infty$. Most importantly, the central assumptions in lemma 4.6 and 4.7 for the accompanying and associated problem of the X° -Hilbert boundary value (which is problem 4.1 with $X^\circ(z) = \mathcal{O}(1)$ at $z = \infty$ and Hölder continuous jumps on all of Γ) are satisfied:

- (a) The accompanying problem has jump $\widehat{X}_+^\circ(z) = \widehat{X}_-^\circ(z)(z - z_0)^{-r}(G(z))^{-1}$, $z \in \Gamma$, so if this problem has a non-trivial solution (that vanishes at infinity) then

$$\widehat{X}(z) := \begin{cases} \widehat{X}^\circ(z), & z \in \Omega^+ \\ (z - z_0)^{-r} \widehat{X}^\circ(z), & z \in \Omega^- \end{cases}$$

solves problem 4.5 and has a zero of order greater than r at $z = \infty$, contradicting our initial choice for r .

- (b) The associated problem has jump $Y_+^\circ(z) = Y_-^\circ(z)(z - z_0)^{-r}(G^\top(z))^{-1}$, $z \in \Gamma$, so if this problem has a non-trivial solution (that vanishes at infinity) then

$$Y(z) := \begin{cases} Y^\circ(z), & z \in \Omega^+ \\ (z - z_0)^{-r} Y^\circ(z), & z \in \Omega^- \end{cases}$$

solves the associated problem (4.7) and vanishes to order greater than r at $z = \infty$. Again a contradiction to our choice of r .

The X° -Hilbert boundary value problem is thus solvable by lemmas 4.6 and 4.7 and by reversing (4.10) we arrive at

Theorem 4.9 ([121, pp 222, 223]). *There exists $r \in \mathbb{Z}_{\geq 1}$ such that problem 4.1 (with assumption 4.3) is solvable for all $\gamma \in \mathbb{C}^{1 \times p}[z]$ with $\deg(\gamma) \leq r$. Moreover, every solution of said problem is of the form*

$$X(z) = \sum_{i=1}^p \gamma_i X_i(z) + \sum_{i=p+1}^m \gamma_i X_i(z), \quad z \in \mathbb{C} \setminus \Gamma, \quad (4.11)$$

with some constants $\gamma_1, \dots, \gamma_m \in \mathbb{C}$ and where $\{X_i(z)\}_{j=1}^m$ are linearly independent particular solutions of problem 4.1 with

$$\lim_{|z| \rightarrow \infty} z^{-r} X^{ij}(z) = \delta_{ij}, \quad i, j = 1, \dots, p; \quad X_i(z) = [X^{ij}(z)]_{j=1}^p \in \mathbb{C}^{1 \times p},$$

and $\{X_i(z)\}_{i=p+1}^m$ (if there are any) have degree strictly less than r at $z = \infty$.

Proof. Clearly, if $\{X_i\}_{j=1}^m$ are any particular solutions of the jump condition in problem 4.1, then any polynomial linear combination of them will also solve the same jump constraint. Thus, choosing $r \in \mathbb{Z}_{\geq 1}$ sufficiently large as indicated in our previous discussion we can always ensure that problem 4.1 is solvable for the collection of constant polynomials $\gamma_i(z) = [\delta_{ij}]_{j=1}^p$

with $i = 1, \dots, p$. Reversing subsequently (4.10) and taking an arbitrary combination of the so obtained solutions yields the first half in (4.11) with the indicated asymptotic behaviour at $z = \infty$. The second sum in (4.11) stems from the complete system of linearly independent solutions of the homogeneous version of problem 4.1, modulo the inversion of (4.10). \square

Although theorem 4.9 solves problem 4.1, the implicit dependence of (4.11) on $r \in \mathbb{Z}_{\geq 1}$ is a drawback, in particular since r affects $X^\circ(z)$ through (4.10) and it affects the parameter m . In order to bypass these deficiencies Plemelj managed then to construct a *fundamental system* $\{F_i\}_{i=1}^p \subset \mathbb{C}^{1 \times p}$ of solutions to problem 4.1 such that every solution of the same problem is a polynomial linear combination of the form

$$X(z) = \sum_{i=1}^p q_i(z)F_i(z), \quad z \in \mathbb{C} \setminus \Gamma; \quad q_i \in \mathbb{C}[z].$$

In order to present his argument, we first note that the system $\{X_i\}_{i=1}^p$ in (4.11) is linearly independent over $\mathbb{C}[z]$ because of its asymptotic normalization. With $r \in \mathbb{Z}_{\geq 1}$ as stated in theorem 4.9, we then observe that among all solutions of the form (4.11) some have the lowest degree at $z = \infty$ since a solution of problem 4.1 cannot vanish to all orders at $z = \infty$. Let $F_1(z)$ denote one of the solutions of lowest degree $-\nu_1 \in \mathbb{Z}$ at $z = \infty$ and $F_2(z)$ one of the solutions of problem 4.1 of lowest degree $-\nu_2 \geq -\nu_1$ at $z = \infty$ such that $F_2 \notin \text{span}\{F_1\}$.⁸ Next, we let $F_3(z)$ denote one of the solutions of problem 4.1 of lowest degree $-\nu_3 \geq -\nu_2$ at $z = \infty$ such that $F_3 \notin \text{span}\{F_1, F_2\}$ and note that this process may be continued until we reach $F_p \notin \text{span}\{F_1, \dots, F_{p-1}\}$ of lowest degree $-\nu_p \geq -\nu_{p-1}$ at $z = \infty$. Indeed, if $F_1(z), \dots, F_k(z)$ with $k < p$ have been constructed in this fashion then we can find yet another solution of problem 4.1 not contained in $\text{span}\{F_1, \dots, F_k\}$. For otherwise, the particular solutions $\{X_i\}_{i=1}^p$ in (4.11) would be contained in $\text{span}\{F_1, \dots, F_k\}$ and thus be linearly dependent over $\mathbb{C}[z]$, contradicting our previous discussion. In summary, the above procedure leads to p fundamental solutions

$$F_1(z), F_2(z), \dots, F_p(z), \tag{4.12}$$

of problem 4.1 with lowest degrees $-\nu_1, -\nu_2, \dots, -\nu_p$ at $z = \infty$ where $\nu_1 \geq \nu_2 \geq \dots \geq \nu_p$ and $F_k \notin \text{span}\{F_1, \dots, F_{k-1}\}$ for $k = 2, \dots, p$. The central properties of the so constructed fundamental system (4.12) are summarized below.

Theorem 4.10 ([121, p 227]). *There exists $r \in \mathbb{Z}_{\geq 1}$ such that problem 4.1 (with assumption 4.3) is solvable for all $\gamma \in \mathbb{C}^{1 \times p}[z]$ with $\text{deg}(\gamma) \leq r$. Moreover, every solution of said problem is of the form*

$$X(z) = \sum_{i=1}^p q_i(z)F_i(z), \quad z \in \mathbb{C} \setminus \Gamma, \tag{4.13}$$

where $\{F_i\}_{i=1}^p$ is a fundamental system of solutions, $q_i \in \mathbb{C}[z]$ and the corresponding fundamental matrix

$$\Phi(z) := [F^{ij}(z)]_{i,j=1}^p, \quad F_i(z) = [F^{ij}(z)]_{j=1}^p \in \mathbb{C}^{1 \times p}, \quad z \in \mathbb{C} \setminus \Gamma$$

satisfies

⁸ All linear combinations are taken over $\mathbb{C}[z]$.

- (a) $\det \Phi(z) \neq 0$ for all $z \in \mathbb{C}$ including $z \in \Gamma$ with the appropriate limiting values $\Phi_{\pm}(z)$.
- (b) There exists a diagonal matrix $\Lambda \in \mathbb{Z}^{p \times p}$ such that $z^{\Lambda} \Phi(z)$ is invertible at $z = \infty$.

Proof. The first statement of the theorem was already established in theorem 4.9 and (4.13) follows from (a) since by construction of the fundamental matrix, $G(z) = (\Phi_{-}(z))^{-1} \Phi_{+}(z)$, $z \in \Gamma$ and thus for any solution $X(z)$ of problem 4.1 (with assumption 4.3),

$$X_{+}(z)(\Phi_{+}(z))^{-1} = X_{-}(z)(\Phi_{-}(z))^{-1}, \quad z \in \Gamma,$$

i.e. $X(z)(\Phi(z))^{-1}$ is an entire function. Hence, by condition (c) in problem 4.1 we find that $X(z)(\Phi(z))^{-1} \in \mathbb{C}^{1 \times p}[z]$, so (4.13) follows and conversely it is clear that the right-hand side in (4.13) solves the Hilbert boundary value problem 4.1 for any $q_i \in \mathbb{C}[z]$. In order to establish (a) we first note that if $X(z)$ is any solution of problem 4.1 of degree strictly less than $-\nu_k \in \mathbb{Z}$ at $z = \infty$ where $k \in \{2, \dots, p\}$, then necessarily

$$X(z) = \sum_{i=1}^{k-1} r_i(z) F_i(z) \quad \text{for some } r_i \in \mathbb{C}[z]. \tag{4.14}$$

For if such X , i.e. of degree strictly less than $-\nu_k$ at $z = \infty$, were not contained in $\text{span}\{F_1, \dots, F_{k-1}\}$, then our previous selection process of the subsequent $F_k \notin \text{span}\{F_1, \dots, F_{k-1}\}$ with lowest degree $-\nu_k \geq -\nu_{k-1}$ would fail. Returning to (a) we now show that the expression

$$F(z) := \sum_{i=1}^p a_i F_i(z), \quad z \in \mathbb{C} \setminus \Gamma, \tag{4.15}$$

with $a_i \in \mathbb{C}$ not all zero, does not vanish at any point $z \in \mathbb{C}$: if $F(z_0) = 0$ for some $z_0 \notin \Gamma$, then

$$(z - z_0)X(z) = \sum_{i=1}^p a_i F_i(z), \quad z \in \mathbb{C} \setminus \Gamma,$$

where $X(z)$ solves problem 4.1. If a_k denotes the last of the coefficients a_1, \dots, a_p which is non-zero, then the degree of $X(z)$ at $z = \infty$ is strictly less than $-\nu_k$, so (4.14) must hold for the same $X(z)$, i.e. we would have

$$(z - z_0) \sum_{i=1}^{k-1} r_i(z) F_i(z) = \sum_{i=1}^k a_i F_i(z).$$

But this is a contradiction to the requirement $F_k \notin \text{span}\{F_1, \dots, F_{k-1}\}$ and thus $F(z) \neq 0$ for all $z \in \mathbb{C} \setminus \Gamma$. If on the other hand $F(z_0) = 0$ for some $z_0 \in \Gamma$, i.e. either $F_{+}(z_0) = 0$ or equivalently $F_{-}(z_0) = F_{+}(z_0)(G(z_0))^{-1} = 0$, then define

$$X(z) := \frac{F(z)}{z - z_0}, \quad z \in \mathbb{C} \setminus \Gamma.$$

This $\mathbb{C}^{1 \times p}$ -valued function is analytic in $\Omega^{\pm} \setminus \{\infty\}$, of finite degree at $z = \infty$ and extends Hölder continuously from either side to the punctured contour $\Gamma \setminus \{z_0\}$. Thus the above $X(z)$ solves the singular integral equation (4.4) for all $z \in \Gamma \setminus \{z_0\}$ and since $F_{\pm}(z), z \in \Gamma$ are Hölder continuous with $F_{\pm}(z_0) = 0$ we conclude that $X_{-}(z), z \in \Gamma$ as well as $X_{+}(z) = X_{-}(z)G(z), z \in \Gamma$ are Hölder continuous on all of Γ . At this point we simply repeat the previous logic

for $z_0 \notin \Gamma$ and conclude $F(z) \neq 0$ for all $z \in \Gamma$. All together, (4.15) does not vanish for any $z \in \mathbb{C}$ (choosing $F_{\pm}(z)$ for $z \in \Gamma$) and thus $\det \Phi(z) \neq 0$ for all $z \in \mathbb{C}$, as claimed. We are left with (b) and thus the point $z = \infty$ at which $\det \Phi(z)$ might have a pole or may remain finite. However, if $-\nu_i \in \mathbb{Z}$ denotes the degree of $F_i(z)$ at $z = \infty$, then

$$\det [z^{\nu_i} F^{ij}(z)]_{i,j=1}^p$$

is certainly finite at $z = \infty$. If it were zero, then we could find $a_1, \dots, a_p \in \mathbb{C}$ not all zero such that

$$a_1 z^{\nu_1} F_1(z) + a_2 z^{\nu_2} F_2(z) + \dots + a_p z^{\nu_p} F_p(z) = \mathcal{O}(z^{-1}), \quad \text{as } z \rightarrow \infty.$$

With a_k denoting the last of the coefficients a_1, \dots, a_p which is non-zero we then have

$$\begin{aligned} X(z) &:= a_1 z^{\nu_1 - \nu_k} F_1(z) + a_2 z^{\nu_2 - \nu_k} F_2(z) + \dots + a_k F_k(z) \\ &= \mathcal{O}(z^{-1 - \nu_k}), \quad z \rightarrow \infty, \end{aligned}$$

i.e. the degree of $X(z)$ at $z = \infty$ is strictly less than $-\nu_k$. But then necessarily (compare above)

$$X(z) = \sum_{i=1}^{k-1} r_i(z) F_i(z) \quad \text{for some } r_i \in \mathbb{C}[z],$$

and in turn $F_k \in \text{span}\{F_1, \dots, F_{k-1}\}$, a contradiction. All together,

$$z^\Lambda \Phi(z) \quad \text{with } \Lambda := \text{diag}[\nu_1, \dots, \nu_p] \in \mathbb{Z}^{p \times p}$$

is invertible at $z = \infty$. This concludes the proof of the theorem. □

Plemelj’s central theorem 4.10 will be the key to solve RHP 2.1 in interpretation (b), provided we can dispose of assumption 4.3. In [121, pp 229–236] this was achieved by relying on the piecewise constant nature of $G(z)$ in problem 4.1. Here, we shall outline a procedure of Vekua [138, chapter 2] which is more widely applicable and which employs multi-valued analytic functions defined in cut planes, equivalently discontinuous single-valued functions. First, we denote with

$$G(a_j \mp 0), \quad j = 1, \dots, n$$

the left, resp. right limits at the point $a_j \in \Gamma$ as we approach a_j along (a_{j-1}, a_j) , resp. (a_j, a_{j+1}) on Γ , compare figure 2. These limits exist by our definition of $G(z)$ and we call $a_\sigma \in \{a_1, \dots, a_n\}$ with $G(a_\sigma - 0) \neq G(a_\sigma + 0)$ a *point of discontinuity*. Second, we use the eigenvalues $\lambda_1^\sigma, \dots, \lambda_p^\sigma$ of the non-singular matrix $G(a_\sigma - 0)(G(a_\sigma + 0))^{-1}$ at a point of discontinuity a_σ written as

$$j = 1, \dots, p: \quad \lambda_j^\sigma = e^{2\pi i \gamma_j^\sigma}, \quad \gamma_j^\sigma \in \mathbb{C}: \quad -\frac{1}{2} < \text{Re } \gamma_j^\sigma \leq \frac{1}{2}.$$

Third, we fix $z_0 \in \Omega^+$ and let L^σ denote the straight line connecting z_0 with a point of discontinuity a_σ and extending to $z = \infty$. With this convention any branch of

$$d_j^\sigma(z) := (z - z_0)^{\gamma_j^\sigma}, \quad z \in \mathbb{C} \setminus L^\sigma, \quad j = 1, \dots, p,$$

is single-valued and we have

$$\forall w \in \Gamma \setminus \{a_\sigma\} : \lim_{\substack{z \rightarrow w \\ z \in \Omega^+}} d_j^\sigma(z) = \lim_{\substack{z \rightarrow w \\ z \in \Omega^-}} d_j^\sigma(z),$$

as well as

$$\frac{d_j^\sigma(a_\sigma - 0)}{d_j^\sigma(a_\sigma + 0)} = e^{2\pi i \gamma_j^\sigma}.$$

We also need an arbitrary branch of the function $(z - a_\sigma)^{\gamma_j^\sigma}$ defined in \mathbb{C} with a cut along $[a_\sigma, \infty) \subset L^\sigma$ and the function $(\frac{z-a_\sigma}{z-z_0})^{\gamma_j^\sigma}$ uniquely defined in \mathbb{C} with a cut on $[z_0, a_\sigma] \subset L^\sigma$ such that $(\frac{z-a_\sigma}{z-z_0})^{\gamma_j^\sigma} \rightarrow 1$ as $z \rightarrow \infty$. We now come to the heart of Vekua’s method: define the diagonal matrices

$$A_\sigma(z) := \bigoplus_{j=1}^p \begin{pmatrix} z - a_\sigma \\ z - z_0 \end{pmatrix}^{\gamma_j^\sigma}, \quad B_\sigma(z) := \bigoplus_{j=1}^p (z - a_\sigma)^{\gamma_j^\sigma}, \quad C_\sigma(z) := \bigoplus_{j=1}^p d_j^\sigma(z),$$

and note that $A_\sigma(z)$ is analytic in Ω^- whereas any branch of $B_\sigma(z)$ is analytic in Ω^+ . But if $X(z)$ solves problem 4.1 then

$$T(z) := X(z) \begin{cases} M_\sigma^{-1}(B_\sigma(z))^{-1}, & z \in \Omega^+ \\ N_\sigma^{-1}(A_\sigma(z))^{-1}, & z \in \Omega^-, \end{cases} \tag{4.16}$$

for properly chosen $M_\sigma, N_\sigma \in GL(p, \mathbb{C})$, will solve a modified problem 4.1 with jump matrix which is Hölder continuous on the whole segment $(a_{\sigma-1}, a_{\sigma+1})$ including the point $z = a_\sigma$. Once a_σ has been removed in this fashion all remaining points of discontinuity can be treated similarly and we eventually arrive at a Hilbert boundary value problem that satisfies assumption 4.3. So how do we choose M_σ and N_σ ? With (4.16) the jump condition on $\Gamma \setminus \{a_1, \dots, a_n\}$ in problem 4.1 gets replaced by

$$\begin{aligned} T_+(z) &= T_-(z) A_\sigma(z) N_\sigma G(z) M_\sigma^{-1} (B_\sigma(z))^{-1} \\ &= T_-(z) (\mathbb{I} + B_\sigma(z) [(C_\sigma(z))^{-1} N_\sigma G(z) - M_\sigma] M_\sigma^{-1} (B_\sigma(z))^{-1}), \end{aligned}$$

and this motivates the requirements

$$(C_\sigma(a_\sigma + 0))^{-1} N_\sigma G(a_\sigma + 0) = M_\sigma = (C_\sigma(a_\sigma - 0))^{-1} N_\sigma G(a_\sigma - 0) \tag{4.17}$$

to ensure that the jump for $T(z)$ is Hölder continuous on $(a_{\sigma-1}, a_{\sigma+1})$. In fact, with (4.17) in place, the matrix $(C_\sigma(z))^{-1} N_\sigma G(z) - M_\sigma$ will be Lipschitz continuous on $(a_{\sigma-1}, a_{\sigma+1})$, see [112, section 6]. In order to make (4.17) more transparent, we use $C_\sigma(a_\sigma - 0)(C_\sigma(a_\sigma + 0))^{-1} = \bigoplus_{j=1}^p \lambda_j^\sigma$ and rewrite (4.17) as

$$\begin{aligned} N_\sigma G(a_\sigma - 0)(G(a_\sigma + 0))^{-1} N_\sigma^{-1} &= \bigoplus_{j=1}^p \lambda_j^\sigma, \\ M_\sigma &= (C_\sigma(a_\sigma + 0))^{-1} N_\sigma G(a_\sigma + 0), \end{aligned}$$

so N_σ and M_σ in (4.17) exist if $G(a_\sigma - 0)(G(a_\sigma + 0))^{-1}$ is diagonalizable. If this is not the case, then one has to properly modify M_σ, N_σ in (4.16) and work with the Jordan normal form

of $G(a_\sigma - 0)(G(a_\sigma + 0))^{-1}$, see [138, pp 87–93] for details. Either way, a finite number of transformations of the type (4.16) lead to a Hilbert boundary value problem that is amenable to our previous analysis, i.e. a problem whose jump matrix is Hölder continuous along all of Γ^9 and which satisfies a normalization of the form (4.1).

At this point we can come full circle and return to the original RHP 2.1: each function of the fundamental system (4.12) can be analytically continued to any point in $\mathbb{C} \setminus \{a_1, \dots, a_n\}$ via any path that does not pass through any of the singularities. Concretely, if τ_{γ_j} denotes the operator of analytic continuation along the fundamental loop γ_j , compare figure 2, then by problem 4.1 condition (b), any fundamental matrix $\Phi(z)$ satisfies

$$\tau_{\gamma_j}(\Phi)(z) = \Phi(z)G_j, \quad j = 1, \dots, n.$$

But the same is also true for the matrix-valued function

$$\Psi(z) := (z - a_1)^\Lambda \Phi(z), \quad z \in \mathbb{CP}^1 \setminus \Gamma \tag{4.18}$$

which, according to theorem 4.10, is invertible and bounded at $z = \infty$. Using Liouville’s theorem and the piecewise constant, that is z -independent, form of $G(z)$ it now follows that the matrix

$$A(z) := \frac{d\Psi}{dz}(z)(\Psi(z))^{-1}, \quad z \in \mathbb{CP}^1 \setminus \{a_1, \dots, a_n\} \tag{4.19}$$

is single-valued on \mathbb{CP}^1 and analytic everywhere except at the singularities $a_1, \dots, a_n \in \mathbb{C}$. Moreover, using problem 4.1, condition (c), (4.18) and our discussion on Vekua’s regularization, it is clear that a_1, \dots, a_n are *regular singular points* of the system (4.19) in the following sense:

Definition 4.11. Given a system (3.1) where $A(z) \in \mathbb{C}^{p \times p}$ is analytic in a punctured disk $\mathbb{D}_r(z_0) \setminus \{z_0\}$ for some $r > 0$, we say that z_0 is a regular singular point of (3.1) if any solution of the system has at most polynomial growth in a vicinity of z_0 .

Since the system (4.19) also has the given monodromy (G_1, \dots, G_n) , the above analysis solves RHP 2.1 in interpretation (b). Precisely

Theorem 4.12 (Plemelj, 1908). Any matrix group with $n \in \mathbb{Z}_{\geq 1}$ generators $G_1, \dots, G_n \in GL(p, \mathbb{C})$ satisfying the constraint $G_1 \cdot \dots \cdot G_n = \mathbb{I}$ can be realized as the monodromy group of a $p \times p$ linear system (3.1) on \mathbb{CP}^1 having only regular singularities.

Plemelj’s 1908 paper does not stop with theorem 4.12: indeed, while all singularities in a Fuchsian system are always regular singular points (this is Sauvage’s 1886 theorem, see [27, theorem 2.1] or [83, theorem 16.10]), the converse statement is in general false. For instance, the non-Fuchsian system

$$\frac{d\Psi}{dz} = \left\{ -\frac{3}{16z^2} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\} \Psi \tag{4.20}$$

has a fundamental solution of the form

$$\Psi(z) = \frac{1}{4} \begin{bmatrix} 4z & 4z \\ 1 & 3 \end{bmatrix} z^{-\frac{1}{4}} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \quad z \in \mathbb{C} \setminus (-\infty, 0],$$

⁹The choice of γ_j^σ intimately ties to the blow up constraint near $z = a_i$ in problem 4.1 (3), see [138, p 95].

and thus $z = 0$ is a regular singular point of system (4.20). Plemelj was aware of the distinction between regular singular and Fuchsian singular points, thus he applied a gauge procedure that took him from (4.19) to another system with equal monodromy and same singular points. The transformed system is Fuchsian for all except perhaps one singular point. This part in Plemelj's work [121, pp 243–244] is rigorous, unfortunately his subsequent claim that also the remaining singular point can be reduced to a Fuchsian one, and so his claim of having solved RHP 2.1 in interpretation (c), is not. This gap went unnoticed for more than 70 years (Plemelj passed away in 1967, likely unaware of it) until it became clear that his argument for Fuchsian systems is valid only with further constraints in place:

Theorem 4.13 (Kohn, 1983). *If at least one of the generators $G_1, \dots, G_n \in GL(p, \mathbb{C})$ is diagonalizable, then RHP 2.1 for Fuchsian systems on \mathbb{CP}^1 , i.e. RHP 3.1, has a positive solution.*

Still, Plemelj's *almost solution* of RHP 3.1 introduced the idea of rephrasing a dynamical system, here (3.1) having only regular singularities, as a Hilbert boundary value problem. In turn, the analysis of the problem's underlying singular integral equations led to valuable information about the dynamical system itself, here theorem 4.12. We will encounter both themes throughout our discussions of OPSFA-type problems after the upcoming section.

5. Hilbert's 21st problem after 1908

Plemelj's 1908 result became widely accepted as providing a positive answer to RHP 3.1. Thus, in the following 75 years, the field connected to Hilbert's 21st problem shifted its focus towards the effective construction of Fuchsian systems with prescribed monodromy groups. Here are the major results of this period:

- In 1913, Birkhoff [14] revisited Plemelj's paper [121] and set out to simplify his argument based on successive approximations while also extending RHP 2.1 to certain difference equations.
- In the late 1920s, Lappo-Danilevskii [104, 105] solved RHP 3.1 constructively, provided all generators G_i are sufficiently close to the identity matrix. His method expressed solutions of a Fuchsian system and their associated monodromy via convergent series of the system's matrix coefficients. The solution of Hilbert's 21st problem subsequently boiled down to the problem of inverting the series and studying its convergence.
- In 1956, Krylov [100] explicitly solved RHP 3.1 for all 2×2 systems with $n = 3$ singular points. His work made crucial use of Gauss hypergeometric functions.
- In 1957, Röhrh [126] introduced a novel set of algebro-geometric ideas in the analysis of Hilbert's 21st problem. This allowed for a reformulation and generalization of RHP 2.1 to holomorphic vector bundles over Riemann surfaces, nowadays summarized under the umbrella of *Riemann–Hilbert correspondences*. Important early contributions to this field were achieved by Deligne [46], Kashiwara [91, 92] and Mebkhout [110, 111].
- In 1979, Dekkers [45] proved the solvability of RHP 3.1 in case $p = 2$ for an arbitrary number of singularities. His work did not directly address Hilbert's 21st problem (which was thought to have been solved by Plemelj), but contains its positive solution for $p = 2$ as special case.
- In 1982, Erugin [55] considered RHP 3.1 for all 2×2 systems with $n = 4$ singular points. His work established a remarkable connection of this specialized problem and the Painlevé-VI equation. Compare [60, chapter 2, section 1.3] for more details and also [88].

- In 1999, Deift *et al* [41] ($p = 2$) and in 2004, Korotin [97] and Enolski and Grava [52] (general p) solved RHP 3.1 for quasi-permutation monodromy matrices by means of algebro-geometric techniques. These works were published after Bolibrukh’s negative solution, to be discussed in the following.

Notice how all of the above works either yielded a positive solution to some special case of RHP 3.1 or were concerned with generalizations of the same problem to more general Riemann surfaces. However, as we already mentioned, RHP 3.1 is in general unsolvable as there are monodromy representations which cannot be representations of any Fuchsian systems. This surprising fact came to light in the important works of Bolibrukh in 1989, see [18, 20]. In more detail, Bolibrukh’s first counterexample concerns the following 3×3 system with $n = 4$ singular points:

$$\frac{d\Psi}{dz} = A(z)\Psi, \quad A(z) = \begin{bmatrix} 0 & a_{12}(z) & a_{13}(z) \\ 0 & B(z) & \\ 0 & & \end{bmatrix}, \tag{5.1}$$

where

$$a_{12}(z) := \frac{1}{z^2} + \frac{1}{z+1} - \frac{1}{z-\frac{1}{2}}, \quad a_{13}(z) := \frac{1}{z-1} - \frac{1}{z-\frac{1}{2}},$$

and we use the 2×2 block

$$B(z) := \frac{1}{z} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{1}{6(z+1)} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} + \frac{1}{2(z-1)} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} + \frac{1}{3(z-\frac{1}{2})} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}.$$

This system has Fuchsian singularities at $a_1 = -1, a_2 = \frac{1}{2}, a_3 = 1$ and a non-Fuchsian singularity at $a_4 = 0$. Although non-Fuchsian, Bolibrukh then showed that those singularities are regular singular points of (5.1) in the sense of definition 4.11. Moreover he proved that system (5.1) has non-trivial monodromy, but *there exists no Fuchsian system with the same monodromy group and encoded singularity locations*. Thus RHP 3.1 is in general unsolvable. The details of Bolibrukh’s argument can be found in the monograph [5, chapter 2] and they are very different from Plemelj’s analysis of the boundary value problem 4.1. Rather than discussing those techniques we only mention two features of Bolibrukh’s counterexample (5.1). One, the sensitive dependence on the location of the singular points: once slightly perturbed, the answer to RHP 3.1 with the same monodromy can become positive, see [18]. Two, the monodromy representation of system (5.1) is reducible, a fact which intimately links to the following general result by Kostov [98] and Bolibrukh [19, 20]:

Theorem 5.1 (Kostov, Bolibrukh, 1992). *For every irreducible representation (3.2), the Riemann–Hilbert problem 3.1 has a positive solution.*

In the years following 1992 and theorem 5.1, Bolibrukh sharpened his results and derived a further series of sufficient conditions for the positive solvability of RHP 3.1, compare [21], these are conditions formulated in the language of holomorphic vector bundles and we refer the reader to the excellent monograph [5] for details.

In closing of this short section, and our content on the original RHP 2.1, we emphasize that Bolibrukh’s application of methods from complex analytic geometry provided a definite, albeit negative, solution to Hilbert’s 21st problem for Fuchsian systems on $\mathbb{C}P^1$. This problem was

formulated back in 1900 and Bolibrukh lectured on his breakthroughs precisely 94 years after Hilbert at the ICM 1994 in Zürich, sadly a few years later he passed away.

6. Developments tangential to Plemelj’s work—five examples

At this point we readjust our focus and turn away from Hilbert’s 21st problem, the original RHP. Instead we highlight certain developments parallel to Plemelj’s 1908 work on Hilbert boundary value problems and associated singular integral equations which found multiple applications in the OPSFA realm. Those appearances gave rise to an analytic apparatus, the *Riemann–Hilbert techniques*, which we are about to highlight through five examples—section 9 contains another example, but it is of different technical nature and more advanced. Overall, it is important to trace several of these techniques back to their origin [121], namely to Plemelj’s work on RHP 2.1.

6.1. The Wiener–Hopf method

The first member of our analytical toolbox, a.k.a. the Riemann–Hilbert techniques, is the *Wiener–Hopf* method (pioneered by Wiener and Hopf in 1931 [140], but applied implicitly by Carleman before 1931). This method is commonly used in hydrodynamics, diffraction or linear elasticity theory and was originally developed for the investigation of the Milne–Schwarzschild integral equation

$$x > 0: \quad \phi(x) = \frac{1}{2} \int_0^\infty E(|x - y|)\phi(y) \, dy, \quad E(x) := \int_x^\infty \frac{e^{-t}}{t} \, dt, \quad (6.1)$$

which is used in models for radiative processes in astrophysics. Rather than discussing the Wiener–Hopf method in the more general setup

$$\lambda\phi(x) + \int_0^\infty k(x - y)\phi(y) \, dy = f(x), \quad x > 0, \quad (6.2)$$

we will simply use the original problem (6.1). How do we solve (6.1) with Riemann–Hilbert techniques? According to the Wiener–Hopf recipe one first extends (6.1) to the full real line via $\phi(x) := 0$ for $x \leq 0$ and obtains

$$x \in \mathbb{R}: \quad \phi(x) = \int_{-\infty}^\infty K(x - y)\phi(y) \, dy + \psi(x), \quad K(x) := \frac{1}{2}E(|x|), \quad (6.3)$$

where $\psi(x) := 0$ for $x > 0$ and otherwise appropriately chosen for $x \leq 0$ to achieve equality in (6.3). Then one takes a formal Fourier transform of (6.3) and finds

$$\Phi_+(z) = \Phi_+(z)L(z) + \Psi_-(z), \quad z \in \mathbb{R}, \quad (6.4)$$

with

$$\Phi_+(z) := \int_0^\infty e^{izx} \phi(x) \, dx, \quad \Psi_-(z) := \int_{-\infty}^0 e^{izx} \psi(x) \, dx$$

and

$$L(z) := \int_{-\infty}^\infty e^{izx} K(x) \, dx.$$

The subscripts \pm in (6.4) indicate our desire that $\Phi_+(z)$, resp. $\Psi_-(z)$ admit analytic continuation to some part of the upper, resp. lower z -plane—in fact we will seek solutions of (6.1) which do *not* exponentially grow at $x = +\infty$, and this implies that $\Phi_+(z)$ is the limiting value of a function $\Phi(z)$ analytic for $\Im z > 0$ and likewise $\Psi_-(z)$ the limiting value of a function $\Psi(z)$ analytic for $\Im z < 0$. Next, by elementary calculus, we find

$$L(z) = \frac{1}{z} \arctan(z) = \frac{i}{2z} \ln \left(\frac{i+z}{i-z} \right), \quad z \in \mathbb{C} \setminus ((-i\infty, -i] \cup [i, i\infty)), \quad (6.5)$$

and now Riemann–Hilbert techniques enter our calculation.

Hilbert Boundary Value Problem 6.1. *Determine $F(z) \in \mathbb{C}$ such that*

- (a) $F(z)$ is analytic in $\mathbb{C} \setminus \mathbb{R}$ and extends continuously from either side to the real axis.
- (b) On the real axis, the pointwise limits

$$F_{\pm}(z) := \lim_{\epsilon \downarrow 0} F(z \pm i\epsilon)$$

satisfy the boundary condition $F_+(z) = F_-(z)G(z)$ with

$$z \in \mathbb{R}: \quad G(z) = (1 - L(z)) \left(1 + \frac{1}{z^2} \right) \in \mathbb{C}.$$

- (c) Near $z = \infty$ we enforce the asymptotic behaviour

$$F(z) = 1 + \mathcal{O}(z^{-1}), \quad |z| \rightarrow \infty.$$

Since $G(z)$ is Hölder continuous on \mathbb{R} , $G(z) \neq 0$ for all $z \in \mathbb{R}$ and $G(z) = 1 + \mathcal{O}(z^{-1})$ as $z \rightarrow \pm\infty$, the unique solution of problem 6.1 is readily seen to be

$$F(z) = \exp \left[\frac{1}{2\pi i} \int_{-\infty}^{\infty} \ln G(w) \frac{dw}{w-z} \right], \quad z \in \mathbb{R} \setminus \mathbb{C}, \quad (6.6)$$

by the Plemelj–Sokhotski formulæ. The point of (6.6) is that we can use the underlying Wiener–Hopf factorization

$$\frac{F_+(z)}{F_-(z)} = (1 - L(z)) \left(1 + \frac{1}{z^2} \right), \quad z \in \mathbb{R}$$

back in (6.4) and obtain

$$\frac{z^2}{z+i} \Phi_+(z) F_+(z) = (z-i) \Psi_-(z) F_-(z), \quad z \in \mathbb{R}.$$

Here, the equation’s left-hand side defines a function which admits analytic continuation to the upper half-plane and the right-hand side a function which admits analytic continuation to the lower half-plane. Hence, there is an entire function $E(z)$, $z \in \mathbb{C}$ such that

$$\Phi(z) = \frac{(z+i)E(z)}{z^2 F(z)}, \quad \Im z > 0; \quad \Psi(z) = \frac{E(z)}{(z-i)F(z)}, \quad \Im z < 0.$$

But $\Phi_+(z)$ is a Fourier integral, so we find $\Phi(z) = \mathcal{O}(z^{-1})$ as $z \rightarrow \infty$, $\Im z > 0$ and thus with problem 6.1 and Liouville’s theorem, $E(z) \equiv ic \in \mathbb{C}$. All together,

$$\Phi(z) = \frac{(z+i)ic}{z^2 F(z)}, \quad \Im z > 0.$$

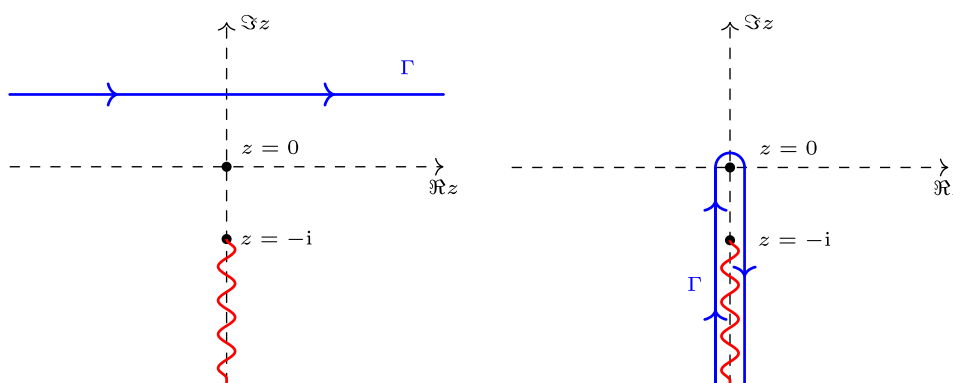


Figure 4. The integration contour Γ for (6.7) shown in blue with branch cut in red. On the left our choice for $x < 0$ and on the right for $x > 0$.

This result at hand, we would now like to return to $\phi(x)$ via an inverse Fourier transform

$$\phi(x) = \frac{1}{2\pi} \int_{\Gamma} \Phi(z) e^{-izx} dz, \tag{6.7}$$

but how do we choose the contour $\Gamma \subset \mathbb{C}$ in (6.7) to achieve equality? Notice that by the Plemelj–Sokhotski formula, $F_+(0) = \frac{1}{\sqrt{3}} \neq 0$, so the continuation of $\Phi(z)$ to the lower half-plane will have a double pole at $z = 0$ and a branch cut extending from $-i$ to $-i\infty$ along the imaginary axis, compare (6.5). Hence, if $x < 0$, we choose Γ as straight line in the upper half-plane and find that (6.7) vanishes by residue theorem. If $x > 0$, then we wrap Γ around the negative imaginary axis and encircle the double pole at $z = 0$ as well as the branch cut, compare figure 4.

Computing the residue and parametrizing the integral along the branch cut, one then obtains the well-known solution formula

$$\begin{aligned} \phi(x) = c \left(\sqrt{3} \left[1 + x - \frac{1}{\pi} \int_0^\infty \left\{ \frac{1}{1 - L(w)} - 1 - \frac{3}{w^2} \right\} \frac{dw}{1 + w^2} \right] \right. \\ \left. + \frac{\pi}{2} e^{-x} \int_0^\infty \frac{e^{-xw}}{F(-i - iw)} \frac{dw}{(2 + w - \frac{1}{2} \ln(\frac{2+w}{w}))^2 + \frac{\pi^2}{4}} \right), \quad x > 0. \end{aligned}$$

In summary of this short subsection, using scalar Hilbert boundary value problems we can solve integral equations of the form (6.2) by the Wiener–Hopf method. This method has found countless applications and is particularly well-suited for PDE problems with semi-infinite boundaries, e.g. the Sommerfeld diffraction problem, see for instance [114] and [33, chapter 16] for further details.

6.2. The integrable systems revolution of the late 1960s

The classical (think of Euler, Hamilton, Jacobi, Liouville, Neumann and Kowaleskaya) field of integrable systems found renewed interest in the late 1960s when it became clear that several nonlinear PDEs in $1 + 1$ dimensions can be *integrated* by the inverse scattering method. Examples are the Korteweg–de Vries equation, the nonlinear Schrödinger equation (NLS), and the sine-Gordon equation, among many others. Important early contributions to this remarkable technique are due to Gardner *et al* [63], due to Lax [106], due to Zakharov and Faddeev [143],

and due to Zakharov and Shabat [144]. For a solid first introduction to the inverse scattering method we refer the interested reader to the monographs [1, 12, 56, 116], here we shall discuss only one example and showcase the appearance of Riemann–Hilbert techniques: consider the defocusing NLS

$$iy_t + y_{xx} - 2|y|^2y = 0, \quad y = y(x, t) : \mathbb{R}^2 \rightarrow \mathbb{C}. \tag{6.8}$$

How do we solve the corresponding initial value problem with Cauchy data $y(x, 0) = y_0(x) \in \mathcal{S}(\mathbb{R})$ in the Schwartz space on the real line? Well, one first computes the reflection coefficient $r(z) \in \mathcal{S}(\mathbb{R})$ associated to y_0 through the direct scattering transform. There is not much freedom in this computation as the map $y_0 \rightarrow r$ is a bijection from $\mathcal{S}(\mathbb{R})$ onto $\mathcal{S}(\mathbb{R}) \cap \{r : \sup_{z \in \mathbb{R}} |r(z)| < 1\}$, see [11]. After that, one considers the so-called inverse scattering problem which is most conveniently formulated as the following Hilbert boundary value problem—this insight grew out of the works by Manakov, Shabat and Zakharov in the mid 1970s:

Hilbert Boundary Value Problem 6.2. *For any $x \in \mathbb{R}$ and $t > 0$, determine $X(z) = X(z; x, t) \in \mathbb{C}^{2 \times 2}$ such that*

- (a) $X(z)$ is analytic in $\mathbb{C} \setminus \mathbb{R}$ and extends continuously from either side to the real axis.
- (b) On the real axis, the pointwise limits

$$X_{\pm}(z) := \lim_{\epsilon \downarrow 0} X(z \pm i\epsilon)$$

satisfy the boundary condition $X_+(z) = X_-(z)G(z)$ with

$$z \in \mathbb{R} : \quad G(z) = \begin{bmatrix} 1 - |r(z)|^2 & -\bar{r}(z)e^{-2i(2tz^2+xz)} \\ r(z)e^{2i(2tz^2+xz)} & 1 \end{bmatrix} \in \text{GL}(2, \mathbb{C}). \tag{6.9}$$

- (c) Near $z = \infty$ we enforce the asymptotic behaviour

$$X(z) = \mathbb{I} + X_1z^{-1} + X_2z^{-2} + \mathcal{O}(z^{-3}), \quad |z| \rightarrow \infty;$$

$$X_k = X_k(x, t) = \left[X_k^{ij}(x, t) \right]_{i,j=1}^2,$$

where $\mathbb{I} \in \text{GL}(2, \mathbb{C})$ is the identity matrix.

Indeed, provided this problem is solvable, its (unique) solution solves the PDE (6.8) with initial condition $y(x, 0) = y_0(x)$ via the formula

$$y(x, t) = 2iX_1^{12}(x, t). \tag{6.10}$$

In order to arrive at (6.10) we use a slight modification of Plemelj’s argument that lead him to the linear system (4.19): the jump matrix $G(z), z \in \Gamma$ in problem 4.1 was piecewise constant whereas (6.9) presents us with the other extreme, namely a jump matrix which depends on z, x and t . However, assuming that problem 6.2 is solvable, we can define

$$Y(z) \equiv Y(z; x, t) := X(z; x, t)e^{-i(2tz^2+xz)\sigma_3}, \quad z \in \mathbb{C} \setminus \mathbb{R},$$

and then conclude that both,

$$\begin{aligned} A(z) &:= \frac{\partial Y}{\partial x}(z)(Y(z))^{-1} \quad \text{and} \quad B(z) := \frac{\partial Y}{\partial t}(z)(Y(z))^{-1}, \\ (z, x, t) &\in \mathbb{C} \times \mathbb{R} \times (0, \infty) \end{aligned} \tag{6.11}$$

are single-valued, entire functions in $z \in \mathbb{C}$. Moreover, using problem 6.2, condition (c) and Liouville’s theorem we obtain the explicit formulæ

$$A(z) \equiv A(z; x, t) = -iz\sigma_3 + 2i \begin{bmatrix} 0 & X_1^{12} \\ -X_1^{21} & 0 \end{bmatrix},$$

$$B(z) \equiv B(z; x, t) = -2iz^2\sigma_3 + 4iz \begin{bmatrix} 0 & X_1^{12} \\ -X_1^{21} & 0 \end{bmatrix} + 4i \begin{bmatrix} -X_1^{12}X_1^{21} & X_2^{12} - X_1^{12}X_1^{22} \\ -X_2^{12} + X_1^{21}X_1^{11} & X_1^{21}X_1^{12} \end{bmatrix},$$

in term of the matrix coefficients $X_k^{ij} = X_k^{ij}(x, t)$ and the third Pauli matrix $\sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. The overdetermined system (6.11) forms one of the celebrated Lax pairs for the PDE (6.8), indeed writing out the Frobenius integrability condition for $\frac{\partial Y}{\partial x} = AY$, $\frac{\partial Y}{\partial t} = BY$, equivalently the compatibility condition

$$AB - BA = \frac{\partial B}{\partial x} - \frac{\partial A}{\partial t},$$

we find entrywise the coupled partial differential equations

$$2i(X_1^{12})_x = 4(X_2^{12} - X_1^{12}X_1^{22}),$$

$$-2i(X_1^{21})_x = 4(X_2^{21} - X_1^{21}X_1^{11}),$$

$$4i(X_2^{12} - X_1^{12}X_1^{22})_x - 2i(X_1^{12})_t = -16X_1^{12}X_1^{21}X_1^{12},$$

$$-4i(X_2^{21} - X_1^{21}X_1^{11})_x + 2i(X_1^{21})_t = -16X_1^{21}X_1^{12}X_1^{21}.$$

However, the coefficients back in condition (c) of problem 6.2 satisfy $\text{tr}X_1 = 0$ and $X_k^{11} = \overline{X_k^{22}}, X_k^{21} = \overline{X_k^{12}}, k \in \mathbb{Z}_{\geq 1}$. Thus the above four coupled equations boil down to

$$y_x = 4(X_2^{12} - X_1^{12}X_1^{22}), \quad 2|y|^2y = 4(X_2^{12} - X_1^{12}X_1^{22})_x + iy_t,$$

where we used (6.10) and so after substitution into one another

$$iy_t + y_{xx} - 2|y|^2y = 0,$$

which is the defocusing NLS (6.8). Summarizing this short computation, provided problem 6.2 is solvable, the linear Hilbert boundary value problem solves the nonlinear evolution equation (6.8) via the formula (6.10). This is a far reaching generalization of Plemelj’s idea to construct solutions of certain linear ODE systems in terms of solutions of Hilbert boundary value problems, nevertheless the common approach to both problems is clearly visible. In section 7 below we will address the solvability question of problem 6.2 which requires a different toolset than the one used by Plemelj in his analysis of problem 4.1.

6.3. Painlevé special function theory

Painlevé functions form a family of special functions which is widely regarded as the substitute for classical special functions in nonlinear mathematical physics (such Airy, Bessel or Hypergeometric functions). Although Painlevé transcendents are non-expressible in terms of a finite number of contour integrals over elementary, elliptic or finite genus algebraic functions,

several of their key analytic and asymptotic properties can be studied with Riemann–Hilbert techniques. These techniques are therefore in a way the analogues of contour integral representations and steepest descent asymptotic methods used in the analysis of classical special functions. For a comprehensive introduction to Painlevé special functions and Riemann–Hilbert techniques associated with them, including several references to their applications in mathematical physics, we recommend the two monographs [60, 67]. Similar to the last subsection we will showcase the Riemann–Hilbert approach to one particular Painlevé equation, namely

$$u_{xx} = xu + 2u^3, \quad u = u(x) : \mathbb{R} \rightarrow \mathbb{R}, \tag{6.12}$$

the homogeneous Painlevé-II equation, which, to a certain extent, is a nonlinear Airy equation. How do we solve the associated initial value problem? Well, in the Riemann–Hilbert approach to (6.12) one parametrizes the solutions $u = u(x)$ of (6.12) *not* in terms of Cauchy data, but in terms of the monodromy data of an associated linear system of ordinary differential equations: precisely, define the generically two-dimensional real manifold

$$\mathcal{S} := \{(s_1, s_2, s_3) \in \mathbb{C}^3 : s_1 - s_2 + s_3 + s_1 s_2 s_3 = 0, \quad s_1 = \bar{s}_3, \quad s_2 = \bar{s}_2\},$$

and introduce with $s_{k+3} = -s_k, k = 1, 2, 3$ for $k = 1, \dots, 6$ the triangular matrices

$$S_k := \begin{bmatrix} 1 & 0 \\ s_k & 0 \end{bmatrix}, \quad k \equiv 1 \pmod 2; \quad S_k := \begin{bmatrix} 1 & s_k \\ 0 & 1 \end{bmatrix}, \quad k \equiv 0 \pmod 2.$$

Now consider the below Hilbert boundary value problem.

Hilbert Boundary Value Problem 6.3. *For any $x \in \mathbb{R}$, determine $X(z) = X(z; x, s) \in \mathbb{C}^{2 \times 2}$ with $s := (s_1, s_2, s_3) \in \mathcal{S}$ such that*

- (a) $X(z)$ is analytic in $\mathbb{C} \setminus \bigcup_{i=1}^6 \Gamma_i$ and continuous on the closed sectors $\bar{\Omega}_i$. Compare figure 5 for the six rays Γ_i and the sectors Ω_i in between them.
- (b) On the rays $\Gamma_i, i = 1, \dots, 6$ the pointwise limits

$$X_{\pm}(z) := \lim_{\substack{w \rightarrow z \\ w \in \Omega_{\pm}}} X(w), \quad \Omega_+ := \Omega_{i+1}, \quad \Omega_- := \Omega_i, \quad \Omega_7 \equiv \Omega_1,$$

satisfy the boundary condition $X_+(z) = X_-(z)G(z)$ with

$$z \in \Gamma_i : \quad G(z) = e^{-\theta(z,x)\sigma_3} S_i e^{\theta(z,x)\sigma_3} \in \text{GL}(2, \mathbb{C}); \quad \theta(z, x) := i \left(\frac{4}{3} z^3 + xz \right).$$

- (c) Near $z = \infty$ we enforce the asymptotic behaviour

$$X(z) = \mathbb{I} + X_1 z^{-1} + X_2 z^{-2} + \mathcal{O}(z^{-3}), \quad |z| \rightarrow \infty;$$

$$X_k = X_k(x, s) = \left[X_k^{ij}(x, s) \right]_{i,j=1}^2.$$

The route between problem 6.3 and (6.12) goes as follows: for any choice of parameters $(s_1, s_2, s_3) \in \mathcal{S}$ the Hilbert boundary value problem 6.3 is meromorphically with respect to x solvable, see [22]. In turn the unique solution to said problem leads to a solution of (6.12) via

$$u(x) \equiv u(x|s) = 2X_1^{12}(x, s), \tag{6.13}$$

which is the analogue of the NLS formula (6.10). Additionally, (6.13) satisfies $\bar{u}(x) = u(\bar{x})$, i.e. the solution is real-valued on the real axis and conversely, every on the real axis real-valued

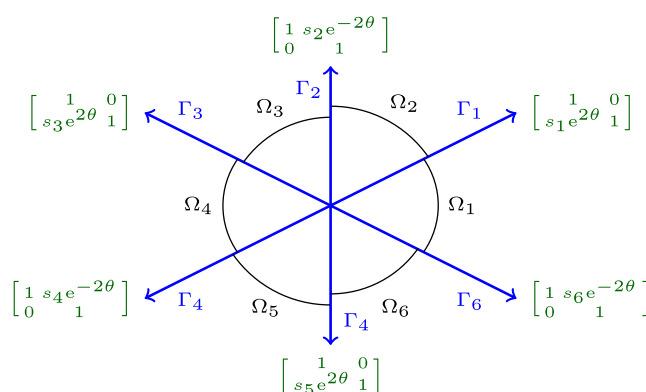


Figure 5. The oriented contour $\bigcup_{i=1}^6 \Gamma_i$ in blue together with the sectors Ω_i in between. The six rays are $\Gamma_i = \{z \in \mathbb{C} : \arg z = \frac{\pi}{6} + \frac{\pi}{3}(i - 1)\}$ and we indicate the values of $G(z)$ on them in green.

solution of (6.12) admits a unique Riemann–Hilbert representation (6.13) for suitable s . In short, we do parametrize solutions of (6.12) in terms of the monodromy data $(s_1, s_2, s_3) \in \mathcal{S}$ using a bijection between the initial value solution space of (6.12) and \mathcal{S} .

The derivation of (6.13), and thus the Riemann–Hilbert approach to the second Painlevé equation is due to Flaschka and Newell [57] and Jimbo *et al* [85–87] in the early 1980s. In a nutshell, we use again Plemelj’s basic idea: suppose problem 6.3 is solvable and set

$$Y(z) \equiv Y(z; s, s) := X(z; x, s)e^{-\theta(z,x)\sigma_3}, \quad z \in \mathbb{C} \setminus \bigcup_{i=1}^6 \Gamma_i.$$

Then both,

$$A(z) := \frac{\partial Y}{\partial z}(Y(z))^{-1} \quad \text{and} \quad B(z) := \frac{\partial Y}{\partial x}(z)(Y(z))^{-1}, \quad (z, x) \in \mathbb{C} \times (\mathbb{R} \setminus D)$$

are single-valued, entire functions in $z \in \mathbb{C}$ away from a discrete set $D \subset \mathbb{R}$. Hence, using condition (c) in problem 6.3 and Liouville’s theorem, we obtain the explicit formulæ

$$\begin{aligned} A(z) \equiv A(z; x, s) &= -4iz^2\sigma_3 + 8iz \begin{bmatrix} 0 & X_1^{12} \\ -X_1^{21} & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} -ix - 8iX_1^{12}X_1^{21} & 8i(X_2^{12} - X_1^{12}X_1^{22}) \\ -8i(X_2^{21} - X_1^{21}X_1^{11}) & ix + 8iX_1^{21}X_1^{12} \end{bmatrix}, \\ B(z) \equiv B(z; x, s) &= -iz\sigma_3 + 2i \begin{bmatrix} 0 & X_1^{12} \\ -X_1^{21} & 0 \end{bmatrix}, \end{aligned}$$

in terms of the matrix coefficients X_k and from those, reading the compatibility condition

$$AB - BA = \frac{\partial B}{\partial z} - \frac{\partial A}{\partial x},$$

entrywise, the following coupled ordinary differential equations,

$$\begin{aligned} (X_1^{12})_x &= -2i(X_2^{12} - X_1^{12}X_1^{22}), \\ (X_1^{21})_x &= 2i(X_2^{21} - X_1^{21}X_1^{11}), \\ -2(X_2^{12} - X_1^{12}X_1^{22})_x &= X_1^{12}(-ix - 8iX_1^{12}X_1^{21}), \\ -2(X_2^{21} - X_1^{12}X_1^{11})_x &= X_1^{21}(ix + 8iX_1^{21}X_1^{12}). \end{aligned}$$

However problem 6.3 also possesses certain implicit symmetries that yield $\text{tr} X_1 = 0$, $X_k^{12} = (-1)^{k-1}X_k^{21}$, $k \in \mathbb{Z}_{\geq 1}$ and which simplify the above four equations to

$$u_x = -4i(X_2^{12} - X_1^{12}X_1^{22}), \quad i(xu + 2u^3) = 4(X_2^{12} - X_1^{12}X_1^{22})_x,$$

where we used (6.13). Substituting both equations into one another we find

$$u_{xx} = xu + 2u^3,$$

and therefore Painlevé-II (6.12). In summary, the Hilbert boundary value problem 6.3 linearizes the nonlinear Painlevé-II ODE and therefore shares common ground with (4.19) and (6.8). Before we move to our next example it will be important to mention that the monodromy data $s = (s_1, s_2, s_3) \in \mathcal{S}$ forms a complete set of first integrals for (6.12). Building on this feature, one can in principle use the highly transcendental equations

$$s_i = s_i(x, u, u_x) \equiv \text{const}, \quad i = 1, 2, 3 \tag{6.14}$$

in the further analysis of the Painlevé-II transcendents. This idea is at the heart of the *isomonodromy method* and the corresponding direct monodromy approach to Painlevé equations as pioneered by Its, Novokshenov, Kapaev and Kitaev in the mid 1980s, see [81] or [60, chapter 7]. Their approach relies on complex WKB techniques applied to problem 6.3 and certain *a priori* assumptions on the behaviour of $u = u(x)$. We will later on highlight a different method, the inverse monodromy approach, a.k.a. Deift–Zhou method [37], which (in the extended form of Deift *et al* [44]) has become extremely popular in nonlinear mathematical physics since its discovery back in 1993. This method is suitable for the analysis of Painlevé functions and other problems that we are about to discuss.

6.4. The Heisenberg antiferromagnet

Our next example is concerned with a quantum mechanical model for an antiferromagnetic spin chain. It was introduced by Lieb *et al* [107] in 1961 and is known as the spin- $\frac{1}{2}$ XY model, equivalently the Heisenberg XX0 antiferromagnet. In this model $\frac{1}{2}$ -spins are situated on a one-dimensional periodic and isotropic lattice, we allow only nearest neighbour interactions between the spins, neighbouring spins tend to point in *opposite* directions and a moderate external transverse magnetic field influences the spins. Subject to these four constraints the Hamiltonian of the model equals, see [107, (2.1)],

$$\mathcal{H} := \sum_{j=1}^N \sigma_j^x \sigma_{j+1}^x + \sum_{j=1}^N \sigma_j^y \sigma_{j+1}^y - h \sum_{j=1}^N \sigma_j^z,$$

where we set the nearest neighbour coupling constant to unity, $0 < h < 2$ denotes the strength of the external magnetic field and

$$\sigma^x \equiv \sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^y \equiv \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^z = \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

are the Pauli spin matrices. The paper [107] computed the model’s ground state, all elementary excitations and the free energy through creation-annihilation operator techniques. Here we shall focus on an important correlation function, the so-called *emptiness formation probability*

$$P_n := \text{Prob} \{ \text{there are } n \in \mathbb{Z}_{\geq 1} \text{ adjacent parallel spins up in the ground state} \},$$

which was evaluated in the thermodynamic limit $N \rightarrow \infty$ for fixed h in [54, (10.2), (10.3)] by the quantum inverse scattering method. The result is the Fredholm determinant formula

$$\begin{aligned} P_n &= \det \left(1 - U_n \upharpoonright_{L^2(J, dx)} \right) \\ &:= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_J \cdots \int_J \det [U_n(x_i, x_j)]_{i,j=1}^k dx_1 \cdots dx_k, \end{aligned} \tag{6.15}$$

where $U_n : L^2(J, dx) \rightarrow L^2(J, dx)$ denotes the trace class operator with kernel

$$U_n(\lambda, \mu) := \frac{1}{2\pi i} \frac{1}{\sinh(\lambda - \mu)} \left\{ 1 - \left(\frac{e^{2\lambda} + i}{e^{2\lambda} - i} \right)^n \left(\frac{e^{2\mu} - i}{e^{2\mu} + i} \right)^n \right\},$$

acting on the interval $J := (-\Lambda, \Lambda) \subset \mathbb{R}$ with $\cosh 2\Lambda = \frac{2}{h} > 1$. The central analytical challenges associated with P_n consist in accessing its large n asymptotic behaviour and in identifying an underlying integrable system for it. Both questions yield physically relevant information for the spin chain model, so how do we go about them? First, one factorizes the operator U_n as

$$U_n = E^{-1} D_n K_n (E^{-1} D_n)^{-1}, \tag{6.16}$$

where D_n denotes multiplication $(D_n f)(z) := z^{-\frac{n}{2}} e^{\lambda(z)} (z + i)$ with the invertible change of variables $\lambda = \lambda(z)$ defined as

$$\begin{aligned} z &:= -i \frac{e^{2\lambda} - i}{e^{2\lambda} + i}, \quad z(J) = \{z \in \mathbb{C} : |z| = 1, \arg \in (\phi, 2\pi - \phi)\}, \\ e^{i\phi} &= z(-\Lambda), \quad \phi \in \left(\frac{\pi}{2}, \pi \right), \end{aligned} \tag{6.17}$$

E is the evaluation $(Ef)(z) := f(\lambda(z))$ and $K_n : L^2(\Gamma, |dz|) \rightarrow L^2(\Gamma, |dz|)$ the trace class operator with kernel

$$K_n(z, w) := \frac{z^{\frac{n}{2}} w^{-\frac{n}{2}} - z^{-\frac{n}{2}} w^{\frac{n}{2}}}{2\pi i (z - w)}, \tag{6.18}$$

acting on the arc $\Gamma := z(J) \subset \mathbb{C}$ defined in (6.17) and shown in figure 6. Second, one notices that (6.18) is a particular instance of an *integrable* kernel, i.e. a kernel of the form

$$K(x, y) = \frac{\sum_{j=1}^N f_j(x) g_j(y)}{x - y}, \quad x, y \in \Gamma \quad \text{with} \quad \sum_{j=1}^N f_j(x) g_j(x) = 0, \quad x \in \Gamma,$$

for some functions $f_j, g_j \in L^\infty(\Gamma, |dz|)$, $1 \leq j \leq N$. The associated integral operator $K : L^2(\Gamma, |dz|) \rightarrow L^2(\Gamma, |dz|)$ has many remarkable properties, most importantly $(1 - K)^{-1}$ (if existent) can be computed in terms of the solution of a naturally associated Hilbert boundary value problem. This insight was formalized in the important 1990 paper [78] by Its *et al*, see [34] for a concise review, and reads in case of (6.18) as follows:

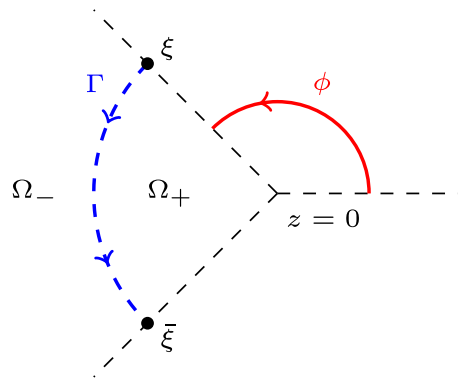


Figure 6. The oriented contour Γ in blue together with the sectors Ω_{\pm} . The endpoints are $\xi := e^{i\phi} = z(-\Lambda)$, $\bar{\xi} = e^{-i\phi}$ and the angle ϕ in (6.17) is shown in red.

Hilbert Boundary Value Problem 6.4. For any $n \in \mathbb{Z}_{\geq 1}$ and $0 < h < 2$, determine $X(z) = X(z; n, h) \in \mathbb{C}^{2 \times 2}$ such that

- (a) $X(z)$ is analytic in $\mathbb{C} \setminus \Gamma$ and extends continuously from either side to $\Gamma \setminus \{\xi, \bar{\xi}\}$. See figure 6 for the jump contour Γ and its endpoints ξ and $\bar{\xi}$.
- (b) On the arc $\Gamma \subset \mathbb{C}$ the pointwise limits

$$X_{\pm}(z) := \lim_{\substack{w \rightarrow z \\ w \in \Omega_{\pm}}} X(w),$$

satisfy the boundary condition $X_+(z) = X_-(z)G(z)$ with

$$z \in \Gamma : \quad G(z) = \begin{bmatrix} 0 & -z^n \\ z^{-n} & 2 \end{bmatrix} \in \text{GL}(2, \mathbb{C}).$$

- (c) In a small neighbourhood of $z = \xi$ and $z = \bar{\xi}$ we impose the blow-up constraint

$$\|X(z)\| \leq C \left| \ln \left| \frac{z - \xi}{z - \bar{\xi}} \right| \right|, \quad C > 0.$$

- (d) Near $z = \infty$ we require the asymptotic normalization

$$X(z) = \mathbb{I} + X_1 z^{-1} + \mathcal{O}(z^{-2}), \quad |z| \rightarrow \infty;$$

$$X_k = X_k(n, h) = \left[X_k^{ij}(n, h) \right]_{i,j=1}^2.$$

Problem 6.4 and the emptiness formation probability (6.15) are related in the following way: as shown in [40, proposition 6.12], the Hilbert boundary value problem 6.4 is uniquely solvable for any $n \in \mathbb{Z}_{\geq 1}$, $h \in (0, 2)$ and its solution leads to the below recursion for P_n ,

$$\frac{P_{n+1}}{P_n} = X^{11}(0; n, h), \tag{6.19}$$

in terms of the (11)-entry of the solution to problem 6.4. This identity is more complicated than the representation formulæ (6.10), (6.13) and its derivation does not use a Lax pair as in the NLS or Painlevé-II example. Nevertheless, (6.19) enables us to use Riemann–Hilbert techniques in the analysis of the spin chain model, i.e. Plemelj’s basic idea occurs once more.

In order to obtain (6.19), we first use the operator factorization (6.16) and Sylvester’s determinant identity,

$$P_n = \det \left(1 - U_n \upharpoonright_{L^2(J, dx)} \right) = \det \left(1 - K_n \upharpoonright_{L^2(\Gamma, |dz|)} \right).$$

Next we realize that

$$K_{n+1}(z, w) = K_n(z, w) + \frac{1}{2\pi i} z^{\frac{n}{2}} w^{-\frac{n}{2}-1} := K_n(z, w) + (\alpha_n \otimes \beta_n)(z, w),$$

where $\alpha_n \otimes \beta_n$ is the finite rank integral operator on $L^2(\Gamma, |dz|)$ with kernel $(\alpha_n \otimes \beta_n)(z, w) = \alpha_n(z)\beta_n(w)$ in terms of $\alpha_n(z) := \frac{1}{2\pi i} z^{\frac{n}{2}}$ and $\beta_n(w) := w^{-\frac{n}{2}-1}$. But [78] showed that problem 6.4 is solvable if and only if $(1 - K_n)^{-1}$ exists, so we can further simplify (6.15),

$$\begin{aligned} P_{n+1} &= \det \left(1 - K_n - \alpha_n \otimes \beta_n \upharpoonright_{L^2(\Gamma, |dz|)} \right) \\ &= P_n \det \left(1 - (1 - K_n)^{-1} (\alpha_n \otimes \beta_n) \upharpoonright_{L^2(\Gamma, |dz|)} \right), \end{aligned}$$

and evaluate the remaining Fredholm determinant with the general theory of finite rank operators,

$$\frac{P_{n+1}}{P_n} = 1 - \int_{\Gamma} \left((1 - K_n)^{-1} \alpha_n \right) (z) z^{-\frac{n}{2}} \frac{dz}{z}. \tag{6.20}$$

At this point we use the general fact, see [78] or (7.5) below, that the solution of problem 6.4 can be expressed in integral form, similar to Plemelj’s quasi-regular equation (4.4),

$$\begin{aligned} X(z) &= \mathbb{I} - \int_{\Gamma} F(w) g^{\top}(w) \frac{dw}{w - z}, \quad z \notin \Gamma; \\ F_i(z) &:= \left((1 - K_n)^{-1} f_i \right) (z), \quad z \in \Gamma, \end{aligned}$$

with $F(z) := (F_1(z), F_2(z))^{\top}$, $g(z) := (g_1(z), g_2(z))^{\top}$ and

$$\begin{aligned} f_1(z) &:= \frac{1}{2\pi i} z^{\frac{n}{2}}, & f_2(z) &:= -\frac{1}{2\pi i} z^{-\frac{n}{2}}, & g_1(z) &:= z^{-\frac{n}{2}}, & g_2(z) &:= z^{\frac{n}{2}}, \\ \arg z &\in (\phi, 2\pi - \phi). \end{aligned}$$

Setting $z = 0$ in the integral formula for $X(z)$ and comparing with (6.20) yields

$$\frac{P_{n+1}}{P_n} = X^{11}(0; n, h),$$

i.e. the representation (6.19). In summary, the Hilbert boundary value problem 6.4 linearizes to a certain extent (note that (6.19) is a formula for the discrete logarithmic derivative of P_n and not P_n itself) the highly transcendental Fredholm determinant (6.15). We will display the usefulness of (6.19) after the next subsection.

6.5. Invariant random matrix models

Our fifth and last example in this section concerns the statistical behaviour of eigenvalues of a random matrix drawn from a particular ensemble. We will focus on an especially well-studied ensemble, the so called *unitary ensemble* $\mathcal{M}(n)$ of $n \times n$ Hermitian matrices equipped with the probability measure

$$P(M) dM = \frac{1}{Z_{n,N}} e^{-N \text{tr } V(M)} dM, \quad Z_{n,N} := \int_{\mathcal{M}(n)} e^{-N \text{tr } V(M)} dM < \infty. \quad (6.21)$$

Here, dM is the Haar measure on $\mathcal{M}(n) \simeq \mathbb{R}^{n^2}$, $N \in \mathbb{Z}_{\geq 1}$ a scaling parameter and $V : \mathbb{R} \rightarrow \mathbb{R}$ assumed to be real analytic satisfying the growth condition

$$\frac{V(x)}{\ln(x^2 + 1)} \rightarrow \infty \quad \text{as } |x| \rightarrow \infty,$$

in order to ensure that (6.21) is a bona fide model. Measures of the form (6.21) appeared seemingly first in the work of Hurwitz in 1897 [73], they appeared in a paper by Wishart in 1928 [142] and in particular in Wigner’s work in the 1950s [141] who used them to model the statistical properties of highly excited energy levels of complex nuclei, see [47] for an excellent historic review of the subject. A classical fact of the setup (6.21) is that the location of the eigenvalues of a matrix drawn from the unitary ensemble form a determinantal point process. This detail allows us to compute several of the key statistical properties of the random matrix model from the underlying Christoffel–Darboux kernel

$$K_{n,N}(x, y) = e^{-\frac{N}{2}V(x)} e^{-\frac{N}{2}V(y)} \sum_{k=0}^{N-1} \frac{1}{h_k^2} p_k(x)p_k(y), \quad x, y \in \mathbb{R}, \quad (6.22)$$

which is defined in terms of the sequence $\{p_n(x)\}_{n=0}^\infty \subset \mathbb{C}[x]$ of monic orthogonal polynomials with respect to the measure $e^{-NV(x)} dx$ supported on \mathbb{R} , i.e.

$$\int_{-\infty}^\infty p_j(x)p_k(x)e^{-NV(x)} dx = h_k^2 \delta_{jk}, \quad p_j(x) = x^j + \mathcal{O}(x^{j-1}).$$

Indeed, focussing on one important statistical property of the model (6.21), the *gap probability*,

$$E_n(s) := \text{Prob} \{M \in \mathcal{M}(n) \text{ has no eigenvalues in the interval } (-s, s) \subset \mathbb{R}, s > 0\},$$

admits the determinantal representation

$$E_n(s) = \det \left(1 - K_{n,N} \upharpoonright_{L^2((-s,s),dx)} \right), \quad (6.23)$$

where $K_{n,N}$ is the finite rank operator with kernel (6.22). Formula (6.23) goes back to the seminal works of Dyson, Gaudin and Mehta in the 1970s and it is the analogue of the emptiness formation probability formula (6.15) in the spin chain model. Thus, in order to compute the likelihood of large eigenvalue gaps from (6.23), we need to carefully analyse the asymptotic behaviour of the kernel (6.22). However, in all but a few cases, the orthogonal polynomials that built the kernel are not known explicitly, so how can we effectively use (6.22)? Well, as it happens, the polynomials $\{p_n(x)\}_{n=0}^\infty$ and the kernel $K_{n,N}(x, y)$ itself can be characterized through a Hilbert boundary value problem. This remarkable fact was discovered by Fokas *et al* [58, 59] in the early 1990s and the details are as follows:

Hilbert Boundary Value Problem 6.5. For any $n \in \mathbb{Z}_{\geq 0}$ and $N \in \mathbb{Z}_{\geq 1}$, determine $X(z) = X(z; n, N) \in \mathbb{C}^{2 \times 2}$ such that

- (a) $X(z)$ is analytic in $\mathbb{C} \setminus \mathbb{R}$ and extends continuously from either side to the real axis.
- (b) On the real axis, the pointwise limits

$$X_\pm(z) := \lim_{\epsilon \downarrow 0} X(z \pm i\epsilon)$$

satisfy the boundary condition $X_+(z) = X_-(z)G(z)$ with

$$z \in \mathbb{R} : G(z) = \begin{bmatrix} 1 & e^{-NV(z)} \\ 0 & 1 \end{bmatrix} \in \text{GL}(2, \mathbb{C}). \tag{6.24}$$

(c) Near $z = \infty$ we enforce the asymptotic behaviour

$$X(z) = (\mathbb{I} + X_1 z^{-1} + \mathcal{O}(z^{-2})) z^{n\sigma_3}, \quad |z| \rightarrow \infty.$$

Fokas, Its and Kitaev showed that this problem is uniquely solvable for a given $n \in \mathbb{Z}_{\geq 0}$ if and only if the n th monic OP $p_n(x)$ exists. Moreover, in case when its solvability is ensured, then we have for any $x, y \in \mathbb{R}$,

$$p_n(x) = X^{11}(x; n, N)$$

and

$$K_{n,N}(x, y) = e^{-\frac{N}{2}V(x)} \left[\frac{(X_+(y))^{-1} X_+(x)}{2\pi i(x-y)} \right]^{21} e^{-\frac{N}{2}V(y)}, \tag{6.25}$$

in terms of the (11)-entry of $X(z)$ for the polynomial and the (21)-entry of the matrix product for the kernel. Here are the details: assuming the solvability of problem 6.5, the jump constraint (6.24) states that the first column of $X(z)$ is an entire function,

$$\begin{aligned} X_+^{11}(z) &= X_-^{11}(z), & X_+^{21}(z) &= X_-^{21}(z), & X_+^{12}(z) &= X_-^{12}(z) + X_-^{11}(z)e^{-NV(z)} \\ X_+^{22}(z) &= X_-^{22}(z) + X_-^{21}(z)e^{-NV(z)}, & z \in \mathbb{R}. \end{aligned} \tag{6.26}$$

But the asymptotic normalization at $z = \infty$ enforces at the same time that

$$X^{11}(z) = z^n + X_1^{11} z^{n-1} + \mathcal{O}(z^{n-2}), \quad X^{21}(z) = X_1^{21} z^{n-1} + \mathcal{O}(z^{n-2}),$$

and

$$\begin{aligned} X^{12}(z) &= X_1^{12} z^{-n-1} + \mathcal{O}(z^{-n-2}), \\ X^{22}(z) &= z^{-n} + X_1^{22} z^{-n-1} + \mathcal{O}(z^{-n-2}). \end{aligned} \tag{6.27}$$

Hence, $X^{11}(z)$ must be a monic polynomial of degree n and $X^{21}(z)$ a polynomial of degree at most $n - 1$ by Liouville’s theorem. Returning with this information to (6.26) and using the Plemelj–Sokhotski formula we then find for $z \in \mathbb{C} \setminus \mathbb{R}$,

$$X^{12}(z) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} X^{11}(w) e^{-NV(w)} \frac{dw}{w-z}$$

and

$$X^{22}(z) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} X^{21}(w) e^{-NV(w)} \frac{dw}{w-z}. \tag{6.28}$$

But geometric progression yields for $z \notin \mathbb{R}$,

$$\begin{aligned} X^{12}(z) &= -\frac{1}{2\pi i} \sum_{k=0}^{n-1} \frac{1}{z^{k+1}} \int_{-\infty}^{\infty} X^{11}(w) w^k e^{-NV(w)} dw \\ &\quad + \frac{z^{-n}}{2\pi i} \int_{-\infty}^{\infty} X^{11}(w) e^{-NV(w)} \frac{w^n}{w-z} dw \end{aligned}$$

and thus after comparison with (6.27)

$$\int_{-\infty}^{\infty} X^{11}(w)w^k e^{-NV(w)} dw = 0, \quad k = 0, \dots, n - 1.$$

Since $X^{11}(z)$ has already been found to be a monic polynomial, it must therefore be the n th monic orthogonal polynomial $p_n(x)$ for the measure $e^{-NV(x)}dx$ on \mathbb{R} , so the first identity in (6.25) follows. The kernel identity is slightly more complicated and involves $X^{21}(z)$: from the integral formula of $X^{22}(z)$ in (6.28), by geometric progression for $z \in \mathbb{C} \setminus \mathbb{R}$,

$$\begin{aligned} X^{22}(z) = & -\frac{1}{2\pi i} \sum_{k=0}^{n-2} \frac{1}{z^{k+1}} \int_{-\infty}^{\infty} X^{21}(w)w^k e^{-NV(w)} dw \\ & + \frac{z^{-n+1}}{2\pi i} \int_{-\infty}^{\infty} X^{21}(w)e^{-NV(w)} \frac{w^{n-1}}{w-z} dw, \end{aligned}$$

so after comparison with (6.27),

$$\begin{aligned} \int_{-\infty}^{\infty} X^{21}(w)w^k e^{-NV(w)} dw = 0, \quad k = 0, \dots, n - 2; \\ \frac{i}{2\pi} \int_{-\infty}^{\infty} X^{21}(w)w^{n-1} e^{-NV(w)} dw = 1. \end{aligned}$$

This shows that $X^{21}(z)$ must be proportional to $p_{n-1}(z)$, i.e. we have for some $\gamma_{n-1} \in \mathbb{C} \setminus \{0\}$,

$$X^{21}(z) = \gamma_{n-1}p_{n-1}(z), \quad z \in \mathbb{C},$$

and thus by orthogonality and the very definition of the sequence $\{p_n(x)\}_{n=0}^{\infty}$,

$$\begin{aligned} -2\pi i &= \int_{-\infty}^{\infty} X^{21}(w)w^{n-1} e^{-NV(w)} dw \\ &= \gamma_{n-1} \int_{-\infty}^{\infty} p_{n-1}^2(w)e^{-NV(w)} dw = \gamma_{n-1}h_{n-1}^2. \end{aligned}$$

Next we note that any solution of problem 6.5 must be unimodular in the entire plane, i.e. $\det X(z) \equiv 1$ for all $z \in \mathbb{C}$. Indeed, the function $\det X(z)$ is analytic for $z \in \mathbb{C} \setminus \mathbb{R}$ and continuous on the closed upper and lower half-planes with

$$\det X_+(z) = \det X_-(z), \quad z \in \mathbb{R}$$

since $\det G(z) \equiv 1$. Hence, $\det X(z)$ is entire and with $\det X(z) \rightarrow 1$ from problem 6.5, condition (c) we indeed arrive at $\det X(z) \equiv 1$. This in mind, we now use the Christoffel–Darboux identity in (6.22),

$$K_{n,N}(x, y) = e^{-\frac{N}{2}V(x)}e^{-\frac{N}{2}V(y)} \frac{1}{h_{n-1}^2} \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x - y} \tag{6.29}$$

and our formulæ $X^{11}(z) = p_n(z)$, $X^{21}(z) = \gamma_{n-1}p_{n-1}(z)$ with $\gamma_{n-1}h_{n-1}^2 = -2\pi i$. A simple matrix multiplication and comparison with (6.29) leads to the second identity in (6.25). Once more we summarize our discussion: the Hilbert boundary value problem (6.5) allows us to access both, orthogonal polynomials and their Christoffel–Darboux kernel. In turn, the Riemann–Hilbert characterization allows us to rigorously analyse relevant statistical quantities in the random

matrix model (6.21), provided we can efficiently derive a large n asymptotic expansion for $X(z; n, N)$. The development of such an efficient scheme has been one of the many highlights over the past 30 years in the Riemann–Hilbert toolbox and we shall return to this important advancement after the next section.

7. Hilbert boundary value problems in L^2 -spaces

The boundary values problems encountered in the last section are different from Plemelj’s initial problem 4.1 in that they include jump contours extending to infinity, jump contours with open ends and jump contours that self-intersect. It is therefore not obvious how Plemelj’s solvability proof can be lifted to the problems of section 6. For this reason we now give a short overview of the relevant theory in which Plemelj’s analysis of (4.4) in a space of continuous functions is extended to spaces of integrable functions—for a more thorough discussion we direct the interested reader to the articles [42, 145] by Zhou and Deift, Zhou. Here are the basic two assumptions of this section:

Assumption 7.1. Let Σ be a contour consisting of a finite union of smooth oriented curves in $\mathbb{C}\mathbb{P}^1$ with finitely many self-intersections.

Assumption 7.2. Let $G : \Sigma \rightarrow GL(p, \mathbb{C})$ be a map such that $G, G^{-1} \in L^\infty(\Sigma, |dz|), G - \mathbb{I} \in L^2(\Sigma, |dz|)$ and G has zero winding.

Now consider the following extended L^2 -Hilbert boundary value problem.

Hilbert Boundary Value Problem 7.3 (The L^2 -problem). Given a pair (Σ, G) subject to assumptions 7.1 and 7.2, determine two $\mathbb{C}^{p \times p}$ -valued functions $X_\pm \in L^2(\Sigma, |dz|)$ such that

(a) There exists a $\mathbb{C}^{p \times p}$ -valued function $H \in L^2(\Sigma, |dz|)$ so that

$$X_\pm(z) - \mathbb{I} = \frac{1}{2\pi i} \int_\Sigma H(w) \frac{dw}{w - z_\pm},$$

where

$$\int_\Sigma H(w) \frac{dw}{w - z_\pm} := \lim_{\lambda \rightarrow z_\pm} \int_\Sigma H(w) \frac{dw}{w - \lambda}$$

denotes the a.e. existent non-tangential limits from the \pm side of Σ .

(b) We have $X_+(z) = X_-(z)G(z)$ a.e. on Σ .

Note how the abstract problem 7.3 captures all our examples in section 6 and Plemelj’s problem 4.1: by condition (a), and the Plemelj–Sokhotski formula in $L^2(\Sigma, |dz|)$, see [132, chapter II], the functions X_\pm are the L^2 -boundary values of the extension

$$X(z) := \mathbb{I} + \frac{1}{2\pi i} \int_\Sigma H(w) \frac{dw}{w - z}, \quad z \in \mathbb{C} \setminus \Gamma. \tag{7.1}$$

But (7.1) is analytic in $\mathbb{C} \setminus \Gamma$, has non-tangential limits $X_\pm(z) \in L^2(\Sigma, |dz|)$ a.e. on Γ , those satisfy $X_+(z) = X_-(z)G(z)$ a.e. on Γ and we have $X(z) \rightarrow \mathbb{I}$ as $z \rightarrow \infty$.

Remark 7.4. If $f \in L^q(\Sigma, |dz|)$ for some $1 \leq q < \infty$, then the Cauchy operators

$$(C_\pm f)(z) := \int_\Sigma f(w) \frac{dw}{w - z_\pm}, \quad z \in \Sigma$$

exist a.e. for $z \in \Sigma$ and if $1 < q < \infty$,

$$\|C_{\pm}f\|_{L^q(\Sigma, |dz|)} \leq c\|f\|_{L^q(\Sigma, |dz|)}, \quad c = c(q) > 0. \tag{7.2}$$

This says that the Cauchy operators are, in particular, bounded linear operators on $L^2(\Sigma, |dz|)$.

Next, we derive the analogue of Plemelj’s singular integral equation (4.4) for problem 7.3: if $X_{\pm}(z)$ solve problem 7.3, then by the Plemelj–Sokhotski formula and condition (b),

$$H(z) = X_+(z) - X_-(z) = X_-(z)(G(z) - \mathbb{I}) \quad \text{a.e. on } \Sigma. \tag{7.3}$$

On the other hand

$$K(z) := \frac{1}{2\pi i} \int_{\Sigma} X_-(w)(G(w) - \mathbb{I}) \frac{dw}{w - z}, \quad z \in \mathbb{C} \setminus \Sigma$$

satisfies, again by Plemelj–Sokhotski,

$$K_{\pm}(z) = -X_-(z) + \frac{1}{2\pi i} \int_{\Sigma} X_-(w)(G(w) - \mathbb{I}) \frac{dw}{w - z_{\pm}} \pm X_{\pm}(z) \quad \text{a.e. on } \Sigma,$$

so by property (a) and (7.3), $\rho(z) := X_-(z)$ solves the singular integral equation

$$\rho(z) = \mathbb{I} + \frac{1}{2\pi i} \int_{\Sigma} \rho(w)(G(w) - \mathbb{I}) \frac{dw}{w - z} \quad \text{a.e. on } \Sigma. \tag{7.4}$$

Moreover, starting from (7.4) we can construct a solution of problem 7.3:

Proposition 7.5. *Every solution pair $X_{\pm} \in L^2(\Sigma, |dz|)$ of problem 7.3 leads to a solution of the singular integral equation*

$$X_-(z) = \mathbb{I} + \frac{1}{2\pi i} \int_{\Sigma} X_-(w)(G(w) - \mathbb{I}) \frac{dw}{w - z}.$$

Conversely, if $\rho(z)$ solves (7.4) with $\rho - \mathbb{I} \in L^2(\Sigma, |dz|)$, then the non-tangential limits $X_{\pm}(z)$ of

$$X(z) := \mathbb{I} + \frac{1}{2\pi i} \int_{\Sigma} \rho(w)(G(w) - \mathbb{I}) \frac{dw}{w - z}, \quad z \in \mathbb{C} \setminus \Sigma \tag{7.5}$$

solve problem 7.3.

Proof. By assumption 7.1, $\rho(G - \mathbb{I}) \in L^2(\Sigma, |dz|)$ if $\rho - \mathbb{I} \in L^2(\Sigma, |dz|)$, thus the non-tangential limits $X_{\pm}(z)$ of (7.5) are easily seen to satisfy condition (a) in problem 7.3. Moreover, by Plemelj–Sokhotski, we find from (7.4) that $\rho(z) = Y_-(z)$ a.e. on Σ and so, also by Plemelj–Sokhotski, from (7.5),

$$X_+(z) - X_-(z) = \rho(z)(G(z) - \mathbb{I}) = X_-(z)(G(z) - \mathbb{I}) \quad \text{a.e. on } \Sigma$$

which yields condition (b) in problem 7.3. This concludes the proof. □

Having established in (7.4) the analogue of Plemelj’s (4.4), how do we solve this equation? In short, we also rely on a Fredholm Alternative argument, similar to lemmas 4.6 and 4.7: let

$$G(z) = (\mathbb{I} - V^-(z))^{-1} (\mathbb{I} + V^+(z)) \quad \text{a.e. on } \Sigma, \tag{7.6}$$

denote a pointwise factorization of G with functions $(\mathbb{I} \pm V^\pm)^{\pm 1} - \mathbb{I} \in L^2(\Sigma, |dz|) \cap L^\infty(\Sigma, |dz|)$. This factorization allows us to rewrite the central integral equation (7.4) in the compact form

$$\begin{aligned} \text{a.e. on } \Sigma : \quad & \rho(z) = \mathbb{I} + (C_w \rho)(z), \\ (C_w f)(z) := & (C_+(fV^-))(z) + (C_-(fV^+))(z) \end{aligned}$$

with the Cauchy operators C_\pm of remark 7.4. In turn we have the following central result which states that the operator $1 - C_w$ is injective if and only if a certain homogeneous version of problem 7.3 is only trivially solvable.

Proposition 7.6 ([145, proposition 3.3]). $\rho \in L^2(\Sigma, |dz|)$ solves the homogeneous equation

$$\rho(z) = (C_w \rho)(z) \quad \text{a.e. on } \Sigma$$

if and only if $X_\pm(z) := \rho(z)(\mathbb{I} \pm V^\pm(z)), z \in \Sigma$ solves the homogeneous version of problem 7.3, i.e. the problem of determining $X_\pm \in L^2(\Sigma, |dz|)$ so that

(a) There exists a $\mathbb{C}^{p \times p}$ -valued function $H \in L^2(\Sigma, |dz|)$ such that

$$X_\pm(z) = \frac{1}{2\pi i} \int_\Sigma H(w) \frac{dw}{w - z_\pm}.$$

(b) We have $X_+(z) = X_-(z)G(z)$ a.e. on Σ .

In order to state the Fredholm alternative needed in the solvability analysis of problem 7.3 we note that all our jump matrices encountered in this note possess much more regularity than what is assumed in assumption 7.2, in fact the matrices are piecewise smooth, admit local analytical continuations and at possible intersection points satisfy a cyclic constraint in the style of (3.4)—for instance in problem 6.3 the total monodromy at $z = 0$ is trivial¹⁰. These properties (see [60, pp 103–104] for a formalization) together with assumptions 7.1 and 7.2 ensure that the following far reaching generalizations of lemmas 4.6 and 4.7 hold true, at least for all matrix-valued Hilbert boundary value problems in this article.

Proposition 7.7 ([145, proposition 4.3]). *If the factors $V^\pm(z) = V^\pm(z; \zeta)$ in (7.6) depend on a parameter ζ analytically, then either $(1 - C_w)^{-1}$ is meromorphic in ζ or $1 - C_w$ is invertible for no ζ .*

Proposition 7.7 is an analytic Fredholm theorem and together with proposition 7.5 yields two central results, given that $1 - C_w$ is a Fredholm operator of index zero, compare remark 4.4.

Theorem 7.8 ([145, proposition 4.2]). *If the factors $V^\pm(z) = V^\pm(z; \zeta)$ in (7.6) depend on a parameter ζ analytically, then either $X(z; \zeta)$ in (7.5) is meromorphic in ζ , or the associated ζ -modified problem 7.3 is not solvable for any ζ .*

Theorem 7.9 (Zhou’s vanishing lemma). *The L^2 -Hilbert boundary value problem 7.3 is solvable if and only if the corresponding homogeneous version of the problem, see proposition 7.6, is only trivially solvable.*

¹⁰The importance of such cyclic constraints, viewed as formal power series identities, cannot be overemphasized, see [12, section 25] for further details and context.

In the usual, smooth, setting of a Hilbert boundary value problem, the homogeneous version of it has the same analyticity and jump behaviour, but the asymptotic normalization at $z = \infty$ is replaced by

$$X(z) = \mathcal{O}(z^{-1}), \quad |z| \rightarrow \infty.$$

This requirement is completely analogous to the behaviours which Plemelj enforced in his accompanying and associated problems, compare for instance problem 4.5.

Before moving to the next section, we will give one application of theorem 7.9 to problem 6.2 and en route verify that the defocusing NLS with Cauchy data $y_0(x) \in \mathcal{S}(\mathbb{R})$ is solvable: suppose $X(z)$ solves problem 6.2 but with condition (c) replaced by

$$X(z) = \mathcal{O}(z^{-1}), \quad |z| \rightarrow \infty. \tag{7.7}$$

Define $Y(z) := X(z)X^\dagger(\bar{z})$ for $z \in \mathbb{C} \setminus \mathbb{R}$ with X^\dagger denoting the conjugate transpose matrix of X . By property (a) in problem 6.2 we find that Y is analytic in the upper z -plane, continuous down to the real line and by (7.7) decays of $\mathcal{O}(z^{-2})$ as $z \rightarrow \infty$ in the upper z -plane. Hence, by Cauchy's theorem, we have $\int_{\mathbb{R}} Y_+(z) dz = 0$, and, adding to this its conjugate transpose, find in turn

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} \left(X_+(z)X_-^\dagger(z) + X_-(z)X_+^\dagger(z) \right) dz \\ &= \int_{-\infty}^{\infty} X_-(z) \left(G(z) + G^\dagger(z) \right) X_-^\dagger(z) dz \\ &= 2 \int_{-\infty}^{\infty} X_-(z) \begin{bmatrix} 1 - |r(z)|^2 & 0 \\ 0 & 1 \end{bmatrix} X_-^\dagger(z) dz. \end{aligned}$$

Now read off the diagonal entries in the last matrix equation and conclude that

$$0 = \int_{-\infty}^{\infty} \left(|X_-^{k1}(z)|^2 (1 - |r(z)|^2) + |X_-^{k2}(z)|^2 \right) dz, \quad k = 1, 2,$$

so by the fact that $r \in \mathcal{S}(\mathbb{R}) \cap \{r : \sup_{z \in \mathbb{R}} |r(z)| < 1\}$, see [11], we have $X_-(z) \equiv 0$ on \mathbb{R} by continuity of $X_-(z)$. In turn, from condition (b) in problem 6.2 also $X_+(z) \equiv 0$ on \mathbb{R} and thus all together $X(z)$ is analytic in the upper z -plane, continuous in the closed upper z -plane and

$$\|X(z)\| \leq C, \quad \Im z > 0, \quad \sup_{z \in \mathbb{R}} \|X_+(z)\| = 0.$$

Using Carlson's theorem [130, theorem 5.1.2 and corollary 5.1.3], this allows us to conclude that $X(z) \equiv 0$ for $\text{Im } z \geq 0$ and with the same logic for $\Im z \leq 0$ all together $X(z) \equiv 0$. By Zhou's vanishing lemma, i.e. theorem 7.9, we then deduce that problem 6.2 is solvable for any $x \in \mathbb{R}$, $t > 0$ and so the solution of the initial value problem (6.8) with $y_0(x) \in \mathcal{S}(\mathbb{R})$ exists.

A similar application of theorem 7.9 to problem 6.3 does not yield the desired solvability result for general choices of $s \in \mathcal{S}$. Indeed, for some choices $s \in \mathcal{S}$, the solution $X(z) = X(z; x, s)$ of problem 6.3 ceases to exist for certain $x \in \mathbb{R}$ and this is because the corresponding Painlevé-II transcendent $u = u(x|s)$ has poles on the real axis, compare subsection 8.1. On the other hand, the solvability of problems 6.4 and 6.5 is always guaranteed and can be established without using theorem 7.9.

8. A Hilbert boundary value problem as Swiss army knife

By now we have seen several Hilbert boundary value problems in OPSFA related problems and we have acquainted ourselves to a certain extent with their abstract solvability theory in sections 4 and 7. Still, besides providing a positive solution to RHP 2.1 in interpretation (b), a reader unaccustomed to Hilbert boundary value problems might not yet grasp their usefulness: after all, although they seem to linearize nonlinear dynamical systems such as (6.8) and (6.12) or rephrase certain Fredholm determinants and orthogonal polynomials, those underlying problems (i.e. problems 6.2–6.5) are in all, but a few trivial, cases *not explicitly solvable*. So what is the whole point of a Hilbert boundary value problem? Well, the last 30 years have clearly shown that Hilbert boundary value problems really possess all the fundamental properties of a contour integral formula known, and appreciated, in classical special function theory. To the point, Hilbert boundary value problems underlie a large class of integrable models and as such allow us to

- systematically derive dynamical systems (continuous ones, discrete ones or hybrids) for the quantities under consideration. This feature was first used by Plemelj in his work [121] on Hilbert's 21st problem. We have seen the same approach in action while discussing (6.8) and (6.12) and will further discuss it in the upcoming subsections. In these modern applications one must mention the pioneering roles of Shabat and Zakharov in the derivation of nonlinear partial differential equations from Hilbert boundary value problems, the role of Krein [99] in obtaining differential equations from integral equations and last, but not least, the role of Its [96, pp 377, 417] and his emphasis on studying correlation functions in statistical mechanics and quantum field theories via Hilbert boundary value problems and their associated differential equations.
- analyse the models asymptotically in their thermodynamical limits. It is this feature which puts Hilbert boundary value problems on the same ground as their linear counterparts, i.e. contour integral formulæ, and which has turned the *Riemann–Hilbert approach* to OPSFA problems into an unprecedented success story over the past 30 years. However the relevant asymptotic techniques did not grow over night: early progress—using Gelfand-Levitan type integral equations—on the asymptotic analysis of nonlinear wave equations that are solvable by the inverse scattering method was achieved in the early 1970s, namely by Šabat [128], Manakov [109] and Ablowitz and Newell [2]. These works were not always fully rigorous but the gaps were covered in the 1980s by Buslaev, Sukhanov, by Novokshenov and by Novokshenov, Sukhanov, see [39, p 182] for references. The first step of using directly a Hilbert boundary value problem for asymptotic questions seems to have originated in the works of Manakov [109] and Its [74] on the NLS equation. Their techniques were subsequently extended by Its, Petrov to the sine-Gordon equation, by Bikbaev, Its to the Landau–Lifshitz equation and by Its, Novokshenov to the modified KdV equation, see again [39, p 182] for references to the relevant papers published in the mid 1980s. However, aside from the NLS case, these works ‘only’ managed to asymptotically localize the initial Hilbert boundary value problems around certain special points. These points are the analogues of stationary phase points in the classical steepest descent method, but the model problems near them could not be explicitly solved, unfortunately. As it turned out, those local Hilbert boundary value problems are precisely the ones one faces in the isomonodromy deformation approach to the Painlevé transcendents as pioneered by Its, Novokshenov, Kapaev and Kitaev, see [81]. Thus, up to the early 1990s, one was able to use Hilbert boundary value problems in the asymptotic analysis of nonlinear wave and Painlevé equations, however one needed certain *a priori* information about the solution's

behaviour. This pitfall was then bypassed in the groundbreaking method of Deift and Zhou [37] in 1993, the *nonlinear steepest descent method*. This method, in the 1997 extended version of Deift, Venakides, Zhou [44], has become the standard tool in the asymptotic analysis of Hilbert boundary value problems and all upcoming asymptotic results in this section have been originally derived from it or re-derived with it.

We will now summarize the technical essence of the Deift–Zhou nonlinear steepest descent method: in complete analogy to the classical steepest descent method, the nonlinear steepest descent method exploits the analytic and asymptotic properties of a Hilbert boundary value problem’s jump matrix $G(z)$. For instance, in case of problem 6.2 the jump matrix in (6.9) is highly oscillatory when x or t tend to infinity. Thus one would like to transform these oscillations to exponentially small contributions through the use of a contour deformation argument. And indeed, noticing that (6.9) admits the factorizations

$$\begin{aligned} z \in \mathbb{R} : \quad G(z) &= \begin{bmatrix} 1 & -\bar{r}(z)e^{-2i(2tz^2+xz)} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ r(z)e^{2i(2tz^2+xz)} & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \frac{r(z)}{1-|r(z)|^2} e^{2i(2tz^2+xz)} & 1 \end{bmatrix} \begin{bmatrix} 1-|r(z)|^2 & 0 \\ 0 & \frac{1}{1-|r(z)|^2} \end{bmatrix} \\ &\quad \times \begin{bmatrix} 1 & -\frac{\bar{r}(z)}{1-|r(z)|^2} e^{-2i(2tz^2+xz)} \\ 0 & 1 \end{bmatrix} \end{aligned}$$

one may now deform the jump contour in problem 6.2 (according to the asymptotic regime at hand while taking into account the analytic properties of the reflection coefficient) and thus localize the problem near the relevant stationary point $z_0 = -\frac{x}{4t}$. Away from that point, the deformed Hilbert boundary value problem will be asymptotically well-behaved in the sense that its jump matrix converges to the identity in $L^1(\Sigma, |dz|) \cap L^2(\Sigma, |dz|) \cap L^\infty(\Sigma, |dz|)$ as the parameters become large. We are thus facing a *small norm problem* away from the stationary point and such a problem is amenable to an iterative solution method. Here is the technical core of the argument:

Theorem 8.1 ([37]). *Assume that the jump matrix $G(z) = G(z; t)$ in problem 7.3 depends on a parameter t and that there exist $\epsilon, t_0, c > 0$ such that*

$$\|G(\cdot; t) - \mathbb{I}\|_{L^1(\Sigma, |dz|) \cap L^2(\Sigma, |dz|) \cap L^\infty(\Sigma, |dz|)} \leq \frac{c}{t^\epsilon} \quad \forall t \geq t_0. \tag{8.1}$$

Then problem 7.3 is uniquely solvable for t sufficiently large, i.e. there exists $t_1 > 0$ such that problem 7.3 is uniquely solvable for all $t \geq t_1$. Moreover, the extension $X = X(z; t)$ (7.5) defined in terms of the problem’s solution satisfies

$$\|X(z; t) - \mathbb{I}\| \leq \frac{C}{(1+|z|)t^\epsilon} \quad \forall t \geq t_1$$

uniformly on compact subsets of $\mathbb{C} \setminus \Sigma$ with $C > 0$.

Proof. Rewrite the integral equation (7.4) as

$$\begin{aligned} \rho(z; t) - \mathbb{I} &= \frac{1}{2\pi i} \int_{\Sigma} (G(w; t) - \mathbb{I}) \frac{dw}{w - z_-} \\ &\quad + \frac{1}{2\pi i} \int_{\Sigma} (\rho(w; t) - \mathbb{I})(G(w; t) - \mathbb{I}) \frac{dw}{w - z_-} \quad \text{a.e. on } \Sigma \end{aligned}$$

and abbreviate (recall remark 7.4) the operator

$$(Tf)(z; t) := \frac{1}{2\pi i} \int_{\Sigma} f(w; t) (G(w; t) - \mathbb{I}) \frac{dw}{w - z_-} \equiv (C_- f(G(\cdot; t) - \mathbb{I}))(z; t).$$

However by assumption (8.1) and (7.2), in $L^2(\Sigma, |dz|)$ operator norm,

$$\|T\| \leq \|C_-\| \|G(\cdot; t) - \mathbb{I}\|_{L^\infty(\Sigma, |dz|)} \leq \frac{c}{t^\epsilon} \quad \forall t \geq t_0,$$

which shows that that the integral equation

$$\begin{aligned} f(z; t) - (Tf)(z; t) &= \frac{1}{2\pi i} \int_{\Sigma} (G(w; t) - \mathbb{I}) \frac{dw}{w - z_-} =: R(z; t) \\ &\equiv (C_-(G(\cdot; t) - \mathbb{I}))(z; t) \quad \text{a.e. on } \Sigma \end{aligned}$$

is solvable in $L^2(\Sigma, |dz|)$ for sufficiently large t by the Neumann series. Moreover, since its right-hand side $R(z; t)$ satisfies

$$\|R(\cdot; t)\|_{L^2(\Sigma, |dz|)} \leq \|C_-\| \|G(\cdot; t) - \mathbb{I}\|_{L^2(\Sigma, |dz|)} \leq \frac{c}{t^\epsilon} \quad \forall t \geq t_0,$$

the unique solution $\rho(\cdot; t) - \mathbb{I} \in L^2(\Sigma, |dz|)$ of (7.4) also satisfies

$$\|\rho(\cdot; t) - \mathbb{I}\|_{L^2(\Sigma, |dz|)} \leq \frac{c}{t^\epsilon} \quad \forall t \geq t_0.$$

In summary, using proposition 7.5, the Hilbert boundary value problem 7.3 is uniquely solvable for sufficiently large t and we find from (7.5) in turn (this is the only time we use the $L^1(\Sigma, |dz|)$ bound in (8.1)),

$$\begin{aligned} \|X(z; t) - \mathbb{I}\| &\leq \frac{1}{2\pi} \left\| \int_{\Sigma} (G(w; t) - \mathbb{I}) \frac{dw}{w - z} \right\| + \frac{1}{2\pi} \\ &\quad \times \left\| \int_{\Sigma} (\rho(w; t) - \mathbb{I})(G(w; t) - \mathbb{I}) \frac{dw}{w - z} \right\| \\ &\leq \frac{1}{2\pi} \frac{1}{\text{dist}(z, \Sigma)} (\|G(\cdot; t) - \mathbb{I}\|_{L^1(\Sigma, |dz|)} + \|\rho(\cdot; t) - \mathbb{I}\|_{L^2(\Sigma, |dz|)}) \\ &\quad \times \|G(\cdot; t) - \mathbb{I}\|_{L^2(\Sigma, |dz|)} \\ &\leq \frac{C}{(1 + |z|)t^\epsilon} \quad \forall t \geq t_1 \geq t_0, \end{aligned}$$

uniformly on compact subsets of $\mathbb{C} \setminus \Sigma$. This concludes the proof of theorem 8.1. □

Near the stationary point(s) non-trivial technicalities of the nonlinear steepest descent method appear: one has to, either explicitly or implicitly, construct appropriate model functions in terms of classical or Painlevé special functions or, more generally, in terms of solutions to other dynamical systems that approximate the Hilbert boundary value problem

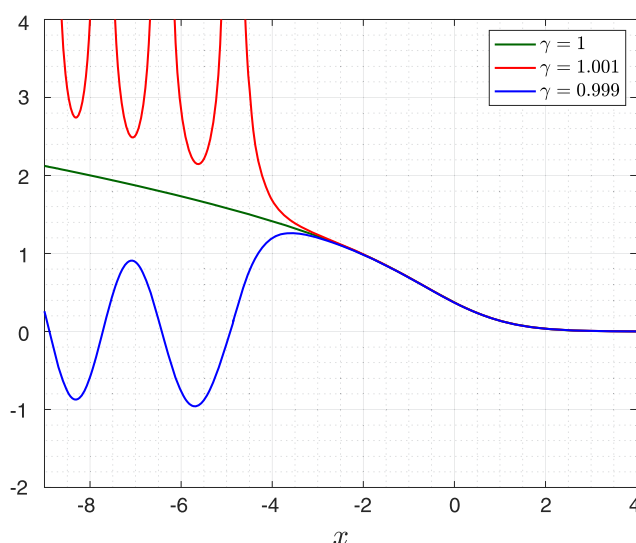


Figure 7. The sensitive dependence of $u = u(x|s)$ for $x \leq 0$ on the monodromy data $s = (-i\sqrt{\gamma}, 0, i\sqrt{\gamma})$ for three values of γ . In green we show $\gamma = 1$, in red $\gamma = 1.001$ and in blue $\gamma = 0.999$.

asymptotically near the stationary point(s). Combining such local constructions with the aforementioned estimates away from the stationary point(s) one then typically arrives at a *global small norm problem* which is amenable to theorem 8.1

Remark 8.2. The above outlined Riemann–Hilbert nonlinear steepest descent techniques have been successfully applied in high performance numerical simulations, see the recent monograph [136] for a summary of research efforts in this direction.

Theory aside, we now discuss a series of impressive results which have been derived with Riemann–Hilbert techniques.

8.1. Connection formulæ for Painlevé-II transcendents

Let us return to (6.12) and focus on the solution family parametrized by the monodromy data

$$s \equiv (s_1, s_2, s_3) = (-i\sqrt{\gamma}, 0, i\sqrt{\gamma}), \quad \gamma \geq 0.$$

This one-parameter family is known as the Ablowitz–Segur family and it plays a dominant role in random matrix theory and integrable probability, see subsection 8.3 below and section 9. Depending on the values of γ , the solution $u = u(x|s)$ of (6.12) has different analytic and asymptotic properties, see figure 7.

For one, the Ablowitz–Segur solution $u = u(x|s)$ is smooth and bounded for $x > 0$ and if $x \rightarrow +\infty$, then

$$u(x|s) \sim \sqrt{\gamma} \frac{x^{-\frac{1}{4}}}{2\sqrt{\pi}} e^{-\frac{2}{3}x^{\frac{3}{2}}}, \tag{8.2}$$

which is proportional to the super exponential decay behaviour of the Airy function on the positive real axis (recall that (6.12) can be viewed as a nonlinear Airy equation). However, things change dramatically on the half ray $x \leq 0$: if $\gamma \in [0, 1)$, then $u(x|s)$ is smooth, bounded and with the following oscillatory behaviour as $x \rightarrow -\infty$,

$$u(x|s) \sim (-x)^{-\frac{1}{4}} \sqrt{-2\beta} \cos\left(\frac{2}{3}(-x)^{\frac{3}{2}} + \beta \ln\left(8(-x)^{\frac{3}{2}}\right) + \phi\right) \tag{8.3}$$

$$\beta = \frac{1}{2\pi} \ln(1 - \gamma), \quad \phi = \frac{\pi}{4} - \arg \Gamma(i\beta).$$

Expansion (8.3) is due to Ablowitz and Segur [3, 4] and was partially proven in [26, 68] by Hastings, McLeod and Clarkson, using Gelfand-Levitan type integral equations. The first complete proof of the leading order in (8.3) is due to Its, Kapaev, Suleimanov and Kitaev [79, 93, 133] via the isomonodromy method. The first nonlinear steepest descent based proof of (8.3) is in the paper [38] by Deift and Zhou. Next, for $\gamma = 1$, the solution $u(x|s)$ is still smooth for $x \leq 0$ but now unbounded, in fact as $x \rightarrow -\infty$

$$u(x|s) \sim \sqrt{-\frac{x}{2}}, \tag{8.4}$$

which, to leading order, can be found in the work of Hastings and McLeod [68] and to all orders in [38], using once more the nonlinear steepest descent method. Finally, if $\gamma > 1$, the qualitative behaviour of $u(x|s)$ changes completely: its smoothness is destroyed at finite $x < 0$ and solutions blow up, compare the below $x \rightarrow -\infty$ asymptotic expansion,

$$u(x|s) \sim \frac{\sqrt{-x}}{\sin\left(\frac{2}{3}(-x)^{\frac{3}{2}} + \widehat{\beta} \ln(8(-x)^{\frac{3}{2}}) + \varphi\right)} \tag{8.5}$$

$$\widehat{\beta} = \frac{1}{2\pi} \ln(\gamma - 1), \quad \varphi = \frac{\pi}{2} - \arg \Gamma\left(\frac{1}{2} + i\widehat{\beta}\right),$$

which is uniform away from the zeros of the trigonometric function appearing in the denominator of (8.5). The singular expansion (8.5) is originally due to Kapaev [90] via isomonodromy techniques and was re-derived in [23] by nonlinear steepest descent techniques. Note that formulae (8.2)–(8.5) solve the connection problem for the underlying Painlevé-II transcendent, namely the problem of completely determining the asymptotic behaviour of $u(x|s)$ as $x \rightarrow -\infty$, given the same description of $u(x|s)$ as $x \rightarrow +\infty$, or vice versa. Such connection problems are standard exercises in classical special function theory and commonly solved based on contour integral formulae. However, the Ablowitz–Segur transcendent $u(x|s)$ does *not* admit any such explicit formulae so it is a great achievement of the Riemann–Hilbert method to be able to solve this genuinely nonlinear connection problem in terms of classical special functions. Even better, one can uniformize the qualitatively different asymptotic behaviours (8.3)–(8.5) at $x = -\infty$ through the use of Jacobi elliptic functions. This recent breakthrough is contained in the paper [24] and was made possible by nonlinear steepest descent techniques.

8.2. The Heisenberg antiferromagnet

We now return to (6.15) and discuss two central results for P_n proven in [40, section 6]. First, the leading order asymptotic behaviour,

$$\ln P_n = n^2 \ln \sin \frac{\phi}{2} + o(n), \quad n \rightarrow \infty,$$

which holds uniformly on compact subsets of $(\frac{\pi}{2}, \pi) \ni \phi$. Hence, an overall Gaussian decay of P_n with a decay rate that decreases with increasing external magnetic field. Second, the emptiness formation probability (6.15) really depends on two variables, n and h , equivalently n and $\xi = e^{i\phi}$, compare (6.17). As it turns out P_n shares an intimate relation with the Painlevé-VI

equation, more precisely P_n is the Jimbo–Miwa–Ueno τ -function [85, section 5] of a particular Painlevé-VI equation when viewed in those variables. Here are the details, see [40, theorem 6.48]: set $t := \bar{\xi}/\xi$, then

$$t \frac{d}{dt} \ln P_n = \frac{t-1}{4y(y-1)(y-t)} \nu^2 + \frac{n}{2y} \nu; \quad \nu := (y-1) \frac{(1-n)y + nt}{t-1} - t \frac{dy}{dt},$$

where $y = y(t, n)$ satisfies the Painlevé-VI equation

$$\begin{aligned} \frac{d^2 y}{dt^2} = & \frac{1}{2} \left(\frac{1}{y} + \frac{1}{y-1} + \frac{1}{y-t} \right) \left(\frac{dy}{dt} \right)^2 - \left(\frac{1}{t} + \frac{1}{t-1} + \frac{1}{y-t} \right) \frac{dy}{dt} \\ & + \frac{y(y-1)(y-t)}{t^2(t-1)^2} \left(\frac{1}{2}(n-1)^2 - \frac{tn^2}{2y^2} + \frac{t(t-1)}{2(y-t)^2} \right). \end{aligned}$$

This impressive connection between P_n and Painlevé SF theory puts the spin chain model firmly on integrable systems ground and as such should be compared with the famous Jimbo and Miwa 1980 result [84] on the diagonal Ising correlation functions and their relation to a σ -version of Painlevé-VI.

8.3. Ulam’s problem

In this section we discuss one of the most beautiful and influential results in mathematical physics that was derived by Riemann–Hilbert techniques over the past 20 years, see [29] for a taste of the following discovery’s impact: Choose a permutation $\pi \in S_n$. For $1 \leq i_1 < i_2 < \dots < i_k \leq n$ we say that $(\pi(i_1), \dots, \pi(i_k))$ is an *increasing subsequence* of π of length k if

$$\pi(i_1) < \pi(i_2) < \dots < \pi(i_k).$$

For instance, if $\pi = (4, 2, 1, 3, 5) \in S_5$, then $(4, 5), (2, 3), (2, 5), (1, 3), (1, 5), (3, 5)$ are subsequences of length 2 and $(2, 3, 5), (1, 3, 5)$ are subsequences of length 3. In turn, the *maximal length* $\ell_n(\pi)$ of all increasing subsequences of π equals $\ell_5(\pi) = 3$ in this example. Now randomize the setup and pick $\pi \in S_n$ with uniform distribution, i.e. we have

$$\text{Prob} \{ \ell_n \leq N \} = \frac{1}{n!} \# \{ \pi \in S_n : \ell_n(\pi) \leq N \}.$$

How does ℓ_n behave statistically for large n ? This question was raised by Ulam in 1961 [134] and he himself conjectured the first moment convergence

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(\ell_n)}{\sqrt{n}} = c \tag{8.6}$$

for some constant c . Ulam’s problem became known as the problem of rigorously verifying the convergence (8.6) and in turn computing the constant. Many people with different methods contributed to this challenge including Erdős, Hammersley, Logan, Shepp and Vershik, Kerov, ultimately settling on $c = 2$, see the recent monograph [9, pp 3–4] or [127, section 1.1] for references. The next statistical quantity of ℓ_n to be considered was its variance. High precision numerical simulations by Odlyzko in the early 1990s indicated that

$$\lim_{n \rightarrow \infty} \frac{\text{Var}(\ell_n)}{\sqrt[3]{n}} = c_0 \approx 0.819 \dots,$$

and from simulations by Odlyzko and Rains

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(\ell_n) - 2\sqrt{n}}{\sqrt[6]{n}} = c_1 \approx -1.758 \dots, \tag{8.7}$$

which constituted a higher order simulation of $\mathbb{E}(\ell_n)$ than (8.6). At this point of the story, in 1999, the Riemann–Hilbert/integrable systems community in the form of Baik *et al* [8] picked up the loose ends and proved the following spectacular result, settling en route (8.6), (8.7) and then some.

Theorem 8.3 ([8, theorems 1.1 and 1.2]). *Let $u(x|s)$ denote the unique solution of (6.12) with the monodromy data $s = (-i, 0, i)$, compare subsection 8.1. Define the distribution function*

$$F(x) := \exp \left[- \int_x^\infty (y - x) u^2(y|s) dy \right], \quad x \in \mathbb{R}. \tag{8.8}$$

Then, for any $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \text{Prob} \left\{ \frac{\ell_n - 2\sqrt{n}}{\sqrt[6]{n}} \leq x \right\} = F(x) \equiv \text{Prob} \{ \chi \leq x \},$$

and for any $k \in \mathbb{Z}_{\geq 1}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}(\chi_n^k) = \mathbb{E}(\chi^k), \quad \chi_n := \frac{\ell_n - 2\sqrt{n}}{\sqrt[6]{n}}.$$

The distribution function $F(x)$ in (8.8), and it really is a distribution function, see our discussion on the analytic and asymptotic properties of $u(x|s)$ in subsection 8.1, had appeared a few years prior to the Baik–Deift–Johansson theorem. Namely in the 1994 work of Tracy and Widom [135] where it was shown to capture the fluctuations of the largest eigenvalue of a matrix drawn from the ensemble (6.21) with quadratic external field, a.k.a. the *Gaussian unitary ensemble* (GUE), in the large n limit. The appearance of a random matrix theory distribution function in a problem of combinatorics was a big surprise at the time, however it turned out not to be an isolated curiosity. The Tracy–Widom distribution $F(x)$ (8.8) (that name stuck after [8]) has become an ubiquitous limiting distribution that made its way to the financial markets, to wireless communication networks, stochastic growth processes and statistical mechanics, just to name a few areas. See [124] for an engaging outreach article on the Tracy–Widom law. In fact, twenty years after [8] it has become clear that the Tracy–Widom distribution is a modern day bell curve in that it describes the random fluctuations of a large class of integrable probabilistic models, see figure 8 for a graph of $F(x)$ and $F'(x)$.

The proof of theorem 8.3 is beyond the material covered in this short note, it is a tour de force journey through (de)-Poissonization arguments, Toeplitz determinants and finally the asymptotic nonlinear steepest descent analysis of orthogonal polynomials on the unit circle. These polynomials can be characterized via a Hilbert boundary value problem in the style of problem 6.5, we refer the interested reader to the monographs [9, 127].

8.4. Universality in invariant random matrix models

Our last discussion in this section concerns the global and local statistical eigenvalue behaviour in the random matrix model (6.21). First, through our assumptions placed on the external field $V : \mathbb{R} \rightarrow \mathbb{R}$, the model’s mean eigenvalue density $\frac{1}{n} K_{n,N}(x, x)$, compare (6.22), has a limit

$$\lim_{\substack{n, N \rightarrow \infty \\ \frac{n}{N} \rightarrow 1}} \frac{1}{n} K_{n,N}(x, x) = \rho_V(x) \geq 0, \tag{8.9}$$

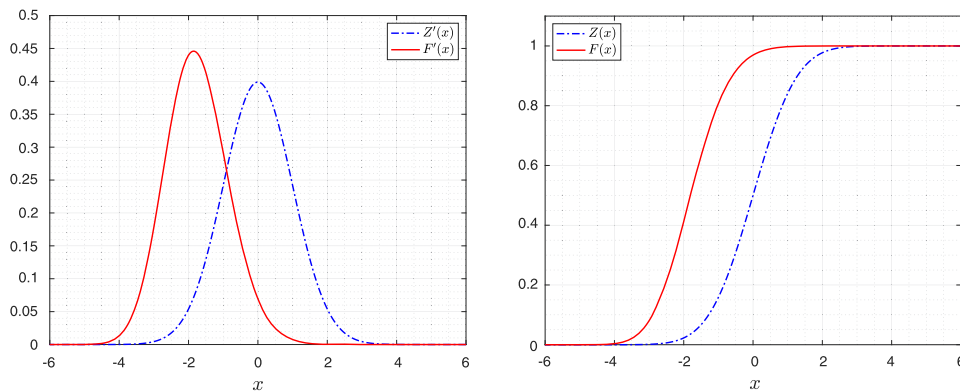


Figure 8. Comparing the Tracy–Widom distribution $F(x)$ in red to the standard normal distribution $Z(x)$ in blue. On the left, probability density functions, on the right cumulative distribution functions. The asymmetry of the Tracy–Widom distribution is clearly visible.

whose support $\Sigma_V := \overline{\{x \in \mathbb{R} : \rho_V(x) > 0\}}$ is a finite union of intervals. The *global* limiting behaviour (8.9) contains the most basic result in random matrix theory as special case, namely the Wigner semicircular law which asserts the almost sure convergence of the rescaled empirical spectral distribution in the GUE to the limiting distribution with density

$$\rho_W(x) = \frac{1}{\pi} \sqrt{(2 - x^2)_+}.$$

While the limiting density ρ_V in (8.9) is determined by the potential V , the *local* eigenvalue statistics are to a large extent independent of V in the large n, N limit. Precisely, and this is the celebrated *universality property*, the local scaling behaviour of $K_{n,N}(x, y)$ is completely determined by the local characteristics of ρ_V . For instance, in case of (6.21), only two qualitatively different scenarios arise generically, compare figure 9 for the GUE: if $x^* \in \Sigma_V$ is such that $\rho_V(x^*) > 0$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n\rho_V(x^*)} K_{n,n} \left(x^* + \frac{\lambda}{n\rho_V(x^*)}, x^* + \frac{\mu}{n\rho_V(x^*)} \right) = \frac{\sin \pi(\lambda - \mu)}{\pi(\lambda - \mu)} \quad (8.10)$$

uniformly on compact subsets of $\mathbb{R} \ni \lambda, \mu$. This is the 1996 bulk universality result of Pastur and Shcherbina [118] which was proven with the resolvent method. Soon after that Bleher and Its [15] and Deift *et al* [43] fine tuned the Deift–Zhou nonlinear steepest descent method of section 8 to problem 6.5 and arrived at the same result. Especially [43] has become the standard reference for nonlinear steepest descent techniques used in the asymptotic analysis of orthogonal polynomials, and it has been extended in various directions. Those directions lead in turn to generalizations of the bulk universality class (8.10) and extensions to other types of orthogonal polynomial random matrix models, see section 10 below for a few references. The second case concerns the scaling limit of $K_{n,n}(x, y)$ at a boundary point $x^* \in \partial\Sigma_V$ where ρ_V typically vanishes square root like, certainly in the GUE. In this case (8.10) is replaced by the limit

$$\lim_{n \rightarrow \infty} \frac{1}{(cn)^{\frac{2}{3}}} K_{n,n} \left(x^* + \frac{\lambda}{(cn)^{\frac{2}{3}}}, x^* + \frac{\mu}{(cn)^{\frac{2}{3}}} \right) = \frac{\text{Ai}(\lambda)\text{Ai}'(\mu) - \text{Ai}'(\lambda)\text{Ai}(\mu)}{\lambda - \mu}, \quad (8.11)$$

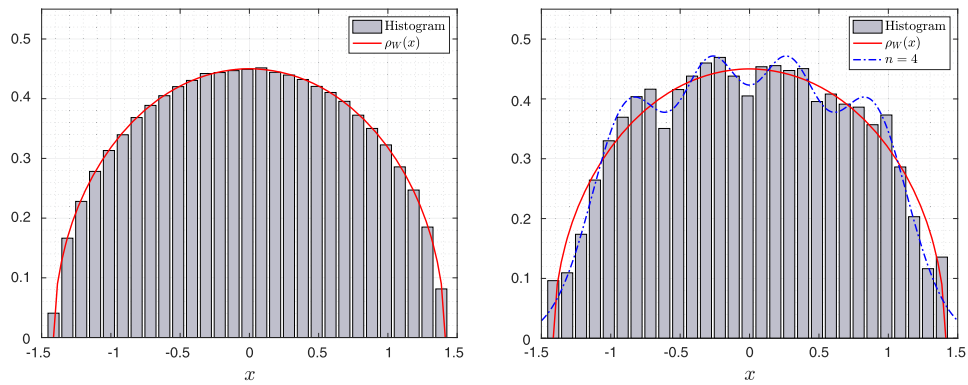


Figure 9. Wigner’s semicircular law for a rescaled 500×500 GUE matrix on the left. Plotted is the rescaled histogram of its 500 eigenvalues, sampled 100 times, against the semicircular density in red. On the right we compare Wigner’s law to the exact eigenvalue density for $n = 4$ in blue and to the associated eigenvalue histogram, again sampled 100 times.

if x^* is a right boundary point and with a sign change $(\lambda, \mu) \mapsto (-\lambda, -\mu)$ in the left-hand side of (8.11) for a left boundary point. The limit is again uniform on compact subsets of $\mathbb{R} \ni \lambda, \mu$, it is expressed in terms of the Airy function $w = \text{Ai}(z)$ and $c > 0$ denotes some constant. The first proof of the soft edge universality (8.11) is due to Deift *et al* [43] and in more explicit form due to Deift and Gioev [36], in both cases based on nonlinear steepest descent techniques.

In summary of this short subsection, the earliest universality results in random matrix theory were almost exclusively derived with Riemann–Hilbert nonlinear steepest descent techniques and thus paved the way for subsequent discoveries. Nowadays a wide array of different techniques is available in the derivation of RMT universality classes, some more suitable to certain ensembles than others. Most recently, the work of Erdős and Yau [53] put forth a toolbox useful in the universality analysis of Wigner random matrices, a broader class of symmetric matrices with independent entries. It is an open question whether the Wigner ensemble is accessible via Riemann–Hilbert techniques.

9. One more example

In this section we will venture into the world of *operator-valued* Hilbert boundary value problems, a topic which has received far less attention so far, especially from a mathematically rigorous perspective. The only exception is the work of Its and Kozłowski [80] and to a certain extent the paper [82] by Its and Slavnov which we will discuss towards the end. Here is our motivation to think about such problems: consider the stochastic heat equation (SHE)

$$\mathcal{Z}_T = \frac{1}{2} \mathcal{Z}_{XX} - \mathcal{Z} \dot{\mathcal{W}}, \quad \mathcal{Z} = \mathcal{Z}(X, T) : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$$

with distributional initial data $\mathcal{Z}(X, 0) = \delta(X)$ and a Gaussian white noise $\dot{\mathcal{W}} = \dot{\mathcal{W}}(X, T)$. Using the Hopf-Cole transformation $\mathcal{H}(X, T) := -\ln \mathcal{Z}(X, T)$ the SHE is mapped to the celebrated Kardar–Parisi–Zhang (KPZ) equation

$$\mathcal{H}_T = \frac{1}{2} \mathcal{H}_{ZZ} - \frac{1}{2} (\mathcal{H}_Z)^2 + \dot{\mathcal{W}} \tag{9.1}$$

which has become the quintessential model for random surface growth processes with several remarkable connections to a number of different physical phenomena, see [28] for a detailed summary. One of the long-standing goals in the analysis of the KPZ equation was achieved in 2011 when Amir *et al* [6] derived the exact probability distribution of its solution with narrow wedge initial data. This allowed in turn to finally verify the celebrated KPZ scaling hypothesis for the KPZ equation itself. Here are the relevant details:

Theorem 9.1 ([6, theorem 1.1]). *For any $X \in \mathbb{R}$ and $T > 0$ the Hopf-Cole solution $\mathcal{H}(X, T) = -\ln \mathcal{Z}(X, T)$ to the KPZ equation (9.1) with initial data $\mathcal{Z}(X, 0) = \delta(X)$ has the probability distribution*

$$F_T(s) := \text{Prob} \left\{ \mathcal{H}(X, T) - \frac{X^2}{2T} - \frac{T}{24} \geq -s \right\}, \quad s \in \mathbb{R},$$

which is independent of $X \in \mathbb{R}$ and which equals

$$F_T(s) = \int_{\Gamma} \det \left(1 - K_{\sigma_{T,z}} \upharpoonright_{L^2(\kappa_T^{-1}s, \infty)} \right) e^{-z} \frac{dz}{z}. \tag{9.2}$$

In (9.2) we choose a Hankel contour Γ around $\mathbb{R}_{\geq 0}$, we abbreviate $\kappa_T := (T/2)^{1/3}$ and K_{σ} is the integral operator with kernel

$$K_{\sigma}(x, y) := \int_{-\infty}^{\infty} \sigma(t) \text{Ai}(x+t) \text{Ai}(y+t) dt \quad \text{and} \quad \sigma_{T,z}(t) := \frac{z}{z - e^{-\kappa_T t}}.$$

The explicit formula (9.2) makes it possible to analyse the statistics of the KPZ equation in the two extreme cases $T \downarrow 0$ and $T \rightarrow \infty$. First, under the KPZ scaling $T^{1/3}$, the statistics converges to the Tracy–Widom distribution (8.8) for large T , see [6, corollary 1.3],

$$\lim_{T \rightarrow \infty} F_T \left((T/2)^{1/3} x \right) = F(x), \quad x \in \mathbb{R}. \tag{9.3}$$

Second, under the scaling $T^{1/4}$, the statistics converges to the standard normal distribution for small T , see [6, proposition 1.8],

$$\lim_{T \downarrow 0} F_T \left(2^{-1/2} (\pi T)^{1/4} (x - \ln \sqrt{2\pi T}) \right) = Z(x), \quad x \in \mathbb{R}. \tag{9.4}$$

Hence, the KPZ equation interpolates between two universality classes and (9.2) is therefore a *crossover distribution*.

From a Riemann–Hilbert/integrable systems viewpoint we are interested in the Fredholm determinant entering the Amir–Corwin–Quastel formula (9.2). Integrating by parts we can rewrite its kernel in the form

$$(x - y) K_{\sigma}(x, y) = \int_{-\infty}^{\infty} (\text{Ai}(x+t) \text{Ai}'(y+t) - \text{Ai}'(x+t) \text{Ai}(y+t)) d\sigma(t) \tag{9.5}$$

with the measure $d\sigma(t) := \sigma'(t)dt$. Besides the particular choice of $\sigma = \sigma_{T,z}$ in (9.2), equality in (9.5) holds as soon as $\sigma : \mathbb{R} \rightarrow \mathbb{C}$ is smooth and

$$\lim_{t \rightarrow -\infty} \sigma(t) = 0, \quad \lim_{t \rightarrow +\infty} \sigma(t) = 1 \quad \text{exponentially fast.}$$

Note that (9.5) includes the Fermi factor choice

$$\sigma(t) = \frac{1}{1 + e^{-\alpha t}}, \quad \alpha > 0,$$

which plays an important role in the analysis of the Moshe–Neuberger–Shapiro random matrix ensemble [89, (14)], [108, (23)] and in the analysis of non-interacting quantum many-body systems [48, (158)]. Now given the formal—think of $\sigma(t) = \chi_{[0,\infty)}(t)$ —similarities between (9.5) and the standard Airy kernel (8.11) one might ask for an analogue of the Tracy–Widom Painlevé-II formula (8.8) for the Fredholm determinant of (9.5) on $L^2(s, \infty)$, perhaps even an underlying Lax pair which would further strengthen the connection between theorem 9.1 and integrable systems theory. The task of finding an analogue of (8.8) has been completed in [6, section 5] using the techniques of [135], see theorem 9.21 below. Here, we will employ Riemann–Hilbert methods and en route reproduce the Amir–Corwin–Quastel formula [6, proposition 5.2] while identifying an underlying Lax pair. Our approach is motivated by the duality between the kernels involved, namely

$$K_{\text{Ai}}(x, y) = \frac{\sum_{i=1}^n f_i(x)g_i(y)}{x - y} \leftrightarrow K_\sigma(x, y) = \frac{\int f(x, t)g(y, t)d\sigma(t)}{x - y},$$

i.e. the matrix-valued methods of subsection 6.4 will be lifted to operator-valued ones.

9.1. Technicalities up front

Throughout we are considering kernels K_σ of the form (9.5) where $d\sigma$ is a positive Borel probability measure on \mathbb{R} for which all moments are finite, i.e.

$$\int_{\mathbb{R}} |t|^k d\sigma(t) < \infty, \quad k \in \mathbb{Z}_{\geq 0},$$

and which is absolutely continuous with respect to the Lebesgue measure.

Definition 9.2. Let $p \in \mathbb{Z}_{\geq 1}$. We require the following abbreviations:

- (a) The direct sum Hilbert space

$$\mathcal{H}_p := \bigoplus_{j=1}^p L^2(\mathbb{R}, d\sigma) = \{f = (f_1, \dots, f_p)^\top \in \mathbb{C}^{p \times 1} : f_j \in L^2(\mathbb{R}, d\sigma)\},$$

equipped with the standard inner product $\langle f, g \rangle_{\mathcal{H}_p} := \sum_{j=1}^p \int_{\mathbb{R}} f_j(t) \overline{g_j(t)} d\sigma(t)$ and induced norm $\|f\|_{\mathcal{H}_p} := \sqrt{\langle f, f \rangle_{\mathcal{H}_p}}$.

- (b) The space $L^2(\mathbb{R}, d\sigma; \mathbb{C}^{p \times p})$ of $p \times p$ matrix-valued functions with entries in $L^2(\mathbb{R}, d\sigma)$, equipped with the norm

$$\|X\|_{L^2(\mathbb{R}, d\sigma; \mathbb{C}^{p \times p})} := \left[\int_{\mathbb{R}} \text{tr} (X(t)X^\dagger(t)) d\sigma(t) \right]^{1/2}, \quad X = [X^{ij}]_{i,j=1}^p.$$

- (c) The space $\mathcal{I}(\mathcal{H}_p)$ of Hilbert–Schmidt integral operators on \mathcal{H}_p of the form

$$(Kf)(x) = \int_{\mathbb{R}} K(x, y)f(y) d\sigma(y),$$

with kernel $K \in L^2(\mathbb{R}^2, d\sigma \otimes d\sigma; \mathbb{C}^{p \times p})$.

- (d) The matrix identity operator \mathbb{I}_p on \mathcal{H}_p .

Next, fix some region $\Omega \subset \mathbb{C}$ and consider $K = K(z) \in \mathcal{I}(\mathcal{H}_p)$ which depends on an auxiliary variable $z \in \Omega$. We now define the notion of an *analytic integral operator*, see [80, p 1781].

Definition 9.3. We say that $K(z) \in \mathcal{I}(\mathcal{H}_p)$ with kernel $K(z|x, y)$ is analytic in $z \in \Omega$, if

- (a) for any $(x, y) \in \mathbb{R}^2$, the map $z \mapsto K(z|x, y) \in \mathbb{C}^{p \times p}$ is analytic in Ω .
- (b) for any $z \in \Omega$, the map $(x, y) \mapsto K(z|x, y)$ is in $L^2(\mathbb{R}^2, d\sigma \otimes d\sigma; \mathbb{C}^{p \times p})$.

Furthermore, if $\Sigma \subset \Omega \subset \mathbb{C}$ is an oriented contour which consists of a finite union of smooth oriented curves in $\mathbb{C}\mathbb{P}^1$ with finitely many self-intersections, we will need boundary values of $K(z) \in \mathcal{I}(\mathcal{H}_p)$, see [80, p 1782].

Definition 9.4. We say that an analytic in $z \in \Omega \setminus \Sigma$ operator $K(z) \in \mathcal{I}(\mathcal{H}_p)$ admits continuous boundary values $K_{\pm}(z) \in \mathcal{I}(\mathcal{H}_p)$ on $\Sigma' \subset \Sigma$ with kernels $K_{\pm}(z|x, y)$ if

- (a) for any $(x, y) \in \mathbb{R}^2$, the map $z \mapsto K_{\pm}(z|x, y) \in \mathbb{C}^{p \times p}$ is continuous on Σ' .
- (b) for any $(x, y) \in \mathbb{R}^2$, the non-tangential limits

$$\lim_{\lambda \rightarrow z} K(\lambda|x, y) = K_{\pm}(z|x, y), \quad \lambda \in \pm \text{ side of } \Sigma' \text{ at } z$$

exist.

At this point we record two basic results about the kernel (9.5).

Lemma 9.5. Let $K_{\sigma} : L^2(s, \infty) \rightarrow L^2(s, \infty)$ be given by

$$(K_{\sigma}f)(x) := \int_s^{\infty} K_{\sigma}(x, y)f(y) dy, \tag{9.6}$$

in terms of the kernel (9.5). Then K_{σ} is trace class for any $s \in \mathbb{R}$.

Proof. If $\sigma(t)$ is the aforementioned Fermi factor, then the claim is proven in [89, proposition 1.1]. For general σ we argue as follows: write

$$K_{\sigma}(x, y) = \int_{-\infty}^{\infty} \left[\int_0^{\infty} \text{Ai}(x+t+u)\text{Ai}(y+t+u) du \right] d\sigma(t),$$

and obtain that for any $x_1, \dots, x_n \in (s, \infty)$ and $z_1, \dots, z_n \in \mathbb{C}$,

$$\sum_{i,j=1}^n z_i \bar{z}_j K_{\sigma}(x_i, x_j) = \int_{-\infty}^{\infty} \left[\int_0^{\infty} \left| \sum_{i=1}^n z_i \text{Ai}(x_i+t+u) \right|^2 du \right] d\sigma(t) \geq 0.$$

Thus, with $K_{\sigma}(x, y) = K_{\sigma}(y, x)$, the function $K_{\sigma}(x, y)$ is Hermitian positive definite, so the claim follows provided we can show that

$$\int_s^{\infty} K_{\sigma}(x, x) dx < \infty, \tag{9.7}$$

see [129, theorem 2.12]. In order to get to (9.7) we compute the elementary integral,

$$\int_s^{\infty} \text{Ai}^2(x+u) du = (\text{Ai}'(x+s))^2 - (x+s)\text{Ai}^2(x+s), \quad x \in \mathbb{R},$$

apply Fubini's theorem,

$$\int_s^{\infty} K_{\sigma}(x, x) dx = \int_{-\infty}^{\infty} \left[\int_{s+t}^{\infty} \{ (\text{Ai}'(x))^2 - x\text{Ai}^2(x) \} dx \right] d\sigma(t),$$

and estimate for any $y \in \mathbb{R}$, see [117, section 9.7],

$$\int_y^\infty \left| (\text{Ai}'(x))^2 - x \text{Ai}^2(x) \right| dx \leq c|y|^{\frac{3}{2}}, \quad c > 0.$$

Hence (9.7) follows from our finite moment assumption placed on the Borel measure $d\sigma$. \square

Next to the trace class property of K_σ we will also need the following result which is well-known at least for the Airy operator induced by the kernel (8.11), i.e. for $\sigma(t) = \chi_{[0,\infty)}(t)$ in (9.5).

Proposition 9.6. *For every $s \in \mathbb{R}$, the self-adjoint operator K_σ on $L^2(s, \infty)$ satisfies $0 \leq K_\sigma \leq 1$. Moreover, $1 - K_\sigma$ is invertible on $L^2(s, \infty)$.*

Proof. We employ the standard Fourier transform trick, see [9, lemma 6.15], and therefore use

$$\text{Ai}(x) = \frac{1}{2\pi} \int_{-\infty}^\infty e^{i(\frac{1}{3}y^3 + xy)} dy, \quad x \in \mathbb{R}.$$

First compute for any $f \in L^2(s, \infty)$,

$$\langle f, K_\sigma f \rangle_{L^2(s,\infty)} = \int_{-\infty}^\infty \left[\int_t^\infty \left| \int_{-\infty}^\infty \text{Ai}(x+u) f_s(x) dx \right|^2 du \right] d\sigma(t) \geq 0, \quad (9.8)$$

with $f_s(x) := f(x)\chi_{(s,\infty)}(x)$ using the characteristic function of $(s, \infty) \subset \mathbb{R}$. But from the Fourier representation, for any $u \in \mathbb{R}$,

$$\int_{-\infty}^\infty \text{Ai}(x+u) f_s(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{i(\frac{1}{3}y^3 + uy)} \check{f}_s(-y) dy = \check{g}(-u),$$

where $\check{f}_s(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty f_s(x) e^{-ixy} dx$ and we set $g(y) := e^{\frac{i}{3}y^3} \check{f}_s(-y)$. Hence, back in (9.8), we find

$$\begin{aligned} 0 \leq \langle f, K_\sigma f \rangle_{L^2(s,\infty)} &= \int_{-\infty}^\infty \left[\int_t^\infty |\check{g}(-u)|^2 du \right] d\sigma(t) \\ &\leq \int_{-\infty}^\infty \left[\int_{-\infty}^\infty |\check{g}(-u)|^2 du \right] d\sigma(t) \end{aligned}$$

and thus, since $d\sigma$ is a probability measure,

$$\begin{aligned} 0 \leq \langle f, K_\sigma f \rangle_{L^2(s,\infty)} &\leq \int_{-\infty}^\infty |\check{g}(-u)|^2 du = \int_{-\infty}^\infty |g(u)|^2 du \\ &= \int_{-\infty}^\infty |f_s(y)|^2 dy = \langle f, f \rangle_{L^2(s,\infty)}, \end{aligned}$$

where we used Plancherel's theorem in the first and second equality. This shows all together that $0 \leq K_\sigma \leq 1$ and so, by self-adjointness, also $\|K_\sigma\| \leq 1$ in operator norm. Finally, since K_σ is compact, it is sufficient to establish that $1 - K_\sigma$ is injective in order to conclude that $1 - K_\sigma$ is invertible on $L^2(s, \infty)$. So assume there is $f \in L^2(s, \infty) \setminus \{0\}$ such that $K_\sigma f = f$. Then all inequalities above must be equalities for this f , i.e. in particular

$$\langle f, K_\sigma f \rangle = \int_{-\infty}^\infty \left[\int_t^\infty |\check{g}(-u)|^2 du \right] d\sigma(t) = \int_{-\infty}^\infty |\check{g}(-u)|^2 du,$$

and which implies that

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^t |\check{g}(-u)|^2 du \right] d\sigma(t) = 0.$$

But $d\sigma$ is a positive Borel probability measure which is absolutely continuous with respect to the Lebesgue measure, so we must have

$$\sigma - \text{a.e.} : \int_{-\infty}^t |\check{g}(-u)|^2 du = 0 \Rightarrow \text{a.e.} : \check{g}(-t) = \int_s^{\infty} \text{Ai}(x+t)f(x) dx = 0,$$

and thus by the analytic properties of the Airy function, $f = 0 \in L^2(s, \infty)$, a contradiction. \square

We now combine lemma 9.5 with proposition 9.6 and conclude that the Fredholm determinant

$$F_{\sigma}(s) := \det \left(1 - K_{\sigma} \upharpoonright_{L^2(s, \infty)} \right) = \exp \left[- \sum_{n=1}^{\infty} \frac{1}{n} \text{tr} K_{\sigma}^n \right], \quad s \in \mathbb{R} \tag{9.9}$$

is well-defined on one hand and on the other hand satisfies $0 < F(s) < 1$ for all $s \in \mathbb{R}$. Even better, since

$$\int_s^{\infty} \left\{ (\text{Ai}'(x))^2 - x\text{Ai}^2(x) \right\} dx > \int_{s'}^{\infty} \left\{ (\text{Ai}'(x))^2 - x\text{Ai}^2(x) \right\} dx$$

whenever $s < s'$, we also have that $F_{\sigma}(s)$ is strictly increasing in s , so $F_{\sigma}(s)$ is a distribution function of some random variable (one also needs Deift’s lemma [6, lemma 2.20] for the regularity of $F_{\sigma}(s)$ in s). Additionally, see [131, theorem 3], there is a determinantal process with correlation kernel K_{σ} . These features of (9.9) and (9.5) are not surprising in light of Johansson’s discussion of the interpolating process in [89, section 1.4]. However we emphasize that [89] uses the Fermi factor choice for σ , here we are in a more general setup.

9.2. Solvability of the operator-valued problem

We now proceed to characterize (9.9) through a naturally associated Hilbert boundary value problem—very much as in subsection 6.4, but the problem is going to be operator-valued from the start. First, write

$$F_{\sigma}(s) = \det \left(1 - K_{\sigma, s} \upharpoonright_{L^2(0, \infty)} \right),$$

where $K_{\sigma, s}$ has kernel $K_{\sigma, s}(x, y) := K_{\sigma}(x + s, y + s)$. Then consider the following operators (we suppress the explicit s -dependence in our notation).

Definition 9.7. Let $M_i(z) \otimes K_j(z) \in \mathcal{I}(\mathcal{H}_1)$, $i, j = 1, 2$ denote the $(0, \infty) \ni z$ -parametric family of rank one integral operators with kernels

$$(M_i(z) \otimes K_j(z))(x, y) := m_i(z|x)k_j(z|y), \quad x, y \in \mathbb{R},$$

defined in terms of the $(0, \infty) \ni z$ -parametric family of functions

$$m_1(z|x) = k_2(z|x) := \text{Ai}'(z + s + x), \quad m_2(z|x) = k_1(z|x) := \text{Ai}(z + s + x). \tag{9.10}$$

Equivalently, $M_i(z)$ are the operators on \mathcal{H}_1 which multiply by the functions $m_i(z|x)$,

$$(M_i(z)f)(x) := m_i(z|x)f(x),$$

and $K_j(z)$ are the integral operators on \mathcal{H}_1 with kernel $k_j(z|y)$,

$$(K_j(z)f)(x) := \int_{-\infty}^{\infty} k_j(z|y)f(y) d\sigma(y).$$

Here is the relevant operator-valued Hilbert boundary value problem.

Hilbert Boundary Value Problem 9.8. For any $s \in \mathbb{R}$, determine $X(z) = X(z; s) \in \mathcal{I}(\mathcal{H}_2)$ such that

- (a) $X(z) = \mathbb{I}_2 + X_0(z)$ and $X_0(z) \in \mathcal{I}(\mathcal{H}_2)$ with kernel $X_0(z|x, y)$ is analytic in $\mathbb{C} \setminus [0, \infty)$.
- (b) $X(z)$ admits continuous boundary values $X_{\pm}(z) \in \mathcal{I}(\mathcal{H}_2)$ on $(0, \infty) \subset \mathbb{R}$, oriented from $-\infty$ to $+\infty$, which satisfy $X_+(z) = X_-(z)G(z)$ with

$$G(z) = \mathbb{I}_2 + 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix}. \tag{9.11}$$

- (c) There exist operators F_i on \mathcal{H}_1 which multiply by z -independent functions $f_i \in L^1(\mathbb{R}, d\sigma) \cap L^\infty(\mathbb{R}, d\sigma)$ such that for $|z| < \epsilon$, in terms of any branch for $\ln : \mathbb{C} \setminus [0, \infty) \rightarrow \mathbb{C}$,

$$X(z) = \mathbb{I}_2 - \begin{bmatrix} F_1 \otimes K_1(0) & -F_1 \otimes K_2(0) \\ F_2 \otimes K_1(0) & -F_2 \otimes K_2(0) \end{bmatrix} \ln z + X_{reg}(z), \tag{9.12}$$

where $X_{reg}(z) \in \mathcal{I}(\mathcal{H}_2)$ with kernel $X_{reg}(z|x, y)$ is analytic at $z = 0$ and satisfies

$$\|X_{reg}(z|x, y)\| \leq c|1 + s + x|^{\frac{3}{4}}|1 + s + y|^{\frac{3}{4}} \quad \text{as } z \rightarrow 0 \tag{9.13}$$

for all $(x, y) \in \mathbb{R}^2$ with $c > 0$.

- (d) For any compact subset $V \subset \mathbb{C}$ with $\text{dist}(0, V) > 0$ there exists $c = c(V) > 0$ such that for $z \in \mathbb{C} \setminus V$,

$$\|X_0(z|x, y)\| \leq \frac{c}{1 + |z|} |s + x|^{\frac{3}{4}} |s + y|^{\frac{3}{4}} \tag{9.14}$$

uniformly in $(x, y) \in \mathbb{R}^2$.

Our first result concerns the unique solvability of problem 9.8

Theorem 9.9. The Hilbert boundary value problem 9.8 is uniquely solvable for any $s \in \mathbb{R}$ and its solution takes the form

$$X(z) = \mathbb{I}_2 + \int_0^\infty \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \frac{dw}{w - z}, \quad z \in \mathbb{C} \setminus [0, \infty), \tag{9.15}$$

where $N_i(w)$ are the operators on \mathcal{H}_1 which multiply by the functions $n_i(w|x)$ determined from the integral equation

$$(1 - K_{\sigma, s} \upharpoonright_{L^2(0, \infty)})n_i(\cdot|x) = m_i(\cdot|x), \quad i = 1, 2, \tag{9.16}$$

with $x \in \mathbb{R}$.

Proof. Suppose first that problem 9.8 is solvable. By condition (a), estimate (9.14) and the moment assumption on $d\sigma$, the Fredholm determinant

$$e(z) := \det(\mathbb{I}_2 + X_0(z) \upharpoonright_{\mathcal{H}_2}), \quad z \in \mathbb{C} \setminus [0, \infty)$$

is well-defined in the indicated z -domain by Hadamard’s inequality (by condition (a) only, $X_0(z)$ is a Hilbert–Schmidt operator, but (d) ensures that the ordinary Fredholm determinant on \mathcal{H}_2 converges). Moreover, by Morera’s and Fubini’s theorem, analyticity of $X_0(z)$ in $\mathbb{C} \setminus [0, \infty)$ implies analyticity of $e(z)$ in $\mathbb{C} \setminus [0, \infty)$. Furthermore, using (9.14) again and the dominated convergence theorem, the boundary values $e_{\pm}(z), z \in (0, \infty)$ satisfy

$$e_{\pm}(z) = \det (X_{\pm} \upharpoonright_{\mathcal{H}_2}), \quad z \in (0, \infty). \tag{9.17}$$

Our immediate goal is to derive a scalar Hilbert boundary value problem for $e(z)$: from condition (b) we have $G(z) = \mathbb{I}_2 + G_0(z)$ where $G_0(z) \in \mathcal{I}(\mathcal{H}_2)$ is trace class on \mathcal{H}_2 and satisfies the estimate, compare definition 9.7 and [117, section 9.7],

$$\|G_0(z|x, y)\| \leq c|z + s + x|^{\frac{1}{4}}|z + s + y|^{\frac{1}{4}}, \quad c > 0,$$

uniformly in $(z, x, y) \in \mathbb{R}^3$. This ensures existence of the determinant $\det(\mathbb{I}_2 + G_0(z) \upharpoonright_{\mathcal{H}_2})$ for $z \in \mathbb{R}$ and since

$$\text{tr}_{\mathcal{H}_2} G_0(z) = 0, \quad (G_0(z))^2 = 0, \quad z \in \mathbb{R},$$

we obtain from the Plemelj–Smithies formula [65, chapter II, theorem 3.1] that $\det(G(z) \upharpoonright_{\mathcal{H}_2}) = 1$ for all $z \in \mathbb{R}$. With this in hand we return to (9.17), apply condition (b) and the multiplicativity property of Fredholm determinants, in turn

$$e_+(z) = e_-(z), \quad z \in (0, \infty).$$

Moreover, estimate (9.14) yields $e(z) \rightarrow 1$ as $z \rightarrow \infty$, so we are left to address the behaviour of $e(z)$ in a vicinity of $z = 0$. By (9.12) and the estimate (9.13) we see that the operator

$$P(z) := \det (\mathbb{I}_2 + X_{\text{reg}}(z) \upharpoonright_{\mathcal{H}_2}) (\mathbb{I}_2 + X_{\text{reg}}(z))^{-1}, \quad |z| < \epsilon$$

on \mathcal{H}_2 is well-defined even if $\det(\mathbb{I}_2 + X_{\text{reg}}(z) \upharpoonright_{\mathcal{H}_2})$ vanishes at some points in the disk $|z| < \epsilon$, see [65, chapter VIII, theorem 1.1]. Hence from (9.12), as $z \rightarrow 0$ and $z \notin [0, \infty)$,

$$e(z) = \det (\mathbb{I}_2 + X_{\text{reg}}(z) \upharpoonright_{\mathcal{H}_2}) - \langle P(z)F, K(0) \rangle_{\mathcal{H}_2} \ln z,$$

with $F = [F_1, F_2]^{\top}, K(0) = [K_1(0), K_2(0)]^{\top}$. This shows that $|e(z)| \leq c|\ln z|$ when $z \rightarrow 0$ and so all together, e is analytic in $\mathbb{C} \setminus \{0\}$ with a possible blow up at the origin that is logarithmic at worst. Hence, $e(z)$ must be entire and with its limiting behaviour at $z = \infty$, thus $e(z) \equiv 1$ for all $z \in \mathbb{C}$. As a consequence of this, we now know that any solution $X(z) \in \mathcal{I}(\mathcal{H}_2)$ of problem 9.8 is invertible for all $z \in \mathbb{C} \setminus [0, \infty)$ and so are its continuous boundary values $X_{\pm}(z), z \in (0, \infty)$. Thus, if $X^{(1)}(z), X^{(2)}(z)$ are two solutions of problem 9.8, then

$$\mathcal{I}(\mathcal{H}_2) \ni Y(z) := X^{(1)}(z)(X^{(2)}(z))^{-1} = \mathbb{I}_2 + Y_0(z), \quad z \in \mathbb{C} \setminus [0, \infty)$$

is analytic in $\mathbb{C} \setminus [0, \infty)$, has continuous boundary values $Y_{\pm}(z)$ on $(0, \infty)$ and it obeys a problem in the style of problem 9.8 with

$$Y_+(z) = Y_-(z), \quad z \in (0, \infty).$$

Moreover, the kernel $Y_0(z|x, y)$ of $Y_0(z)$ has at worst a square logarithmic singularity at $z = 0$, so the map $z \mapsto Y_0(z|x, y)$ is entire $(\sigma \otimes \sigma)$ -almost everywhere. But $Y_0(z|x, y) \rightarrow 0$ as $z \rightarrow \infty$

also $(\sigma \otimes \sigma)$ -almost everywhere, so by Liouville’s theorem we conclude $Y(z) = \mathbb{I}_2$ and which proves that problem 9.8, if solvable, is uniquely solvable.

Second, we prove solvability of problem 9.8 and recall to this end that $1 - K_{\sigma,s}$ is invertible on $L^2(0, \infty)$ for any $s \in \mathbb{R}$, see proposition 9.6. Thus the integral equation (9.16) is uniquely solvable and by the boundedness of the resolvent we find for its solution, uniformly in $x \in \mathbb{R}$,

$$\|n_i(\cdot|x)\|_{L^2(0,\infty)} \leq c\|m_i(\cdot|x)\|_{L^2(0,\infty)}, \quad c > 0, \quad i = 1, 2. \tag{9.18}$$

Now consider $X(z)$ as in (9.15) and note that the kernel of $X_0(z)$ (the integral piece) is built out of the functions

$$X_0^{ij}(z|x, y) = \int_0^\infty n_i(w|x)k_j(w|y)\frac{dw}{w - z}, \quad z \notin [0, \infty), \quad (x, y) \in \mathbb{R}^2.$$

However, using (9.18) and standard estimates for Airy functions, see [117, section 9.7] we obtain by Cauchy–Schwarz

$$\left|X_0^{ij}(z|x, y)\right| \leq \frac{c}{\text{dist}(z, \mathbb{R}_{\geq 0})} |s + x|^{\frac{3}{4}}|s + y|^{\frac{3}{4}}, \tag{9.19}$$

uniformly in $(x, y) \in \mathbb{R}^2$. Thus $(x, y) \mapsto X_0(z|x, y)$ is in $L^2(\mathbb{R}^2, d\sigma \otimes d\sigma; \mathbb{C}^{2 \times 2})$ for $z \in \mathbb{C} \setminus [0, \infty)$. Moreover, by the standard regularity properties of the resolvent operator, we also find from (9.16) that $z \mapsto X_0^{ij}(z|x, y)$ is Hölder continuous on $\mathbb{C} \setminus [0, \infty)$ for any $(x, y) \in \mathbb{R}^2$ and $i, j = 1, 2$. Thus, by the Plemelj–Sokhostki formulæ, $z \mapsto X_0(z|x, y)$ is analytic in $\mathbb{C} \setminus [0, \infty)$ for any $(x, y) \in \mathbb{R}^2$ and so the right-hand side in (9.15) analytic in $z \in \mathbb{C} \setminus [0, \infty)$ according to definition 9.3. Having just established condition (a) in problem 9.8, we now turn to (b), (c) and (d): First, estimate (9.14) follows from (9.19) and the Plemelj–Sokhotski formulæ. Second, $z \mapsto X_0^{ij}(z|x, y)$ is Hölder continuous on $[0, \infty)$ for any $(x, y) \in \mathbb{R}^2$, thus $X(z)$ admits continuous boundary values $X_\pm(z) \in \mathcal{I}(\mathcal{H}_2)$ on $(0, \infty)$ according to definition 9.4 by the Plemelj–Sokhotski formulæ. Now verify (9.11), so first compute from (9.15),

$$X_+(z) - X_-(z) = 2\pi i \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix}, \quad z \in (0, \infty). \tag{9.20}$$

On the other hand, by (9.12) and (9.15),

$$\begin{aligned} X_-(z)G(z) &= X_-(z) \left\{ \mathbb{I}_2 + 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \right\} \\ &= X_-(z) + 2\pi i \left\{ \mathbb{I}_2 + \int_0^\infty \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \right. \\ &\quad \left. \times \frac{dw}{w - z_-} \right\} \circ \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix}, \\ & \quad z \in (0, \infty), \end{aligned} \tag{9.21}$$

and since, using here the general $(N_i \otimes K_j)(M_j \otimes K_\ell) = \langle K_j, M_j \rangle_{\mathcal{H}_1} (N_i \otimes K_\ell)$ fact,

$$\begin{aligned} &\begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \\ &= (w - z)K_{\sigma,s}(w, z) \begin{bmatrix} N_1(w) \otimes K_1(z) & -N_1(w) \otimes K_2(z) \\ N_2(w) \otimes K_1(z) & -N_2(w) \otimes K_2(z) \end{bmatrix}, \end{aligned}$$

$$(w, z) \in (0, \infty), \times (0, \infty),$$

the chain of equalities (9.21) simplifies further

$$\begin{aligned} X_-(z)G(z) = & X_-(z) + 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \\ & + 2\pi i \int_0^\infty K_{\sigma,s}(w, z) \begin{bmatrix} N_1(w) \otimes K_1(z) & -N_1(w) \otimes K_2(z) \\ N_2(w) \otimes K_1(z) & -N_2(w) \otimes K_2(z) \end{bmatrix} dw, \\ & z \in (0, \infty). \end{aligned} \tag{9.22}$$

But (9.16) is equivalent to the operator equation

$$N_i(z) - \int_0^\infty K_{\sigma,s}(z, w)N_i(w) dw = M_i(z), \quad z \in (0, \infty),$$

which, by symmetry of $K_{\sigma,s}$, simplifies (9.22) to

$$\begin{aligned} X_-(z)G(z) = & X_-(z) + 2\pi i \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix} \\ & \stackrel{(9.20)}{=} X_+(z), \quad z \in (0, \infty). \end{aligned}$$

Condition (b) in problem 9.8 has therefore been established and we are left with the singular behaviour (9.12) and (9.13) near $z = 0$. Fix $z \notin [0, \infty)$, $0 < \epsilon < 1$ and rewrite (9.15) as

$$\begin{aligned} X(z) = & \mathbb{I}_2 + \ln \left(\frac{z - 2\epsilon}{z} \right) \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix} \\ & + \int_0^{2\epsilon} \left\{ \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \right. \\ & \left. - \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix} \right\} \frac{dw}{w - z} \\ & + \int_{2\epsilon}^\infty \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \frac{dw}{w - z}. \end{aligned} \tag{9.23}$$

The remaining two integrals are operators in $\mathcal{I}(\mathcal{H}_2)$ which are analytic in the disk $|z| < \epsilon$ and the same is true for the operator multiplied by the logarithm: indeed we already showed that

$$X_-(z)G(z) = X_-(z) + 2\pi i \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix}, \quad z \in (0, \infty), \tag{9.24}$$

and since the trace class $G(z) \in \mathcal{I}(\mathcal{H}_2)$ is invertible with

$$(G(z))^{-1} = \mathbb{I}_2 - 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix}, \quad z \in \mathbb{R},$$

we also compute (as done previously for $X_-(z)G(z)$),

$$\begin{aligned}
 X_+(z)(G(z))^{-1} &= X_+(z) \left\{ \mathbb{I}_2 - 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \right\} \\
 &= X_+(z) - 2\pi i \left\{ \mathbb{I}_2 + \int_0^\infty \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \frac{dw}{w - z_+} \right\} \\
 &\quad \circ \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \\
 &= X_+(z) - 2\pi i \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix}, \quad z \in (0, \infty). \tag{9.25}
 \end{aligned}$$

Hence, combining (9.24) and (9.25) with the explicit form (9.11), we deduce

$$\begin{aligned}
 X_\pm(z) &\begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \\
 &= \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix}, \quad z \in (0, \infty),
 \end{aligned}$$

and from this, for $z \in (0, \infty)$, the following operator composition formula for $N_i(z)$,

$$N(z) = X(z)M(z); \quad N(z) := [N_1(z), N_2(z)]^\top, \quad M(z) := [M_1(z), M_2(z)]^\top, \tag{9.26}$$

which is independent of the \pm choice for the boundary values of $X(z)$ on $(0, \infty)$. Now use (9.26) back in (9.23) and conclude that $N(z)$ near $z = 0$ is of the form

$$N(z) = (\mathbb{I}_2 + N_{\text{reg}}(z)) M(z), \quad N_{\text{reg}}(z) \in \mathcal{I}(\mathcal{H}_2) \text{ is analytic at } z = 0,$$

since $(N_i(z) \otimes K_1(z))M_1(z) - (N_i(z) \otimes K_2(z))M_2(z) = N_i(z) \otimes (M_1(z)K_1(z) - M_2(z)K_2(z)) = 0$, i.e. since the logarithmic term drops out. All together, we can expand $N_i(z)$ near $z = 0$, which verifies the existence of the multiplication operators F_i in (9.12) with their indicated properties, and estimate the kernel

$$\begin{aligned}
 &\left| \frac{1}{w - z} \left\{ \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \right. \right. \\
 &\quad \left. \left. - \begin{bmatrix} N_1(z) \otimes K_1(z) & -N_1(z) \otimes K_2(z) \\ N_2(z) \otimes K_1(z) & -N_2(z) \otimes K_2(z) \end{bmatrix} \right\} (x, y) \right| \\
 &\leq c|1 + s + x|^{\frac{3}{4}}|1 + s + y|^{\frac{3}{4}}, \quad c > 0,
 \end{aligned}$$

uniformly for $w \in [0, \epsilon] \subset \mathbb{R}$, $|z| < \epsilon$ and any $(x, y) \in \mathbb{R}^2$. All combined in (9.23) we have now established (9.12) together with (9.13). This completes our proof. \square

During our upcoming computations we will use the following consequences of theorem 9.9.

Corollary 9.10. *Let $X(z)$ denote the unique solution of problem 9.8. Then*

$$\begin{aligned}
 (X(z))^{-1} &= \mathbb{I}_2 - \int_0^\infty \begin{bmatrix} M_1(w) \otimes L_1(w) & -M_1(w) \otimes L_2(w) \\ M_2(w) \otimes L_1(w) & -M_2(w) \otimes L_2(w) \end{bmatrix} \frac{dw}{w - z}, \tag{9.27} \\
 z &\in \mathbb{C} \setminus [0, \infty),
 \end{aligned}$$

where $L_i(w)$ are the integral operators on \mathcal{H}_1 with kernel $\ell_i(z|y)$ determined from the integral equation

$$(1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)})\ell_i(\cdot|y) = k_i(\cdot|y), \quad i = 1, 2, \tag{9.28}$$

with $y \in \mathbb{R}$. Moreover, for any $z \in (0, \infty)$,

$$N(z) = X(z)M(z), \quad L(z) = K(z)(X(z))^{-1} \tag{9.29}$$

with $L(z) := [L_1(z), -L_2(z)]$, $K(z) := [K_1(z), -K_2(z)]$ and near $z = 0$,

$$N(z) = (\mathbb{I}_2 + N_{reg}(z)) M(z), \quad L(z) = K(z) (\mathbb{I}_2 + L_{reg}(z)),$$

where $N_{reg}(z) \in \mathcal{I}(\mathcal{H}_2)$ and $L_{reg}(z) \in \mathcal{I}(\mathcal{H}_2)$ are analytic at $z = 0$.

Proof. Let $Y(z)$ denote the right-hand side of (9.27) and $\bar{Y}_0(z)$ the underlying integrand without the Cauchy kernel. With (9.15) we then compute ($\bar{X}_0(z)$ being similarly defined)

$$\begin{aligned} X(z)Y(z) = & \mathbb{I}_2 + \int_0^\infty (\bar{X}_0(w) - \bar{Y}_0(w)) \frac{dw}{w-z} \\ & - \int_0^\infty \int_0^\infty \bar{X}_0(w)\bar{Y}_0(\lambda) \frac{dw}{w-z} \frac{d\lambda}{\lambda-z}, \quad z \notin [0, \infty). \end{aligned}$$

But

$$\bar{X}_0(w)\bar{Y}_0(\lambda) = (w - \lambda)K_{\sigma,s}(w, \lambda) \begin{bmatrix} N_1(w) \otimes L_1(\lambda) & -N_1(w) \otimes L_2(\lambda) \\ N_2(w) \otimes L_1(\lambda) & -N_2(w) \otimes L_2(\lambda) \end{bmatrix},$$

so using partial fractions in the double integral and both integral equations (9.16) and (9.28),

$$\begin{aligned} \int_0^\infty (\bar{X}_0(w) - \bar{Y}_0(w)) \frac{dw}{w-z} - \int_0^\infty \int_0^\infty \bar{X}_0(w)\bar{Y}_0(\lambda) \frac{dw}{w-z} \frac{d\lambda}{\lambda-z} = & 0, \tag{9.30} \\ z \notin [0, \infty), \end{aligned}$$

which shows that $X(z)Y(z) = \mathbb{I}_2$ for $z \notin [0, \infty)$. But any solution $X(z) \in \mathcal{I}(\mathcal{H}_2)$ of problem 9.8 is invertible, so a one-sided inverse is a two-sided inverse and thus identity (9.27) holds for $z \notin [0, \infty)$. Moving ahead, the first half of (9.29) and the associated local regularity statement have already been proven above. Focussing thus on the second half, the right-hand side in (9.27) is such that by the Plemelj–Sokhotski formula

$$(X(z))_+^{-1} - (X(z))_-^{-1} = -2\pi i \begin{bmatrix} M_1(z) \otimes L_1(z) & -M_1(z) \otimes L_2(z) \\ M_2(z) \otimes L_1(z) & -M_2(z) \otimes L_2(z) \end{bmatrix}, \tag{9.31}$$

$$z \in (0, \infty),$$

as well as

$$\begin{aligned} (G(z))^{-1}(X(z))_-^{-1} = & \left\{ \mathbb{I}_2 - 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \right\} (X(z))_-^{-1} \\ \stackrel{(9.27)}{=} & (X(z))_-^{-1} - 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \\ \circ & \left\{ \mathbb{I}_2 - \int_0^\infty \begin{bmatrix} M_1(w) \otimes L_1(w) & -M_1(w) \otimes L_2(w) \\ M_2(w) \otimes L_1(w) & -M_2(w) \otimes L_2(w) \end{bmatrix} \frac{dw}{w-z_-} \right\}. \end{aligned}$$

But since

$$\begin{aligned} & \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \begin{bmatrix} M_1(w) \otimes L_1(w) & -M_1(w) \otimes L_2(w) \\ M_2(w) \otimes L_1(w) & -M_2(w) \otimes L_2(w) \end{bmatrix} \\ &= (z - w)K_{\sigma,s}(z, w) \begin{bmatrix} M_1(z) \otimes L_1(w) & -M_1(z) \otimes L_2(w) \\ M_2(z) \otimes L_1(w) & -M_2(z) \otimes L_2(w) \end{bmatrix}, \\ & (w, z) \in (0, \infty) \times (0, \infty), \end{aligned}$$

we find for any $z \in (0, \infty)$

$$\begin{aligned} (G(z))^{-1}(X(z))_+^{-1} &\stackrel{(9.28)}{=} (X(z))_+^{-1} - 2\pi i \\ &\times \begin{bmatrix} M_1(z) \otimes L_1(z) & -M_1(z) \otimes L_2(z) \\ M_2(z) \otimes L_1(z) & -M_2(z) \otimes L_2(z) \end{bmatrix}, \end{aligned} \tag{9.32}$$

and similarly

$$\begin{aligned} G(z)(X(z))_+^{-1} &= \left\{ \mathbb{I}_2 + 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \right\} (X(z))_+^{-1} \\ &\stackrel{(9.27)}{=} (X(z))_+^{-1} + 2\pi i \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} \\ &\circ \left\{ \mathbb{I}_2 - \int_0^\infty \begin{bmatrix} M_1(w) \otimes L_1(w) & -M_1(w) \otimes L_2(w) \\ M_2(w) \otimes L_1(w) & -M_2(w) \otimes L_2(w) \end{bmatrix} \frac{dw}{w - z_+} \right\} \\ &= (X(z))_+^{-1} + 2\pi i \begin{bmatrix} M_1(z) \otimes L_1(z) & -M_1(z) \otimes L_2(z) \\ M_2(z) \otimes L_1(z) & -M_2(z) \otimes L_2(z) \end{bmatrix}, \quad z \in (0, \infty). \end{aligned}$$

Comparing this last identity with (9.32), we deduce

$$\begin{aligned} & \begin{bmatrix} M_1(z) \otimes K_1(z) & -M_1(z) \otimes K_2(z) \\ M_2(z) \otimes K_1(z) & -M_2(z) \otimes K_2(z) \end{bmatrix} (X(z))_+^{-1} \\ &= \begin{bmatrix} M_1(z) \otimes L_1(z) & -M_1(z) \otimes L_2(z) \\ M_2(z) \otimes L_1(z) & -M_2(z) \otimes L_2(z) \end{bmatrix}, \quad z \in (0, \infty), \end{aligned}$$

which is once more independent of the \pm choice in $(X(z))_+^{-1}$. Reading the last operator equality entry wise we then find

$$[L_1(z), -L_2(z)] = [K_1(z), -K_2(z)](X(z))_+^{-1}, \quad z \in (0, \infty),$$

as claimed in (9.29). Finally, the expansion of $L(z)$ near $z = 0$ follows as in the proof of theorem 9.9 by splitting the formula (9.27) as in (9.23) and using that $K_1(z)(M_1(z) \otimes L_i(z)) - K_2(z)(M_2(z) \otimes L_i(z)) = (K_1(z)M_1(z) - K_2(z)M_2(z)) \otimes L_i(z) = 0$. This concludes our proof. \square

Another useful consequence of theorem 9.9 is the following result

Corollary 9.11. *Let $(1 - K_{\sigma,s}|_{L^2(0,\infty)})^{-1} = 1 + R_{\sigma,s}|_{L^2(0,\infty)}$, then the resolvent kernel is of the form*

$$(x - y)R_{\sigma,s}(x, y) = \int_{-\infty}^\infty (\ell_1(x|t)n_1(y|t) - \ell_2(x|t)n_2(y|t)) \, d\sigma(t), \quad x, y > 0 \tag{9.33}$$

with the functions defined in (9.16) and (9.28). In particular, for any $x > 0$,

$$\int_{-\infty}^{\infty} (\ell_1(x|t)n_1(x|t) - \ell_2(x|t)n_2(x|t)) \, d\sigma(t) = 0.$$

Proof. The representation formula (9.33) is a general fact of integrable operators, see [34, (6), (7)]. In our case the finite sum in [34, (1)] is naturally replaced by a weighted integral but this has no significant impact on the formula for the resolvent kernel. In detail, since $K_{\sigma,s}$ is a Hilbert-Schmidt operator on $L^2(0, \infty)$, compare lemma 9.5, so is $R_{\sigma,s} = (1 - K_{\sigma,s})^{-1}K_{\sigma,s}$ and therefore $R_{\sigma,s}$ has a kernel $R_{\sigma,s}(x, y)$. Next, with M denoting multiplication by the independent variable, we check that the commutators $[M, K_{\sigma,s}]$, resp. $[M, R_{\sigma,s}]$ have kernels $(x - y)K_{\sigma,s}(x, y)$, resp. $(x - y)R_{\sigma,s}(x, y)$. On the other hand, on $L^2(0, \infty)$,

$$\begin{aligned} [M, R_{\sigma,s}] &= [M, (1 - K_{\sigma,s})^{-1} - 1] = [M, (1 - K_{\sigma,s})^{-1}] \\ &= (1 - K_{\sigma,s})^{-1} [M, K_{\sigma,s}] (1 - K_{\sigma,s})^{-1}, \end{aligned}$$

which yields for the underlying kernels, by (9.10), (9.16) and (9.28),

$$(x - y)R_{\sigma,s}(x, y) = \int_{-\infty}^{\infty} (\ell_1(x|t)n_1(y|t) - \ell_2(x|t)n_2(y|t)) \, d\sigma(t), \quad x, y > 0.$$

□

The last corollary concludes our content on the operator-valued problem 9.8. We will now use this problem in the derivation of an integrable system for the Fredholm determinant $F_{\sigma}(s)$.

9.3. An operator-valued Lax pair

From now on it will be convenient, but not essential, to view the aforementioned multiplication operators $M_i(z)$, see definition 9.7, and $N_i(w)$, see theorem 9.9, as integral operators on \mathcal{H}_1 with appropriate distributional kernels. To the point, we replace

$$\begin{aligned} m_i(z|x) &\mapsto m_i(z|x, y) := m_i(z|x)\delta(x - y)(\sigma'(y))^{-1}, \\ n_i(w|x) &\mapsto n_i(w|x, y) := n_i(w|x)\delta(x - y)(\sigma'(y))^{-1}, \end{aligned}$$

for $(x, y) \in \mathbb{R}^2$ with the (positive) Radon–Nikodym derivative $\sigma'(t) := \frac{d\sigma}{dt}$ and where, by definition,

$$\int_{-\infty}^{\infty} \delta(x - y)(\sigma'(y))^{-1} f(y) \, d\sigma(y) := f(x), \quad f \in L^2(\mathbb{R}, d\sigma).$$

Subject to this convention we then compute, recall (9.26),

$$M_z(z|x, y) = \begin{bmatrix} 0 & z + s + x \\ 1 & 0 \end{bmatrix} M(z|x, y), \quad M_z(z|x, y) \equiv \frac{\partial}{\partial z} M(z|x, y),$$

which is the standard Airy system for the kernels and which translates into the integral operator identity

$$M_z(z) = (zA_1 + A_2)M(z), \quad z \in \mathbb{R}. \tag{9.34}$$

Here, A_i are z -independent integral operators on \mathcal{H}_2 with distributional kernels

$$A_1(x, y) := \delta(x - y) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} (\sigma'(y))^{-1},$$

$$A_2(x, y) := \delta(x - y) \begin{bmatrix} 0 & s + x \\ 1 & 0 \end{bmatrix} (\sigma'(y))^{-1}.$$

Now return to corollary 9.10 and z -differentiate the first identity in (9.29),

$$\frac{\partial N}{\partial z}(z) \stackrel{(9.34)}{=} (X_z(z)(X(z))^{-1} + X(z)(zA_1 + A_2)(X(z))^{-1}) N(z),$$

$$z \in \mathbb{C} \setminus [0, \infty),$$

where $X(z)$ is the unique solution of problem 9.8. As it turns out the right-hand side in the last equation gives rise to an integral operator in $\mathcal{I}(\mathcal{H}_2)$ which is analytic for $z \neq 0$.

Proposition 9.12. *Define the integral operator*

$$Y(z) := X_z(z)(X(z))^{-1} + X(z)(zA_1 + A_2)(X(z))^{-1}, \quad z \in \mathbb{C} \setminus [0, \infty) \tag{9.35}$$

on \mathcal{H}_2 . Then $Y(z) \in \mathcal{I}(\mathcal{H}_2)$ is analytic for $z \in \mathbb{C} \setminus \{0\}$.

Proof. Note that $Y(z) \in \mathcal{I}(\mathcal{H}_2)$ as consequence of theorem 9.9, corollary 9.10 and a direct computation of the kernel of $X(z)(zA_1 + A_2)(X(z))^{-1}$. Moreover, $Y(z)$ admits continuous boundary values $Y_{\pm}(z) \in \mathcal{I}(\mathcal{H}_2)$ by consequence of theorem 9.9 and corollary 9.10, and we compute them directly,

$$Y_+(z) = (X_{z-}(z)G(z) + X_-(z)G_z(z))(G(z))^{-1}(X_-(z))^{-1}$$

$$+ X_-(z)G(z)(zA_1 + A_2)(G(z))^{-1}(X_-(z))^{-1}, \quad z > 0. \tag{9.36}$$

However, from (9.11), we derive the kernel identity

$$G_z(x|x, y) = \int_{-\infty}^{\infty} \{ (zA_1(x, t) + A_2(x, t)) G_0(z|t, y) - G_0(z|x, t)(zA_1(t, y) + A_2(t, y)) \} d\sigma(t)$$

with $G(z) = \mathbb{I}_2 + G_0(z)$ and this leads us to the operator commutator identity

$$G_z(z) = [zA_1 + A_2, G(z)] \in \mathcal{I}(\mathcal{H}_2). \tag{9.37}$$

Inserting (9.37) into (9.36), after simplification,

$$Y_+(z) = X_{z-}(z)(X_-(z))^{-1} + X_-(z)(zA_1 + A_2)(X_-(z))^{-1} = Y_-(z),$$

$$z > 0,$$

so $z \mapsto Y(z|x, y)$ is in fact continuous across $(0, \infty) \ni z$ for any $(x, y) \in \mathbb{R}^2$. But $(x, y) \mapsto Y_{\pm}(z|x, y)$ is also in $L^2(\mathbb{R}^2, d\sigma \otimes d\sigma; \mathbb{C}^{2 \times 2})$ for any $z \in (0, \infty)$, so $Y(z)$ is analytic for $z \in \mathbb{C} \setminus \{0\}$ according to definition 9.3. This completes our proof. \square

Summarizing our computations so far, we have obtained the operator-valued ODE system

$$\frac{\partial N}{\partial z}(z) = Y(z)N(z), \quad z \in \mathbb{C} \setminus \{0\},$$

where $Y(z)$ is analytic for $z \in \mathbb{C} \setminus \{0\}$. This is the analogue of Plemelj’s (4.19) in our context and we will now evaluate the ‘coefficient’ operator $Y(z)$ from Liouville’s theorem. First, return

to the explicit formula (9.15), write $\frac{1}{w-z} = -\frac{1}{z} + \frac{w}{z(w-z)}$ and obtain

$$X(z) = \mathbb{I}_2 - \frac{1}{z} \int_0^\infty \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} dw + X_{\text{asy}}(z),$$

$$z \in \mathbb{C} \setminus [0, \infty),$$

where $X_{\text{asy}}(z) \in \mathcal{I}(\mathcal{H}_2)$ and for any compact $K \subset \mathbb{C}$ with $\text{dist}(0, K) > 0$ there exists $C > 0$ such that for $z \in \mathbb{C} \setminus K$,

$$\|X_{\text{asy}}(z|x, y)\| \leq \frac{C}{1 + |z|^2} |s + x|^{\frac{7}{4}} |s + y|^{\frac{7}{4}} \tag{9.38}$$

uniformly in $(x, y) \in \mathbb{R}^2$. Similarly, from (9.27),

$$(X(z))^{-1} = \mathbb{I}_2 + \frac{1}{z} \int_0^\infty \begin{bmatrix} M_1(w) \otimes L_1(w) & -M_1(w) \otimes L_2(w) \\ M_2(w) \otimes L_1(w) & -M_2(w) \otimes L_2(w) \end{bmatrix} dw$$

$$+ \widehat{X}_{\text{asy}}(z), \quad z \in \mathbb{C} \setminus [0, \infty),$$

and $\widehat{X}_{\text{asy}}(z) \in \mathcal{I}(\mathcal{H}_2)$ behaves similarly to (9.38). Inserting these two formulæ for $X(z)$ and $(X(z))^{-1}$ into (9.35) we obtain the exact operator identity

$$Y(z) = zA_1 + A_2 + A_1 \int_0^\infty \begin{bmatrix} M_1(w) \otimes L_1(w) & -M_1(w) \otimes L_2(w) \\ M_2(w) \otimes L_1(w) & -M_2(w) \otimes L_2(w) \end{bmatrix} dw$$

$$- \int_0^\infty \begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} dw A_1 + Y_\infty(z),$$

$$z \in \mathbb{C} \setminus [0, \infty),$$

with $Y_\infty(z) \in \mathcal{I}(\mathcal{H}_2)$. Observe that for any compact $K \subset \mathbb{C}$ with $\text{dist}(0, K) > 0$ there exists $C > 0$ such that for $z \in \mathbb{C} \setminus K$,

$$\|Y_\infty(z|x, y)\| \leq \frac{C}{1 + |z|} |s + x|^{\frac{7}{4}} |s + y|^{\frac{7}{4}}, \quad \forall (x, y) \in \mathbb{R}^2. \tag{9.39}$$

Lemma 9.13. *We have the finite rank integral operator identity*

$$\int_0^\infty M_i(w) \otimes L_j(w) dw = \int_0^\infty N_i(w) \otimes K_j(w) dw,$$

valid for any $i, j = 1, 2$.

Proof. This follows from (9.30) by multiplying through with z and sending $z \rightarrow \infty$ with $\Im z > 0$, say. The double integral term vanishes in this limit by the dominated convergence theorem and the first term yields the stated identity by the same reasoning. \square

The symmetry lemma 9.13 allows us to simplify our expression for $Y(z)$, we have now

$$Y(z) = zA_1 + A_0 + Y_\infty(z), \quad z \in \mathbb{C} \setminus [0, \infty), \tag{9.40}$$

where A_0 is the integral operator on \mathcal{H}_2 with kernel $A_0(x, y) = [A_0^{ij}(x, y)]_{i,j=1}^2$ and

$$A_0^{11}(x, y) = \int_0^\infty (N_2(w) \otimes K_1(w))(x, y) dw = -A_0^{22}(x, y),$$

$$\begin{aligned}
 A_0^{21}(x, y) &= \delta(x - y)(\sigma'(y))^{-1}, \\
 A_0^{12}(x, y) &= \delta(x - y)(s + x)(\sigma'(y))^{-1} \\
 &\quad - \int_0^\infty (N_2(w) \otimes K_2(w) + N_1(w) \otimes K_1(w))(x, y) dw.
 \end{aligned}$$

The representation (9.40) mimics a Poincaré asymptotic expansion of $Y(z)$ at $z = \infty$. We will now combine it with a corresponding representation at $z = 0$ which is computed from (9.23) and its counterpart for (9.27) when inserted into (9.35). The result equals

$$Y(z) = -\frac{1}{z} \begin{bmatrix} N_1(0) \otimes L_1(0) & -N_1(0) \otimes L_2(0) \\ N_2(0) \otimes L_1(0) & -N_2(0) \otimes L_2(0) \end{bmatrix} + Y_{\text{ori}}(z), \quad 0 < |z| < \epsilon, \tag{9.41}$$

with $Y_{\text{ori}}(z) \in \mathcal{I}(\mathcal{H}_2)$ analytic at $z = 0$ and where we have used the identity

$$\begin{aligned}
 &\begin{bmatrix} N_1(w) \otimes K_1(w) & -N_1(w) \otimes K_2(w) \\ N_2(w) \otimes K_1(w) & -N_2(w) \otimes K_2(w) \end{bmatrix} \begin{bmatrix} M_1(\lambda) \otimes L_1(\lambda) & -M_1(\lambda) \otimes L_2(\lambda) \\ M_2(\lambda) \otimes L_1(\lambda) & -M_2(\lambda) \otimes L_2(\lambda) \end{bmatrix} \\
 &= (w - \lambda)K_{\sigma,s}(w, \lambda) \begin{bmatrix} N_1(w) \otimes L_1(\lambda) & -N_1(w) \otimes L_2(\lambda) \\ N_2(w) \otimes L_1(\lambda) & -N_2(w) \otimes L_2(\lambda) \end{bmatrix}, \quad w, \lambda > 0,
 \end{aligned}$$

to ensure that all logarithmic terms cancel out in our computation of (9.41). At this point we combine proposition 9.12 with the representations (9.40), (9.41) and deduce from Liouville’s theorem that

$$Y(z) = zA_1 + A_0 + \frac{1}{z}A_{-1}, \quad z \in \mathbb{C} \setminus \{0\}, \tag{9.42}$$

where $A_{-1} \in \mathcal{I}(\mathcal{H}_2)$ has kernel $A_{-1}(x, y) = \left[A_{-1}^{ij}(x, y) \right]_{i,j=1}^2$ with

$$A_{-1}^{ij}(x, y) = (-1)^j (N_i(0) \otimes L_j(0))(x, y), \quad (x, y) \in \mathbb{R}^2.$$

Remark 9.14. The operator representation (9.42) is the analogue of [75, (9.4.28)] in the derivation of the Tracy–Widom Painlevé-II formula [135, (1.17)] through matrix-valued Riemann–Hilbert techniques. And, as in [75, (9.4.28)], the operator A_{-1} is nilpotent as a consequence of (9.33),

$$(A_{-1})^2 = 0, \quad \text{tr}_{\mathcal{H}_2} A_{-1} = 0.$$

Furthermore by (9.28), (9.16), for any $z > 0$,

$$\begin{aligned}
 L_2(z) &= \left(1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)} \right)^{-1} K_2(z) \\
 &\stackrel{(9.10)}{=} \left(1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)} \right)^{-1} M_1(z) = N_1(z), \\
 N_2(z) &= \left(1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)} \right)^{-1} M_2(z) \stackrel{(9.10)}{=} \left(1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)} \right)^{-1} K_1(z) \\
 &= L_1(z),
 \end{aligned}$$

and which should be compared with the parametrization of the nilpotent matrix in [75, (9.4.30)].

Before continuing, let us summarize our computations as they will form the first half of the anticipated Lax pair.

Proposition 9.15. *Let $N(z), z \in (0, \infty)$ denote the integral operator on \mathcal{H}_2 as in (9.29). Then*

$$\frac{\partial N}{\partial z}(z) = A(z)N(z), \quad A(z) \equiv A(z; s) = zA_1 + A_0 + \frac{1}{z}A_{-1}, \quad z \in \mathbb{C} \setminus \{0\}, \quad (9.43)$$

where the integral operators $A_k = \left[A_k^{ij} \right]_{i,j=1}^2$ on \mathcal{H}_2 have the kernels $A_k^{ij}(x, y)$ stated above.

The second half of the Lax pair

$$\frac{\partial N}{\partial s}(z) = B(z)N(z)$$

is computed in the same fashion, throughout replacing z -derivatives with s -derivatives. The main change occurs near $z = 0$ since the analogue of (9.35) with s -partial derivatives is analytic at $z = 0$. We thus have the simpler result

Proposition 9.16. *Let $N(z), z \in (0, \infty)$ denote the integral operator on \mathcal{H}_2 as in (9.29). Then*

$$\frac{\partial N}{\partial s}(z) = B(z)N(z), \quad B(z) \equiv B(z; s) = zB_1 + B_0, \quad z \in \mathbb{C}, \quad (9.44)$$

where the integral operators B_k on \mathcal{H}_2 equal $B_1 := A_1$ and $B_0 := A_0$.

Remark 9.17. Our derivation of (9.43) and (9.44) is inspired by the non-rigorous computations in [95, section 5] and in [96, chapter XV.6].

Propositions 9.15 and 9.16 complete our derivation of the operator-valued Lax pair. We now proceed and show that this pair indeed produces an integrable system for the Fredholm determinant $F_\sigma(s)$, exactly the integro-differential Painlevé-II equation of Amir–Corwin–Quastel.

9.4. Derivation of the integro-differential equation

We first establish a connection between $F_\sigma(s)$ and problem 9.8.

Proposition 9.18. *For any $s \in \mathbb{R}$,*

$$\frac{d}{ds} \ln F_\sigma(s) = \operatorname{tr}_{\mathcal{H}_1} \int_0^\infty N_2(w) \otimes K_1(w) dw = -\operatorname{tr}_{\mathcal{H}_1} A_0^{22} \quad (9.45)$$

in terms of the integral operator A_0 back in (9.40).

Proof. We begin with the standard Jacobi identity

$$\frac{d}{ds} \ln F_\sigma(s) = -\operatorname{tr}_{L^2(0,\infty)} \left(\left(1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)} \right)^{-1} \frac{dK_{\sigma,s}}{ds} \right),$$

and compute the kernel derivative

$$\begin{aligned} \frac{dK_{\sigma,s}}{ds}(x, y) &\stackrel{(9.5)}{=} - \int_{-\infty}^{\infty} \text{Ai}(x + s + t)\text{Ai}(y + s + t) d\sigma(t) \\ &\stackrel{(9.10)}{=} - \int_{-\infty}^{\infty} m_2(x|t)k_1(y|t)d\sigma(t). \end{aligned}$$

Hence,

$$\begin{aligned} \frac{d}{ds} \ln F_{\sigma}(s) &= - \int_0^{\infty} \int_0^{\infty} (1 - K_{\sigma,s}|_{L^2(0,\infty)})^{-1}(x, y) \frac{dK_{\sigma,s}}{ds}(y, x) dx dy \\ &\stackrel{(9.16)}{=} \int_{-\infty}^{\infty} \left[\int_0^{\infty} (N_2(x) \otimes K_1(x))(t, t) dx \right] d\sigma(t) \\ &= \text{tr}_{\mathcal{H}_1} \int_0^{\infty} N_2(x) \otimes K_1(x) dx, \end{aligned}$$

and the second equality in (9.45) follows from the explicit formulæ for the entries of A_0 . \square

At this point we use the Lax system (9.43), (9.44), i.e.

$$\frac{\partial N}{\partial z}(z) = A(z)N(z), \quad \frac{\partial N}{\partial s}(z) = B(z)N(z)$$

and write out its compatibility condition

$$AB - BA = \frac{\partial B}{\partial z} - \frac{\partial A}{\partial s},$$

keeping in mind that the entries of A and B are operators and thus do not commute in general. The result equals

$$(A_0)_s = A_1 + A_1A_{-1} - A_{-1}A_1, \quad (A_{-1})_s = A_0A_{-1} - A_{-1}A_0,$$

or more explicit, using partial information for the A_0^{ij} operators ($A_1^{12} = A_0^{21} = \mathbb{I}_1, A_0^{22} = -A_0^{11}$),

$$(A_0^{22})_s = -A_{-1}^{21}, \quad (A_0^{12})_s = \mathbb{I}_1 + A_{-1}^{22} - A_{-1}^{11}, \tag{9.46}$$

which follows from the equation on $(A_0)_s$ and from the equation on $(A_{-1})_s$,

$$\begin{cases} (A_{-1}^{11})_s = -A_0^{22}A_{-1}^{11} + A_{-1}^{11}A_0^{22} + A_0^{12}A_{-1}^{21} - A_{-1}^{12}, \\ (A_{-1}^{21})_s = A_{-1}^{11} - A_{-1}^{22} + A_0^{22}A_{-1}^{21} + A_{-1}^{21}A_0^{22}, \\ (A_{-1}^{12})_s = -A_0^{22}A_{-1}^{12} - A_{-1}^{12}A_0^{22} + A_0^{12}A_{-1}^{22} - A_{-1}^{11}A_0^{12}, \\ (A_{-1}^{22})_s = A_0^{22}A_{-1}^{22} - A_{-1}^{22}A_0^{22} - A_{-1}^{21}A_0^{12} + A_{-1}^{12}. \end{cases} \tag{9.47}$$

The coupled system (9.46), (9.47) is more complicated than its analogue [75, (9.4.34)] because of the aforementioned non-commutativity, still both systems share certain similarities. For instance

Proposition 9.19. *The integral operator*

$$I := A_0^{12} + A_{-1}^{21} + (A_0^{22})^2 - A_{-1}^{12}$$

on \mathcal{H}_1 is a conserved quantity for the system (9.46), (9.47). In fact, $I = 0$ is the zero integral operator on \mathcal{H}_1

Proof. We have

$$I_s = (A_0^{12})_s + (A_{-1}^{21})_s + (A_0^{22})_s A_0^{22} + A_0^{22} (A_0^{22})_s - (A_2^{12})_s = 0$$

by (9.46), (9.47) and with $(A_2^{12})_s = \mathbb{I}_1$. Next, for any $(x, y) \in \mathbb{R}^2$,

$$I(x, y) = -n_2(0|x)\ell_1(0|y) + \int_{-\infty}^{\infty} \left[\int_0^{\infty} n_2(w|x)k_1(w|t) \, dw \right] \\ \times \left[\int_0^{\infty} n_2(\lambda|t)k_1(\lambda|y) \, d\lambda \right] d\sigma(t),$$

so letting $s \rightarrow +\infty$ with the help of (9.10) and (9.16), using a Neumann series argument for $n_i(\cdot|x)$ we find $\lim_{s \rightarrow +\infty} I(x, y) = 0$ pointwise in $(x, y) \in \mathbb{R}^2$ by the dominated convergence theorem. But I was already shown to be s -independent, thus it must be the zero operator as claimed. \square

At this point we s -differentiate the equation for A_{-1}^{21} in (9.47) one more time and use (9.46), the equations for $(A_{-1}^{11})_s, (A_{-1}^{22})_s, (A_{-1}^{21})_s$ from (9.47) as well as the above integral of motion in the ensuing simplification. The result equals

$$(A_{-1}^{21})_{ss} = 2 \left(A_{-1}^{11} A_0^{22} - A_0^{22} A_{-1}^{22} - A_{-1}^{12} - 2(A_{-1}^{21})^2 + A_0^{22} A_{-1}^{21} A_0^{22} \right) \\ + A_2^{12} A_{-1}^{21} + A_{-1}^{21} A_2^{12}, \tag{9.48}$$

and we now introduce (recall our notational convention to suppress all s -dependence throughout)

$$u(s|t) := n_2(0|t) \stackrel{(9.16)}{=} \left((1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)})^{-1} m_2(\cdot|t) \right) (0), \quad s, t \in \mathbb{R} \tag{9.49}$$

which will ultimately solve the anticipated integro-differential Painlevé-II equation. In order to get there, evaluate all kernels in (9.48) on their diagonal, i.e. on the left-hand side of (9.48) we obtain

$$(A_{-1}^{21})_{ss}(t, t) = -(N_2(0) \otimes L_1(0))_{ss}(t, t) = -(N_2(0) \otimes N_2(0))_{ss}(t, t) \\ \stackrel{(9.49)}{=} - \left(u^2(s|t) \right)_{ss} = -2(u_s(s|t))^2 - 2u(s|t)u_{ss}(s|t) \tag{9.50}$$

through the use of remark 9.14. On the right-hand side of (9.48) instead,

$$\left(A_2^{12} A_{-1}^{21} + A_{-1}^{21} A_2^{12} - 4(A_{-1}^{21})^2 \right) (t, t) \\ = -2(s+t)u^2(s|t) - 4u^2(s|t) \int_{-\infty}^{\infty} u^2(s|\lambda) d\sigma(\lambda), \tag{9.51}$$

and

$$\left(A_{-1}^{11} A_0^{22} - A_0^{22} A_{-1}^{22} - A_{-1}^{12} + A_0^{22} A_{-1}^{21} A_0^{22} \right) (t, t) \\ = \left(2A_{-1}^{11} A_0^{22} - A_{-1}^{12} + A_0^{22} A_{-1}^{21} A_0^{22} \right) (t, t) \\ = - \left[n_1(0|t) + \int_{-\infty}^{\infty} n_2(0|u) A_0^{22}(u, t) d\sigma(u) \right]^2, \tag{9.52}$$

where we used the following fact in the first and second equality (as well as remark 9.14):

Proposition 9.20. *We have for any $(x, y) \in \mathbb{R}^2$ that $A_0^{22}(x, y) = A_0^{22}(y, x)$.*

Proof. Simply compute

$$\begin{aligned} A_0^{22}(x, y) &= - \int_0^\infty n_2(w|x)k_1(w|y) \, dw \\ &\stackrel{(9.16)}{=} - \int_0^\infty \left((1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)})^{-1} m_2(w) \right) (x)k_1(w|y) \, dw \\ &\stackrel{(9.10)}{=} - \int_0^\infty \left((1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)})^{-1} k_1(w) \right) (x)m_2(w|y) \, dw \\ &\stackrel{(9.28)}{=} - \int_0^\infty \ell_1(w|x)m_2(w|y) \, dw \\ &= - \int_0^\infty (M_2(w) \otimes L_1(w)) (y, x) \, dw = A_0^{22}(y, x), \end{aligned}$$

with lemma 9.13 in the last equality. □

However, returning to the equation for $(A_{-1}^{21})_s$ in (9.47), its kernel evaluated on the diagonal yields (we use again proposition 9.20 and remark 9.14),

$$\begin{aligned} -2u(s|t)u_s(s|t) &= (A_{-1}^{21})_s(t, t) \\ &= -2u(s|t) \left(n_1(0|t) + \int_{-\infty}^\infty n_2(0|u)A_0^{22}(u, t) \, d\sigma(u) \right), \end{aligned}$$

and since the zeros of the Airy function are discrete and the resolvent analytic in $s \in \mathbb{R}$, therefore

$$u_s(s|t) = n_1(0|t) + \int_{-\infty}^\infty n_2(0|u)A_0^{22}(u, t) \, d\sigma(u). \tag{9.53}$$

At this point we are left to substitute (9.50)–(9.53) into (9.48),

$$u_{ss}(s|t) = \left[s + t + 2u(s|t) \int_{-\infty}^\infty u^2(s|\lambda) \, d\sigma(\lambda) \right] u(s|t), \quad s, t \in \mathbb{R}.$$

This is exactly the integro-differential Painlevé-II equation of Amir *et al*, see [6, proposition 1.2], subject to the boundary data

$$\begin{aligned} u(s|t) &\stackrel{(9.49)}{=} \left((1 - K_{\sigma,s} \upharpoonright_{L^2(0,\infty)})^{-1} m_2(\cdot|t) \right) (0) \sim m_2(0|t) \\ &\stackrel{(9.10)}{=} \text{Ai}(s + t), \quad s \rightarrow +\infty, \end{aligned} \tag{9.54}$$

since $K_{\sigma,s} \rightarrow 0$ in operator norm as $s \rightarrow \infty$. We are now left to derive the relevant Tracy–Widom formula for $F_\sigma(s)$. To this end s -differentiate (9.45)

$$\begin{aligned} \frac{d^2}{ds^2} \ln F_\sigma(s) &= - \text{tr}_{\mathcal{H}_1} (A_0^{22})_s \stackrel{(9.46)}{=} \text{tr}_{\mathcal{H}_1} A_{-1}^{21} \\ &= - \int_{-\infty}^\infty (n_2(0|t))^2 \, d\sigma(t) \stackrel{(9.49)}{=} - \int_{-\infty}^\infty u^2(s|t) \, d\sigma(t), \end{aligned}$$

and now integrate twice while respecting the boundary condition (9.54),

$$\begin{aligned} \frac{d}{ds} \ln F_\sigma(s) &= \int_s^\infty \left[\int_{-\infty}^\infty u^2(x|t) d\sigma(t) \right] dx, \\ \ln F_\sigma(s) &= - \int_s^\infty \left(\int_y^\infty \left[\int_{-\infty}^\infty u^2(x|t) d\sigma(t) \right] dx \right) dy. \end{aligned}$$

Finally rearrange the outer two integrals and summarize,

Theorem 9.21 ([6, proposition 5.2]). *Let $d\sigma$ be a positive Borel probability measure on \mathbb{R} with finite moments to all orders and which is absolutely continuous with respect to the Lebesgue measure. Then the Fredholm determinant $F_\sigma(s)$ of the operator K_σ with kernel (9.5) on $L^2(s, \infty)$ equals*

$$F_\sigma(s) = \exp \left[- \int_s^\infty (x - s) \left(\int_{-\infty}^\infty u^2(x|t) d\sigma(t) \right) dx \right], \quad s \in \mathbb{R},$$

where $u = u(s|t) : \mathbb{R}^2 \rightarrow \mathbb{R}$ solves the integro-differential Painlevé-II equation

$$u_{ss}(s|t) = \left[s + t + 2u(s|t) \int_{-\infty}^\infty u^2(s|\lambda) d\sigma(\lambda) \right] u(s|t), \tag{9.55}$$

subject to the boundary condition $u(s|t) \sim \text{Ai}(s + t)$ as $s \rightarrow +\infty$ and $t \in \mathbb{R}$ fixed.

We emphasize again that theorem 9.21 was first proven in [6] through an application of the Tracy–Widom method [135]. Here we obtained the same result from the operator-valued Lax pair (9.43), (9.44) and the underlying Hilbert boundary value problem 9.8. Note that our method, en route, established the following novel features of $u(s|t)$ —which should be compared with the ones for the Ablowitz–Segur transcendent $u(x) = u(x|(-i, 0, i))$ in subsection 8.1.

- (a) the above boundary value problem for (9.55) is uniquely solvable, as $u(s|t) = n_2(0|t)$ is naturally constructed from (9.26), i.e. from the solution of the uniquely solvable Hilbert boundary value problem 9.8.
- (b) the map $\mathbb{R} \ni s \mapsto u(s|t)$ is smooth for any $t \in \mathbb{R}$ in light of (9.49) and the analyticity properties of the resolvent. Equivalently because of (9.26) and the analyticity properties of $X(z) = X(z; s)$, the solution of problem 9.8.

Remark 9.22. A final remark is in order concerning the possibility to derive asymptotic expansions of $F_\sigma(s)$ as $s \rightarrow \pm\infty$ directly from problem 9.8: while it is now fair to say that the operator-valued problem is well amenable to the rigorous derivation of dynamical systems, the development of a robust nonlinear steepest descent approach is work in progress by the author. As of now, the only rigorous operator-valued asymptotic techniques available are in [80, 82], for two particular integral kernels. In that sense, operator-valued Hilbert boundary value problems are not yet on the same level as their matrix-valued counterparts. However, if one accepts other techniques, then certain asymptotic information for $F_\sigma(s)$ has been derived recently, albeit for very specific measures $d\sigma$, see [25, 30, 137].

10. Summary and further reading

This article is meant to provide a short survey of the theory of Riemann–Hilbert problems and their original appearances in mathematics. Particular emphasis is placed on Plemelj’s

1908 work [121] on Hilbert's 21st problem and his use of Hilbert boundary value problems. Although Plemelj did not solve the problem for Fuchsian systems on $\mathbb{C}\mathbb{P}^1$, his use of singular integral equations in the derivation of dynamical systems foreshadowed one of the key aspects of modern Riemann–Hilbert theory. Indeed, a Hilbert boundary value problem in modern terms is a Swiss army knife: once we know how a problem's observables can be characterized through a Hilbert boundary value problem then it presents us with a systematic tool to derive dynamical systems for the observables and with an efficient way to analyse them asymptotically—overall in complete analogy to classical contour integral methods. Still, the identification of a Riemann–Hilbert approach to a problem is non-systematic and we have displayed six different applications in the second half of the text. Much more than that has been done over the past decades and we would now like to give some guidance to the relevant Riemann–Hilbert literature—naturally based on the author's personal preferences: the monograph [60] is the standard reference concerning the Riemann–Hilbert approach to Painlevé equations and as a first taste to this subject we recommend the review articles [75, 76] by Its. For a first introduction to OP Riemann–Hilbert methods in random matrix theory we single out Deift's notes [35], the notes by Bleher [16], by Its [77] and by Kuijlaars [101]. For continued reading on orthogonal polynomial methods in more general ensembles than (6.21) we recommend Kuijlaars' survey [102], the review article on universality [103] by the same author and the chapter on multi-matrix models by Bertola [13]. The standard reference to integrable operators is [34], the paper [40] and the somewhat non-rigorous monograph [96]. Regarding Ulam's problem and other combinatorial problems amenable to Riemann–Hilbert methods, the recent monograph [9] is likely going to become standard. Finally, and this is a topic which we did not touch upon, Hilbert boundary value problems have been identified as central tools in Gromov–Witten theories, i.e. in pure algebraic geometry. These theories aim to describe certain geometrical aspects of projective varieties in terms of invariants, suitably defined as intersection numbers on moduli spaces of curves. The invariants themselves are encoded in their generating functions whose convergence domains carry rich geometrical structures, nowadays known as Dubrovin–Frobenius manifolds. It was Dubrovin [49], see also [50], who first recognized the possibility to reconstruct such a manifold via a Hilbert boundary value problem. Further refinements of his procedure are given in the recent works of Cotti *et al* [31, 32], sadly some of the last papers coauthored by Boris Dubrovin.

ORCID iDs

Thomas Bothner  <https://orcid.org/0000-0001-8300-7467>

References

- [1] Ablowitz M J and Clarkson P A 1991 *Solitons, Nonlinear Evolution Equations and Inverse Scattering* (London Mathematical Society Lecture Note Series vol 149) (Cambridge: Cambridge University Press)
- [2] Ablowitz M J and Newell A C 1973 The decay of the continuous spectrum for solutions of the Korteweg–de Vries equation *J. Math. Phys.* **14** 1277–84
- [3] Ablowitz M J and Segur H 1976/77 Asymptotic solutions of the Korteweg–de Vries equation *Stud. Appl. Math.* **57** 13–44
- [4] Ablowitz M J and Segur H 1981 Asymptotic solutions of nonlinear evolution equations and a Painlevé transcendent *Phys. D* **3** 165–84

- [5] Anosov D V and Bolibruch A A 1994 *The Riemann–Hilbert Problem (Aspects of Mathematics, E22)* (Braunschweig: Vieweg)
- [6] Amir G, Corwin I and Quastel J 2011 Probability distribution of the free energy of the continuum directed random polymer in $1 + 1$ dimensions *Commun. Pure Appl. Math.* **64** 466–537
- [7] Arnol'd V I and Il'yashenko Y S 1988 *Ordinary Differential Equations (Encyclopaedia of Mathematical Sciences vol 1)* (Berlin: Springer) Translated from the Russian by E R Dawson and D O'Shea
- [8] Baik J, Deift P and Johansson K 1999 On the distribution of the length of the longest increasing subsequence of random permutations *J. Amer. Math. Soc.* **12** 1119–78
- [9] Baik J, Deift P and Suidan T 2016 *Combinatorics and Random Matrix Theory (Graduate Studies in Mathematics vol 172)* (Providence, RI: American Mathematical Society)
- [10] Balsler W 2000 *Formal Power Series and Linear Systems of Meromorphic Ordinary Differential Equations* (New York: Springer)
- [11] Beals R and Coifman R R 1984 Scattering and inverse scattering for first order systems *Commun. Pure Appl. Math.* **37** 39–90
- [12] Beals R, Deift P and Tomei C 1988 *Direct and Inverse Scattering on the Line (Mathematical Surveys and Monographs vol 28)* (Providence, RI: American Mathematical Society)
- [13] Bertola M 2011 Two-matrix models and biorthogonal polynomials *The Oxford Handbook of Random Matrix Theory* (Oxford: Oxford University Press) pp 310–28
- [14] Birkhoff G D 1913 The generalized Riemann problem for linear differential equations and the allied problems for linear difference and Q -difference equations *Proc. Am. Acad. Arts Sci.* **49** 521–68
- [15] Bleher P and Its A 1999 Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model *Ann. Math.* **150** 185–266
- [16] Bleher P M 2011 *Lectures On Random Matrix Models: The Riemann–Hilbert Approach (CRM Series in Mathematical Physics)* (New York: Springer)
- [17] Bolibruch A A 1989 The Riemann–Hilbert problem on the complex projective line *Mat. Zametki* **46** 118–20
- [18] Bolibruch A A 1990 The Riemann–Hilbert problem *Uspekhi Mat. Nauk* **45** 3–47
- [19] Bolibruch A A 1992 Sufficient conditions for the positive solvability of the Riemann–Hilbert problem *Mat. Zametki* **51** 156
- [20] Bolibruch A A 1993 *Hilbert's Twenty-First Problem for Fuchsian Linear Systems* (London: Chapman and Hall)
- [21] Bolibruch A A 1999 On sufficient conditions for the existence of a Fuchsian equation with prescribed monodromy *J. Dynam. Control Systems* **5** 453–72
- [22] Bolibruch A A, Its A R and Kapaev A A 2004 On the Riemann–Hilbert–Birkhoff inverse monodromy problem and the Painlevé equations *Algebra i Analiz* **16** 121–62
- [23] Bothner T and Its A 2012 The nonlinear steepest descent approach to the singular asymptotics of the second Painlevé transcendent *Phys. D* **241** 2204–25
- [24] Bothner T 2017 Transition asymptotics for the Painlevé II transcendent *Duke Math. J.* **166** 205–324
- [25] Cafasso M and Claeys T 2019 A Riemann–Hilbert approach to the lower tail of the KPZ equation (arXiv:1910.02493)
- [26] Clarkson P A and McLeod J B 1988 A connection formula for the second Painlevé transcendent *Arch. Rational Mech. Anal.* **103** 97–138
- [27] Coddington E A and Levinson N 1955 *Theory of Ordinary Differential Equations* (New York: McGraw-Hill)
- [28] Corwin I 2012 The Kardar–Parisi–Zhang equation and universality class *Random Matrices: Theory Appl.* **01** 1130001
- [29] Corwin I 2018 Commentary on 'longest increasing subsequences: from patience sorting to the Baik–Deift–Johansson theorem' by David Aldous and Persi Diaconis *Bull. Amer. Math. Soc.* **55** 363–74
- [30] Corwin I and Ghosal P 2018 Lower tail of the KPZ equation (arXiv:1804.05721)
- [31] Cotti G, Dubrovin B and Guzzetti D 2019 Isomonodromy deformations at an irregular singularity with coalescing eigenvalues *Duke Math. J.* **168** 967–1108
- [32] Cotti G, Dubrovin B and Guzzetti D 2020 Local moduli of semisimple Frobenius coalescent structures *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications* **16** 040
- [33] Davies B 2002 *Integral Transforms and Their Applications (Third, Texts in Applied Mathematics, vol 41)* (New York: Springer)

- [34] Deift P 1999 *Integrable Operators, Differential Operators and Spectral Theory* (Providence, RI: American Mathematical Society) pp 69–84
- [35] Deift P A 1999 *Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach (Courant Lecture Notes in Mathematics vol 3)* (New York, Providence, RI: New York University, Courant Institute of Mathematical Sciences, American Mathematical Society)
- [36] Deift P and Gioev D 2007 Universality at the edge of the spectrum for unitary, orthogonal, and symplectic ensembles of random matrices *Commun. Pure Appl. Math.* **60** 867–910
- [37] Deift P and Zhou X 1993 A steepest descent method for oscillatory Riemann–Hilbert problems. Asymptotics for the MKdV equation *Ann. Math.* **137** 295–368
- [38] Deift P A and Zhou X 1995 Asymptotics for the painlevé II equation *Commun. Pure Appl. Math.* **48** 277–337
- [39] Deift P, Its A R and Zhou X 1993 *Long-Time Asymptotics for Integrable Nonlinear Wave Equations, Important Developments in Soliton Theory* (Berlin: Springer) pp 181–204
- [40] Deift P A, Its A R and Zhou X 1997 A Riemann–Hilbert approach to asymptotic problems arising in the theory of random matrix models, and also in the theory of integrable statistical mechanics *Ann. Math.* **146** 149–235
- [41] Deift P, Its A, Kapaev A and Zhou X 1999 On the algebro-geometric integration of the Schlesinger equations *Commun. Math. Phys.* **203** 613–33
- [42] Deift P and Zhou X 2002 A priori L^p -estimates for solutions of Riemann–Hilbert problems *Internat. Math. Res. Notices* **2002** 2121–54 (to appear in print)
- [43] Deift P, Kriecherbauer T, McLaughlin K T-R, Venakides S and Zhou X 1999 Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory *Commun. Pure Appl. Math.* **52** 1335–425
- [44] Deift P, Venakides S and Zhou X 1997 New results in small dispersion KdV by an extension of the steepest descent method for Riemann–Hilbert problems *Int. Math. Res. Not.* **1997** 285–99
- [45] Dekkers W 1979 *The matrix of a connection having regular singularities on a vector bundle of rank 2 on $P^1(C)$ (Lecture Notes in Mathematics vol 712)* (Berlin: Springer)
- [46] Deligne P 1970 *Équations Différentielles à Points Singuliers Réguliers (Lecture Notes in Mathematics vol 163)* (Berlin: Springer)
- [47] Diaconis P and Forrester P J 2017 Hurwitz and the origins of random matrix theory in mathematics *Random Matrices: Theory Appl.* **06** 1730001
- [48] Doussal P L, Majumdar S N and Schehr G 2017 Periodic Airy process and equilibrium dynamics of edge fermions in a trap *Ann. Phys., NY* **383** 312–45
- [49] Dubrovin B 1998 Geometric and analytic theory of Frobenius manifolds *Documenta Mathematica* vol 2 pp 315–26
- [50] Dubrovin B 1999 *Painlevé Transcendents in Two-Dimensional Topological Field Theory (CRM Series in Mathematical Physics)* (Berlin: Springer)
- [51] Encyclopedia of Mathematics 2020 Hilbert problems https://www.encyclopediaofmath.org/index.php/Hilbert_problems#References accessed March 31
- [52] Enolski V and Grava T 2004 Singular Z_N -curves and the Riemann–Hilbert problem *Internat. Math. Res. Notices* **2004** 1619–83
- [53] Erdős L and Yau H-T 2017 *A Dynamical Approach to Random Matrix Theory (Courant Lecture Notes in Mathematics vol 28)* (New York, Providence, RI: Courant Institute of Mathematical Sciences, American Mathematical Society)
- [54] Eßler F H L, Frahm H, Izergin A G and Korepin V E 1995 Determinant representation for correlation functions of spin-1/2 XXX and XXZ Heisenberg magnets *Commun. Math. Phys.* **174** 191–214
- [55] Erugin N P 1982 *Problema Rimana (Lecture Notes in Mathematics vol 936)* (Minsk: “Nauka i Tekhnika”)
- [56] Faddeev L D and Takhtajan L A 2007 *Hamiltonian Methods in the Theory of Solitons, English (Classics in Mathematics)* (Berlin: Springer) Translated from the 1986 Russian original by Alexey G Reyman
- [57] Flaschka H and Newell A C 1980 Monodromy- and spectrum-preserving deformations I *Commun. Math. Phys.* **76** 65–116
- [58] Fokas A S, Its A R and Kitaev A V 1990 An isomonodromy approach to the theory of two-dimensional quantum gravity *Uspekhi Mat. Nauk* **45** 135–6
- [59] Fokas A S, Its A R and Kitaev A V 1991 Discrete Painlevé equations and their appearance in quantum gravity *Commun. Math. Phys.* **142** 313–44

- [60] Fokas A S, Its A R, Kapaev A A and Novokshenov V Y 2006 *Painlevé Transcendents (Mathematical Surveys and Monographs vol 128)* (Providence, RI: American Mathematical Society) The Riemann–Hilbert approach
- [61] Fredholm I 1903 Sur une classe d'équations fonctionnelles *Acta Math.* **27** 365–90
- [62] Gakhov F D 1949 On Riemann's boundary problem for a system of n pairs of functions *Doklady Akad. Nauk SSSR (N.S.)* **67** 601–4
- [63] Gardner C S, Greene J M, Kruskal M D and Miura R M 1967 Method for solving the Korteweg-deVries equation *Phys. Rev. Lett.* **19** 1095–7
- [64] Giraud G 1939 Sur une classe d'équations linéaires ou figurent des valeurs principales d'intégrales simples *Ann. Sci. École Norm. Sup.* **56** 119–72
- [65] Gohberg I, Goldberg S and Nahum K 2000 *Traces and Determinants of Linear Operators (Operator Theory: Advances and Applications vol 116)* (Basel: Birkhäuser)
- [66] Gontsov R R and Poberezhnyi V A 2008 Various versions of the Riemann–Hilbert problem for linear differential equations *Russ. Math. Surv.* **63** 603–39
- [67] Gromak V I, Laine I and Shimomura S 2002 *Painlevé Differential Equations in the Complex Plane* vol 28 (Berlin: de Gruyter & Co) viii+303
- [68] Hastings S P and McLeod J B 1980 A boundary value problem associated with the second painlevé transcendent and the Korteweg-de Vries equation *Arch. Rational Mech. Anal.* **73** 31–51
- [69] Hilbert D 1900 Mathematische Probleme *Göttinger Nachrichten* **3** 253–97
- [70] Hilbert D 1902 Mathematical problems *Bull. Amer. Math. Soc.* **8** 437–79
- [71] Hilbert D 1904 *Über eine Anwendung der Integralgleichungen auf ein Problem der Funktionentheorie, Verhandl. des III. International Mathematiker Kongresses* (Leipzig: Teubner)
- [72] Hilbert D 1912 *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen* (Leipzig: Teubner)
- [73] Hurwitz A 1897 Über die Erzeugung der Invarianten durch Integration *Nachr. Ges. Wiss. Göttingen* **1897** 71–90
- [74] Its A R 1981 Asymptotic behaviour of the solutions to the nonlinear Schrödinger equation, and isomonodromic deformations of systems of linear differential equations *Dokl. Akad. Nauk SSSR* **261** 14–8
- [75] Its A R 2011 Painlevé transcendents *The Oxford Handbook of Random Matrix Theory* (Oxford: Oxford University Press) pp 176–97
- [76] Its A R 2003 The Riemann–Hilbert problem and integrable systems *Noties of the American Mathematical Society* **50** 1389–400
- [77] Its A R 2011 *Large N Asymptotics in Random Matrices: The Riemann–Hilbert Approach (CRM Series in Mathematical Physics)* (New York: Springer)
- [78] Its A R, Izergin A G, Korepin V E and Slavnov N A 1990 Differential equations for quantum correlation functions *Proceedings of the Conference on Yang–Baxter Equations, Conformal Invariance and Integrability in Statistical Mechanics and Field Theory* pp 1003–37 (unpublished)
- [79] Its A R and Kapaev A A 1987 The method of isomonodromic deformations and relation formulas for the second Painlevé transcendent *Izv. Akad. Nauk SSSR Ser. Mat.* **51** 878–92
- [80] Its A R and Kozłowski K K 2016 Large-analysis of an operator-valued Riemann–Hilbert problem *Int Math Res Notices* **2016** 1776–806 (to appear in print)
- [81] Its A R and Novokshenov V Y 1986 *The Isomonodromic Deformation Method in the Theory of Painlevé Equations (Lecture Notes in Mathematics vol 1191)* (Berlin: Springer)
- [82] Its A R and Slavnov N A 1999 On the Riemann problem method for the asymptotic analysis of the correlation functions of the quantum nonlinear Schrödinger equation. The case of interacting fermions *Teoret. Mat. Fiz.* **119** 179–248
- [83] Ilyashenko Y and Yakovenko S 2008 *Lectures on Analytic Differential Equations (Graduate Studies in Mathematics vol 86)* (Providence, RI: American Mathematical Society)
- [84] Jimbo M and Miwa T 1980 Studies on holonomic quantum fields, XVII *Proc. Japan Acad. A* **56** 405–10
- [85] Jimbo M, Miwa T and Ueno K 1981 Monodromy preserving deformation of linear ordinary differential equations with rational coefficients *Phys. D* **2** 306–52
- [86] Jimbo M and Miwa T 1981 Monodromy perserving deformation of linear ordinary differential equations with rational coefficients. II *Phys. D* **2** 407–48
- [87] Jimbo M and Miwa T 1981/82 Monodromy preserving deformation of linear ordinary differential equations with rational coefficients. III *Phys. D* **4** 26–46

- [88] Jimbo M 1982 Monodromy problem and the boundary condition for some Painlevé equations *Publ. Res. Inst. Math. Sci.* **18** 1137–61
- [89] Johansson K 2007 From Gumbel to Tracy–Widom *Probab. Theory Relat. Fields* **138** 75–112
- [90] Kapaev A 1992 Global asymptotics of the second Painlevé transcendent *Phys. Lett. A* **167** 356–62
- [91] Kashiwara M 1980 *Faisceaux constructibles et systèmes holonomes d'équations aux dérivées partielles linéaires à points singuliers réguliers* (Palaiseau: École Polytech.)
- [92] Kashiwara M 1984 The Riemann–Hilbert problem for holonomic systems *Publ. Res. Inst. Math. Sci.* **20** 319–65
- [93] Kitaev A V 1989 *The Justification Of Asymptotic Formulas that can be Obtained by the Method of Isomonodromic Deformations (Zap. Nauchn. Sem. 179)* (Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)) no. Mat. Vopr. Teor. Rasprostr. Voln. 19, 101–109, 189–190 (unpublished)
- [94] Kohn A T 1983 Un résultat de Plemelj *Mathematics and Physics* (Paris) pp 307–12 1979/1982 (unpublished)
- [95] Kojima T, Korepin V E and Slavnov N A 1997 Completely integrable equation for the quantum correlation function of nonlinear Schrödinger equation *Commun. Math. Phys.* **189** 709–28
- [96] Korepin V E, Bogoliubov N M and Izergin A G 1993 *Quantum Inverse Scattering Method and Correlation Functions (Cambridge Monographs on Mathematical Physics)* (Cambridge: Cambridge University Press)
- [97] Korotkin D 2004 Solution of matrix Riemann–Hilbert problems with quasi-permutation monodromy matrices *Mathematische Annalen* **329** 335–64
- [98] Kostov V P 1992 Fuchsian linear systems on $\mathbb{C}P^1$ and the Riemann–Hilbert problem *C. R. Acad. Sci. Paris Sér. I Math.* **315** 143–8
- [99] Kreĭn M G 1954 On integral equations generating differential equations of 2nd order *Doklady Akad. Nauk SSSR (N.S.)* **97** 21–4
- [100] Krylov B L 1956 The solution in a finite form of the Riemann problem for a Gauss system *Trudy Kazan. Aviat. Inst.* **31** 203–445
- [101] Kuijlaars A B J 2003 *Riemann–Hilbert Analysis for Orthogonal Polynomials (Lecture Notes in Mathematics vol 1817)* (Berlin: Springer)
- [102] Kuijlaars A B J 2010 Multiple orthogonal polynomials in random matrix theory *Proceedings of the International Congress of Mathematicians vol III* (New Delhi: Hindustan Book Agency) pp 1417–32
- [103] Kuijlaars A B J 2011 *Universality, the Oxford Handbook of Random Matrix Theory* (Oxford: Oxford University Press) 103–34
- [104] Lappo-Danilevskii I A 1957 *Application of Matrix Functions to the Theory of Linear Systems of Ordinary Differential Equations* ed V I Smirnov (Moscow: Gosudarstv. Izdat. Tehn.-Teor. Lit.) with an introductory essay
- [105] Lappo-Danilevskii I A 1953 *Mémoires sur la théorie des systèmes des équations différentielles linéaires* (New York: Chelsea Publishing Co.)
- [106] Lax P D 1968 Integrals of nonlinear equations of evolution and solitary waves *Commun. Pure Appl. Math.* **21** 467–90
- [107] Lieb E, Schultz T and Mattis D 1961 Two soluble models of an antiferromagnetic chain *Ann. Phys., NY* **16** 407–66
- [108] Liechty K and Wang D 2020 Asymptotics of free fermions in a quadratic well at finite temperature and the Moshe–Neuberger–Shapiro random matrix model *Ann. Inst. H. Poincaré Probab. Statist.* **56** 1072–98
- [109] Manakov S V 1973 Nonlinear Fraunhofer diffraction *Zh. Eksp. Teor. Fiz.* **65** 1392–8 (Russian)
- [110] Manakov S V 1974 *Soviet Physics JETP* **38** 693–6 Engl. transl.
- [111] Mebkhout Z 1980 Sur le problème de Hilbert–Riemann *C. R. Acad. Sci. Paris Sér. A-B* **290** A415–7
- [112] Mebkhout Z 1984 Une autre équivalence de catégories *Compositio Math* **51** 63–88
- [113] Muskhelishvili N I 1992 *Singular integral equations Boundary Problems of Function Theory and Their Application to Mathematical Physics* (New York: Dover Publications, Inc.) Translated from the second (1946) Russian edition and with a preface by J R M Radok, Corrected reprint of the 1953 English translation
- [114] Muskhelishvili N I and Vekua N P 1943 Riemann's boundary value problem for several unknown functions and its application to systems of singular integral equations *Trav. Inst. Math. Tbilissi [Trudy Tbiliss. Mat. Inst.]* **12** 1–46

- [114] Noble B 1958 *Methods Based on the Wiener–Hopf Technique for the Solution of Partial Differential Equations (International Series of Monographs on Pure and Applied Mathematics vol 7)* (Oxford: Pergamon)
- [115] Noether F 1920 Über eine Klasse singulärer Integralgleichungen *Math. Ann.* **82** 42–63
- [116] Novikov S, Manakov S V, Pitaevskii L P and Zakharov V E 1984 *Theory of Solitons, Contemporary Soviet Mathematics* (New York: Consultants Bureau) The inverse scattering method, Translated from the Russian
- [117] Olver F W J, Lozier D W, Boisvert R F and Clark C W (ed) 2010 *NIST Handbook of Mathematical Functions* (Washington, DC, Cambridge: U.S. Department of Commerce, National Institute of Standards and Technology, Cambridge University Press) with 1 CD-ROM (Windows, Macintosh and UNIX)
- [118] Pastur L and Shcherbina M 1997 Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles *J. Stat. Phys.* **86** 109–47
- [119] Plemelj J 1904 Zur theorie der Fredholmschen Funktionalgleichung *Monatsh. f. Mathematik und Physik* **15** 93–128
- [120] Plemelj J 1908 Ein Ergänzungssatz zur Cauchyschen Integraldarstellung analytischer Funktionen, Randwerte betreffend *Monatsh. f. Mathematik und Physik* **19** 205–10
- [121] Plemelj J 1908 Riemannsche Funktionenscharen mit gegebener Monodromiegruppe *Monatsh. f. Mathematik und Physik* **19** 211–45
- [122] Plemelj J 1964 Problems in the sense of Riemann and Klein *Interscience Tracts in Pure and Applied Mathematics, No. 16* ed J R M Radok (New York: Interscience Publishers Wiley Inc.)
- [123] Poincaré H 1884 Sur les groupes des équations linéaires *Acta Math.* **4** 201–312
- [124] Wolchover N 2020 At the far ends of a new universal law <https://www.quantamagazine.org/beyond-the-bell-curve-a-new-universal-law-20141015/> accessed July 22
- [125] Riemann B 1876 *Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen komplexen Grösse* (Leipzig: Werke) pp 3–43
- [126] Röhl H 1957 Das Riemann–Hilbertsche problem der theorie der linearen differentialgleichungen *Math. Ann.* **133** 1–25
- [127] Romik D 2015 *The Surprising Mathematics of Longest Increasing Subsequences (Institute of Mathematical Statistics Textbooks vol. 4)* (Cambridge: Cambridge University Press)
- [128] Sabat A B 1973 The Korteweg–de Vries equation *Dokl. Akad. Nauk SSSR* **211** 1310–3
- [129] Simon B 2005 *Trace Ideals and Their Applications (Second, Mathematical Surveys and Monographs vol 120)* (Providence, RI: American Mathematical Society)
- [130] Simon B 2015 *Basic Complex Analysis, A Comprehensive Course in Analysis, Part 2A* (Providence, RI: American Mathematical Society)
- [131] Soshnikov A 2000 Determinantal random point fields *Uspekhi Mat. Nauk* **55** 107–60
- [132] Stein E M 1970 *Singular Integrals and Differentiability Properties of Functions (Princeton Mathematical Series, No. 30)* (Princeton, NJ: Princeton University Press)
- [133] Suleimanov B I 1987 The connection of asymptotics on various infinities of solutions of the second Painleve equation *Dif. Urav.* **23** 834–42 (Russian)
- [134] Ulam S M 1961 *Monte Carlo Calculations in Problems of Mathematical Physics (Modern Mathematics for the Engineer Second Series)* (New York: McGraw-Hill) pp 261–81
- [135] Tracy C A and Widom H 1994 Level-spacing distributions and the Airy kernel *Commun. Math. Phys.* **159** 151–74
- [136] Trogdon T and Olver S 2016 *Riemann–Hilbert Problems, Their Numerical Solution, and the Computation of Nonlinear Special Functions* (Philadelphia, PA: SIAM)
- [137] Tsai L C 2018 Exact lower tail large deviations of the KPZ equation (arXiv:1809.03410)
- [138] Vekua N P 1967 *Systems of Singular Integral Equations* ed J H Ferziger (Groningen: P. Noordhoff, Ltd.) Translated from the Russian by A G Gibbs and G M Simmons
- [139] Walsh J L 1933 The Cauchy–Goursat theorem for rectifiable Jordan curves *Proc. Natl Acad. Sci.* **19** 540–1
- [140] Wiener N and Hopf E 1931 *Über eine Klasse singulärer Integralgleichungen* (Berlin: Sitzungber. Akad. Wiss.) pp 696–706
- [141] Wigner E P 1957 Gatlinburg conference on neutron physics Oak Ridge National Laboratory Report ORNL vol 2309 p 59
- [142] Wishart J 1928 The generalised product moment distribution in samples from a normal multivariate population *Biometrika* **20A** 32–43

- [143] Zakharov V E and Faddeev L D 1971 The Korteweg–de Vries equation is a fully integrable Hamiltonian system *Funkcional. Anal. i Priložen.* **5** 18–27
- [144] Zakharov V E and Shabat A B 1971 Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media *Ž. Ėksper. Teoret. Fiz.* **61** 118–34 (Russian, with English summary)
Zakharov V E and Shabat A B 1972 *Soviet Physics JETP* **34** 62–9 MR 0406174 (Engl. transl.)
- [145] Zhou X 1989 The Riemann–Hilbert problem and inverse scattering *SIAM J. Math. Anal.* **20** 966–86