Contemporary Mathematicians

F. Alberto Grünbaum
Pierre van Moerbeke
Victor H. Moll
Editors

Henry P. McKean Jr.

Selecta

Birkhäuser

Birkhäuser

Contemporary Mathematicians

Joseph P.S. Kung

Editor

More information about this series at http://www.springer.com/series/4817

F. Alberto Grünbaum •
Pierre van Moerbeke •
Victor H. Moll
Editors

Henry P. McKean Jr. Selecta

Birkhäuser

Editors
F. Alberto Grünbaum
Department of Mathematics
University of California
Berkeley
California, USA

Pierre van Moerbeke
Institut de Recherche en Mathématique at Physique
Université catholique de Louvain
Louvain-la-Neuve, Belgium

Department of Mathematics
Brandeis University
Waltham
Massachusetts, USA

Victor H. Moll
Tulane University
New Orleans
Louisiana, USA

Contemporary Mathematicians
ISBN 978-3-319-22236-3 ISBN 978-3-319-22237-0 (eBook)
DOI 10.1007/978-3-319-22237-0

Library of Congress Control Number: 2015960201

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.birkhauser-science.com)

To my teachers Will Feller, Kiyoshi Ito, Mark Kac, Norman Levinson and Gretchen Warren

Contents

1 Personal Recollections

Henry P. McKean Jr.[1]

1.1. Norman Levinson

Thinking of Norman Levinson, I remember how much I learned from him as a very young and inexperienced person, and how much I found to admire, both in his mathematical work and in himself, as a man.

Looking back, his choice of mathematical questions seems memorable enough: mostly close to applications, rich in their details, suggestive of general phenomena, as in the wonderful papers on the forced vander Pol equation, etc. foreshadowing the current vogue of attractors, chaos, and all that.

What I could better appreciate then was his mastery of the kind of hard analysis such questions require, the kind in which every equality costs you two opposing inequalities. When we got stuck, working together, he'd always take an example, and he'd estimate things with an understanding and a speed that impressed me equally, and soon we'd be back on solid ground. It was excitingly easy, Norman doing all the hard work, as I understood later. *Gap and Density Theorems* (Chap. 8, vol. 2) and the extraordinary papers on Riemann's zeta function (Chap. 11, vol. 2) is where you can see this expertise at its best.

Other lessons I was not so ready to digest though happy to benefit from. I mean his unobtrusive, remarkably effective administrative style, the way he seemed to run the department with the back of his hand, and, on the personal side, his patience and gentle encouragement.

He was shy. Fagi said: "Norman, he's terrible. He never wants to go out. He's afraid he'll meet somebody he doesn't know." I was shy, too, but slowly we got to know each other a little, coming from as different backgrounds as you could imagine, and I thought myself lucky when he told me bits about his early life. He said: "We were very poor, but we didn't think of ourselves as poor." I take the liberty to transpose that and to say he was a rich man in his particular way, spreading about his riches quietly, with an open hand.

Henry McKean, New York, May 1997.

1.2. Will Feller

Will Feller was born in Zagreb, Yugoslavia on July 7, 1906, the ninth of 12 children of a well to-do family. They named him Willy in the then popular German style. This changed to William upon his coming here (1939), but everybody called him plain Will. His early studies called him from Zagreb (1923–1925) to Gottingen (1925–1928) where he got his degree in 1926, aged 20! Then to Kiel as Privatdozent (1928–1933). Nazi times: Will taught a class on the new ideas in probability of Kolmogorov, etc., attended, by chance, by a person of some importance in the SS. One day, this person and some

[1]Courant Institute, New York University, New York, NY, USA, mckean@cims.nyu.edu.

F.A. Grünbaum et al. (eds.), *Henry P. McKean Jr. Selecta*, Contemporary Mathematicians, DOI 10.1007/978-3-319-22237-0_1

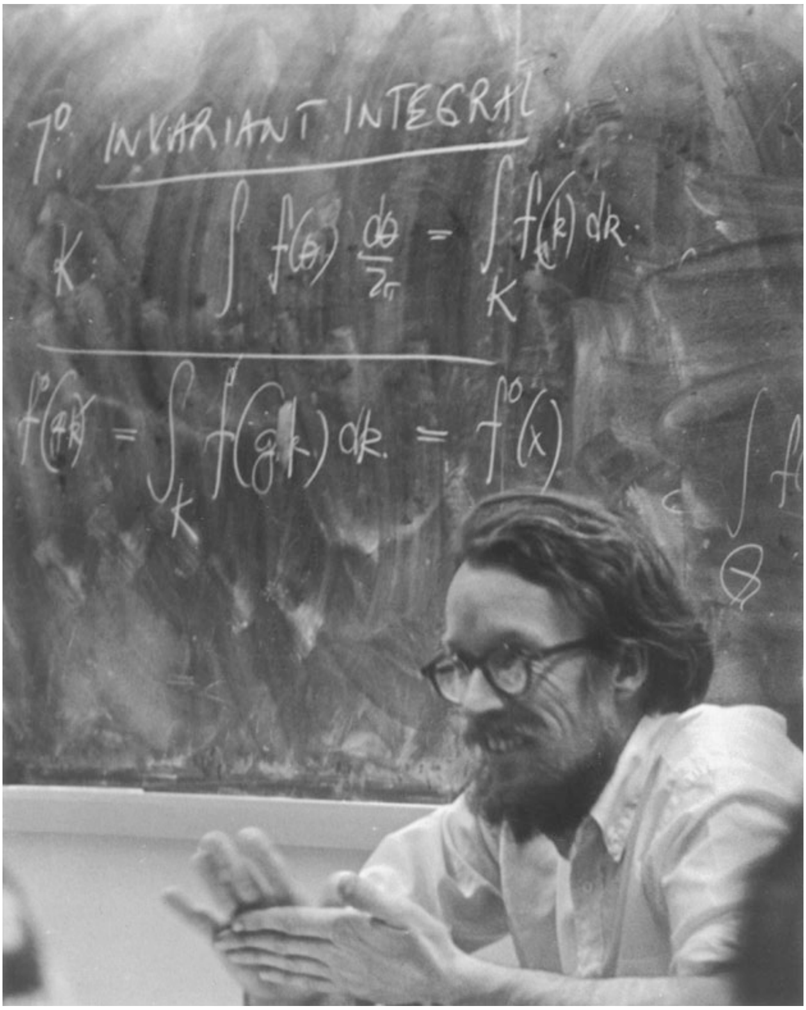

FIGURE 1.1. Henry P. McKean Jr.

two or three of his men present themselves at Will's apartment. Will lets them in in fear and trembling, whereupon the boss says how much he loves Will's lectures and if there is somebody Will would like them to beat up, just to say the word. A courtesy call I suppose. Will declined this civility and after a subsequent refusal to sign a Nazi oath, packed his bags for Copenhagen where he stayed a year (1933). Then to Stockholm (1934–1939), Providence (1939–1944), Ithaca (1945–1950), and Princeton (1950–) which is where I first knew him (1953). He died, with difficulty, in New York on January 14, 1970, aged 63.

I think it is fair to say that Kolmogorov, Paul Lévy, and Will made probability an honest woman. They are the people chiefly responsible for its rise from a not quite respectable rule of thumb to the ubiquitous, precise, intuitively appealing subject it is today. I think that, of the three, Will had the wider view. He understood Kolmogorov's mostly analytical way and also Paul Lévy's way with sample paths, and was a master of both, as can best be seen in his splendid book, *An Introduction to Probability Theory* (John Wiley & Sons, 1950) and its subsequent amplifications and revisions (1957/1968 and 1966/1971). Here you can see him endlessly

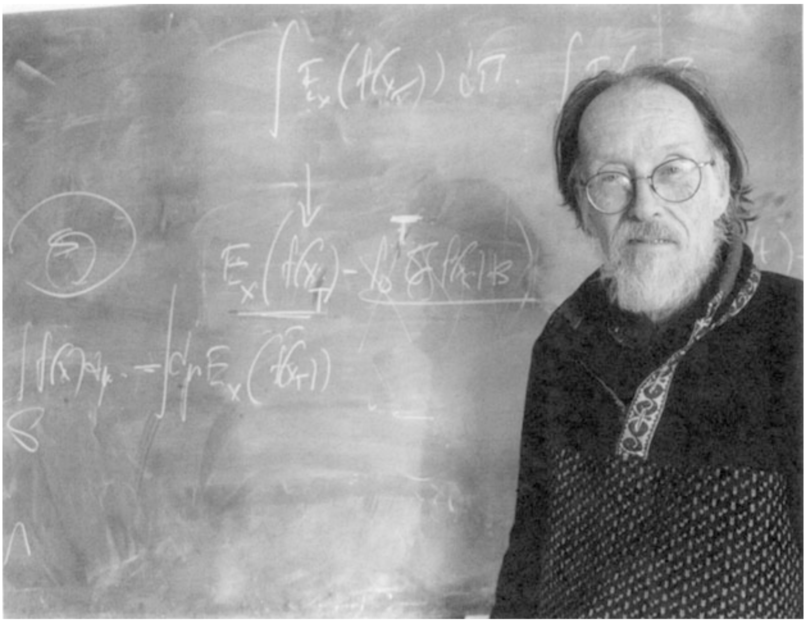

FIGURE 1.2. Henry P. McKean Jr.

perfecting proofs and pictures. I point, for example, to his simple way with supposedly hard Tauberian theorems and to his so appealing presentation of Sparre-Andersen's combinatorial way with random walks. This is *the* book for beginners. It is so full of interesting things, both mathematical and practical, looked at from an astonishing number of aspects, and full of Will's own self: his enthusiasm, his high standards, his indefatigable desire to make you understand "what's really going on".

That was also his watch-word when he lectured. He would get quite excited, his audience in his hand, and come (almost) to the point. Then the hour would be over, and he would promise to tell us "what's really going on" next time. Only next time the subject would be not quite the same, and so a whole train of things was left hanging, somewhat in the manner of Tristram Shandy. But it didn't matter. We loved it and couldn't wait for the next (aborted) revelation.

I was too young to appreciate my luck in coming so accidentally into his orbit, but soon realized that I had not only a teacher but a friend whose generosity and sense of the ridiculous in the gloomy moments of the young could be counted on for sure. I remember a note he sent me at such a moment in which he said: "You were made for success and the Lord God himself is kicking Himself that you do not understand his good intentions". (Will's capitals). That's how he was with me and many others, too: good company, a raiser of spirits, smart, kind.

I learned so much from him as a man, and also in mathematics, from him and others of my elders and betters: D. Ray, G. Hunt, and K. Itô.

A bit about Will's one-dimensional "diffusions" which he was in the middle of when we were together (1953/1957). It exemplifies his simultaneous desire for generality and simplicity, taking what people thought to be quite complicated and making it obvious, as all good mathematics should be. In short, he reduced the general diffusion to the simplest one (Brownian motion) by: (1) a change of "scale" to make the motion appear unbiased, and (2) a change of "clock" or "speed" to make the (local, Gaussian) fluctuations the same at any place. This is easy enough in simple cases with stringent technical conditions imposed. But what Will realized is that, effectively, no technical conditions are needed at all; in the "natural" language, the technicalities evaporate and simple, perfectly general expressions for objects of interest are found. Here is an example. Let $\mathfrak{G} \equiv \frac{1}{2}e^2(x)\partial^2/\partial x^2 + f(x)\partial/\partial x$ be the infinitesimal operator of a one-dimensional diffusion $x(t) : t \geq 0$. Here, f specifies the "drift" or "bias" and e specifies the "fluctuations". Introduce the new "scale"

$$\int_0^x dy \exp\left[-\int_0^y (2f/e^2)\right]$$

and the "speed measure"

$$dm(x) \equiv 2\frac{dx}{e^2(x)} \exp\left[+\int_0^x (2f/e^2)\right],$$

in terms of which $\mathfrak{G} = (d/dm)(d/dx)$. Take $a < x < b$, start the diffusion at $x(0) = x$ and let T be the "exit time" $\min(t : x(t) = a \text{ or } b)$. Then in the new scale,

$$P\big[x(T) = a\big] = \frac{b-x}{b-a},\ P\big[x(T) = b\big] = \frac{x-a}{b-a},$$

$$\text{and } E(T) = \int_a^b G(x,y)dm(y)$$

with the (symmetric) Green's function $G(x,y) = (x-a)(b-y)/(b-a)$ for $x \leq y$. What could be simpler?

Back to Will himself. He was short; compact, with a mop of wooly gray hair; irrepressible. In conversation quick, always ready with an opinion (or two), addicted to exaggeration. If you knew the code, you applied the "Feller factor" (discount by 90 %). If you didn't, it could be awkward, as with the immigration official at Providence when they came to the question: "Do you advocate bigamy?" Will delivered a lengthy opinion on the distinction between practice and advocacy which, he said, must surely obtain in this great free land about to be his own. The official was not amused. So he could seem opinionated, even rude if you didn't know him. But the real Will took a wide view, meeting life with enthusiasm and good cheer. I think of him often, hearing his voice, remembering him so full of fun.

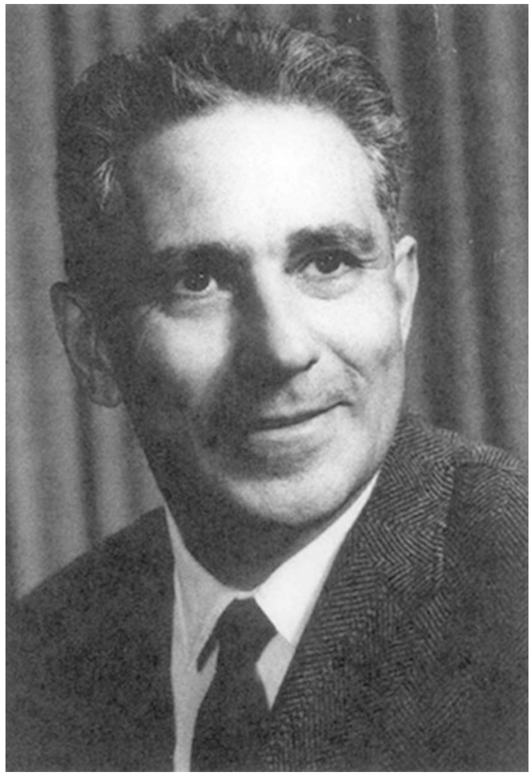

FIGURE 1.3. N. Levinson

Brooklyn,
April 15, 2005

1.3. Kiyoshi Itô: recollections of Kyôto 1957/1958

The following recollections formed a little talk I spoke in Kyôto on the occasion of K. Itô's 88th birthday. They bring the memory of the happy times we had together and congratulations on his Gauss Prize. The Gauss Committee will not find it easy to keep to the standard they have set.

Kyôto 1957/1958 It's a pleasure to think back for a little while to the happy days my family and I spent in Kyôto in 1957/1958.

I met Itô-*sensei* in Princeton 1954 where Kosaku Yosida also came for a shorter time. Dan Ray was there also Hunt & Trotter, and of course Feller who was the activator, the regisseur of it all. Feller had just formulated his ideas on diffusion and I was helping (feebly to be sure) to bring them together into a little

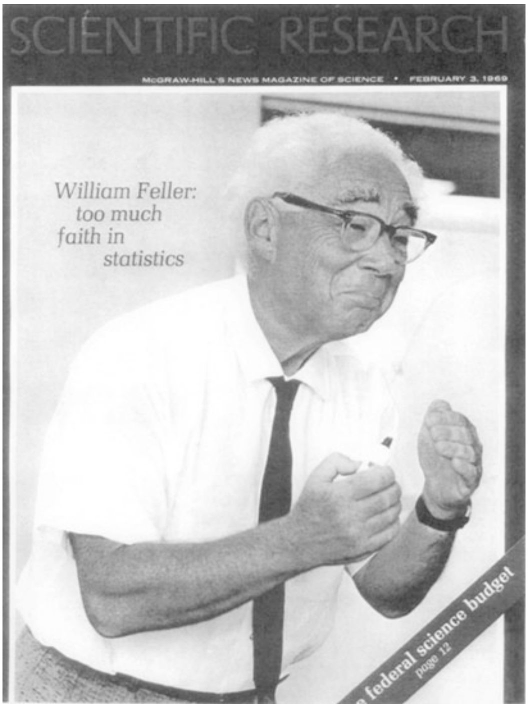

FIGURE 1.4. Will Feller

FIGURE 1.5. K. Ito
Picture by Konrad Jacobs. With friendly permission from the Archives of the Mathematisches Forschungsinstitut Oberwolfach

book. At the same time we were just hearing about Dynkin's ideas on stopping times so we had a seminar to digest these things. It was lucky for me, being so young and largely uneducated, to find myself at a moment when the understanding of diffusion was taking a big jump. I think I never worked so hard (without fatigue) or learned so much so fast (without tears) as I did then, and from two such patient kindly teachers as Feller and Itô.

Then Itô and I began to combine all that with ideas of Paul Lévy, especially his "measure du voisinage" or "local time" as we called it in English, and we understood quite quickly how this local time is a sort of 1/2-dimensional measure on the zeros of the Brownian path, how it could be used to implement the elastic Brownian motion, and so on. I say "we" understood. I should say Itô understood, and since he was patient and I was pretty quick though ignorant, pretty soon I understood, too. But invariably, at that happy moment when you say to yourself "I see", it was Itô who saw the whole. It was my first real taste of how mathematics is done and I cannot think of it now without excitement and gratitude. So far Princeton.

I stayed on another year or two and then to Japan (1957/1958) where we continued to work on our book. Never ever write a book before everything is proved and never ever let the junior partner write it! It would have been a better book had it been written half as many times at twice the length. But never mind: Regrets are not very interesting, what's more interesting is that year 1957/1958 in Kyôto.

It was a long trip by a little Japanese freighter from Los Angeles, up past the Aleutians where it was cold and rough with a sick child of 2 years, but after a couple of weeks we arrived at Yokohama to be met by Itô and

Itô-*san no okusan* and Junko, and after that all was well, though naturally new and more than a little foreign.

I might tell you that I already knew a little about Japan as my great-grandfather in Boston was an enthusiastic collector of Japanese art. That was in the days of Fenellosa who first made such things known in America and who was a friend of that grandfather, so the chance to see some of the treasures of that marvelous art with my own eyes and to do mathematics with Itô was a combination not to be resisted. Now back to Kyôto.

Itô had rented us a spacious house right on Takanogawa in the Shimogamo district, just a short walk from the university. I still see in my mind's eye the lovely printed cloths washing in the current of the river and the moon coming up from behind the Hieisan, and I hear in my mind's ear the call of the noodle-seller at night. Now the cloths and the noodle-seller are gone but the river is sparkling clean, which it was not then. I went to the university every day walking, and back at night. I had a vast office with a couch and many chairs and tables and a huge glass-topped desk, and a coal stove for the winter which smoked horribly when the wind was wrong and the kind old librarian (whose name I regret to have forgotten) would try to make it work. I had a sink, too, to wash my hands, and a big safe—was it to keep my theorems in? I never figured out. And after the morning, we had lunch all seated about a long table, or else if the weather was fine, we went to a little out door restaurant a few steps away and had some kind of *domburi* and a beer.

Once a week, we had a seminar with so many quick and eager people, then very young like me, now not so young like me: Nobuyuki Ikeda was there and Hiroshi Tanaka, and Shinzo Watanabe and Takesi Watanabe, Minoru Motoo, Tunekiti Sirao, Tadashi Ueno, Takeyuki Hida, Makiko Nisio. I must have forgotten some names. If so please forgive me. Itô said that we were "sowing the seeds of diffusion in the mathematical fields of Japan". It was all very welcoming and exciting and I loved it.

To come back for a moment to our joint work and the way Itô taught me. I remember we were flying to some place (Fukuoka I think) and trying to understand Feller's most general boundary conditions for Brownian motion on a half-line. Itô sat beside me drawing pictures of sample paths. These did not please him until he got to the right one (it didn't take him very long) and as soon as he showed it to me I understood perfectly, but lacking his experience and deep feeling could not have thought of it myself. Some things we missed entirely like the deep facts about the spatial dependence of stopped local time. We had prototype formulas in front of us, all perfectly explicit, and never imagined what they meant: That the stopped local time was itself a diffusion in its spatial parameter. Oh well.

Leaving mathematics, I remember a continual attentive kindness from Itô and Itô-*san no okusan*. I remember happy suppers at their house when Itô, knowing next to nothing about cooking, would explain what his wife had placed before us by saying: "You take it and you put it and then it's very good." I remember also excursions near and far: To Shugakuin, Hieisan, Kokedera, Nara, and so on, with Itô always on the lookout that we should be comfortable and at ease. Once at the zoo, when the little Japanese children were staring at my odd-looking American children, Itô said to them sharply: "You are here to look at the animals!" and they did that.

So now you can see what a load of obligation (of *on* or *giri* if you like) I must carry, but that is not Itô's way. Itô's father was a very traditional, correct man who kept a record of every kindness done to himself and to his family over the years. When he died, Itô discharged each recorded debt meticulously, as his father would have wished him to do. After that, he went his own way, marrying his dear Shizue by inclination and for love, giving always freely: to myself, to many of you in this room, and I must suppose to many many others unknown to me.

Arigato gozaimashita, arigato gozaimasu.

FIGURE 1.6. M. Kac
Picture by Konrad Jacobs. With friendly permission from the Archives of the Mathematisches Forschungsinstitut Oberwolfach

1.4. Mark Kac, August 16, 1914–October 25, 1984

Poland. Mark Kac was born "to the sound of the guns of August on the 16th day of that month, 1914," in the town of Krzemieniec—then in Russia, later in Poland, now in the Soviet Ukraine (1985, 1, p. 6). In this connection Kac liked to quote Hugo Steinhaus, who, when asked if he had crossed the border replied, "No, but the border crossed me."

In the early days of the century Krzemieniec was a predominantly Jewish town surrounded by a Polish society generally hostile to Jews. Kac's mother's family had been merchants in the town for three centuries or more. His father was a highly educated person of Galician background, a teacher by profession, holding degrees in philosophy from Leipzig, and in history and philosophy from Moscow.

As a boy Kac was educated at home and at the Lycée of Krzemieniec, a well-known Polish school of the day. At home he studied geometry with his father and discovered a new derivation of Cardano's formula for the solution of the cubic—a first bite of the mathematical bug that cost Kac *père* five Polish zlotys in prize money. At school, he obtained a splendid general education in science, literature, and history. He was grateful to his early teachers to the end of his life.

In 1931 when he was 17, he entered the John Casimir University of Lwów, where he obtained the degrees M. Phil. in 1935 and Ph.D. in 1937.

This was a period of awakening in Polish science. Marian Smoluchowski had spurred a new interest in physics, and mathematics was developing rapidly: in Warsaw, under Waclaw Sierpinski, and in Lwów, under Hugo Steinhaus. In his autobiography (1985, 1, p. 29), Kac called this renaissance "wonderful." Most wonderful for him was the chance to study with Steinhaus, a mathematician of perfect taste, wide culture, and wit; his adored teacher who became his true friend and introduced him to the then undigested subject of probability. Kac would devote most of his scientific life to this field and to its cousin, statistical mechanics, beginning with a series of papers prepared jointly with Steinhaus on statistical independence (1936, l–4; and 1937, l–2).

Kac's student days saw Hitler's rise and consolidation of power, and he began to think of quitting Poland. In 1938 the opportunity presented itself in the form of a Polish fellowship to Johns Hopkins in Baltimore. Kac was 24. He left behind his whole family, most of whom perished in Krzemieniec in the mass executions of 1942–1943. Years later he returned, not to Krzemieniec but to nearby Kiev. I remember him rapt, sniffing about him and saying he had not smelled such autumn air since he was a boy. On this trip he met with a surviving female cousin who asked him, at parting, "Would you like to know how it was in Krzemieniec?" then added, "No. It is better if you don't know".

These cruel memories and their attendant regrets surely stood behind Kac's devotion to the plight of Soviet *refusniks* and others in like distress. His own life adds poignancy to his selection of the following quote from his father's hero, Solomon Maimon: "In search of truth I left my people, my country and my family. It is not therefore to be assumed that I shall forsake the truth for any lesser motives" ([1], p. 9).

America. Kac came to Baltimore in 1938 and wrote of his reaction to his new-found land:

> "I find it difficult... to convey the feeling of decompression, of freedom, of being caught in the sweep of unimagined and unimaginable grandeur. It was life on a different scale with more of everything—more air to breathe, more things to see, more people to know. The friendliness and warmth from all sides, the ease and naturalness of social contacts. The contrast to Poland... defied description."

After spending 1938–1939 in Baltimore, Kac moved to Ithaca, where he would remain until 1961. Cornell was at that time a fine place for probability: Kai-Lai Chung, Feller, Hunt, and occasionally the peripatetic Paul Erdös formed, with Kac, a talented and productive group. His mathematics bloomed there. He also courted and married Katherine Mayberry, shortly finding himself the father of a family. So began, as he said, the healing of the past.

From 1943 to 1947 Kac was associated off and on with the Radiation Lab at MIT, where he met and began to collaborate with George Uhlenbeck. This was an important event for him. It reawakened his interest in statistical mechanics and was a decisive factor in his moving to be with Uhlenbeck at The Rockefeller University in 1962. There Detlev Bronk, with his inimitable enthusiasm, was trying to build up a small, top-flight school. While this ideal was not fully realized either then or afterwards, it afforded Kac the opportunity to immerse himself in the statistical mechanics of phase transitions in the company of Ted Berlin and Uhlenbeck, among others. Retiring in 1981, Kac moved to the University of Southern California, where he stayed until his death on October 25, 1984, at the age of seventy.

I am sure I speak for all of Kac's friends when I remember him for his wit, his personal kindness, and his scientific style. One summer when I was quite young and at loose ends, I went to MIT to study mathematics, not really knowing what that was. I had the luck to have as my instructor one M. Kac and was enchanted not only by the content of the lectures but by the person of the lecturer. I had never seen mathematics like that nor anybody who could impart such (to me) difficult material with so much charm.

As I understood more fully later, his attitude toward the subject was in itself special. Kac was fond of Poincaré's distinction between God-given and man-made problems. He was particularly skillful at pruning away superfluous details from problems he considered to be of the first kind, leaving the question in its simplest interesting form. He mistrusted as insufficiently digested anything that required fancy technical machinery—to the extent that he would sometimes insist on clumsy but elementary methods. I used to kid him that he had made a career of noting with mock surprise that $e^x = 1 + x + x^2/2 +$ etc. when the whole thing could have been done without expanding anything. But he did wonders with these sometimes awkward tools. Indeed, he loved computation (*Desperationmatemtik* included) and was a prodigious, if secret, calculator all his life.

I cannot close this section without a Kac story to illustrate his wit and kindness. Such stories are innumerable, but I reproduce here a favorite Kac himself recorded in his autobiography:

> "The candidate [at an oral examination] was not terribly good—in mathematics at least. After he had failed a couple of questions, I asked him a really simple one... to describe

> the behavior of the function $1/z$ in the complex plane. ‘The function is analytic, sir, except at $z = 0$, where it has a singularity,’ he answered, and it was perfectly correct. ‘What is the singularity called?’ I continued. The student stopped in his tracks. ‘look at me,’ I said. ‘What am I?’ His face lit up. ‘A simple Pole, sir,’ which was the correct answer.”

[1] *Enigmas of Chance*. (Autobiography). New York: Harper and Row.

1.5. Gretchen Warren

I add to these recollections another, from my boyhood. The first time I met Gretchen Warren I must have been 10 or 12. It was her custom then to spend the summer at the house of my cousin Eleo (much older than me), looking out over the New England sea going all the way to Portugal, which I imagined I could perceive faintly, far away. Two more different women can hardly be imagined: Eleo the complete sports-woman, devoted to swimming, tennis, 100-mile walks, horses, gossip, handsome men and women. Grechen intellectual, learned in literature, philosophy and myth, especially the old things of East and West: the Icelandia Sagas, Chanson de Roland, Homer, Villon, the Bhagavad-Gita, Plotinus, the Bible, in no particular order; loving music and also Natural History; a friend of Santayana, Henry George, and A.K. Coomaraswami; a mixture you might say, of Emerson and Agassiz. How these two became (and stayed) friends I do not know, but they did.

Gretchen was of another generation, maybe 60 then or more, but she spoke to me as an equal in a way I have never forgotten. She took my childish love of Natural History with perfect seriousness, sharing her books and her marvelous collection of shells: the violet snail, thinner than paper, making a raft of foam to carry her eggs, far out in the uttermost parts of the ocean; Cuban land snails with their wonderful variegated colors; things I still keep, making me think of her. She introduced me to other things she loved like Homer and Plotinus, not then but later on. She encouraged me to think that I might do something of my own, in science perhaps, or writing. And nothing heavy here, only that serious attention to a little child which was her great charm and kind gift to me. She was a beautiful woman. You may have seen her at the Fine Arts in Boston in Sargent’s painting: Mrs. Warren and her Daughters. We met less often as time went by. I saw her last in Boston, Beacon Hill, in 1949 or so. Then she died, leaving these memories.

1.6. Acknowledgements

It is my pleasant duty to thank F. Alberto Grünbaum, Pierre van Moerbeke, and Victor H. Moll for putting all this together. It has been a long job, not without my grateful acknowledgement which I offer here. Thanks also to David Williams and Hermann Flaschka for kind words and true understanding of what I have tried imperfectly to do.

My Debt to Henry P. McKean Jr. 2

David Williams[1]

I was very privileged to have had as research supervisors, David Kendall and Harry Reuter. I learnt a great deal from them and from Eugene Dynkin, André Meyer, and of course, Paul Lévy. But it has been to Henry McKean that I have most often turned for inspiration.

There is a simple reason for this. If anyone else writes a book on Stochastic Integrals, Fourier Series and Integrals, or Elliptic Curves, they might produce a fine book on the topic. But Henry (either alone, or with Harry Dym, or with Victor H. Moll) writes *Mathematics*, not mere exposition of a topic. One is left awestruck by the rich interconnectedness of the subject as evidenced by a dazzling array of examples and (usually very challenging) exercises. What a great antidote to the too prevalent 'elegant abstraction' culture in which (for example) Number Theory is OK provided one keeps away from those common-or-garden numbers, and in which even the Generalized Riemann Hypothesis, astoundingly deep though it is, is perhaps rather closer to the ground than one should be flying.

A few years ago, I had to have a brain tumour removed in something of an emergency. It was explained to me that the operation might seriously impede my ability to understand Mathematics. (I had great surgeons, and I don't think it has!) How glad I was that I had McKean and Moll [14] with me for what might have been my last few hours of Mathematics! Yes, there are a few slips in the book, but these are fussed over only by those who could never write anything a tenth as inspirational.

I started my research career on Markov-chain theory, and soon became haunted by the then recent paper by William Feller and McKean [4]. This showed that there exists a chain with all states instantaneous, counter to what Lévy had thought, though it was he who then gave the beautiful probabilistic construction of the F-M chain. From the viewpoint of the time, the F-M chain was even more amazing because all its off-diagonal jump rates are zero. I became rather obsessed by the Q-matrix problem of characterizing what could be the off-diagonal jump rates of a totally instantaneous chain.

When I realized that I could then make no progress with this problem, I decided to switch fields and to read the great Itô-McKean book on diffusion processes. Again I had to work very hard to do the exercises. I felt that Itô and McKean had calculated everything there is to calculate about Brownian motion. (This was in the days before Marc Yor and coworkers had found, and solved, lots more explicit problems.)

When it came to the famous Section 2.8 of Itô-McKean on local time, I despaired of

[1]Wales Institute of Mathematical and Computational Sciences (WIMCS), Swansea University, Singleton Park, Swansea SA2 8PP, UK, dw@reynoldston.com.

F.A. Grünbaum et al. (eds.), *Henry P. McKean Jr. Selecta*, Contemporary Mathematicians, DOI 10.1007/978-3-319-22237-0_2

having a full understanding. I therefore tried hard to decompose the problem into simpler ones. Henry included a nice exposition of my efforts as part of his paper [13].

What I had never expected was that thinking hard about Brownian local time would lead me to solve that Q-matrix problem. (I also made heavy use of ideas of Kendall, Reuter, Jacques Neveu and Lèvy.) Things really are interconnected!

For many years, I had been intrigued by McKean's paper [9] on a winding problem driven by white noise. In this, he looks at windings around the origin of the two-dimensional process with Brownian motion as one component and its integral up to current time as the other. The winding process is not at all easy to analyze. However, in this case, the joint process is Gaussian, and this allows one to describe a key distribution by an integral equation. In a typical *tour de force* of transform theory, McKean obtained the explicit solution as an unfamiliar distribution, and derived striking probabilistic consequences. This was the first paper on what came to be known as Markovian Wiener-Hopf theory.

I became interested in more general Wiener-Hopf winding problems in which one component of the two-dimensional process is a Markov process, and the other a fluctuating additive functional of that process. One of the simplest problems required calculation of a jump distribution from 0 of an induced Brownian motion on a half-line of the type studied by Feller and, more fully, by Ito and McKean [5]. When Henry visited Swansea, I told him that I conjectured that the jump distribution in the W-H context must be totally monotone. A day later (I recall Michael Atiyah's saying to me that Henry thinks at a million miles per hour!), Henry told me that he had proved both this conjecture and, by using Krein theory as in Dym and McKean, that all totally monotone jump laws arise this way. See London, McKean, Rogers and Williams [6].

My most recent paper (`http://arxiv.org/abs/1011.6513`), which I hope to be an amusing *divertissement*, is on the simplest non-linear version of Markovian W-H problems. Its W-H aspects can all be traced back to McKean on windings.

The paper's non-linear aspects can be traced back to McKean's paper on travelling waves and the KPP equation [12] which prompted massive developments on branching diffusions, measure-valued processes, and the like. How often he has sparked off new fields of investigation! He was way ahead of the field even in financial mathematics [10].

McKean has also been interested in winding problems of non-W-H type: in particular, topological problems on windings of the Brownian path. See, in particular, the papers [7, 15] by McKean with Lyons and with Sullivan, a formidable trio indeed, and doing Mathematics, not mere Probability.

His expertise in Gaussian processes was also used to great effect in his paper [8] on Lèvy's Brownian motions in multi-dimensional (and even Hilbert-space) time. There is profound work here on splitting fields, etc., and there are really surprising results. See also papers [2, 3] with Dym.

In 1980, I organized a conference at Durham in which the then-brand-new Malliavin calculus played a large part. I decided to write an introduction to the proceedings, and found McKean's paper [11] on the geometry of differential space invaluable for this. The fundamental Malliavin process is Henry's Brownian motion on an infinite-dimensional sphere of radius the square-root of infinity. (Surreal?!)

Though perhaps primarily interested in seeing principles put to good use in the concrete, he is a master of the abstract too. His paper [1] with Blumenthal and Getoor, proves that two Markov processes with identical hitting distributions are time-changes of each other. No result in Markov-process theory is deeper than this.

The above is just a hint as to how Henry's work has enriched the life of one probabilist. I have concentrated on books and papers which, as it were, have become part of me:

ones on which I do not need to refresh my memory. Other, better, probabilists could say much more.

I know that Henry's work is regarded with the same admiration and gratitude by people working in differential equations and in other fields.

If there is an explicit solution to be found, then Henry is the man to find it. But his almost unique skill at calculation is always combined with deep new insights into the underlying principles.

My sincere thanks, and my very best wishes, Henry!

Bibliography

[1] R. M. Blumenthal, R. K. Getoor, and H. P. McKean. Markov processes with identical hitting distributions. *Illinois J. Math.*, 6:402–420, 1962.

[2] H. Dym and H. P. McKean. Applications of the de Branges spaces of integral functions to the prediction of stationary Gaussian processes. *Illinois J. Math.*, 14:299–343, 1970.

[3] H. Dym and H. P. McKean. Extrapolation and interpolation of stationary Gaussian processes. *Ann. Math. Stat.*, 41:1819–1844, 1970.

[4] W. Feller and H. P. McKean. A diffusion equivalent to a countable Markov chain. *PNAS*, 42:351–354, 1956.

[5] K. Itô and H. P. McKean. Brownian motion on a half line. *Ill. J. Math.*, 7:181–231, 1963.

[6] R. R. London, H. P. McKean, L. C. G. Rogers, and D. Williams. A martingale approach to some Wiener-Hopf problems. *Sem. Prob., Lecture Notes in Math.*, 920:41–90, 1982.

[7] T. J. Lyons and H. P. McKean. Winding of the plane Brownian motion. *Adv. Math.*, 51:212–225, 1984.

[8] H. P. McKean. Brownian motion with a several-dimensional time. *Teor. Verojatnost. i Primenen.*, 8:357–378, 1963.

[9] H. P. McKean. A winding problem for a resonator driven by a white noise. *J. Math. Kyôto Univ.*, 2:228–295, 1963.

[10] H. P. McKean. A free boundary problem for the heat equation arising from a problem in mathematical economics. *Ind. Management Rev.*, 6:32–39, 1965.

[11] H. P. McKean. Geometry of differential space. *Ann. Prob.*, 1:197–206, 1973.

[12] H. P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piscounov. *Comm. Pure Appl. Math.*, 28:323–331, 1975.

[13] H. P. McKean. Brownian local times. *Adv. Math.*, 16:91–111, 1975.

[14] H. P. McKean and V. H. Moll. *Elliptic Curves: Function Theory, Geometry, Arithmetic*. Cambridge University Press, New York, 1997.

[15] D. Sullivan and H. P. McKean. Brownian motion and harmonic functions on the class surface of the thrice-punctured sphere. *Adv. Math.*, 51:203–211, 1984.

Henry P. McKean Jr. and Integrable Systems

3

Hermann Flaschka[1]

Henry McKean's first contribution to integrable systems appeared in 1975 [15]. Over a period of more than 30 years since, in some 50 papers, he has explored integrable systems from uniquely original points of view. His selecta could have included the pioneering work, with Airault and Moser, on the time-dynamics of poles of meromorphic solutions of KdV [2]; or the series on invariant measures for wave equations, integrable or otherwise; or one of the seminal papers with Trubowitz [12, 13] that created a theory of infinite-genus hyperelliptic curves (= Riemann surfaces) and infinite-dimensional Jacobian varieties, which they applied to KdV under periodic boundary conditions, or subsequent extensions of that theory motivated by the desire to understand all the iconic integrable partial differential equation on the circle.

There is another project, of great scope, in which "Geometry of KdV (1)" [7], in this volume, is the first step. Read in isolation, without appreciation of the developments in papers [8] and [10] at least, this paper is incomplete. It sets out the vocabulary for the striking results to follow. I thought that Henry's invention of unimodular spectral classes and additive classes would serve as exemplar of his imagination and technical virtuosity (and fearlessness, one might add). Moreover, there are many questions to be thought about; some are natural, though probably hard, but the most difficult task is to think of workable examples that will reveal something new.

The papers [7, 8, 10] are themselves building blocks of a sweeping conjecture. Henry proposes to interpret the spectral theory of differential operators[2] $Q = -D^2 + q(x)$ as a reflection of infinite-dimensional algebraic geometry in the space of all such operators. The space is stratified into classes labeled by spectral data of some sort, and parametrized by an additive group; the classes are Jacobian varieties of objects resembling, somehow, quadratic algebraic curves with perhaps a continuum of branch cuts or singular points; and the coefficients $q(x)$ and properly normalized eigenfunctions $\mathfrak{e}(x, \lambda)$ on each class are represented by an object resembling, somehow, a Riemann theta function, and these representations, quoting from [10], "may be viewed as uniformizing, class by class, the eigenvalue problem $Q\mathfrak{e} = \lambda\mathfrak{e}$. It seems that such a geometrical attitude would be new to spectral theory".

Substantial evidence suggests that yes, "something must be going on". Supporting examples draw on quantum scattering, analytic functions, algebraic curves, and the geometry of infinite-dimensional manifolds, all intertwined and placed into a framework of the infinite-dimensional dynamical systems

[1]Department of Mathematics, University of Arizona, Tucson, AZ, USA, flaschka@math.arizona.edu.

[2]D stands for the operator d/dx, and dot and prime will indicate time and space derivatives.

F.A. Grünbaum et al. (eds.), *Henry P. McKean Jr. Selecta*, Contemporary Mathematicians, DOI 10.1007/978-3-319-22237-0_3

sometimes called "integrable". This name is often used more as a suggestion than a definition, and equally often merely indicates the origin of a piece of mathematics that has little to say about dynamical systems. The historical context, however, is still relevant to the story.

The modern theory of integrable systems was born in 1965, when Gardner, Greene, Kruskal, and Miura [4] invented a nonlinear version of the Fourier transform, based on the inverse scattering method from quantum mechanics, in order to solve the KdV equation. This is the partial differential equation $\dot{q} = -q''' + 6qq'$; it governs a certain asymptotic regime of the nonlinear motion of waves on water, or in a plasma, or in an atomic lattice. A few years later, Zakharov and Faddeev [16] interpreted the inverse scattering solution of KdV as a transformation of an infinite-dimensional Hamiltonian system to action-angle coordinates, thereby translating a successful, but unmotivated and mysterious, technique into the more familiar language of classical mechanics. As the supply of integrable Hamiltonian ordinary and partial differential equations grew, so did the scope of their applications, in mathematics and physics, and the breadth of techniques brought to bear on the analysis of their dynamics. In the early years, it was not terribly misleading to refer to this corner of mathematical physics as "the inverse scattering method", or "soliton theory", or "KdV theory". However, much as Fourier series, invented for the purpose of solving the heat equation, evolved into the all-encompassing paradigm called harmonic analysis, so has the invention of the inverse scattering method for the solution of KdV transformed mathematics through the creation of a new paradigm; one that has no official name, but "integrable systems" is as good a label as any, since, like "harmonic", it reminds one of the origin of the ideas that have emerged from the first examples.

I would have liked to spend time on theta functions and their incarnation in scattering theory [3] and in the quantum harmonic oscillator [14], since algebraic geometry is an integral part of the broad picture, but my description of the background and implications of [7] is of substantial length already. In any case, I could not improve on Henry's exposition, in [9], of the evidence for a theory of Riemann surfaces of countably and uncountably infinite genus.

I start with a summary of that part of KdV and the inverse scattering method that is needed for an appreciation of Henry's ideas, and only that part. Maybe it is of use to a non-expert. KdV (1)–(3) enter in Section 3, where I explain the content of [7]. Section (4) deals with a remarkable result, from [8], about the class of operators known as "finite-gap" operators. It says that if the orbit of a certain group of transformations acting on the space of all $-D^2 + q$ is finite dimensional, then q is an Abelian function.

3.1. KdV Manifolds in the Scattering Class

The inverse scattering method was created to solve KdV with localized initial conditions, and this is also the starting point here; the coefficient q is taken to be in the space $C^\infty_\downarrow$ of rapidly decreasing smooth functions. The spectrum of $-D^2 + q$ then consists of a continuous part, $0 \leq \lambda < \infty$, and possibly a finite number of negative eigenvalues. These are of great importance in applications of KdV, but complicate the analysis. Therefore it is assumed throughout that there is no discrete spectrum.[3] The collection $\mathcal{S}$ of such Q is called the *scattering class*. The KdV equation defines a flow on $\mathcal{S}$.

The first order of business is to introduce an equivalence relation that stratifies $\mathcal{S}$ into infinite-dimensional tori[4] invariant under the

[3]This assumption leads to incorrect conclusions in the Hamiltonian approach to KdV, but that difficulty is ignored.

[4]It would be (usually) possible to give precise characterizations, in terms of analyticity, growth, smoothness, etc., of the function classes introduced below. Statements about geometry, for instance, that something is "stratified" or a "manifold" or a "torus", are more problematical; the intuition is very important, but there is often no proof or even precise formulation. I do not try to separate fact from useful fiction.

KdV flow. These are called *KdV manifolds*. They are labeled by a kind of integral transform $Z(k)$ of $q(x)$ constructed from the generalized eigenfunctions of Q. The functions Z are known as *action variables* in Hamiltonian mechanics.

KdV is in fact only one of an infinite family of commuting vector fields on $\mathcal{S}$ that are tangent to the KdV manifolds and span the tangent spaces. The next step will be to understand their flows; they can be interpreted as action of a huge additive group of operators, the so-called group $\mathfrak{A}$ of *additions*. It generates the KdV manifold from a reference potential. The coordinates on the tori introduced by $\mathfrak{A}$ are (related to) the *angle variables* of Hamiltonian mechanics.

Finally, one needs an algorithm to pass from the torus coordinates Z and $A \in \mathfrak{A}$ back to q. This is implemented by an operator determinant indexed by the torus label Z. Schematically, $q(x) = \det(A(x); Z)$, where $A(x)$ is a 1-parameter subgroup of $\mathfrak{A}$.

This picture is a rewording of the inverse scattering solution of KdV, formulated with an eye towards generalizations to q that are not of scattering class. The geometric interpretation is enriched by incorporation of ideas from the theory of algebraic curves. That comes later.

3.1.1. Model: A Geometric Picture of Linear KdV. The KdV manifolds are infinite-dimensional, except for the operator $Q_0 := -D^2$ which is a fixed point–the only one in $\mathcal{S}$–of KdV. One can linearize KdV, and in fact the whole (yet to be explained) construction at Q_0; the tangent space $T_{Q_0}\mathcal{S}$ is naturally identified with $C^\infty_\downarrow$, and the linearized KdV manifolds are infinite-dimensional tori indexed by the modulus of the Fourier transform. This familiar setting affords a convenient starting point. It is meant to introduce the idea of stratification by tori, their labeling, and coordinates on them.

Consider the KdV equation linearized at $q = 0$,

(LKdV)
$$\dot{q} = M_0 q := -D^3 q, \quad q(x,0) \text{ given}, \ x \in \mathbb{R}.$$

It is diagonalized by the Fourier transform, which we define by

$$\hat{q}(k) = \int q(x)\mathrm{e}^{-2ikx}\, dx. \tag{3.1.1}$$

The Fourier transformed LKdV equation,

$$\dot{\hat{q}}(k,t) = 8ik^3\hat{q}(k,t), \tag{3.1.2}$$

is solved by inversion. Evidently, $|\hat{q}(k,t)|^2$ is independent of time t, signifying that the energy in the k-th Fourier mode is conserved by LKdV. More generally, every equation of the form

$$\dot{q} = D\Omega(-D^2)q, \ \text{with transform}$$
$$\dot{\hat{q}} = 2ik\Omega(4k^2)\hat{q}, \tag{3.1.3}$$

preserves modal energy; here Ω denotes a real function of slow growth. The PDEs

$$\dot{q} = D(-D^2)^m q, \ m = 1, 2, 3, \dots,$$

belong to this class. They are the linearizations, about $q = 0$, of the equations in the so-called *KdV hierarchy*, which will be introduced shortly.

Equations (3.1.3) define commuting vector fields $\mathbb{Y}_\Omega$ on $C^\infty_\downarrow$. They are tangent to the subsets of $C^\infty_\downarrow$ characterized by common modal energies. Following [3], the sets of q with Fourier transforms $\hat{q} = |\hat{q}_0|\mathrm{e}^{i\psi}$ whose modulus $|\hat{q}_0|$ is fixed and phase $\psi(k)$ is a slowly increasing odd function are called *LKdV manifolds* and denoted by $\mathbb{J}(|\hat{q}_0|)$. An LKdV manifold is shaped like an infinite-dimensional torus. It is a product of continuum many circles, one for each k; the radii are labeled by $|\hat{q}_0(k)|$, and the circumferences are coordinatized by $\exp(i\psi(k))$.

Even though LKdV appears initially to be the central object, the goal, were we to continue in the linear approximation, would really be to understand the totality of vector fields $\mathbb{Y}_\Omega$, their interaction with the LKdV manifolds, and the stratification of $C^\infty_\downarrow$ into tori. We now begin this program in the nonlinear setting.

The review of scattering theory contains nothing that is not known to experts. It is in the nature of an appendix, but placed where it should logically appear.

3.1.2. Scattering Theory: A Nonlinear Fourier Transform. The KdV equation is a nonlinear modification of LKdV:

(KdV)
$$\dot{q} = -q''' + 6qq' = (M_0 + 2(qD + Dq))\, q := Mq.$$

It, too, can be diagonalized, in a basis formed by eigenfunctions Φ, not quite of M but of the pseudodifferential operator $\mathcal{L} := D^{-1}M$ and its adjoint. There is a Φ-transform of q, denoted by $R(k)$, whose time dependence is governed by the same decoupled system, (3.1.2), as the linearized equation. However, because M depends on q, the expansion basis $\{\Phi\}$ will also depend on q, and to invert this Φ-transform one must reconstruct both q and $\{\Phi\}$ from R.

The KdV miracles happen because M is a very special object. In differential Galois theory, it is known as the *2nd symmetric power* of the Schrödinger operator $Q := -D^2 + q(x)$. The name signifies that products $\Phi = f^2, fg, g^2$ of two solutions f, g of $Qy = k^2y$ satisfy the generalized eigenvalue equation

$$M\Phi = 4k^2\Phi', \tag{3.1.4}$$

Therefore, to understand the Fourier-like expansions in the squared eigenfunctions Φ, one must first study the solutions of $Qy = k^2y$.

This reasoning "explains" why the Schrödinger operator Q is so fundamental in KdV theory.[5]

3.1.2.1. *Scattering Matrix.* Since $-D^2+q$ is approximately $-D^2$ for large $|x|$, there are two solutions $f_\pm$ of $Qy = k^2y$ normalized as shown in the table.

	$x \to \infty$	$x \to -\infty$
$f_+(x,k)$	$T_+(k)\exp(ikx)$	$\exp(ikx)$ $+R_-(k)\exp(-ikx)$
$f_-(x,k)$	$\exp(-ikx)$ $+R_+(k)\exp(ikx)$	$T_-(k)\exp(-ikx)$

Incoming plane waves $\exp(\pm ikx)$ are scattered by a "potential" $q(x)$. A portion carrying energy $|T_\pm(k)|^2$ is transmitted, and $|R_\pm(k)|^2$ worth is reflected. The *scattering matrix*

$$S(k) = \begin{bmatrix} T_+(k) & R_-(k) \\ R_+(k) & T_-(k) \end{bmatrix} \tag{3.1.5}$$

is unitary. The condition $|R_\pm|^2 + |T_\pm|^2 = 1$ signifies conservation of energy at each wavenumber k. One finds that $T_+ = T_-$; the common value is the transmission coefficient T. We will rarely need R_-, and write R for R_+. It is the (right) reflection coefficient.

The reflection coefficient $R(k)$ determines $q(x)$. This crucial fact will be taken for granted. The inversion formula is stated below in the section on the quantum harmonic oscillator.

3.1.2.2. *Tools from Complex Function Theory.* Analyticity properties of $f_\pm$ and T in the upper half plane $\operatorname{Im} k > 0$ play an essential role. Key pieces of the evidence for Henry's conjecture, presented in [8], are essentially theorems about analytic functions. To give a flavor of the tools required, I list some essential facts that are used in many proofs.

Analyticity of $f_\pm(x,k)$ for $\operatorname{Im} k > 0$, for each fixed x, is a simple side product of the Neumann series argument that proves existence of these functions. The Wronskian of $f_\pm$ is $-2ik/T(k)$, so T is also analytic. The reflection coefficient R is rapidly decreasing as $|k| \to \infty$ on the real axis.

More careful analysis shows that the functions

$$k \mapsto e_\pm(x,k) := \frac{1}{T(k)} f_\pm(x,k) \mathrm{e}^{\mp ikx} \tag{3.1.6}$$

are outer functions in the Hardy space $(1 + H^{2+}) \cap H^{\infty+}$, and that $T(k)$ is determined for $\operatorname{Im} k > 0$ by its values along the real axis by the Poisson formula

(3.1.7)
$$\ln T(k) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\ln |T(k')|^2}{k' - k}\, dk', \ \operatorname{Im} k > 0.$$

The reflected functions $e^*_\pm(x,k) = e_\pm(x,-k)$ belong to $(1 + H^{2-}) \cap H^{\infty-}$. The

[5]The squared eigenfunction approach to integrable PDEs and the interpretation of inverse scattering as nonlinear Fourier transform were introduced in the early days of soliton theory in the seminal paper by Ablowitz et al. [1].

pairs $e_\pm^*$ and $e_\pm$ are patched across the real axis according to

$$e_\pm^* + R_\pm \mathrm{e}^{\mp 2ikx} e_\pm = T e_\mp. \tag{3.1.8}$$

The Fourier transform of this relation amounts to an integral equation from which q can be determined. This is inverse scattering.

3.1.2.3. *Squared Eigenfunction Expansion.* The so-called *squared eigenfunctions* $\Phi_\pm := f_\pm^2$ and $\Phi_0 := f_+ f_-$, are solutions of the pseudodifferential eigenvalue problem obtained by integrating (3.1.4),

$$D^{-1}M\Phi := \mathfrak{L}\Phi = 4k^2\Phi, \text{ where}$$
$$\mathfrak{L} = -D^2 + 4q - 2D^{-1}q'. \tag{3.1.9}$$

Their derivatives satisfy the adjoint equation, $\mathcal{L}^*\Phi'_\pm = 4k^2\Phi'_\pm$.

The product Φ_0 is Green's function G_{xy} for $x = y$, up to factor. The other two squared eigenfunctions $\Phi_\pm$ are of decisive importance for (at least) two reasons.

They are derivatives, with respect to $q(x)$, of the reflection coefficient $R(k)$. When one leaves the scattering class, and normalization at $|x| = \infty$ is no longer possible, the gradients of other spectral quantities are often natural substitutes (cf. the quantum harmonic oscillator, below). They always satisfy $M\Psi = 4\lambda\Psi'$.

The sets $\Phi_\pm$ and $\Phi'_\mp$ form the biorthogonal bases used to diagonalize KdV. The representation of q itself is simple and elegant:

$$\tilde{q}(k) = \int q\,\Phi'_-\,dx = 4k^2 R(k) \tag{3.1.10a}$$

$$q(x) = \frac{2i}{\pi}\int \frac{kR}{T^2}\,\Phi_+\,dk. \tag{3.1.10b}$$

This takes a familiar form when q is "small". To first order, $\Phi_\pm(x,k) \sim \mathrm{e}^{\pm 2ikx}$. The small scatterer q reflects but a small portion of the incoming wave, whence $T \sim 1$. The orthogonality relations between Φ and Φ' become the standard ones for exponentials, and the relations (3.1.10) reduce to

$$R(k) \sim -i\hat{q}(k)/2k.$$

The reflection coefficient is hereby revealed to be a nonlinearization of the Fourier transform. A number of later formulas are profitably understood as perturbations of familiar Fourier facts.

The orthogonality of Φ, Φ' can also be seen as consequence of a geometric property: the Φ' are tangent, and the Φ are normal, to the nonlinear analogs of the LKdV manifolds; these will now be introduced.

3.1.3. The KdV Manifold. In Sect. 3.1.1 we defined an LKdV manifold as a set of q with prescribed modal energies $|\hat{q}(k)|^2$. A KdV manifold[6] consists of the set of q with the *reflected* and *transmitted* energy in each mode, $|R(k)|^2$ and $|T(k)|^2$, prescribed (they add to 1).

Let $\mathbb{J}(r)$ denote such a set. It consists of all q whose reflection coefficient has the form $R(k) = r(k)\mathrm{e}^{i\psi(k)}$ with fixed r and odd function ψ of slow growth. Since $|R|$ determines T, one may also write $\mathbb{J}(T)$.

In complete analogy with the linear case (3.1.3), there is a distinguished family of vector fields tangent to $\mathbb{J}(r)$. Take the evolution equation (a priori nonlocal)

$$\dot{q} = \mathbb{Y}_\Omega q := D\Omega(\mathcal{L})q. \tag{3.1.11}$$

Substitution of (3.1.10b) for q and application of (3.1.9) lead to

$$\dot{R}(k,t) = 2ik\Omega(4k^2)R(k,t). \tag{3.1.12}$$

Thus, $|R(k)|^2$ is independent of time and the flows (3.1.11) preserve the KdV manifolds. Furthermore, they are simultaneously diagonalized, and so commute.

The vector fields

(3.1.13)
$$\dot{q} = \mathbb{X}_m q := D\mathcal{L}^m q = (-1)^m D^{2m+1} q$$
$$+ \text{nonlinear terms}, m = 0, 1, 2, 3, \ldots,$$

are known as the *KdV hierarchy* ($m = 0$ is translation of the potential, $\dot{q} = q'$, and $m = 1$ is KdV). The iterated antiderivatives in $D\mathcal{L}^m$ magically cancel, and the $\mathbb{X}_m$ are polynomial in q and its derivatives.

The nonlocal equation defined by the resolvent of $\mathcal{L}$, namely $\Omega(\mathcal{L}) = -2(\mathcal{L}+4\kappa_0^2)^{-1}$, is an *infinitesimal addition.* It will soon assume a place of prominence.

[6] The "KdV" indicates no more than invariance under the flows of a family of commuting vector fields that includes KdV; often, KdV per se has nothing to do with the matters of interest.

3.1.4. KdV as Integrable Hamiltonian System. At this point, we know[7] that the scattering class $\mathcal{S}$ is stratified by infinite-dimensional tori that are orbits of a huge group denoted by $\mathfrak{A}$ above. This is the group of translations of angles in the phase functions, generated by the vector fields $\mathbb{Y}_\Omega$. The torus-angle picture is familiar in classical Hamiltonian mechanics, and it will be very convenient to express KdV geometry with a Hamiltonian vocabulary. The list of definitions and facts is relegated to an appendix.

3.1.4.1. *Complete Integrability of KdV.* Hamiltonian systems are a skew analog of gradient systems. In Euclidean space, they have the form

$$\dot{\mathbf{x}} = P \operatorname{grad} H(\mathbf{x}).$$

For KdV, the so-called Poisson operator P is the derivative $D = \frac{d}{dx}$. Its non-invertibility causes complications, but these are ignored.

The Poisson bracket and Hamiltonian vector field have the form

$$\{F, G\}(q) = \int (\operatorname{grad} F)\,(\operatorname{grad} G)',$$

$$\dot{q} = \mathbb{X}_H q = (\operatorname{grad} H(q))'. \tag{3.1.14}$$

The Hamiltonian for KdV is $\frac{1}{2}\int (q')^2 + 2q^3$; the quadratic term by itself is the Hamiltonian for LKdV.

The KdV manifolds are interpreted as follows. They are common level sets $\mathbb{J}(r) = \cap_{k\in\mathbb{R}}\{q \mid |R(k)| = r(k)\}$ of the family of functions $q \mapsto |R(k)|$, indexed by k. These functions Poisson commute; for every pair k, ℓ one has

$$\{|R(k)|, |R(\ell)|\} = 0. \tag{3.1.15}$$

Remarkably, this geometric property is merely a rewording of bi-orthogonality of the squared eigenfunctions. The gradients of the right and left reflection coefficients are

$$\operatorname{grad} R_\pm = \frac{1}{2ik}\,\Phi_\mp,$$

and, up to factors,

$$\{R_-(k), R_+(\ell)\} = \int \Phi_+(x,k)\Phi'_-(x,\ell)\,dx = \delta(k-\ell). \tag{3.1.16}$$

Equation (3.1.15) for the moduli is then derived by appeal to unitarity of the scattering matrix (3.1.5).

The KdV manifolds $\mathbb{J}(|R|)$ are continuum tori labeled by $|R|$ and parametrized by $\arg R$; a little rewriting turns these coordinates into canonical variables of action-angle type,[8]

$$Z(k) = -\frac{k}{\pi}\ln|T(k)|^2 \quad\text{and}\quad \theta(k) = \arg\frac{R(k)}{T(k)}. \tag{3.1.17}$$

3.1.4.2. *Tangent Spaces to* $\mathbb{J}(r)$. The functions $T(i\kappa), \kappa > 0$, are in involution since the Poisson representation (3.1.7) determines T in the upper half plane from R on the real axis. The gradient of $T(i\kappa)$ is Green's function G_{xy} on the diagonal and generates the Hamiltonian vector field

$$\dot{q} = \mathbb{X}_\kappa q := (\operatorname{grad} T(i\kappa))' = -G'_{xx}(-\kappa^2|Q). \tag{3.1.18}$$

The $\mathbb{X}_\kappa$ span the tangent spaces to the KdV manifold.

The transform of $\mathbb{X}_\kappa$ to $R(k,t)$, which is needed later, can be computed in action-angle variables from the Poisson representation (3.1.7) of $T(i\kappa)$:

$$\dot{R}(k,t) = -\frac{ik}{k^2+\kappa^2}R(k,t). \tag{3.1.19}$$

According to (3.1.12), the vector field (3.1.18) can also be represented in terms of the resolvent of $\mathcal{L}$, namely, $\dot{q} = -2D(\mathcal{L}+4\kappa^2)^{-1}q$.

3.1.4.3. *KdV Hierarchy.* The role of the KdV hierarchy of local equations, (3.1.13), deserves special emphasis. Green's function satisfies $MG_{xx} = 4\lambda G'_{xx}$; the coefficients of the asymptotic expansion of $G'_{xx}(k^2|Q)$ in inverse powers of k can be determined recursively, and are nothing but the KdV vector fields $\mathbb{X}_m$:

$$G'_{xx}(k^2|Q) \sim \frac{1}{2}q'\frac{1}{k^3} + \frac{1}{8}(6qq' - q''')\frac{1}{k^5} + \mathbb{X}_3\frac{1}{k^7} + \cdots \tag{3.1.20}$$

Thus, Eq. (3.1.18) contain the KdV hierarchy, albeit in a very implicit manner.

---

[7]Once more: this is the guiding picture; what we *really* know may be a lot or a little.

[8]For small q, they reduce to $Z_0(k) = \frac{1}{4\pi k}|\hat{q}(k)|^2$ and $\theta_0(k) = \arg\hat{q}(k),\ k > 0$.

3.2. Geometry of KdV: Additive and Unimodular Classes

We now leave the scattering class $\mathcal{S}$ and only ask that q be a smooth function and that the spectrum of Q be positive. The class of such operators is denoted by $\mathfrak{Q}$.

The steps to be accomplished are:

(a) to find an equivalence relation on $\mathfrak{Q}$ whose equivalence classes in the scattering case are the tori $\mathbb{J}(r)$;
(b) to generate an equivalence class from one of its members by a procedure that generalizes translation of R by all possible phase functions;
(c) to find a family of commuting vector fields that span the tangent spaces of the equivalence classes and, in the scattering case, contains the KdV hierarchy.

Paper (1) in the *Geometry of KdV* series proposes solutions of these three problems. The equivalence classes are the *unimodular spectral classes* of the title. The analogs of the phase functions are generated by *additions*. The tangent vectors are *infinitesimal additions*, which generalize the vector fields (3.1.18) with Hamiltonians $T(i\kappa_0)$.

Paper (2) verifies that the scattering class, the class of operators with periodic potential (the Hill operators), and a certain finite-dimensional subclass thereof, all fit the framework. It will become clear that this is a highly nontrivial result.

Paper (3) integrates a large subset of the vector fields from (c), by the "simple" expedient of exhibiting an explicit solution as infinite determinant.

Some of these results, particularly (c), were motivated by earlier work [14] of McKean and Trubowitz[9] on the quantum mechanical harmonic oscillator, $Q_0 := -D^2 + q_0$ with $q_0 = x^2 - 1$. This operator is the complete opposite of the scattering type (and Hill's operator); it has pure discrete spectrum $\lambda_n = 2n, n \geq 0$, and the KdV vector field $-q''' + 6qq'$ cannot be integrated because the x^3 growth of the nonlinearity is not balanced by the linear terms. The analog of the KdV manifold is parametrized by the exponential map of Hamiltonian vector fields, $\mathbb{X}_n = (\operatorname{grad}\lambda_n)'$, that are no longer local in q,

$$(t_0, t_1, t_2, \dots) : Q_0 \mapsto \exp(\sum_0^\infty t_n \mathbb{X}_n) \cdot Q_0,$$

In the generality of class $\mathfrak{Q}$, a continuum version of this map will be required; loosely speaking, it is a superposition of vector fields $\mathbb{X}_\lambda$ of infinitesimal additions, smeared by a measure μ,

$$t(\lambda) : Q_{\text{reference}} \mapsto \exp\left(\int_0^\infty t(\lambda)\mathbb{X}_\lambda \, d\mu(\lambda)\right) \cdot Q_{\text{reference}}. \tag{3.2.1}$$

For the oscillator, μ is concentrated on the nonnegative integers.

3.2.1. The Harmonic Oscillator and Paired Additions. The goal is to describe the spectral manifold[10] $\mathbf{Q}$ of Q_0, meaning the set of operators of the form $Q = -D^2 + q$ whose eigenvalues are also $\lambda_n = 2n$, for q in the space $x^2 - 1 + C_\downarrow^\infty$. It is the analog of the KdV manifold in the scattering class (but recall that KdV has no meaning in $\mathbf{Q}$).

The spectral manifold $\mathbf{Q}$ will be imagined as submanifold of an ambient space of "all" operators Q that have eigenvalues $\lambda_n = \lambda_n(Q)$ "near" those of the harmonic oscillator. We picture $\mathbf{Q}$ as one of a stack of level sets $\mathbf{Q_c}$ of operators with all $\lambda_n(Q) = c_n$ prescribed. The geometry of $\Phi_\pm$ and $\Phi'_\pm$ is replayed simply and cleanly. The normal and tangent directions of $\mathbf{Q}$ are spanned, respectively, by[11] $\operatorname{grad}\lambda_n = e_n^2$ and $(e_n^2)'$. Bi-orthogonality translates into $\{\lambda_m, \lambda_n\} = 0$. The Hamiltonian vector fields $\mathbb{X}_n = (\operatorname{grad}\lambda_n)'$ commute, and it is possible to solve the equations

$$\frac{\partial q}{\partial t_m} = (e_m^2)', \; 0 \leq m \leq n, \tag{3.2.2}$$

[9] As far as I know, this is still the only paper to analyze an operator from outside the traditional KdV world.

[10] This time it really is a manifold.

[11] e_n is the normalized eigenfunction corresponding to λ_n.

simultaneously. Miraculously, there is an explicit formula:

(3.2.3)
$$q(x;t_0,t_1,\dots,t_n) = q_0 - 2\frac{d^2}{dx^2}\left(\ln\det\left[\delta_{k\ell} + (e^{t_k}-1)\int_x^\infty e_k^0 e_\ell^0,\ 0\le k,\ell\le n\right]\right)$$

(superscript zero denotes initial values). One can even take $n\to\infty$ as long as $\mathbf{t}=\{t_m\}_0^\infty$ decreases rapidly. The sequences $\mathbf{t}$ form a coordinate grid on the spectral manifold $\mathbf{Q}$, or equivalently, on the group $\mathfrak{A}$ of *additions* mentioned at the beginning of Section 2.

3.2.1.1. *Additions in General.* Additions are substitutes for the local flows, like KdV, that may no longer exist in the generality of the class $\mathfrak{Q}$. They are special instances of a classical transformation of 2nd-order ODEs, the Darboux transformation,[12] which takes a zero-free solution y_1 of $Q_1 y = -y'' + q_1 y = \mu y$ as input and creates a new operator according to

(3.2.4)
$$q_2 = q_1 - 2(\ln y_1)'', \quad \text{and} \quad Q_2 = -D^2 + q_2.$$

The map[13] $P : y \mapsto y_1^{-1} W(y, y_1)$ sends solutions of $Q_2 y = \lambda y$ to solutions of $Q_1 y = \lambda y$, for $\lambda \ne \mu$; for $\lambda = \mu$, the new solutions are

$$y_1^{-1}\Big(a_1 + a_2\int^x y_1^2\Big). \tag{3.2.5}$$

Thus, everything about the new operator is known. Moreover, Darboux transformations commute and preserve the Poisson bracket.

The determinant (3.2.3) is constructed from repeated additions that start at the initial condition q_0 and use the data λ_m and e_m^0. However, because e_m^0 has zeros when $m > 0$, so that the new potential (3.2.4) has poles, the additions must be done in pairs. If y_1 is not zero-free, do the first Darboux transformation as above, simply ignore the singular nature of the resulting q_2, and perform a second transformation with the solution (3.2.5). The result of this paired addition will be the 1×1 version of (3.2.3),

(3.2.6)
$$\text{new } q := q_1 - 2\frac{d^2}{dx^2}\left(\ln[1 + a\int_x^\infty y_1^2]\right).$$

It is rather surprising that the new potential is smooth (when $a \ge 0$). The determinant in (3.2.3) is built by iterating paired additions. This procedure is not peculiar to the harmonic oscillator; if an operator Q has simple eigenvalues $\lambda_0 < \lambda_1 < \lambda_2 < \cdots$, then the solution of (3.2.2) is given by the same formula.

3.2.1.2. *The Group of Additions Acts Freely and Transitively.* It is shown in [14] that the map $\mathbf{t} \mapsto \exp(\sum t_m \mathbb{X}_m) q_0$ is 1:1 onto $\mathbf{Q}$, by a method reminiscent of inverse scattering. For the normalized eigenfunctions of $-D^2 + q(x;\mathbf{t})$, define

$$c_\pm^n(\mathbf{t}) := \frac{e_n(x=\pm\infty,\mathbf{t})}{e_n^0(x=\pm\infty,\mathbf{t}=\mathbf{0})}.$$

These asymptotic data satisfy

$$c_+^n c_-^n = 1, \qquad \frac{c_+^n}{c_-^n} = e^{t_n}. \tag{3.2.7}$$

When a $\tilde q \in \mathbf{Q}$ is given, and is to be written as $\exp(\sum t_m \mathbb{X}_m) q_0$, look at the ratios (3.2.7) built from its eigenfunctions to determine what $\mathbf{t}$ must be, and then verify (difficult) that the formula (3.2.3) in fact reproduces the given $\tilde q$.

3.2.1.3. *The Dyson Determinant in Inverse Scattering.* The beautiful representation (3.2.3) of the solution of the system (3.2.2) is universal.

In [10], Henry shows that the commuting Hamiltonian flows $\dot q = (\operatorname{grad} T(i\kappa))'$ from the scattering class retain meaning in $\mathfrak{Q}$, and that their solutions have the form (3.2.3), but with a regularized continuum limit of the $\mathbb{Z}_{\ge 0}\times\mathbb{Z}_{\ge 0}$ determinant.

In the Hill class, there is a similar expression,

$$q(x) = -2\frac{d^2}{dx^2}\ln\Theta(A(x);Z). \tag{3.2.8}$$

[12] [6].

[13] W is the Wronskian.

The function Θ has two arguments. The second one labels the KdV manifolds; Z denotes a Riemann surface of (generically) infinite genus. The first one is a one-parameter subgroup of $\mathfrak{A}$, which is the group of translations on the Jacobian variety of Z. A summary and references may be found in [11].

In the scattering class, the function Θ is the Fredholm determinant[14] of an integral operator,

(3.2.9)
$$\Theta(\theta; Z) = \det\left[I + \frac{1}{2\pi}\int e^{ik(x+y)}R(k)\,dk,\ 0 \le x, y\right].$$

The notation indicates that the dependence of Θ on R should be thought of as split into an angle part (the additions) and an action part (the unimodular invariant), as in (3.1.17). The inversion formula that recovers the potential from the scattering data again has the form (3.2.8), with $A(x) = \arg(\exp(ikx)R(k))$.

3.2.2. The Unimodular Class.

3.2.2.1. *Unimodular: Definition.* We now turn to the definition of KdV manifolds for operators in the class $\mathfrak{Q}$. By the spectral theorem, all self-adjoint Q give rise to an expansion in eigenfunctions, generalized or square integrable. The measure that appears in the inversion formula will be the replacement for the scattering matrix.

Fix an operator[15] $Q_0 = -D^2 + q_0$, and let $E_0(x, \lambda)$ be the fundamental system satisfying $[E_0, E_0'] = 2 \times 2$ identity at $x = 0$. In the resolution of the identity[16]

$$\delta(x - y) = \frac{1}{2\pi}\int_0^\infty E_0(x, \lambda)^\ddagger d\mathsf{F}_0(\lambda) E_0(y, \lambda),$$

the matrix $d\mathsf{F}_0(\lambda)$ is the *spectral weight*. If Q_0 and Q_1 are unitarily equivalent, their spectral weights are related by

$$d\mathsf{F}_1(\lambda) = \mathsf{G}(\lambda)\, d\mathsf{F}_0(\lambda)\mathsf{G}(\lambda)^\ddagger. \tag{3.2.10}$$

Conversely, this relation between spectral weights implies unitary equivalence.

Now define the *unimodular class* $\mathbb{U}(Q_0)$ to be the set of all $Q = -D^2 + q$ for which $\mathsf{G}(\lambda) \in SL(2)$, or equivalently, for which $\det d\mathsf{F} = \det d\mathsf{F}_0$.

The unimodular classes are Henry's substitute for the KdV manifolds. Some motivation is suggested in the next article.

For this proposal to be sensible, a unimodular class must coincide with a KdV manifold in the scattering case; i.e. if Q_0 has transmission coefficient T_0, then

$$\mathbb{J}(T_0) = \mathbb{U}(Q_0)$$

should be true. Now, for Q of scattering class the spectral weight can be computed in terms of the scattering matrix; in particular,

$$\det d\mathsf{F}/d\lambda = |T|^2. \tag{3.2.11}$$

For two operators Q_0, Q_1, relation (3.2.10) implies $|T_1|^2 = (\det \mathsf{G})^2 |T_0|^2$; if $Q_1 \in \mathbb{J}(Q_0)$, then $T_1 = T_0$ and so $\det \mathsf{G} = 1$. Hence a KdV manifold is contained in a unimodular class. The converse is much harder and is addressed in Sect. 3.2.4.

3.2.2.2. *Unimodular: Interpretation.* The condition $\det \mathsf{G} = 1$ is not as different from the parametrization of KdV manifolds by transmission coefficients as it appears on first glance. All operators of scattering class are unitarily equivalent, with identical continuous spectrum $\{\lambda \ge 0\}$. For every pair there is a scattering matrix $S(\lambda; Q_0, Q_1)$ that encodes the relation between their plane wave solutions ($\lambda = k^2$). When $Q_0 = -D^2$, this is our standard scattering matrix (3.1.5).

The scattering matrix relating operators in the same KdV manifold is unimodular: $\det S(\lambda; Q_0, Q_1) = 1$. This property has a very pretty and suggestive translation into an eigenvalue perturbation picture.

Let $d\mathfrak{P}^1_\lambda$ and $d\mathfrak{P}^0_\lambda$ be the spectral projections of $Q_1 = -D^2 + q_1$ and $Q_0 = -D^2$. The increment (Tr is the operator trace)

$$d\xi(\lambda) := \mathrm{Tr}(d\mathfrak{P}^1_\lambda - d\mathfrak{P}^0_\lambda)$$

should be the difference between the number of eigenvalues of Q_0, Q_1 in an infinitesimal interval about ω. This trace does not exist, but one can convert the formula into something meaningful. The *spectral shift function* $\xi(\lambda; Q_0, Q_1)$, introduced by Kreĭn, quantifies the displacement of "virtual eigenvalues" of

[14] Introduced into inverse scattering by Dyson.
[15] Note that Q_0 is not $-D^2$.
[16] The dagger $\ddagger$ denotes transpose.

Q_0 to "virtual eigenvalues" of Q_1. The scattering matrix $S(\lambda; Q_0, Q_1)$ instead looks at the "rotation" of the eigenfunctions at *fixed* eigenvalue λ. A general theorem of Birman and Kreĭn relates the two:

$$\det S(\lambda; Q_0, Q_1) = \mathrm{e}^{-2\pi i \xi(\lambda; Q_0, Q_1)}. \tag{3.2.12}$$

The determinant of the standard scattering matrix (3.1.5) is $T(k)/\overline{T}(k)$, and is the same throughout a KdV manifold $\mathbb{J}$. Then according to (3.2.12), the spectral shift from $-D^2$ remains the same as well. Put differently, the operators in $\mathbb{J}$ have identical "virtual" eigenvalues, and the pairwise scattering matrices are unimodular.

The condition $\det \mathsf{G} = 1$ similarly restricts the deformation of a basis at fixed λ, but I do not know a more transparent interpretation. It would be very pleasant if it were equivalent, for an interesting class of operators, to the rigidity of virtual eigenvalues, whatever that might mean.

3.2.3. Additive Class. The KdV manifolds in the scattering case were interpreted as tori of an integrable Hamiltonian system, labeled by $\det S(k)$. Instead of tori there are now unimodular strata inside a unitary equivalence class, labeled by $\det \mathsf{G}$. The KdV manifolds were swept out by the group of translations acting on the angular coordinate, $\arg R(k)$. We need a replacement for the angles. It was suggested earlier that the unimodular manifolds are generated by the action of a big additive group $\mathfrak{A}$. These are the additions, which now need to be described in a little more detail, first for the scattering class, followed by the generalization.

3.2.3.1. *Additions in Scattering.* Let $\lambda_0 = -\kappa_0^2 < 0$. Let $\mathfrak{p}$ denote the pair $(\lambda_0, +)$ or $(\lambda_0, -)$, and let $e(x, \mathfrak{p})$ be f_+ or f_- in accordance with that choice. Define *addition of* $\mathfrak{p}$ as before,

$$A^{\mathfrak{p}} : Q \mapsto Q^{\mathfrak{p}} := Q - 2(\ln e(x, \mathfrak{p}))''. \tag{3.2.13}$$

Repeated additions change Q by a Wronskian determinant,

$$A^{\mathfrak{p}_1} \cdots A^{\mathfrak{p}_n} Q = Q - 2\frac{d^2}{dx^2} \ln W(e(x, \mathfrak{p}_1), \ldots, e(x, \mathfrak{p}_n)).$$

From this it is clear that additions commute and that $A^{\mathfrak{p}} A^{-\mathfrak{p}} =$ identity.

Additions change R by a factor of modulus 1,

$$A^{\pm\mathfrak{p}} : R(k) \mapsto \frac{\kappa \mp ik}{\kappa \pm ik} R(k), \ \lambda = -\kappa^2; \tag{3.2.14}$$

they preserve $\mathbb{J}(|R|)$ and translate the coordinate $\arg R$ by an odd function.

An *infinitesimal addition* is the derivative of $A^{\mathfrak{p}}$ with respect to $\mathfrak{p}$. Set $\mathfrak{p} = (\lambda, +)$ and $\mathfrak{p}' = (\lambda + \Delta\lambda, +)$ and work out $A^{\mathfrak{p}'} A^{-\mathfrak{p}} Q$. The $\mathfrak{p}'$ factor has an f_+, the $-\mathfrak{p}$ has an f_-, and together they give $f_+ f_-$, which is Green's function:

$$A^{\mathfrak{p}'} A^{-\mathfrak{p}} Q = Q - 2G'_{xx}(\lambda | Q)\Delta\lambda + \cdots. \tag{3.2.15}$$

It is now clear why infinitesimal additions are so important: they are precisely the Hamiltonian vector fields $(\operatorname{grad} T(i\kappa))'$, which were denoted by $\mathbb{X}_\lambda$ in (3.1.18) (up to irrelevant factor). They span the tangent spaces to the KdV tori (redundantly), and generate the big additive group $\mathfrak{A}$ that acts on the tori. Integration of an infinitesimal addition vector field does not produce a global addition (3.2.14). To get one of those, one must piece together segments of integral curves of many $\mathbb{X}_\lambda$, along the lines of

$$\mathrm{e}^{\int t(\lambda)\mathbb{X}_\lambda \, d\mu(\lambda)} q_0.$$

3.2.3.2. *Additions in General.* No major adjustments are needed to define additions for general Q because there are natural substitutes for $f_\pm$. Take $\lambda_0 < 0$, to the left of the spectrum of Q. There exist solutions $h_\pm$ of $Qy = -\lambda^2 y$ characterized by[17]

$$h_\pm \in L^2(\mathbb{R}_\pm) \text{ and } h_\pm \notin L^2(\mathbb{R}_\mp)$$

and a normalization at $x = 0$. In the scattering class, $h_\pm$ are proportional to $f_\pm$. The vector fields of infinitesimal addition, $\mathbb{X}_\lambda := -2G'_{xx}(\lambda|Q)$, are everywhere defined, since every Q has a Green function; they include KdV when it makes sense, because the KdV vector field appears in the asymptotic expansion of $-2G'_{xx}(\lambda|Q)$; they preserve the unimodular class (this is not as obvious as in the scattering case, cf. (3.2.14)); and finally, remarkably,

[17] $\mathbb{R}_+ = [0, \infty)$ etc.

the flows of $\mathbb{X}_\lambda$ exist for all time, as mentioned in Sect. 3.2.1.3. More on this below.

3.2.4. Unimodular Class = Additive Class. Additions will be an acceptable substitute for the nonexistent angles only if every unimodular class $\mathbb{U}(Q)$ is the closure[18] of an orbit of $\mathfrak{A}$. Such a set is called the *additive class of Q* [7], written $\mathbb{A}(Q)$.

Conjecture: $\mathbb{U}(Q_0)$ = unimodular class = additive class = $\mathbb{A}(Q_0)$.

The conjecture is verified for three examples in [8]. I will indicate briefly how one goes about proving it in one case, just to show how intricate and difficult such an argument turns out to be.

Fix a Q_0 of scattering class. Additions preserve $\mathbb{J}(|R_0|)$ and generate all $\arg R$. Hence $\mathbb{A}(Q_0) = \mathbb{J}(|R_0|)$. It is to be shown that the unimodular class is not larger. Suppose then that $Q = -D^2 + q \in \mathbb{U}(Q_0)$; we know nothing about it, other than $\det \mathsf{G} = 1$, so for instance the potential q could be unbounded. It is to be shown that in fact $q \in C^\infty_\downarrow$, which implies that Q belongs to $\mathbb{J}(|R_0|)$.

Since q may not decay, one cannot normalize eigenfunctions at infinity, and works with $h_\pm$ instead. There is a still a kind of scattering matrix connecting the $x < 0$ behavior with the $x > 0$ behavior:

$$\begin{aligned} \overline{h}_+ &= r_{11}h_- + r_{12}h_+ \\ \overline{h}_- &= r_{21}h_- + r_{22}h_+. \end{aligned} \tag{3.2.16}$$

To prove that Q is of scattering class, one must use the r_{ij} to define functions R and T that possess the numerous decay and analyticity properties required of realistic scattering data, use them to get an operator $\tilde{Q}$ by the Dyson determinant formula (3.2.9), and then verify that the $\tilde{Q}$ so constructed is the same as the starting Q.

If Q were of scattering type, the $f_\pm$ (which we don't really have) and the $h_\pm$ (which we do have) would be proportional, say $f_+ = ch_+$ for an unknown normalizing function $c(k)$; further, there would be four relations between r_{ij} and the scattering coefficients $R_\pm, T$ and the factor $c(k)$. Two of them say that scattering data, if they exist, must satisfy

$$r_{21} = R_-\bar{c}/c \quad\text{and}\quad 2ikr_{22} = |c|^2. \tag{3.2.17}$$

They are to be solved for R_- and c.

The unimodularity condition must of course be used somewhere, and it enters right at the start. The spectral weight has an expression in terms of $h_\pm$, and from it one computes that

$$1 - |r_{12}|^2 = \det \frac{dF}{d\lambda},$$

According to (3.2.11) and unimodularity,

$$\det \frac{dF}{d\lambda} = \det \frac{dF^0}{d\lambda} = |T^0|^2.$$

But because T^0 comes from a scattering class potential, one knows that its modulus belongs to $1 + C^\infty_\downarrow$ (qua function of k), and thus $|r_{12}|^2 \in C^\infty_\downarrow$. If r_{12} passes all the other tests required of scattering data, its rapid decay translates into rapid decay of q (as it would in Fourier theory).

The rest is exploitation of analyticity properties of $h_\pm(0,\lambda)$ and the induced analyticity properties of r_{ij} in the upper $k = \sqrt{\lambda}$ plane. Here is a sample.

The normalizing coefficient $c(k), \operatorname{Im} k > 0$, can be reconstructed from its modulus $|c| = \sqrt{2ikr_{22}}$. Then the Hilbert transform produces the phase $\bar{c}/c$, so that R_- can be determined from the first equation in (3.2.17). Next, one defines $|T|^2 = 1 - |R_-|^2$ and recovers T by the Poisson formula (3.1.7). This has produced candidates R and T, and one now goes on to check a list of properties of realistic scattering data, and so on.

3.2.5. Integration of Smeared Addition Flows. The vector fields of infinitesimal addition were claimed to be superior to the more famous KdV hierarchy, because the latter might fail to exist, while the former was of controllable size. This assertion is proved to be true in [10]. Henry considers not only $\dot{q} = -2G'_{xx}(\lambda|Q)$, the infinitesimal addition, but more generally $\dot{q} = -2H'_{xx}(\lambda|Q)$, where H_{xy} is the kernel of a fairly general function $H(Q)$ of Q, and he is able to write global solutions as regularized determinants.

[18] There is not yet a precise definition of "closure".

These determinants are honest versions of the symbolic superposition of paired additions in (3.2.1). We first reinterpret the matrix (3.2.3) as a Fredholm determinant. Take[19] $f = c_0e_0 + c_1e_1 + c_2e_2 + \cdots$, and map it by the row-wise action on the coefficients

$$c_k \mapsto c_k + (\mathrm{e}^{t_k} - 1)\sum_{\ell-0}^{\infty}\int_x^{\infty} e_k e_\ell c_\ell.$$

The effect on f is

$$f \mapsto f + \sum_{k=0}^{\infty}(\mathrm{e}^{t_k} - 1)e_k \otimes e_k P_x f,$$

where $P_x f$ is projection $f \mapsto 1_{[x,\infty)}f$.

If now $H(Q)$ is defined on the spectrum of Q by $H(\lambda_k) = t_k$ then (3.2.3) may be written in the compact form

$$q = q_0 - 2\left(\ln\det[I + (\mathrm{e}^{H(Q)} - I)P_x]\right)''.$$

The formula extends to more general H satisfying certain (natural) technical conditions. The regularized determinant

$$\Theta(H|Q) = \det\frac{I + (\mathrm{e}^{H(Q)} - I)P_x}{I + (\mathrm{e}^{H(Q)} - I)P_0}.$$

is shown to exist and be smooth in x. The vector field[20]

$$\mathbb{X}_H = -2H'_{xx}(Q)$$

has integral curves

$$Q \mapsto Q - 2\frac{d^2}{dx^2}\ln\Theta(H|Q).$$

Ordinary infinitesimal addition at parameter λ' is $-2G_{xx}(\lambda'|Q)$; this corresponds to $H(\lambda) = 1/(\lambda - \lambda')$.

3.3. Finite-Gap Operators

The KdV manifolds $\mathbb{J}(|R|)$, alias additive classes, for operators of scattering type are all infinite dimensional.[21] When is an additive class a finite-dimensional manifold? How do you translate this *geometric property* in the space of all operators into *analytic properties* of individual operators? Remarkably, this question has a complete answer.

The operators are of the type known as *finite gap* or *algebro-geometric*; the names refer, respectively, to a characteristic property of the spectrum of Q and to the techniques by which the explicit form of the potentials is obtained. The connected components of the finite-dimensional additive classes will be products $\mathbb{R}^k \times (S^1)^{g-k}$, the case in "general position" being a pure torus $(S^1)^g$.

Much is known about finite-gap operators. According to McKean-van Moerbeke [15] (this volume) and Its-Matveev [5], these two conditions are equivalent:

(a) The equation $M\Phi = -\Phi''' + 4q\Phi' + 2q'\Phi = 4\lambda\Phi'$ has a solution that is a polynomial of degree g in λ, with x-dependent coefficients;

(b) the spectrum of Q consists of g intervals, with some perhaps shrunk to a point, and one infinite interval extending to $+\infty$; the spectral weight $d\mathsf{F}$ is singular at the points, and $\det(d\mathsf{F}/d\lambda) = 1$ on the intervals;

and, as shown in [8], both are equivalent to

(c) the additive class $\mathbb{A}(Q)$ is a smooth, finite-dimensional manifold of dimension g.

The potential q will then be quasiperiodic in x. It is obtained from a multiply periodic function of g complex variables, say $P(z_1, \ldots, z_g)$, by evaluation along a line $z_j = c_j x + d_j$. This P is an Abel-Jacobi function,[22] meaning that its periods are loop integrals of holomorphic differentials on a Riemann surface.

It is remarkable that such complete information can be deduced from the finite-dimensionality of an additive class, with no assumptions about any operator in the class.[23] Matters were quite different in the scattering setting. It was shown that the additive (= unimodular) class *of a given operator* consists of operators with similar decay properties and spectral invariants,

[19] The superscript zero is dropped for convenience.

[20] H_{xy} is the kernel of the integral operator $H(Q)$.

[21] But for $Q = -D^2$.

[22] In the terminology of Siegel, *Topics in Complex Function Theory, v. 2.*

[23] Other than the standing hypothesis of semi-boundedness and smoothness of q.

not that *an* additive class possessing such-and-such geometric properties was forced to contain only Q of such-and-such type. Additive classes of Hill operators are likewise characterized with reference to a given operator, but in constrast to the scattering case, where the KdV manifolds are still rather amorphous objects, the Hill additive classes have a very clearcut structure. They are generically smooth and compact infinite-dimensional tori admitting a certain kind of complexification; one may conjecture that this geometry forces the potentials to be (evaluations of) Abel-Jacobi functions associated with infinite genus curves.[24] I believe this question has not been investigated.

3.3.1. Jacobi Inversion. Hill operators are a chapter in the theory of the multivalued function $\sqrt{\prod(\lambda-\lambda_m)}$, the product being finite or infinite. The most direct introduction to this circle of ideas starts with condition (a).

So suppose we seek a polynomial solution $\Phi(x,\lambda) = \lambda^g + A_{g-1}(x)\lambda^{g-1} + \cdots$ of $M\Phi = 4\lambda\Phi'$. The coefficients can be determined recursively, down to the constant term, which does not vanish automatically. It will be zero if q satisfies an ordinary differential equation of order $2g+1$. Repeated integrations reduce its order until only g unknowns remain, which are then found by inversion of integrals of algebraic functions.

The steps are implemented by a standard trick. The equation $M\Phi - 4\lambda\Phi' = 0$ is multiplied by Φ to produce a perfect x-derivative (of the left side in (3.3.1) just below). The integration constant is a polynomial in λ, independent of x, that may be prescribed arbitrarily. So we have

(3.3.1)
$$\Phi\Phi'' - \frac{1}{2}(\Phi')^2 - 2(q+\lambda)\Phi^2 = \ell(\lambda)^2 = -2\prod_0^{2g}(\lambda-\lambda_j).$$

The g roots of $\Phi(x,\lambda) = \prod_1^g(\lambda-\mu_j(x))$ are now introduced as new variables. Substitute the product into (3.3.1) and successively set $\lambda = \mu_j(x)$. The resulting equations,

(3.3.2)
$$\frac{1}{2}(-\mu_j'(x)\prod_{i,\neq j}(\mu_i(x)-\mu_j(x)))^2 = \ell(\mu_j(x))^2,$$

are then solved for μ_j'. Because of the ambiguity of sign of the square root, there must be a signature to indicate the sheet on which μ_j is found. Write $\mathfrak{p}_j = (\mu_j, \sqrt{\Delta(\mu_j)^2-1})$. Now (3.3.2) can be organized into a system of Abelian integrals

$$\text{(3.3.3)} \qquad \sum_{j=1}^g \int_{\mathfrak{p}^0}^{\mathfrak{p}_j} \omega_k = \zeta_k,\ \ k=0,\dots,g-1,$$

where $\omega_k = \frac{s^k\,ds}{\ell(s)}$ and $\zeta_k = 2x\delta_{k,g-1}$. The potential q can be found from the so-called trace formula,

$$\text{(3.3.4)} \qquad q(x) = \sum_{m=0}^{2g}\lambda_m - 2\sum_{j=1}^g \mu_j(x),$$

which follows from comparison of the λ^{2g} coefficients in (3.3.1), provided we are able to express the symmetric function $\sum\mu_j$ of the upper limits in terms of the right side. Riemann's theory of his theta function was created to solve precisely this *Jacobi inversion* problem. A. Its and V. B. Matveev exploited the classical theory to derive a representation for q:

(3.3.5)
$$q(x) = \text{const} - 2\frac{d^2}{dx^2}\left(\ln\theta(Vx+W)\right).$$

Some elaboration will come later; for now, it is enough to know that θ is an entire function on $\mathbb{C}^g$, and that $V, W \in \mathbb{C}^g$ are parameters depending on the roots of $\ell(\lambda)^2$ and the initial values.

3.3.2. The Unimodular and Additive Finite-Gap Classes. The theta function vanishes on a $(g-1)$-dimensional set. If the λ_m and the initial conditions $\mu_j(0)$ are chosen carelessly, then q will have poles. The correct disposition of these parameters is identified by use of the spectral theory of the operator Q with smooth potential q. Here are the essentials.

[24] See [11] and references therein.

The spectrum of Q is determined by the λ_m. It consists of g possibly degenerate finite intervals called bands,[25] $[\lambda_{2j}, \lambda_{2j+1}]$, which are separated by g "gaps" $[\lambda_{2j+1}, \lambda_{2j+2}]$, for $j = 0, 1, \ldots, g-1$, and to the right an infinite band $[\lambda_{2g}, \infty)$. The bands that are shrunk to a point form the discrete spectrum. Gaps are assumed to be nondegenerate. They are the branch cuts of the algebraic function $\sqrt{\Delta(\lambda)^2 - 1}$.

The unimodular class $\mathbb{U}(Q)$ is characterized by the band structure and the nature of the spectral weight $d\mathsf{F}(\lambda)$. The latter is computed (by general spectral theory) to be singular on degenerate bands, absolutely continuous with $\det(d\mathsf{F}/d\lambda) = 1$ on the other bands, and zero elsewhere.

The $\mathfrak{p}_j$ provide coordinates on $\mathbb{U}(Q)$. There is exactly one per gap. As function of x, each moves in a circle around the branch cut, turning around when it hits the bordering band; if that band happens to be degenerate, then $\mathfrak{p}_j(x)$ moves towards it in infinite time.

Thus, the general $q(x)$ is built from bounded oscillations of different frequencies, and localized waves that asymptote to a constant at $|x| = \infty$. As manifold, this set of Q is diffeomorphic to a disjoint union of cylinders over tori, $\mathbb{R}^k \times (S^1)^{g-k}$.

3.3.3. Finite-Dimensional Additive Classes. Now suppose there is an additive class $\mathbb{A}$, of operators with yet unknown properties, that is a finite-dimensional manifold.

Henry proves this striking result: the class $\mathbb{A}$ is a manifold of operators as just described, and therefore the potentials are expressed in terms of Abelian functions by the Its-Matveev formula.

I want at least to indicate how such a proof gets under way. What is the immediate consequence of finiteness?

One has to get hold of $d\mathsf{F}$ somehow. It is obtained as boundary value of a matrix $M(\lambda)$, the *Weyl matrix*, built from the solutions $h_\pm$ introduced earlier:

$$d\mathsf{F}(\lambda = a) = \lim_{\epsilon \downarrow 0} \operatorname{Im} M(a + i\epsilon)\, da.$$

[25]The term comes from solid state physics.

The $(1,1)$ entry of M is the most important. It is $m_{11} = 2h_- h_+$ at $x = 0$, so $G_{00}(\lambda|Q)$ up to factor, and $d\mathsf{F}_{11}$ is essentially a jump in Green's function.

Since $\mathbb{A}$ has dimension g, any $g+1$ infinitesimal additions are linearly dependent; if $\mathbb{X}_k$ is addition with λ'_k, then $\mathbb{X} := \sum_0^g c_k \mathbb{X}_k = 0$. Next, $\mathbb{X} m_{11}(\lambda)$ is computed, and its vanishing leads to identities on the Weyl matrix that imply

$$\sqrt{d\mathsf{F}(\lambda)} = \sqrt{r(\lambda)} \times d\mathsf{F}_{11},$$

where r is a rational function of the form $[\sum_0^g \frac{a_k}{\lambda - \lambda'_k}]^{-1}$. (The fact that $\mathbb{X}_k = -G'_{xx}(\lambda'_k|Q)$ is reflected in the factor $1/(\lambda - \lambda'_k)$.)

The proof continues with careful exploitation of the properties of $\sqrt{r(\lambda)}$, which follow from known facts about the behavior of $h_\pm$. Sign changes under the radical introduce the band-gap structure. One then makes contact with Sect. 3.3.1 by deriving a representation

$$2h_+ h_- = m_{11} = \frac{\Pi_1^g (\lambda - \mu_j)}{\ell(\lambda)},$$

with one μ_j per gap. These are the $\mu_j(x)$ from earlier, evaluated at $x = 0$. They can be moved to arbitrary positions in the gaps by the flows of vector fields of infinitesimal addition; by this procedure one can reach the whole set of quasiperiodic potentials contained in the trace formula (3.3.4).

3.3.4. Hill Operators. The extension of the finite-gap picture to operators with arbitrary smooth periodic potential requires substantial new ideas. There are now infinitely many bands and gaps, infinitely many $\mathfrak{p}_j$, and a theta function in infinitely many complex variables, see [11] for more information.

The conjecture $\mathbb{U}(Q_0) = \mathbb{A}(Q_0)$ is verified in [8]. The unimodular class is characterized by absolutely continuous spectrum on a set of intervals (of certain asymptotic spacing and length), on which $\det d\mathsf{F} = 1$. As in Sect. 3.2.4, suppose $Q \in \mathbb{U}(Q_0)$. At first, nothing–aside from smoothness and semicoundedness–is assumed about Q. Combining function theory and spectral theory in a masterful way, Henry succeeds in proving that Q is a Hill operator, which means: it has a period 1 potential whose divisor $\{\mathfrak{p}_j\}$ is distributed one per gap.

3.3.5. Other Operators. Modifications and combinations of inverse scattering and the band-gap spectral theory have been worked out for potentials with rapid decay to nonzero constants at $|x| = \infty$ and for rapidly decaying perturbations of finite-gap potentials, but not, I believe, for local perturbations of general Hill operators. The characterization of unimodular and additive classes for potentials of the first two types would be of interest; perhaps the requisite techniques can be extracted from [8]. In [7], Henry makes brief mention of preliminary results about operators with almost periodic potentials. One may expect that progress, for instance a log det representation of the potentials, characterization of unimodular classes, equality of unimodular and additive, etc., will require entirely new ideas.

Afterword

Early in my years at Arizona, a group of enthusiastic aspiring probabilists organized an ambitious reading seminar, meeting, I recall, twice a week without fail. During one particularly intense semester, we worked through every line of Henry's recently published *Stochastic Integrals*. My colleagues continued to careers in probability and statistics, whereas I was introduced to, and seduced by, integrable systems. After my U-turn away from stochastic matters, I published a note about Hill's equation. Just in time to insert an "added in proof" amendment, I discovered the paper "The Spectrum of Hill's Equation", and was astonished to find that Henry had made the U-turn with me (better, I made it with him). Thereupon, during another intense semester, I worked through every line of that paper. Over decades since then, I have continued to learn from Henry, sometimes in person, more often by poring over his writings. I am glad to have been given a reason to study the KdV papers, which I really did not appreciate 25 years ago. Today, I am closer to the goal. It is always exciting to come to understand his ideas, even if imperfectly and slowly. Thank you, Henry, for making mathematics so much fun.

Appendix: Hamiltonian Mechanics

The prototypical Hamiltonian system in Euclidean space has the form

$$\dot{\mathbf{x}} = P \operatorname{grad} H(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^N, \tag{A-1}$$

where P is a constant, skew-symmetric linear operator called *Poisson operator* and H is a given function called the *Hamiltonian* of (A-1). For simplicity, we assume P to be invertible, which forces N to be even, say $N = 2n$.

The vector field defined by the right side of (A-1) is written $\mathbb{X}_H$. The commutator of Hamiltonian vector fields is again Hamiltonian, $[\mathbb{X}_H, \mathbb{X}_G] = \mathbb{X}_F$, where F, called *Poisson bracket* of G and H, is given by the dot product

$$F = \{G, H\} := \operatorname{grad} G \cdot P \operatorname{grad} H.$$

When $\{G, H\} = 0$, one says that G, H are *in involution* or *Poisson commute*. Equivalently, $[\mathbb{X}_G, \mathbb{X}_H] = 0$, which also implies that G is constant along integral curves of $\mathbb{X}_H$, and vice versa.

Suppose $C_1, \dots, C_n$ are (independent) functions in involution; there can be no more because their gradients span an isotropic subspace of P. The commuting vector fields $\mathbb{X}_{C_j}$ leave the n-dimensional level manifolds[26] $\mathbb{J}(\mathbf{c}) := \cap_j \{C_j = c_j\}$ invariant. If these are compact they must be tori. After a choice of origin, the flows of $\mathbb{X}_{C_j}$ define a transformation group $(t_1, \dots, t_n) \mapsto \mathbf{x}(t_1, \dots, t_n)$ on each torus.

On an open set filled by n-dimensional tori $\mathbb{J}(\mathbf{c})$ one can introduce special coordinates, the *action-angle coordinates*, also called *action-angle variables*, Z_j, θ_j. The actions Z_j are functions of the c_k and label the tori, and the angles go around the cycles of the torus. These coordinates are *canonical*, meaning that all Poisson brackets vanish except for $\{Z_j, \theta_1\} = 1$.

[26] $\mathbf{c} = (c_1, \dots, c_n\}$ is given.

If a function H depends only on the actions Z_j, the Hamiltonian system takes the canonical form

$$\dot{\theta}_j = \frac{\partial H}{\partial Z_j},\ \dot{Z}_j = -\frac{\partial H}{\partial \theta_j} = 0, \tag{A-2}$$

and since $\dot{\theta}_j$ is a constant, call it ω_j, the solution is

$$Z_j(t) \equiv Z_j(0),\ \theta_j(t) = \omega_j \cdot t + \theta_j(0).$$

Thus, in action-angle coordinates, the integral curves of the vector field $\mathbb{X}_{\tilde{H}}$ are exhibited as straight lines on the covering spaces of the tori.

Bibliography

[1] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur. The inverse scattering transform. Fourier analysis for nonlinear problems. *Stud. Appl. Math.*, 53:249–315, 1974.

[2] H. Airault, H. P. McKean, and J. Moser. Rational and elliptic solutions of the Korteweg-de Vries equation and a related many-body problem. *Comm. Pure Appl. Math.*, 30:95–148, 1977.

[3] N. Ercolani and H. P. McKean. Geometry of KdV (4). Abel sums, Jacobi variety, and theta function in the scattering case. *Invent. Math.*, 99(3):483–544, 1990.

[4] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Method for solving the Korteweg-de Vries equation. *Phys. Rev. Letters*, 19:1095–1097, 1967.

[5] A. R. Its and V. B. Matveev. Hill operators with a finite number of lacunae. *Funkcional. Anal. i Priložen.*, 9(1):69–70, 1975.

[6] V. B. Matveev. *Darboux Transformations in Nonlinear Dynamics.* Springer Series in Nonlinear Dynamics. Springer-Verlag, 1991.

[7] H. P. McKean. Geometry of KdV (1): addition and the unimodal spectral class. *Rev. Math. Iberoamer.*, 2:235–261, 1986.

[8] H. P. McKean. Geometry of KdV (2): three examples. *J. Stat. Phys.*, 46:1115–1143, 1987.

[9] H. P. McKean. Is there an Infinite-Dimensional Algebraic Geometry? Hints from KDV. In L. Ehrenpreis and R. C. Gunning, editors, *Proceedings of Symposia in Pure Mathematics*, volume 49(1), pages 27–37. American Mathematical Society, 1989.

[10] H. P. McKean. Geometry of KdV (3): Determinants and Unimodular Isospectral Flows. *Comm. Pure Appl. Math.*, 45:389–415, 1992.

[11] H. P. McKean. Book reviews: *Riemann surfaces of infinite genus* by J. Feldman, H. Knorrer and E. Trubowitz. *Bull. Amer. Math. Soc.*, 42:79–87, 2005.

[12] H. P. McKean and E. Trubowitz. Hill's operator and hyperelliptic function theory in the presence of infinitely many branch points. *Comm. Pure Appl. Math.*, 29:143–226, 1976.

[13] H. P. McKean and E. Trubowitz. Hill surfaces and their theta functions. *Bull. Amer. Math. Soc.*, 84:1042–1085, 1978.

[14] H. P. McKean and E. Trubowitz. The spectral class of the quantum-mechanical harmonic oscillator. *Comm. Math. Phys.*, 82(4):471–495, 1981/82.

[15] H. P. McKean and P. van Moerbeke. The spectrum of Hill's equation. *Inv. Math.*, 30:217–274, 1975.

[16] V. E. Zakharov and L. D. Faddeev. The Korteweg-de Vries equation: a completely integrable Hamiltonian system. *Funct. Anal. Applic.*, 5:280–287, 1971.

4 Some Comments

Henry P. McKean Jr.[1]

4.1. Options

McK [49, 'A free boundary problem for the heat equation arising from a problem in mathematical economics'] is my one and only excursion into "mathematical finance". It is an appendix to P. Samuelson [90, 'Rational theory of warrant pricing'] in which the correct recipe for pricing an American put option is worked out. My contribution was to reduce the recipe to a free boundary problem for a (backward) diffusion equation related to the "geometrical Brownian motion" $x(t) = \exp[\sigma b(t) + \delta t]$ and to compute, concretely, as much as I could. I believe this the worst written paper I ever did (though the version printed here has been cleaned up) and refer you to P. Myneni [83, 'The pricing of an american option'] for a much better account. Anyhow, what we did was a novelty then (1965), so I include it here. The work of Black-Scholes [4, 'The pricing of options and corporate liabilities'] and Merton [82, 'Theory of rational option pricing'] is similar, but with them the derivation of the mathematical problem is done, more transparently, in terms of the individual Brownian path (Itô's Lemma) and the principle of "no free lunch", and would seem to lie deeper from a financial point of view. This is an important difference.

4.2. Geometry and the Laplacian

The main paper here, McK-Singer [76, 'Curvature and the eigenvalues of the Laplacian] has to do with the spectrum of the Laplacian on a manifold. Kac's celebrated paper [32, 'Can you hear the shape of a drum?'] got me started. There, he had made a beautiful proof of H. Weyl's estimate for the eigenvalues $0 > \lambda_1 \geq \lambda_2 \geq \lambda_3$ etc.[2] of the ordinary Laplacian acting on functions that vanish at the boundary of a nice domain $D \subset \mathbb{R}^d$ of volume V:

$$-\lambda_n \simeq \left(2\pi\left(\frac{d}{2}\right)!\right)^{d/2} \times \left(\frac{n}{V}\right)^{2/d} \quad \text{for } n \uparrow \infty,$$

or what is the same,

$$Z = \text{trace } e^{t\Delta} = \sum_{1}^{\infty} \exp(t\lambda_n)$$
$$\simeq (4\pi t)^{-d/2} \times V \text{ for } t \downarrow 0.$$

This he made obvious: The Brownian motion, started inside D and killed at the boundary ∂D, does not know that death is around the corner, so its transition density $p(t, x, y)$, expressive of $e^{t\Delta/2}$, imitates the density for the

[1] Courant Institute, New York University, New York, NY, USA, mckean@cims.nyu.edu.

[2] The numbers $w = \sqrt{-\lambda}$ are the "fundamental tones" of a drum-head spanning ∂D, whence Kac's title.

F.A. Grünbaum et al. (eds.), *Henry P. McKean Jr. Selecta*, Contemporary Mathematicians, DOI 10.1007/978-3-319-22237-0_4

free Brownian density on the diagonal, i.e. inside D, $p(t,x,x) \simeq (2\pi t)^{-d/2}$ for $t \downarrow 0$. But then $Z = \int p(2t,x,x)dx \simeq (4\pi t)^{-d/2} \times V$. That's it! Kac went further in dimension 2, conjecturing that for smooth ∂D,

$$Z = \frac{\text{area } D}{4\pi t} - \frac{\text{length } \partial D}{\sqrt{4\pi t}} + \frac{1}{6}(1 - \text{the number of holes}) + o(1).$$

This is true. Singer and I proved it together with a general extension to manifolds of higher dimension, both closed and with boundary, and to other elliptic operators. Kac is asking: Does the spectrum of Δ determine the region, up to a rigid motion, say? His hunch was *no*, and that's correct. Very different iso-spectral regions were found by Gordon, Webb and Wolpert [23, 'Isospectral plane domains and surfaces via Riemannian manifold']; see also Buser, Conway, Doyle and Semmler [7, 'Some planar isospectral domains'] for a whole gallery of such, and Sarnak [91, 'Determinants of Laplacians, heights and finiteness'] who proves that isospectral classes of (compact) manifolds are compact. Now the $\frac{1}{6}(1-h)$ in Kac's formula can be traced back to the Gauss-Bonnet formula for the manifold (with crease) made by doubling D: indeed, you can "hear" the Euler number of any closed manifold, as de Bruijn and Arnold had already proved but not published. It appears in the form of an integral over the manifold of a complicated expression in the curvature, hopefully reducible to Chern's integrand, but we could not see into the necessary cancellations. Patodi [85, 'Curvature and the eigenforms of the Laplace operator'] did so; see Gilkey [21, 'Curvature and eigenvalues of the Laplacian for elliptic operators'] for a much simplified version and more; see also Sakharov [104] where the method is applied to the physical problem of polarization.

McK [53, 'Selberg's trace formula as applied to a compact Riemann surface'] is mostly expository. It explains the application of the marvelous trace formula of Selberg [94, 'Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series'] to a compact Riemann surface $\mathbf{X}$ of genus $g \geq 2$, equipped with its natural Laplacian. For genus 1, $\mathbf{X}$ is the quotient of the plane by a lattice $\mathbb{Z} + \omega\mathbb{Z}$ with $\omega = a + \sqrt{-1}b$ in the upper half-plane, and Poisson's summation formula shows that you can hear the area b and also $a^2 + b^2$, so that the fundamental cell is known up to a reflection $a \to -a$ and $\mathbf{X}$, itself, is known up to a triviality. Selberg's formula reveals that knowing spec Δ is the same as to know, in each (free) deformation class of closed paths of $\mathbf{X}$, the length of the shortest one. Remarkable! Then I could prove with help of Fricke-Klein that any isospectral class is finite, improving upon a remark of Gelfand [20, 'Automorphic functions and the Theory of Representations']. Later M. F. Vigneras [99, 'Variétés riemanniennes isospectrales et non isométriques'] produced conformally inequivalent isospectral surfaces, showing that my result can't be improved. There is a blunder on pp. 236–237 which purports to prove that the top eigenvalue is $\leq -1/4$. This is not so, as many more knowledgeable people told me: already for $g = 2$, it can be as close to 0 as you like; see [93, 'Geometric bounds on the low eigenvalues of a compact manifold'].

McK [52, 'An upper bound of the spectrum of Δ on a manifold of negative curvature'] is, in fact, an afterthought to [53, 'Selberg's trace formula…']. There, I had noticed that for smooth, compact f on the half-line $y > 0$,

$$\int f'^2 \int \frac{f^2}{y^2} \geq \left(\int \frac{ff'}{y} \right)^2 = \frac{1}{4} \int \left(\frac{f^2}{y^2} \right)^2,$$

from which it is easy to see that, if f is a smooth compact function on the hyperbolic upper half-plane, then

$$\frac{1}{4} \int\int f^2 \frac{dxdy}{y^2} \leq \int\int f\,(-\Delta f)\,\frac{dxdy}{y^2}$$

with the hyperbolic Laplacian $\Delta = y^2(\partial^2/\partial x^2 + \partial^2/\partial y^2)$, *i.e.* spec $\Delta \leq -1/4$. Now the top of spec Δ regulates how fast the Brownian motion runs off to ∞, as does the curvature, so I wondered how general this type of thing might be. McKean [52, 'An upper bound of the spectrum ...'] is a home-made proof that if M is a smooth, simply-connected Riemannian manifold with all its sectional curvatures less than a fixed number $-k < 0$, then spec $\Delta \le \frac{1}{4}(n-1)^2 k$. R. Bishop kindly sent me the much nicer proof which I reproduce here. My ignorance of differential geometry being all but complete, I did not know that on such a manifold with its geodesic polar coordinates, the determinant g of the fundamental form satisfies

$$\frac{\partial}{\partial r} \ln \sqrt{g} \ge (n-1)\left(\kappa \frac{\cosh \kappa r}{\sinh \kappa r} - \frac{1}{r}\right)$$
$$\text{with } \kappa^2 = -k.$$

Notice that the right side is $\simeq (n-1)\kappa$ if the origin is far away. Take f smooth and compact and place the origin far from where f lives. Then

$$\int_0^\infty \left(\frac{\partial f}{\partial r}\right)^2 \sqrt{g}\, dr \int_0^\infty f^2 \sqrt{g}\, dr$$
$$\ge \frac{1}{4}\left(\int f^2 \frac{\partial \sqrt{g}}{\partial r} dr\right)^2 \quad \text{much as before}$$
$$\simeq \frac{1}{4}(n-1)^2 \kappa^2 \left(\int_0^\infty f^2 \sqrt{g}\, dr\right)^2.$$

Now proceed, canceling one $\int f^2 \sqrt{g}$, integrating out the angles, and so on.

4.3. Geometry and ODEs

McK-Scovel [75, 'Geometry of some simple non-linear differential operators'] concerns an entirely different sort of geometry. Ambrosetti-Prodi [1, 'On the inversion of some differentiable mappings...'] and Berger-Church [3, 'Complete integrability and perturbation of a nonlinear Dirichlet problem'] had proved that if D is a nice region in $\mathbb{R}^d$, if $-\Delta$ is the standard Laplacian acting on functions vanishing at ∂D with eigenvalues $0 < \lambda_1 < \lambda_2$ etc., and if $K : R \to R$ is convex with $-\infty < K'(-\infty) < \lambda_1 < K'(+\infty) < \lambda_2$, then the map $f \to \Delta f + K(f)$ is a "fold", meaning that, in suitable coordinates $x = (x_1, x_2, \dots)$, it looks like $x \to (x_1^2, x_2, \dots)$. What Clint Scovel and I did was to work out the geometry of the simplest problem of this type when K' crosses the *whole* spectrum of $-\Delta$, namely $d = 1$, $D = (0, 1)$, and $K(f) = f^2/2$, i.e. the map $f \to -f'' + f^2/2$, subject to $f(0) = f(1) = 0$. The energy levels are now $n^2\pi^2 : n \ge 1$ and K' crosses them all. The "singular set" = the locus where the gradient $F = -D^2 + f$ has a null-function is comprised of disjoint sheets $M_n : n = 1, 2, 3$, etc., determined by the vanishing of the successive eigenvalues $\lambda_n(f)$ of F, sitting each below its predecessors. M_1 is convex, the rest have each 1 more principal direction of negative curvature than the one before. M_1 maps to a convex surface and what lies above this image is at once the 1:1 image of what lies above M_1, and the full range of the map. As the image $g = -f'' + f^2/2$ rises, its preimages f proliferate: for example, if f lies below M_n, then there are $2n$ of these or more; in fact, if g is constant and if the lowest f is still above M_{n+1}, then the count is exact. It would be amusing to find a sharp approximate count for $g + c$ and $c \uparrow \infty$. This is not known to me, but you will find the details for $g = 0$ worked out in [75, 'Geometry of some simple ...']. It is remarkable that such counts can be made at all. The trick is the fact that $-f_1'' + f_1^2/2$ and $-f_2'' + f_2^2/2$ coincide only if $f = \frac{1}{2}(f_1 + f_2)$ lies on a singular sheet and $e = \frac{1}{2}(f_1 - f_2)$ is proportional to the singular direction at f. For more information, see McK [61, 'Curvature of an ∞-dimensional manifold related to Hill's equation'] where the sectional curvatures of the image of M_1 are computed pretty explicitly, and also McK [63, 'Geometry of KdV (2): three examples'] which purports to describe the singularities (fold, cusp, and so on) of the singular sheets, but is spoiled by a mistake: on p. 101, the evaluation $Z_{n+1} = 1$ (line 25) is wrong, as B. Ruf notified me.

McK [69, 'A quick proof of Riemann's mapping theorem'] is just what it says. The idea is simple: if you can map the half-line to your favorite polygon, then you will be (almost) home. I learned from M. Hausner and P. Lax that this had been done as early as 1874 by Schläfli [92]. The present proof (just a page and a half) wins in respect to brevity if not priority. The proof is cute and seems to have dropped out of common knowledge.

McK [66, 'How real is resonance?'] suggests that the conventional wisdom, that resonance obstructs the smoothness of the change of coordinates $x \to y$ which converts $x^\bullet = A(x)$ into $y^\bullet = (dA(0) = B)y$ in the vicinity of a resting point $x = 0$, is (sometimes) only a prejudice. The map takes a patch of $\mathbb{R}^d$ onto a patch of an (auxiliary) copy of $\mathbb{R}^d$, but all you really need is a map into some, possibly *non*-linear space that simplifies the flow. Insisting that it be $\mathbb{R}^d$ could *invite* resonance to come in. The example presented is a caricature of Burger's equation $\partial v/\partial t = \partial^2 v/\partial x^2 + 2v\partial v/\partial x$. It reads

$$x^\bullet + \Delta^+ e^x + \Delta^- e^{-x} \simeq \Delta^2 x + \tfrac{1}{2}(\Delta^- + \Delta^+)x^2$$

in which $x = (x_1, \ldots, x_d)$ is considered to be periodic ($x_0 = x_d$ and so on), Δ^+/Δ^- is the forward/backward difference, and $\Delta^2 = \Delta^-\Delta^+$. The Cole-Hopf-type map $x_n \to y_n = \exp(x_1 + \ldots + x_n)$ is not periodic if the mean $m = d^{-1}(x_1 + \ldots + x_d)$ does not vanish but becomes so when regarded as a map to the projective cone $(P\mathbb{R}^d)^+$. It converts the flow into $y^\bullet = \Delta^2 y$, considered projectively. Then the inverse map $\Delta^- \ell n$ washes out any ambiguous projective multiplier, so everything is fine and even as smooth as you could want. What is startling is that no conventional change of coordinates can be smooth, even for $d = 2$. There, the mean $m = \frac{1}{2}(x_1 + x_2)$ is a constant of motion (also for $d \geq 3$) and y_1 is of the form

$$y_1(x) = A(m) + B(m)|\tanh \tfrac{1}{4}(x_1 - x_2)|^{1/\cosh m},$$

which is not even of class C^1 at the diagonal. Here, and more generally, the moral is this: if you want to solve non-linear problems, do not insist upon linear methods. Let the problem tell you what to do.

4.4. Gaussian Processes

About 1950/1, I read (with tears) Wiener's *Cybernetics* [101] about prediction, feed-back, etc. for shift invariant Gaussian processes $\mathfrak{x}(t) : -\infty < t < \infty$. It was way over my head. By 1958 when I came to MIT, I was better educated and understood what he had done.

Here are the basic facts: *either* the remote past determines the whole path $\mathfrak{x}(t)$ *or else* the spectral measure has a density Δ with $\int \ln \Delta'(1 + \gamma^2)^{-1} d\gamma > -\infty$, in which case Δ' factors as $|h|^2$ with a Hardy function $h \in H^{2+}$, and if k is the inverse Fourier transform of h, then $\mathfrak{x}(t) = \int_{-\infty}^t k(t-t')db(t')$ with a standard Brownian motion b; in particular, the field of $\mathfrak{x}(t') : t' \leq t$ is contained in the corresponding Brownian field, and may be made to match it by taking h to be outer, as can always be done. Then the best prediction of the future observation $\mathfrak{x}(T)$ knowing the past $\mathfrak{x}(t) : t \leq 0$ is simply

$$E[\mathfrak{x}(T)| \text{ the past }] = \int_{-\infty}^0 k(T-t)db(t').$$

I wanted to understand more about the past, the future, the projection of the future on the past (splitting), the germ $= \bigcap_{T>0}$ span$[\mathfrak{x}(t) : |t| \leq T]$, etc. and found myself deep into Hardy functions and entire functions of exponential type. Pretty much at sea, I went to Norman Levinson, knowing that he had done an engineer's version of Wiener's recipe and was expert in complex function theory. Look at his proof in [39], that more than 1/3 of the roots of Riemann's zeta function lie on the line 1/2, if you want to see what "expert" means. As I had hoped, he took a fancy to my stuff, and we worked it out in quite a lot of detail in [40, 'Weighted trigonometrical approximation ...'] below. Harry Dym and I made further progress in two papers [12, 'Applications of the de Branges spaces of integral functions to prediction of stationary Gaussian processes'] and [13] with the added help of M.G. Krein's weighted strings, which proved to be just the right tools; see, also, our joint book [14, 'Gaussian processes'] for the complete story (in overwhelming detail). The non-stationary case is *much* much more complicated, but see P. Lévy [42, 'Random functions: general theory ...'] for lots of information about it.

I do want to comment on the curious nature of this whole subject. It's a lot of fun if you like Fourier and complex function-theory but very unsatisfactory from a statistical standpoint. Here's the simplest example of what I'm thinking of. The remote past is trivial if (and only if) $H = \int \ln \Delta'(1+\gamma^2)^{-1}d\gamma > -\infty$; otherwise, the whole path $\mathfrak{x}$ is determined by it. Now $H = -\infty$ if Δ' vanishes far out or decays fast. Then $\mathfrak{x}(t)$ is (real) analytic in $-\infty < t < +\infty$ and the fact is obvious. But what if Δ' vanishes for $|\gamma| \leq 1$, say, or has too hard a root at $\gamma = 0$. H diverges once more, so the prediction of, e.g. $\mathfrak{x}(1)$ from $\mathfrak{x}(t) : t \leq T$ is perfect for any $T \leq 0$, but why? I mean statistically speaking. It is only that the path lacks a small band of frequencies, and how could that matter? Besides, you know how to predict if $H > -\infty$, but nobody really knows how to do it in the contrary case. Worse still, I have to say that the *path-wise* behavior of $\mathfrak{x}$ is practically unknown, but see S.O. Rice [88, 89, 'Mathematical analysis of random noise'] and also Kac and Slepian [33, 'Large excursions of a Gaussian process'].

Back-up to [40] for a bit: Norman and I had asked what kind of Markovian property $\mathfrak{x}$ might have, i.e. how much information is needed to make past and future conditionally independent. This is the "splitting" alluded to before. For example, we found that $\mathfrak{x}$ splits over its germ = the intersection over $T > 0$ of the field of $\mathfrak{x}(t) : |t| \leq T$ if (and only if) $\Delta' = |E|^{-2}$ with an entire function E of minimal exponential type. McK [29] studied splitting for P. Lévy's Brownian motion with a several-dimensional time. This is the Gaussian "field" with mean 0 and correlation

$$E[\mathfrak{x}(a)\mathfrak{x}(b)]=\frac{1}{2}(|a|+|b|-|a-b|) \quad \text{for } a, b \in \mathbb{R}^d.$$

Fix a smooth closed surface dividing $\mathbb{R}^d$ into an inside and an outside and consider the past/future to be all information from the inside/outside. Then, if d is odd, past and future split, conditional upon the knowledge of $\mathfrak{x}$ and its partial derivatives of degree $\leq [d/2] + 1$ taken *on* the surface. Contrariwise, if d is even, no amount of surface or interior information short of the whole will do, *i.e.* $\mathfrak{x}$ has no Markovian property at all. This is a kind of Huygens' principle, conjectured by Lévy, himself; compare G.M. Golchan [22] and, for Gaussian fields more generally, L. Pitt [86].

4.5. Mostly Brownian Motion

This was my first love in mathematics ever since I began to learn probability from M. Kac (MIT, summer 1950) and to read P. Lévy [41], from which I first understood the central place of the individual Brownian path. In 1954/6, Itô and I were mostly occupied with Feller's new picture of 1-dimensional diffusion that he was perfecting just then, and with Dynkin's strict Markov property, which everybody sort of knew in special cases, but was now comprehensively stated and skillfully used by him.

4.6. Local Times

But it was P. Lévy's "mesure du voisinage", translated as "local time" by us, that really caught our eye. This refers to the standard 1-dimensional Brownian motion $\mathfrak{x}(t) : t \geq 0$, starting at $\mathfrak{x}(0) = 0$, say, and is defined as follows:

$$\mathfrak{t}(t,x) = \lim_{h\downarrow 0} \frac{1}{2h} \text{measure } (t' \leq t : x \leq \mathfrak{x}(t') < x + h);$$

see Lévy [41], p. 228. It is the density of the Brownian occupation times:

$$\text{measure } (t' \leq t : a \leq \mathfrak{x}(t') < b) = 2\int_a^b \mathfrak{t}(t,x)dx \text{ for any } a < b,$$

with apologies for the nuisance factor 2. H. Trotter [97] proved its continuity in t and x; see McK [47] for my refinement using Tanaka's (unpublished) formula

$$\mathfrak{t}(t,x) = (\mathfrak{x}(t) - x)^+ - (\mathfrak{x}(0) - x)^+ - \int d\mathfrak{x}(t') \text{ taken for } t' \leq t \text{ with } \mathfrak{x}(t') \geq x.$$

The local time has marvelous properties, many due to Lévy himself, some to Itô and me, deeper ones due to D.B. Ray [87] and F.B. Knight [34] and to D. Williams, the virtuoso of the Brownian path, in [102] and [103]. These deeper results reveal that the local time is Markovian in respect to x(!) quite unforeseen by Itô and myself, though we had a special case in front of us and didn't have the wit to *see* what we were looking at. McK [57] is a comprehensive review of the whole subject.

4.7. Boundary Conditions

It was one of Feller's most intuitive remarks that the infinitesimal operator $\mathfrak{G}$ of a diffusion is really instructions to the particle: *what to do next.* You see at once that boundary conditions are nothing special: They only spell out the instructions at a barrier or at ∞; in short, they are at an aspect of $\mathfrak{G}$ like any other! Then it was a nice problem to find (and to interpret) the boundary conditions for, say, Brownian motion on the half-line $[0, \infty)$, subject to the strict Markov property. Feller [18] had found them all. Inside $[0, \infty), \mathfrak{G}$ is (of course) $\frac{1}{2}D^2$ acting on $C^2(0, \infty)$, but then it must be restricted, as in

$$p_1 f(0) - p_2 f^+(0) + p_3 \frac{1}{2} f''(0+) = \int_0^\infty [f(x) - f(0)] dp_4(x)$$

with non-negative p_1, p_2, p_3, dp_4, subject to

$$p_1 + p_2 + p_3 + \int_{0+}^\infty (x \wedge 1) dp_4(x) = 1.$$

This was easy to understand in simple cases: $f(0) = 0$ means killing, $f^+(0) = 0$ means reflection, $f''(0+) = 0$ means the particle sticks, and the general condition with $(p_4 = 0)$ is a mixture. As to the paths, Feller suggested that the local time at $x = 0$ should enter in the case $p_1 f(0) - p_2 f^+(0) = 0$ with $p_1 p_2 > 0$ (the elastic barrier, so called), namely that the Brownian particle should reflect off $x = 0$ until its local time builds up to the value of an independent exponential holding time, whereupon the particle is killed, and we checked that, Itô and I. The role of p_4 was also easy to understand if $p_4[0, \infty) < +\infty$: Then it is proportional to the distribution of jumps from $x = 0$ back into the interior $x > 0$. But what if $p_4[0, \infty) = +\infty$? That was mysterious. Luckily, Itô saw at once that it must describe "jumps" of a new kind, produced by the increasing "differential" process with infinitesimal operator

$$f \to \int_0^\infty [f(x + h) - f(x)] dp_4(h),$$

and as we were flying one day, to Fukuoka I think, Itô kept drawing pictures, one after another, trying to see how these jumps could be interlaced with the Brownian path. After a while he got it; after a longer while I got it, too, and the rest was plain sailing, described in Itô-McK [29] below.

4.8. Clocks

Another incisive idea of Feller's was to write the infinitesimal operator of a 1-dimensional diffusion in the form $\mathfrak{G} = (d/dm)(d/ds)$. The "scale" s describes exit probabilities: if $\mathfrak{x}(t) : t \geq 0$ is the path and T_x is the passage time $\min(t : \mathfrak{x}(t) = x)$, then

$$P_x(T_b < T_a) = \frac{s(x) - s(a)}{s(b) - s(a)} \quad \text{for any } a < x < b.$$

It is the road map so to say. The "speed measure" m tells how fast you go:

$$E_x(T_a \wedge T_b) = \int_a^b G(x, y) dm(y)$$

with the (symmetric) Green's function $G(x, y) = (s(x) - s(a)) \times (s(b) - s(x))$ over $s(b) - s(a)$. Now, in that scale s, $\mathfrak{x}$ has the same road map as the Brownian motion, and if you reduce to this case $(s = x)$ and suppose that m has a smooth, positive density $2/\sigma^2$, then $\mathfrak{G}$ appears as $\frac{1}{2}\sigma^2 \frac{d^2}{dx^2}$, i.e. the "drift" is removed. Think next, for the utmost simplicity, that σ^2 is a constant $c^2 > 0$. Then $x = c\mathfrak{n}$ with a standard Brownian motion $\mathfrak{n}$ or, what is the same by Brownian scaling, $\mathfrak{x}(t) = \mathfrak{n}(c^2 t)$ with a new Brownian motion $\mathfrak{n}$, suggesting that the general $\mathfrak{x}$ is just Brownian motion run with a new "clock". With the help of Hale Trotter, Itô and I found the right recipe: $\mathfrak{x}(t) = \mathfrak{n}[T^{-1}(t)]$, in which T^{-1} is the clock inverse to $T(t) = \int \mathfrak{t}(t, x) dm(x)$ or, what is the same, $T(t) = 2 \int_0^t \sigma^{-2}[\mathfrak{n}(t')] dt'$ if m is smooth. Volkonskii [100] had proved it, too, and at about the same time. H. Tanaka and I, in [96] extended the recipe to diffusions in $\mathbb{R}^d$ with the same road map as the d-dimensional Brownian motion. A further extension was made by Blumenthal, Getoor, and myself: any two diffusions with the same road map differ by a change of clock [5]. McK [70] is a review of this stuff for a financial audience; see also [44], in which fluctuating clocks of the form

$$T(t) = \text{measure } (t' \le t : \mathfrak{n}(t') \ge 0) - \int_{-\infty}^{0} \mathfrak{t}(t,x)dm(x)$$

are used to reproduce a special class of Feller's Brownian motions on the half-line.

4.9. Potentials and Capacity

The now standard connection of Brownian motion to harmonic functions, (electrostatic) potentials, and capacity, clarified by J. Doob [11] and G. Hunt [26, 27] was current news in 1956. The paper Itô-McK [28] is expository, aiming to explain these ideas in the simplest format, *viz.* the standard random walk in $d \ge 3$ dimensions. Then what was intriguing to Itô and me was the connection between the equilibrium charge distribution e for compact $K \subset R^3$ and the last leaving-time from K of the Brownian motion on its way to ∞. We understood it in part. I made a little progress in [50], but the real picture is due to K.L. Chung [9]: The total charge of e is the capacity $C(K)$, so e/C is a probability distribution, and any well brought up probabilist must ask: What is it the distribution of? Chung's beautiful answer is this: The 3-dimensional Brownian motion runs off to ∞, so if it visits K at all, it has a last leaving-time $\mathfrak{f} = \max(t : \mathfrak{x}(t) \in K)$, and with the standard Green's function $G = 1/4\pi r$, $G(x,y)de(y)$ is the distribution of the last leaving place for paths starting at x, *to wit*, $P_x[\mathfrak{x}(\mathfrak{f}) \in dy, \mathfrak{f} > 0]$.

4.10. Differential Space

This was Norbert Wiener's name for the space of Brownian paths. McK [54] takes off from the observation that if $x = (x_1, \ldots, x_n)$ is uniformly distributed on the $(n-1)$-dimensional sphere of radius $\sqrt{n}$, then for fixed m,

$$\lim_{n \uparrow \infty} P[\bigcap_{i=1}^{m} (a_i \le x_i < b_i)] = \prod_{i=1}^{m} \int_{a_i}^{b_i} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx,$$

a fact often attributed to Poincaré, but due, in fact, to Mehler [81] who knew a lot of what is said in [54]. Anyhow, if you think this way and reflect that the Brownian path may be synthesized out of an infinite stock of independent Gaussian variables, then you will see that Wiener measure is (or ought to be) the round measure on the admittedly fabulous sphere $S^\infty(\sqrt{\infty})$, and a lot of things fall into place. For example, there is a nice rotation group acting on the sphere and a perfectly respectable commuting Laplacian Δ = a sum of uncoupled Hermite operators; also pleasing spherical harmonics in the guise of products Hermite polynomials, spanning Wiener's "polynomial chaos". There is even a Brownian motion (of Brownian motion!) with infinitesimal operator $\mathfrak{G} = \frac{1}{2}\Delta$, this being the ∞-dimensional Ornstein-Uhlenbeck process at the bottom of P. Malliavin's calculus in the Brownian path space [46]. Here's an instructive picture in this style: If $f : A \to B$ is a smooth map between manifolds, then the pushforward (to B via f) of a smooth measure on A is also smooth. Now the sphere $S^\infty(\sqrt{\infty})$ is eminently smooth – how should it be otherwise as Feller would say – and its round measure is as smooth as they come. Map the Brownian motion $\mathfrak{n}$ to the diffusion $\mathfrak{x}$ with smooth $\mathfrak{G} = \frac{1}{2}\sigma^2 D^2 + mD$ via Itô's equation $d\mathfrak{x} = \sigma(\mathfrak{x})d\mathfrak{n} + m(\mathfrak{x})dt$. Then $\mathfrak{x}$ must be smoothly distributed, e.g. it must have a smooth transition density. This looks naive but is not: Malliavin has confirmed it all by his "integration by parts" in the Brownian path space. The expository McK [55] belongs to the same circle of ideas. It is an account of Wiener's "polynomial chaos" and what he hoped it could do. I have to say that his attempt was a gallant failure, but it's still interesting.

4.11. Fisher's Equation

McK [56] refers to the equation $\partial u/\partial t = \frac{1}{2}\partial^2 u/\partial x^2 + u^2 - u$ of Kolmogorov, Petrovskii and Piscounov [35]. Actually, it was introduced by the great English statistician R.A. Fisher [19], but no matter. Its solution can be neatly expressed as the expectation

$$u(t, x) = E_x[f(\mathfrak{x}_1(t)) \ldots f(\mathfrak{x}_\#(t))]$$

in which $f = u(0+, \cdot)$ and the $\mathfrak{x}$'s are the positions of the $\#(t)$ particles produced, up to time $t \geq 0$, by a branching Brownian motion, rooted at x at time $t = 0$. I wanted to use this picture to reproduce the result of KP^2: that if f is the indicator of the half-line $x \geq 0$, then the solution u, tracked at speed $\sqrt{2}$ (with small corrections), tends to the traveling wave w of that speed, this being the only *stable* wave, rising from 0 to 1, that the equation supports. The paper is spoiled by two mistakes, but it is still a favorite of mine since such a use of Brownian motion was new then. I comment on it here with apologies and some account of how it all came out OK. The trouble starts with the ineffective proof of 3′). That's easy, but in §7 comes a serious mistake. At $+\infty$, the wave w looks like $1 - xe^{-\sqrt{2}x}$, not $1 - e^{-\sqrt{2}x}$ as I said, so what I wrote is simply wrong. Happily, the idea was salvaged by S.P. Lally and T. Selke [36]. They prove that $w(x) = E[\exp(-Ze^{-\sqrt{2}x})]$ in which $0 \leq Z < \infty$ is the limit of

$$\sum_{n=1}^{\#}[\sqrt{2}t - \mathfrak{x}_n(t)]e^{\sqrt{2}\mathfrak{x}_n(t)}.$$

To return to KP^2, the tracking of u for $f =$ the indicator of $x \geq 0$ is done by centering at the median m where $u(t, x) = 1/2$, and it's easy to see that $m \simeq \sqrt{2}t$. The refinement $m = \sqrt{2}t - (3 - 2^{-3/2})\ln t + O(1)$ of Bramson [6] is really hard work.

4.12. Oscillators

Kac [31] had studied the motion of a damped spring driven by white noise. In McK [48], I took the simplest case and looked at the winding of $(\mathfrak{x}, \mathfrak{x}^\bullet)$ about the origin of the phase plane. I was entertained to find coming in the Kontorovich-Lebedev transform:

$$\begin{aligned} f \quad &\to \quad \hat{f}(\gamma) = \int_0^\infty f(x) K_{\sqrt{-1}\gamma}(x) \frac{dx}{x} \\ \hat{f} \quad &\to \quad f(x) = \int_0^\infty \hat{f}(\gamma) K_{\sqrt{-1}\gamma}(x) \\ &\qquad \frac{2}{\pi^2}\gamma sh(\pi\gamma) d\gamma. \end{aligned}$$

see Bateman and Erdélyi [2]. Unfortunately, I was not as clever with them as I thought. P.G. Gait's letter to me pointed out that the derivation of 6 in §3 is obscure to put it kindly, and he was not the only one to be put off. E. Wong and Per Ch. Hemmer sent still formal but neat proofs. Finally, I fixed it up in a pretty laborious way, but it seems this was never published.

4.13. Winding of Plane Brownian Motion

This refers to the winding of $BM(2)$ about a couple of punctures, 0 and 1, say. It is easy to see that this is a big mess from the viewpoint of homotopy, i.e. the path gets inextricably tangled up as time passes. This is because the (lifted) Brownian motion on the universal cover of the twice-punctured plane is transient, the cover being the (hyperbolic) half-plane. But what about the viewpoint of homology, i.e. is the joint motion $(\mathfrak{x}, n_0, n_1)$ of position and winding numbers recurrent or does it wander off to ∞. This is the Brownian motion on the "class surface" of the punctured plane, which looks like a parking garage of infinitely many floors connected by (equivalent) helical staircases located at $2\pi\sqrt{-1}\mathbb{Z}$. I thought I had a proof of the its recurrence, but that was wrong, as Dennis Sullivan noticed: a certain Poincaré sum did not diverge as I had believed. T.J. Lyons and McKean [45], corrects this mistake: the two winding numbers we found to behave like $\mathfrak{n}_0 - \mathfrak{n}_\infty$ & $\mathfrak{n}_1 - \mathfrak{n}_\infty$ with three independent Cauchy processes $\mathfrak{n}_0, \mathfrak{n}_1, \mathfrak{n}_\infty$, run by a common clock related to Brownian motion on the Riemann sphere. Sullivan and McKean [95], is a simplified, more geometrical proof of the transience on the class surface per se.

4.14. Nonlinear Markov Processes

McK [51] is about a simple idea: to permit a diffusion to be "guided by its present distribution, i.e. to permit drift and any Gaussian fluctuations to be so influenced. Then the corresponding (forward) equation for p is *non*-linear, as in $\partial n/\partial t = \partial^2(\ln n)/\partial x^2$ which comes up in connection with a "central limit theorem" for Carleman's model of a gas; see McK [58]. I have said more about this in [60] and [59].

4.15. Invariant Distributions for Diffusion

McKean [74] deals with the following question: If a several-dimension diffusion *has* an invariant distribution, what then is it the distribution *of*? A complete picture is obtained only in dimension 1. There is an entertaining interplay between three related processes: The original diffusion $\mathfrak{x}^{\uparrow}$, $\mathfrak{x}^{\#} = \mathfrak{x}^{\uparrow}$ with its drift reversed, and $\mathfrak{x}^{\downarrow} =$ (at each time) the inverse of the map $x \to \mathfrak{x}^{\#}(t,x)$, *aka* $\mathfrak{x}^{\uparrow}$ with its driving Brownian motion reversed. The (statistical) stabilization of $\mathfrak{x}^{\uparrow}$ is related to the path-wise existence of $\mathfrak{x}^{\downarrow}(\infty)$ which is distributed by the invariant density, and this, in turn, is related to the focusing of $\mathfrak{x}^{\uparrow}$. It is striking that in dimension 1, this focusing takes place exponentially fast, spectral gap or no. McK [71] is an ∞-dimensional example: Burgers equation $\partial v/\partial t + v\partial v/\partial x = \frac{1}{2}\partial^2 v/\partial x^2 + f$ with a "white" force f. The proof of the existence of the invariant distribution, based on ideas of Döblin ("loops") and Feller, is simpler than the others which had been put forward.

4.16. KdV and All That

One morning, spring 1973, Pierre van Moerbeke came and told me that KdV: $\partial v/\partial t = 3v\partial v/\partial x - \frac{1}{2}\partial v^3/\partial x^3$ on the circle $0 \le x < 1$ had a traveling wave $2\wp(x - ct) - c/3$, $\wp$ being Weierstrass's function. I'd never heard of KdV before, but here was an attractive problem (dispersion completing with shocks), and as I was a fan of both elliptic functions and special non-linear PDEs, I took notice. Hochstadt [24, 25] and Lax [37, 38] helped us to figure out, soon enough how to solve when the number of "gaps" is finite [79] printed below, with revisions of the part on theta functions. Unknown to us, S.P. Novikov [84] had already done this. Anyhow, it was obvious then how to do it generally, for an infinite number of gaps, though there was still a lot of technical work to do; see McK-Trubowitz [77] and also [78]. This work was my real introduction to Hamiltonian mechanics with many commuting constants of motion (integrability), to theta functions, and to curves of infinite genus.

The machinery for solving KdV is described in my review of the remarkable book of Feldman, Knörrer and Trubowitz [17] also printed below, so I won't say more about it here. Ercolani-McK [16] is a companion to McK-van Moerbeke [79] and Dyson [15] had expressed the solution of KdV in $C^{\infty}_{\downarrow}(\mathbb{R})$ as a Fredholm determinant:

$$v(t,x) = -2\frac{\partial^2}{\partial x^2}\ln\det[I + w(t, x+\xi+\eta) \\ : \xi,\eta \ge 0]$$

in which

$$w(t,x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\sqrt{-1}kx} s_{21}(k) e^{4\sqrt{-1}k^3 t}\,dk$$

and s_{21} is the (right-hand) reflection coefficient of $v(0+,\cdot)$, this w being a solution of $\partial w/\partial t = 4\partial^3 w/\partial x^3$, or as you may say, of KdV with the non-linearity crossed out. Here, $\Theta = \det[I + w(\xi+\eta) : \xi,\eta \ge 0]$ has to be some kind of theta function (compare Its-Matveev [30]), and we wanted to known how much of the geometry of Hill's curves carries over. Venakides [98] had already found that if $v \in \mathbb{C}^{\infty}_{\downarrow}(\mathbb{R})$ is periodized as in $\sum_{\mathbb{Z}} v(x+np)$ with period p, then the associated Riemann theta function tends to Dyson's determinant as $p \uparrow \infty$. Ercolani and I did that in a more laborious way, obtaining, as by-product, nice analogues of everything: curve, divisor, DFK, Jac, the works. The part I like best is the description (§ 6) of the (complex) theta divisor for Dyson's determinant, reproducing, almost word for word, the known structure of that divisor for classical hyperelliptic curves. I believe there is a moral here, *to wit*: complex algebraic-geometrical objects are very robust and may be recognized, if you come without prejudice, in many objects that seem at first glance to be of purely analytic type: for example, if you think this way, then objects like $\int_0^\infty (\lambda - \mu)^{-1} d\sigma(\mu)$ will put you in mind of Jacobi's ellipsoidal coordinates.

McK [62] introduces a new theme, "addition", based upon Jacobi's presentation of divisors on a hyperelliptic curve of genus $g < \infty$, adapted here to Hill's curves with $g = \infty$ compare McK [80] for background. Take Hill's operator $H = -D^2 + v$ and λ below spectrum. Then $He = \lambda e$ has two positive "multiplicative" solutions, e_- with $\int_{-\infty}^0 e_-^2 < \infty$ and e_+ with $\int_0^\infty e_+^2 < \infty$. "Addition" is the map $v \to v - 2[\ln e(\cdot, \mathfrak{p})]''$ in which $e = e_\pm$, and the point $\mathfrak{p}$ on the multiplier curve records the value of λ and the signature of the radical, this being used to specify which function, e_+ or e_-, is employed. The reason for the name (addition) is that, in the case $g < \infty$, the pole divisor $\mathfrak{p}_1 + \ldots + \mathfrak{p}_g$ of e is changed by addition to a new divisor $\mathfrak{p}_1' + \ldots + \mathfrak{p}_g'$ according to the rule

$$\begin{aligned} \mathfrak{p}_1 + \ldots + \mathfrak{p}_g &\to -\mathfrak{p} + \mathfrak{p}_1 + \ldots + \mathfrak{p}_g \\ &\equiv \mathfrak{p}_1' + \ldots + \mathfrak{p}_g' + \infty, \end{aligned}$$

in which the $\equiv$ means equivalence in the Jacobi variety, i.e. the two divisors are, *resp.* the roots and poles of a function of rational character. In the present paper, addition is extended to any reasonable v on the line for which $\text{spec}(H)$ does not extend to $-\infty$ (so that you can get below it). In this generality, the several additions commute, infinitesimal additions produce the whole KdV hierarchy, and (most important) addition is not only isospectral but respects a foliation which I called "unimodular isospectrality". I explain. H has a spectral resolution involving a 2×2 spectral weight $dF(\lambda) : \lambda \in \mathbb{R}$, and two such operators H_1 & H_2 are isospectral if and only if $dF_1 = G^{-1} dF_2 G$ with 2×2 function $G : R \to GL(2, R)$. What is striking now is that, under addition $dF \to G^{-1} dF G$ with $G \in SL(2, \mathbb{R})$, whence the term *unimodular*. Then for any H with $\text{spec}(H)$ bounded away from $-\infty$, you have (1) its KdV-invariant manifold (this can be made sense of in general), (2) the class of operators produced from one operator of that class by repeated additions (plus a closing up), and (3) its unimodular spectral class, each included in the one before, and examples indicate that these should always be the same. See [63, 65].

McK [64, 'Curvatura integra, handle number, and genus of transcendental curves'] is my attempt to extend the idea of genus to transcendental curves. It was inspired by the example of Hill's curves (see KdV & all that) but seemed best standing alone.

McK [72] printed here deals with the equation of Camassa-Holm [8]:

$$\text{CH} : \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} + \frac{\partial p}{\partial x} = 0 \text{ with "pressure"}$$
$$p = (1 - D^2)^{-1} (v^2 + \frac{1}{2} v'^2),$$

which is, to me, the nicest of the shallow-water equations. Unlike KdV which is only a leading edge approximation, CH is reminiscent of Euler in its form. Besides, it has solitons that can travel in both directions, right and left; the wave can break; it has extra (non-commuting) constants of motion reminiscent of (but not all that much like) circulations in a real fluid; and with the right machinery, it's easy to work with, easier that KdV I'd say.

McK-Constantin [10] did it on the circle, mostly for $g < \infty$. There is only one surprise after KdV. You go to the curve and over to its Jacobian variety where you find straight-line motion (good) at non-constant speed (bad). A double cover of the curve cures that–why I have never understood. The present paper solves CH on the line provided $m = v - v''$ is summable. The "curve" is purely singular and the solution is found in terms of three interconnected theta-like functions ϑ_-, ϑ_+, and ϑ plain, expressed by Fredholm determinants, mediated by the Lagrangian scale $\bar{x}$, descriptive of the displacement of the fluid particle. I remind you: $\partial \bar{x}(t, x)/\partial t = v(t, \bar{x})$ with $\bar{x}(0+x) \equiv x$. This is the first (and very valuable) appearance of $\bar{x}$ in "KdV & all that". Now $\vartheta_\pm$

cannot vanish and $v(t,x) = \partial/\partial t[\ln(\vartheta_-/\vartheta_+)]$ makes sense for all time! The "breaking" is seen only in v' which comes with a ϑ downstairs, and *that* function can and will vanish for some $t \geq 0$ if (and only if) $m = v - v''$ takes a positive value to the left of some negative value: positive/negative stuff likes to move right/left, and trouble comes when they collide. I had made a laborious proof of that [67], corrected in [68]. The present machinery produced an easy proof, reported here as well. McK [73] deals mostly with CH on the circle with $m > 0$. Then the Liouville correspondence

$$V(X) = \frac{1}{4}\frac{1}{m(x)} + \frac{1}{4}\frac{m'(x)}{m^2(x)} - \frac{5}{16}\frac{m'^2(x)}{m^3(x)} \text{ with}$$
$$X = \int_0^x \sqrt{m}$$

converts the spectral problem $-f'' + \frac{1}{4}f = \lambda m f$ for CH into the spectral problem for KdV: $-F'' + VF = \lambda F$ with $F(X) = \sqrt[4]{m(x)}f(x)$. V. Fock suggested that this must imply a correspondence between the CH and KdV hierarchies, and we worked at this on a pleasant afternoon in Providence, but it wouldn't come out. We forgot the Lagrangian scale. Of course, such correspondences must be fairly common: CH with $m > 0$ goes right as does KdV; once that is acknowledged, there's no big surprise: everything reduces to straight-line motion at constant speed on a torus, and that always looks the same. The trick is to express the correspondence in an intelligible way. Still a word about the Lagrangian scale $\bar{x}$: The extra constants of motion alluded to above, are $m(\bar{x})(\bar{x}')^2$ with $m = v - v''$ as before, one such to each value of x. They do not commute either with themselves or with the CH hierarchy. I wonder what they are doing here and whether such objects appear commonly, perhaps in classical mechanics already. Anyhow, that's what I've done about CH, but I have still to cite the beautiful paper of L.-C. Li [43]. There, the solution is expressed in way reminiscent of the Toda lattice. It's not as explicit as my way but much more elegant.

Selected publications of H. P. McKean

General lectures

- "Probability and its applications." pp. 90–95. *The Mathematical Sciences after the Year 2000*, Beirut, Lebanon. World Scientific, Singapore, 2000.
- "Jürgen K. Moser (part 3 of 4)." *Notices Amer. Math. Soc. 47* (2000).
- "Kyoto 1957/58"
- "Norbert Wiener." *American Scientist 65* (1977) 220–221.
- Marc Kac, 1914–1984. "Biographical Memoirs," *Natl. Acad. Sci. USA 59* (1990) 215–235.
- "Onsager and Machlup's work on fluctuations in statistical physics." Collected Works of Lars Onsager, ed. Hemmer et al., pp. 769–771. World Scientific, Singapore, 1996.
- Foreword to: *Selected Papers of Norman Levinson.* Birkhauser Boston, 1997.

Diffusion Theory

- "Potentials and the random walk" (with K. Ito) *J. Math. 4* (1960) 119–132.
- "Brownian motion on the 3-dimensional rotation group" *Mem. Coll. Sci. Kyoto 33* (1960) 25–38.
- "The Bessel motion and a singular integral equation" *Mem. Coll. Sci. Kyoto 33* (1960) 317–322.
- "Additive functionals of the Brownian path" (with H. Tanaka) *Mem. Coll. Sci. Kyoto 33* (1961) 479–506.
- "A Hölder condition for Brownian local time" *J. Math. Kyoto 1* (1962) 197–201.
- "Markov processes with identical hitting distributions" (with R. Blumenthal and R. Getoor) *Ill. J. Math. 6* (1962) 402–420.
- "Excursions of a nonsingular diffusion" *Z. Wahrschein 1* (1963) 230–239.
- "A winding problem for a resonator driven by a white noise" *J. Math. Kyoto 2* (1963) 228–235.
- "Brownian motions on a half-line" (with K. Ito) *Ill. J. Math. 7* (1963) 181–231.

- "A. Skorohod's stochastic integral equation for a reflecting barrier diffusion" *J. Math. Kyoto 3* (1963) 85–88.
- "Brownian motion with a several-dimensional time" *Theor. Prob. Appl. 8* (1963) 335–354.
- "Weighted trigonometrical approximation on R with application to the germ field of a stationary Gaussian noise" (with N. Levinson) *Acta Math. 112* (1964) 99–143.
- "A probabilistic interpretation of equilibrium charge distributions" *J. Math. Kyoto 4* (1965) 617–625.
- "Geometry of differential space" *Ann. Prob. 1* (1973) 197–206.
- "Wiener's theory of nonlinear noise" *SIAM-AMS Symp. Appl. Math. 6* (1973) 191–209.
- "Brownian local times" *Adv. Math. 16* (1975) 91–111.
- "Application of Brownian motion to the equation of Kolmogorov–Petrovski–Piscounov" *Comm. Pure Appl. Math. 28* (1975) 323–331.
- "Two queues in parallel" (with L. Flatto) *Comm. Pure Appl. Math. 30* (1977) 255–263.
- "Winding of the plane Brownian motion" (with T. J. Lyons) *Adv. Math.* (1981) 212–225.
- "Brownian motion and harmonic functions on the thrice-punctured sphere" (with D. Sullivan) *Adv. Math. 51* (1981) 203–211.
- "A martingale approach to some Wiener–Hopf problems" (with R. R. London, L. C. G. Rogers, and D. Williams) Séminaire de probabilités XVI, 1980/1981, *Lect. Notes in Math. 920* (1982) 41–90.

Financial Mathematics

- "A free boundary problem for the heat equation arising from a problem in mathematical economics" *Ind. Management Rev. 6* (1965) 32–40.
- "Brownian motion and the general diffusion: Scale & Clock" *Mathematical Finance-Bachelier Congress* June 2000, pp. 75–84, Springer–Verlag, Berlin, 2001.

Gaussian processes

- "Application of de Branges spaces of integral function to the prediction of stationary Gaussian process" (with H. Dym) *Ill. J. Math. 14* (1970) 299–343.
- "Extrapolation and interpolation of stationary Gaussian processes" (with H. Dym) *Ann. Math. Stat. 41* (1970) 1817–1844.

Statistical Mechanics

- "Entropy is the only increasing functional of Kac's one-dimensional caricature of a Maxwellian gas" *Z. Wahrschein 2* (1963) 167–172.
- "Kramers–Wannier duality for a 2-dimensional Ising model as an instance of the Poisson summation formula" *J. Math. Phys. 5* (1964) 775–776.
- "Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas" *Arch. Rat. Mech. Anal. 21* (1966) 343–367.
- "A class of Markov processes associated with nonlinear parabolic equations" *Proc. Nat. Acad. Sci. 56* (1966) 1907–1911.
- "An exponential formula for solving Boltzmann's equation for a Maxwellian gas" *J. Comb. Th. 2* (1967) 358–382.
- "Chapman–Enskog–Hilbert expansion for a class of solutions of the telegraph equation" *J. Math. Phys. 8* (1967) 547–552.
- "A simple model of the derivation of fluid mechanics from the Boltzmann equation" *Bull. Amer. Stat. 41* (1970) 1817–1844.
- "The central limit theorem for Carleman's equation" *Israel J. Math. 21* (1975) 54–92.

- "Fluctuations in the kinetic theory of gases" *Comm. Pure Appl. Math. 28* (1975) 435–455.
- "Statistical mechanics of nonlinear wave equations" (with K. Vaninsky) *Trends and Perspectives Appl. Math.* ed. L. Sirovich (1994) 239–264, Springer–Verlag, New York; see also *Proc. Symp. Pure Math. AMS 57* (1995) 457–463.
- "Brownian motion with restoring drift: the micro-canonical ensemble" (with K. Vaninsky) *Comm. Math. Phys. 160* (1994) 615–630.
- "Turbulence without pressure: existence of the invariant measure" *Math. Appl. Anal. 9* (2002) 463–468.
- "Cubic Schrödinger: the petit canonical ensemble" *Adv. Studies Pure Math. 4* (2004) to appear.
- "The spectral class of the quantum-mechanical oscillator" (with E. Trubowitz) *Comm. Math. Phys. 82* (1982) 471–495.

Differential geometry

- "Curvature and the eigenvalues of the Laplacian" (with I. M. Singer) *J. Diff. Geom. 1* (1967) 43–69.
- "An upper bound to the spectrum of Δ on a manifold of negative curvature" *J. Diff. Geom. 4* (1970) 359–366.
- "Selberg's trace formula as applied to a compact Riemann surface" *Comm. Pure Appl. Math. 25* (1972) 225–246.

Analysis

- "Nagumo's equation" *Adv. Math. 4* (1970) 209–223.
- "Stabilization of solutions of a caricature of the Fitzhugh–Nagumo equation" *Comm. Pure Appl. Math. 36* (1983) 291–324.
- "Variation on a theme of Jacobi" *Comm. Pure Appl. Math. 38* (1985) 669–678.
- "A threshold for a caricature of the nerve equation" (with V. H. Moll) *Bull. Amer. Math. Soc. 12* (1985) 255–259.
- "Stabilization to the standing wave in a simple caricature of the nerve equation" (with V. H. Moll) *Comm. Pure Appl. Math. 39* (1986) 485–529.
- "Geometry of some nonlinear differential operators" *Ann. Scuola Norm. Sup. Pisa 13* (1986) 299–346.
- "Singularities of a simple elliptic operator" *J. Diff. Geom. 25* (1987) 157–165.
- "How real is resonance?" *Comm. Pure Appl. Math. 50* (1997) 317–322.

Integrable equations and non-linear partial differential equations

- "Spectrum of Hill's equation" (with P. van Moerbeke) *Invent. Math. 30* (1975) 217–274.
- "Hill's operator and hyperelliptic function theory in the presence of infinitely many branch points" (with E. Trubowitz) *Comm. Pure Appl. Math. 29* (1976) 143–276.
- "Δ plus a bad potential" *J. Math. Phys. 18* (1977) 1277–1279.
- "Rational and elliptic solutions of the Korteweg–de Vries equation and a related many-body problem" (with H. Airault and J. Moser) *Comm. Pure Appl. Math. 30* (1977) 95–148.
- "Integrable systems and algebraic curves" *Lect. Notes Math. 775* (1979) 83–200.
- "Hill's surfaces and their theta functions" (with E. Trubowitz) *Bull. Amer. Math. Soc. 84* (1978) 1042–1085.
- "Boussinesq's equation as a Hamiltonian system" *Adv. Math. Suppl. Studies 3* (1978) 217–226.
- "Algebraic curves of infinite genus arising in the theory of nonlinear waves" *Proc. Int. Cong. Math., Helsinki* (1978) 777–784.
- "Theta functions, solitons and singular curves" *Partial Differential Equations and Geometry* C. Brynes. M. Dekker (ed.), New York (1979) 237–254.
- "Integrable systems and algebraic curves" *Lect. Notes in Math. 755* (1980) 83–200.

- "Boussinesq's equation on the circle" *Joint US–USSR Conf. on Solitons and Turbulence*, Kiev, Sept. 1979.
- "The sine-Gordon and Sinh-Gordon equation on the circle" *Comm. Pure Appl. Math. 34* (1980) 197–257.
- "Boussinesq's equation on the circle" *Comm. Pure Appl. Math. 34* (1981) 599–691.
- "Boussinesq's equation: how it blows up" *J. C. Maxwell Sesquicentennial Symp.*, ed. M. Berger, North-Holland (1984).
- "Geometry of KdV (1): addition and the unimodular spectral classes" *Rev. Mat. Iberoamer. 2* (1986) 235–261.
- "Geometry of KdV (2): three examples" *J. Stat. Phys. 46* (1987) 1115–1143.
- "Is there an infinite-dimensional algebraic geometry? Hints from KdV" *AMS Summer Conf. on Theta Functions*, Bowdoin College, New Brunswick, Maine (1987).
- "Geometry of KdV (4): Abel sums, Jacobi variety, and theta function in the scattering case" (with N. Ercolani) *Invent. Math. 99* (1990) 483–544.
- "Geometry of KdV (5): scattering from the Grassmannian point of view" *Comm. Pure Appl. Math. 42* (1989) 687–701.
- "The nonlinear Schrödinger equation and the nonlinear heat equation: reduction to linear form" (with J. Shatah) *Comm. Pure Appl. Math. 44* (1991) 1057–1066.
- "Curvatura integra, handle-number, and genus of transcendental curves" *Comm. Pure Appl. Math. 44* (1991) 1057–1066.
- "Geometry of KdV (3): determinants and unimodular isospectral flows" *Comm. Pure Appl. Math. 45* (1992) 389–415.
- "Compatible brackets in Hamiltonian mechanics" *Important Developments in Soliton Theory* ed. A. Fokas and V. Zakharaov, Springer–Verlag, New York, 1993.
- "A quicker proof of Fay's identity" *Comm. Pure Appl. Math. 46* (1993) 501–503.
- "Baker–Akhiezer functions in the scattering case" *Comm. Pure Appl. Math. 46* (1993) 783–785.
- "The rigidity of sine-Gordon breathers" (with B. Birnir and A. Weinstein) *Comm. Pure Appl. Math. 47* (1994) 1043–1051.
- "A limit law for the ground state of Hill's equation" *J. Stat. Phys. 74* (1994) 1227–1232.
- "Statistical mechanics of nonlinear wave equations (4): cubic Schrödinger" *Comm. Math. Phys. 168* (1995) 479–491. Erratum: *173* (1995) 675.
- "Action-angle variables for the cubic Schrödinger equation" (with K. Vaninsky) *Comm. Pure Appl. Math. 50* (1997) 489–562.
- "Cubic Schrödinger: the petit canonical ensemble in action-angle variables" (with K. Vaninsky) *Comm. Pure Appl. Math. 50* (1997) 593–622.
- "A novel aspect of action-angle variables" *Proc. Symp. Pure Math. AMS 65* (1999) 155–160.
- "A quick proof of Riemann's mapping theorem" *Comm. Pure Appl. Math. 52* (1999) 405–409.
- "A shallow water equation on the circle" (with A. Constantin) *Comm. Pure Appl. Math. 52* (1999) 947–982.
- "Variations on a theme of M. G. Krein" *Comm. Pure Appl. Math. 52* (1999) 453–459.
- "The ground state eigenvalue of Hill's equation with white noise potential" (with S. Cambronero) *Comm. Pure Appl. Math. 52* (1999) 1277–1294.
- "Addition for the acoustic equation" *Comm. Pure Appl. Math. 54* (2001) 1271–1288.
- "Fredholm determinants and the Camassa–Holm hierarchy" *Comm. Pure Appl. Math. 56* (2003) 638–680.
- "Liouville's correspondence between the Korteweg–de Vries and the Camassa–

Holm hierarchies" *Comm. Pure Appl. Math. 56* (2003) 998–1015.
- "Breakdown of the Camassa–Holm equation" *Comm. Pure App. Math. 57* (2004) 416–418.
- "Fredholm Determinants" *Cent. Eur. J. Math. 9* (2011), 205–243.

Bibliography

[1] A. Ambrosetti and G. Prodi. On the inversion of some differentiable mappings with singularities between Banach spaces. *Ann. Mat. Pura Appl.*, 93:231–246, 1972.

[2] A. Bateman, H. Erdélyi. *Higher Transcendental Functions*, volume I. McGraw-Hill, New York/London, 1953.

[3] M. Berger and P. Church. Complete integrability and perturbation of a nonlinear Dirichlet problem. *Indiana Univ. Math. J.*, 28:935–952, 1979.

[4] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Political Econ.*, 81:637–654, 1973.

[5] R. M. Blumenthal, R. K. Getoor, and H. P. McKean. Markov processes with identical hitting distributions. *Illinois J. Math.*, 6:402–420, 1962.

[6] M. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, 31:531–581, 1978.

[7] P. Buser, J. Conway, P. Doyle, and K. Semmler. Some planar isospectral domains. *Internat. Math. Res. Notices*, 9:301–400, 1994.

[8] R. Camassa and D. Holm. An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett.*, 71(11):1661–1664, 1993.

[9] K. L. Chung. Probabilistic approach in potential theory to the equilibrium problem. *Ann. Inst. Fourier, Grenoble*, 23:313–322, 1973.

[10] A. Constantin and H. P. McKean. A shallow water equation on the circle. *Comm. Pure Appl. Math.*, 52:949–982, 1999.

[11] J. Doob. Semi-martingales and subharmonic functions. *Trans. AMS*, 77:86–121, 1954.

[12] H. Dym and H. P. McKean. Applications of the de Branges spaces of integral functions to the prediction of stationary Gaussian processes. *Illinois J. Math.*, 14:299–343, 1970.

[13] H. Dym and H. P. McKean. Extrapolation and interpolation of stationary Gaussian processes. *Ann. Math. Stat.*, 41:1819–1844, 1970.

[14] H. Dym and H. P. McKean. *Gaussian Processes*. Academic Press, 1976.

[15] F. J. Dyson. Fredholm determinants and inverse scattering. *Comm. Math. Phys.*, 47:171–183, 1976.

[16] N. Ercolani and H. P. McKean. Geometry of KdV (4). Abel sums, Jacobi variety, and theta function in the scattering case. *Invent. Math.*, 99(3):483–544, 1990.

[17] J. Feldman, H. Knörrer, and E. Trubowitz. *Riemann surfaces of infinite genus*, volume 2 of CRM. AMS, Providence, 2002.

[18] W. Feller. Generalized second order differential operators and their lateral conditions. *Illinois J. Math.*, 1:459–564, 1957.

[19] R. A. Fisher. The wave of advance of advantageous genes. *Ann. Eugenics*, 7:353–369, 1937.

[20] I. M. Gelfand. Automorphic functions and the Theory of Representations. In *Proc. Inter. Congress Math.*, volume 1, pages 74–85. Institut Mittag-Leffler, 1962.

[21] P. Gilkey. Curvature and eigenvalues of the Laplacian for elliptic operators. *Adv. Math.*, 10:344–382, 1973.

[22] H. P. Golchan. On some problems concerning Brownian motion in Lèvy's sense. *Teor. Veroyatnost. i Primen*, 12:682–690, 1967.

[23] C. Gordon, D. Webb, and S. Wolpert. Isospectral plane domains and surfaces via Riemannian orbofolds. *Invent. Math.*, 110:1–22, 1992.

[24] H. Hochstadt. Function-theoretic properties of the discriminant of Hill's equation. *Math. Zeit.*, 82:237–242, 1963.

[25] H. Hochstadt. On the determination of Hill's equation from its spectrum. *Arch. Rat. Mech. Anal.*, 19:353–362, 1965.

[26] G. Hunt. Markoff processes and potentials, I. *Illinois J. Math.*, 1:44–93, 1957.

[27] G. Hunt. Markoff processes and potentials, III. *Illinois J. Math.*, 2:151–213, 1958.

[28] K. Itô and H. P. McKean. Potentials and the random walk. *Ill. J. Math.*, 4:119–132, 1960.

[29] K. Itô and H. P. McKean. Brownian motion on a half line. *Ill. J. Math.*, 7:181–231, 1963.

[30] A. R. Its and V. B. Matveev. Hill operators with a finite number of lacunae. *Funkcional. Anal. i Priložen.*, 9(1):69–70, 1975.

[31] M. Kac. Probability theory: its role and impact. *SIAM Review*, 4:1–11, 1962.

[32] M. Kac. Can you hear the shape of a drum? *Amer. Math. Monthly*, 73:1–23, 1966.

[33] M. Kac and D. Slepian. Large excursions of a Gaussian process. *Ann. Math. Stat.*, 30:1215–1228, 1959.

[34] F. B. Knight. Random walks and a sojourn density process for Brownian motion. *Trans. AMS*, 167:56–86, 1963.

[35] A. Kolmogorov, I. Petrovskii, and N. Piscounov. Etude de l'èquation de la diffusion avec croissance de la quantitè de la matière. *Bull. Math. Soc. Moscou*, 1:1–35, 1937.

[36] S. P. Lally and T. Selke. A conditional limit theorem for the frontier of a branching Brownian motion. *Ann. Prob.*, 15:1052–1061, 1987.

[37] P. Lax. Integrals of non-linear equations and solitary waves. *Comm. Pure Appl. Math.*, 21:467–490, 1968.

[38] P. Lax. Periodic solutions of the Korteweg-de Vries equation. *Comm. Pure Appl. Math.*, 28:141–188, 1975.

[39] N. Levinson. *Selected Papers of Norman Levinson, volume 2.* Birkhauser-Verlag, Boston, 1998.

[40] N. Levinson and H. P. McKean. Weighted trigonometrical approximation on $\mathbb{R}^1$ with application to the germ field of a stationary gaussian noise. *Acta Math.*, 112:99–143, 1964.

[41] P. Lévy. *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, Paris, 1948.

[42] P. Lévy. Random functions: general theory with special reference to Laplacian random functions. *Univ. Cal. Pub. Stat.*, 1:331–390, 1953.

[43] L. C. Li. Factorization problems on the Hilbert-Schmidt group and the Camassa-Holm equation. *Comm. Pure Appl. Math.*, 56:186–209, 2008.

[44] R. R. London, H. P. McKean, L. C. G. Rogers, and D. Williams. A martingale approach to some Wiener-Hopf problems. *Sem. Prob., Lecture Notes in Math.*, 920:41–90, 1982.

[45] T. J. Lyons and H. P. McKean. Winding of the plane Brownian motion. *Adv. Math.*, 51:212–225, 1984.

[46] P. Malliavin. *Stochastic Analysis*, volume 313 of *Grundlehren der Math. Wiss.* Springer-Verlag, 1997.

[47] H. P. McKean. A Hölder condition for Brownian local time. *J. Math. Kyôto Univ.*, 1:195–201, 1962.

[48] H. P. McKean. A winding problem for a resonator driven by a white noise. *J. Math. Kyôto Univ.*, 2:228–295, 1963.

[49] H. P. McKean. A free boundary problem for the heat equation arising from a problem in mathematical economics. *Ind. Management Rev.*, 6:32–39, 1965.

[50] H. P. McKean. A probabilistic interpretation of equilibrium charge distributions. *J. Math. Kyôto Univ.*, 4:617–625, 1965.

[51] H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *PNAS*, 56:1907–1911, 1966.

[52] H. P. McKean. An upper bound of the spectrum of Δ on a manifold of negative curvature. *J. Diff. Geom.*, 4:359–366, 1970.

[53] H. P. McKean. Selberg's trace formula as applied to a compact Riemann surface. *Comm. Pure Appl. Math.*, 25:225–246, 1972.

[54] H. P. McKean. Geometry of differential space. *Ann. Prob.*, 1:197–206, 1973.

[55] H. P. McKean. Wiener's theory of non-linear noise. *SIAM-AMS Symp. Appl. Math.*, 6:191–209, 1973.

[56] H. P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piscounov. *Comm. Pure Appl. Math.*, 28:323–331, 1975.

[57] H. P. McKean. Brownian local times. *Adv. Math.*, 16:91–111, 1975.

[58] H. P. McKean. The central limit theorem for Carleman's equation. *Israel J. Math.*, 21:54–92, 1975.

[59] H. P. McKean. An exponential formula for solving Bolzmann's equation for a Maxwellian gas. *J. Comb. Theory.*, 2:358–382, 1975.

[60] H. P. McKean. Fluctuations in the kinetic theory of gases. *Comm. Pure Appl. Math.*, 28:435–455, 1975.

[61] H. P. McKean. Curvature of an ∞-dimensional manifold related to Hill's equation. *Jour. Diff. Geom.*, 17:523–529, 1982.

[62] H. P. McKean. Geometry of KdV (1): addition and the unimodal spectral class. *Rev. Math. Iberoamer.*, 2:235–261, 1986.

[63] H. P. McKean. Geometry of KdV (2): three examples. *J. Stat. Phys.*, 46:1115–1143, 1987.

[64] H. P. McKean. Curvatura integra, handle number, and genus of transcendental curves. *Comm. Pure Appl. Math.*, 44:1057–1066, 1991.

[65] H. P. McKean. Geometry of KdV (3): Determinants and Unimodular Isospectral Flows. *Comm. Pure Appl. Math.*, 45:389–415, 1992.

[66] H. P. McKean. How real is resonance? *Comm. Pure Appl. Math.*, 50:317–322, 1997.

[67] H. P. McKean. Breakdown of a shallow water equation. *Asian J. Math.*, 2:867–874, 1998.

[68] H. P. McKean. Correction to "Breakdown of a shallow water equation". *Asian J. Math.*, 3:3, 1999.

[69] H. P. McKean. A quick proof of Riemann's mapping theorem. *Comm. Pure Appl. Math.*, 52:405–409, 1999.

[70] H. P. McKean. Brownian motion and the general diffusion: scale and clock. *Bachelier Congress*, pages 15–84, 2000.

[71] H. P. McKean. Turbulence without pressure: existence of the invariant measure. *Math. Appl. Anal*, 9:463–468, 2002.

[72] H. P. McKean. Fredholm determinants and the Camassa-Holm hierarchy. *Comm. Pure Appl. Math.*, 56:638–680, 2003.

[73] H. P. McKean. The Liouville correspondence between the Korteweg-de Vries and the Camassa-Holm hierarchies. *Comm. Pure Appl. Math.*, 56:998–1015, 2003.

[74] H. P. McKean. Book reviews: *Riemann surfaces of infinite genus* by J. Feldman, H. Knorrer and E. Trubowitz. *Bull. Amer. Math. Soc.*, 42:79–87, 2005.

[75] H. P. McKean and C. Scovel. Geometry of some simple non-linear differential operators. *Ann. Scuola Norm. Sup. Pisa*, 13:299–346, 1986.

[76] H. P. McKean and I. M. Singer. Curvature and the eigenvalues of the Laplacian. *J. Diff. Geom.*, 1:43–69, 1967.

[77] H. P. McKean and E. Trubowitz. Hill's operator and hyperelliptic function theory in the presence of infinitely many branch points. *Comm. Pure Appl. Math.*, 29:143–226, 1976.

[78] H. P. McKean and E. Trubowitz. Hill surfaces and their theta functions. *Bull. Amer. Math. Soc.*, 84:1042–1085, 1978.

[79] H. P. McKean and P. van Moerbeke. The spectrum of Hill's equation. *Inv. Math.*, 30:217–274, 1975.

[80] H. P. McKean and P. van Moerbeke. Hill and Toda curves. *Comm. Pure Appl. Math.*, 33:23–42, 1980.

[81] F. G. Mehler. Über die Entwicklungen einer Funktion von beliebig vielen Variabeln nach Laplaceschen Funktionen höherer Ordnung. *J. Reine Angew. Math.*, 66:161–176, 1866.

[82] R. C. Merton. Theory of rational option pricing. *Bell J. Econ. Managenent Sci.*, 4:141–183, 1973.

[83] P. Myneni. The pricing of an american option. *Ann. Appl. Prob.*, 2:1–23, 1992.

[84] S. P. Novikov. The periodic problem for the Korteweg-de Vries equation. *Funct. Anal. Appl.*, 8:236–246, 1974.

[85] V. K. Patodi. Curvature and the eigenforms of the Laplace operator. *Jour. Diff. Geom.*, 5:233–249, 1971.

[86] L. Pitt and L. Tran. Local sample path properties of Gaussian fields. *The Annals of Probability*, 7:477–493, 1979.

[87] D. B. Ray. Sojourn times of a diffusion process. *Illinois J. Math.*, 7:615–630, 1963.

[88] S. O. Rice. Mathematical analysis of random noise. *Bell System Tech. J.*, 23:282, 1944.

[89] S. O. Rice. Mathematical analysis of random noise. *Bell System Tech. J.*, 24:46, 1945.

[90] P. Samuelson. Rational theory of warrant pricing. *Ind. Management Rev.*, 6:13–32, 1965.

[91] P. Sarnak. Determinants of Laplacians, heights and finiteness. In P. Rabinowitz, editor, *Analysis Etc.*, pages 601–622. Academic Press, 1990.

[92] H. Schläfli. Ueber die allgemeine Möglichkeit der conformen Abbildung einer von Geraden begrenzten ebenen Figur in eine Halbebene. *J. reine angew. Math.*, 78:63–80, 1874.

[93] R. Schoen, S. T. Yau, and S. Wolpert. Geometric bounds on the low eigenvalues of a compact manifold. *Proc. Symp. Pure Math.*, 36:279–285, 1980.

[94] A. Selberg. Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Indian Math. Soc.*, 20:47–87, 1956.

[95] D. Sullivan and H. P. McKean. Brownian motion and harmonic functions on the class surface of the thrice-punctured sphere. *Adv. Math.*, 51:203–211, 1984.

[96] H. Tanaka and H. P. McKean. Additive fractionals of the Brownian paths. *Mem. Coll. Sci. Kyôto*, 33:479–506, 1961.

[97] H. F. Trotter. A property of Brownian motion paths. *Illinois J. Math.*, 2:425–433, 1958.

[98] S. Venakides. The infinite period limit of the inverse formalism for periodic potentials. *Comm. Pure Appl. Math.*, 41:3–17, 1988.

[99] M. F. Vigneras. Variétés riemanniennes isospectrales et non isométriques. *Ann. Math.*, 112:21–32, 1980.

[100] V. A. Volkonskii. Random time substitution in strong Markov processes. *Teor. Veroyatnost. i Prim.*, 3:332–350, 1958.

[101] N. Wiener. *Cybernetics.* John Wiley & Sons, New York, 1948.

[102] D. Williams. Markov properties of Brownian local time. *Bull. Amer. Math. Soc.*, 75:1035–1036, 1964.

[103] D. Williams. Decomposition of the Brownian path. *Bull. Amer. Math. Soc.*, 70:871–973, 1970.

[104] V. E. Zakharov. *Collected Works.* Dekker, New York & Basel, 1990.

5 Some Words from Three Students

F. Alberto Grünbaum,[1] Pierre van Moerbeke,[2,3] and Victor H. Moll[4]

5.1. My Own Debt to Henry, and Others: *F. Alberto Grünbaum*

Henry writes that he has been very lucky in his mathematical life.

I am sure that the sentiment is shared by the many youngsters who, like myself a long time back, some way or another ended up becoming one of his graduate students.

When I was in my third year of college in Córdoba, Argentina, having just decided to switch from physics to mathematics, we had a short visit from A.P. Calderón. A couple of us took him out for a cup of coffee to get advice as to what to do after graduation. I told him very openly that I wanted to pursue some combination of analysis, physics, and (my new love) probability theory. I had just discovered the book by W. Feller. Calderón was not an impulsive man, but it took him about one second to say "I know the guy for you: Henry McKean".

Two years went by, I had already written a couple of papers with E. Zarantonello and had gone to Buenos Aires to start working with Mischa Cotlar. Cotlar had suggested extending to the multidimensional case some work of I.Gohberg and M.G.Krein. It was just too hard for me and I was going nowhere. Another visitor showed up in Buenos Aires in May '66 : J. J. Kohn. I sat in a class he was teaching and told him about my serious doubts about my progress on this problem as well as my interest in the three topics I had mentioned to Calderón some years earlier. Kohn immediately said "I know the guy for you: Henry McKean".

I remember, or maybe I only made this up, but I think I went to my wife Loli that night and said: either I drop out of mathematics or I try to follow the advice of these two guys. Luck intervened again and in September '66 I arrived at Rockefeller University where Henry had just moved from MIT.

In a million years I could not have made a better choice. For quite a while I was (almost) the only student that Henry had at his new place and going to him with a question usually turned into a few hours of his explaining to me all sorts of things. I was too shy to address him as Henry, or to tell him that I needed to take a break.

I could go on for pages on the many ways in which Henry was an exceptional teacher.

[1]Department of Mathematics, University of California, Evans Hall 970, 94720-3840 Berkeley, CA, USA, grunbaum@berkeley.edu.

[2]Institut de Recherche en Mathématique at Physique, Université catholique de Louvain, Louvain-la-Neuve, Belgium, pierre.vanmoerbeke@uclovain.be.

[3]Department of Mathematics, Brandeis University, Waltham, Massachusetts, USA, vanmoerbeke@geom.ucl.ac.be.

[4]Tulane University, Gibson Hall 404, 70118 New Orleans, Louisiana, USA, vhm@math.tulane.edu.

F.A. Grünbaum et al. (eds.), *Henry P. McKean Jr. Selecta*, Contemporary Mathematicians, DOI 10.1007/978-3-319-22237-0_5

I will just mention two: I was rather impatient, trying to finish my thesis quickly and asked him very early on for an open problem. He suggested looking at the harmonic oscillator forced by white noise, a problem that went back to George Uhlenbeck, the father figure of physics at Rockefeller U. I managed to extend something that Henry had done for the free particle, but just then Henry got a preprint from Eugene Wong from EECS in Berkeley, who had done just what I had done (but better). I felt quite bad but Henry just said: do not worry, work on something else and if nothing works out come back to what you have done and we will make this into a thesis.

He then did something fantastic; he gave me a thick notebook full of problems, one in each page. I could take it home and pick something I liked, with no pressure to work in any particular area.

After a few days I started focusing on the Boltzmann equation and invented my own problem: using the ideas of H. Poincare to relate the full nonlinear flow with its linearized version. This worked out and I sort of had a thesis.

I was also tempted by a problem that I seem to recall was in the notebook: the approach of Mark Kac to the equation of Boltzmann by means of his "propagation of chaos". I ended up writing sort of a second thesis on a rather general way of formulating this problem, but in the end there were two "technical conditions" that I could not verify. Henry could have insisted that I keep working on this, but he had a better suggestion: write this up making it very clear that there were gaps to be filled.

It was clear to me that these were highly technical points that I was most likely never going to make progress on. So with the thesis finished thanks to Henry allowing for this gap to be left open, I left this nice program of Kac alone and went to do very different things, never looking back.

I am incredibly happy to report to Henry (and to myself too) that this gap in my approach has been recently taken care of by S. Mischler and C. Mouhot in the paper "Kac's program in kinetic theory", Invent. Math. (2013) 193:1–47. These young authors, strongly influenced by P.L. Lions and C. Villani, may be giving me too much credit in their paper. But for someone who had left the field a long time ago, this is very nice feeling.

Coming back to the matter of luck, Henry was not too wise or lucky when a (not so) few years ago, having been approached by some publisher to put some of his papers together, he asked me to take charge of this project. After a year or so of going nowhere, it occurred to me to ask Pierre van Moerbeke to join me in this adventure. We moved very slowly for a very long time. About July 2013 I had another good idea: I asked Victor H. Moll to join us in this effort, and now less than a year later we are done. I hope that Henry is happy that we have finished, with the great contributions of two people that care deeply about him: David Williams and Hermann Flaschka. He knows that I am at least as happy as he is. Apologies for being so slow.

Readers of this collection may wonder how the selection of papers was made. The answer is simple: after a few arguments involving Pierre, Henry and myself, we usually left it to Henry to decide what to put in. For someone with such a rich collection of gems having to settle on a few papers was not easy at all.

I hope that people looking at this book will get a glimpse of the breadth, depth and joy that is typical of Henry's work. For the lucky ones, like me, the real pleasure has always been talking to Henry and trying to share and practice his vision of mathematics as the engine and the language for the many very different fields he has enriched. Equally important are the many lives he has molded in very positive ways. I consider myself incredibly lucky to be able to call him my teacher and my friend. I am sure this is true for the many people that love him.

5.2. Reminiscences: *Pierre van Moerbeke*

I remember very vividly my first encounter with Henry McKean on the ninth floor of South Lab (Rockefeller University) summer 1968 on my first visit to New York and the US. I had been assigned to make notes of lectures given by Mark Kac at the CRM in Montreal that same summer. Henry's unusual mathematical style, taste for good problems, originality and kindness attracted me. My decision to stay on was made instantaneously! Me, coming from a Bourbaki stronghold: woe betide you for non-orthodoxy!

I started my PhD-work with Henry on a famous appendix he wrote to a paper of Paul Samuelson on the warrant pricing problem. Henry taught me a great deal on Brownian motion, local time, Green functions, PDE's and functional analysis, etc. ..., but never instructed me what to do, except occasionally what not to do. He also taught me the value of alternative views; he was always supportive, because for a young apprentice there are so many hurdles that depression and loss of motivation are a constant danger. I was fortunate to have met such an astonishingly generous advisor. Henry's office door on the 9th floor was always open; he was always at work and available, he seemed undisturbed in spite of the many blasts next door, rocking the whole building.

Having written a thesis on a free boundary problem, which was rooted in a financial question, I was predestined to start a career in financial mathematics. But then Ludwig Faddeev and others came to explain to us, at Rockefeller, about KdV and the motion on spectral data for the one-dimensional Schrodinger equation. Soon after, Henry and I embarked into an exciting project on the resolution of the periodic KdV equation, in terms of hyperelliptic theta functions; in more geometrical terms: the periodic KdV equation with finite band initial condition can be linearized on the Jacobi variety of a hyperelliptic Riemann surface. That year 1973 I commuted regularly between Princeton and the Courant Institute.

This collaboration was for me a fantastic learning experience with Henry in the field of Riemann surfaces and in hands-on algebraic geometry: it kept me from pursuing a career in financial math. Rather it launched me in an exciting integrable adventure which via 'Random Matrices' brought me back in recent years- to probabilistic models utilizing the tools of my youth: it gave me again the joy of reading 'stochastic integrals', the 'Ito-McKean' and Henry's beautiful papers on diffusion processes. Henry's style and taste left on us (and in turn on the next generation) an indelible mark!

I have unforgettable memories from Henry's logging big timber at South-Landaff, New Hampshire; dinners at 186 E 93rd street; candlelight math discussions at the Hotel des Grandes Ecoles; cheese fondue at Lenk, with Henry playing the blues on his guitar; canoeing in the Pine Barrens; garden suppers at Orange street; exploring the wetlands and bird watching at Essex,... Henry made us discover the beauty of New-England!

This volume contains only a few of Henry McKean's works. There are many other McKean gems lying out there, which do not appear in these limited Selecta. This volume truly reflects Henry's exquisite choice of problems, broad and inspirational insights, terrific mathematical power and groundbreaking results, occasionally surrealistic.

These Selecta, a 7-year project, was to be published in the series of Drs Klaus and Alice Peters. After the unfortunate merger with Taylor and Francis, Anne Boutet de Monvel came at our rescue and kindly put us in contact with Executive Editor Dr Thomas Hempfling at Birkhäuser-Springer, who very enthusiastically encouraged us to publish this volume in the present series. Thank you, Anne!

We are very grateful to our co-editor Victor H. Moll who put in an enormous amount of work wrapping up this project in such a short amount of time.

I also would like to thank David Williams charming tribute to Henry McKean and also Hermann Flaschka for a sweeping essay on McKean's huge contribution to integrable systems.

5.3. More Than Words Can Express: *Victor H. Moll*

The setting: September 16, 1980, my first class of graduate Complex Variables at the Courant Institute. Enters Henry McKean.

The beginning of my graduate education was not very standard. A couple of years before I had finished my undergraduate degree at Universidad Santa Maria, Valparaiso, Chile. At that time it was still possible to obtain a position in a regional university without a graduate degree. So I did. My advisor, Luis Salinas C., always insisted that I should go abroad for a Ph.D. During the chilean winter, Eugene Trubowitz visited Chile and, being able to speak English, I became part of the local welcoming committee. My visit to the Courant was arranged between Luis and Eugene. At that time I was interested in Number Theory, so going to Courant was not the obvious choice. Being so ill-informed is sometimes a blessing.

I thought that my background was very good and I specially liked Complex Variables. So when Henry taught for the first class (in the usual Courant schedule: one 2-h class per week), I enjoyed the review. The next week, when the material somehow newer and the speed did not change, it was time to panic. My favorite recollection of that class was when Henry gave us the midterm exam. It contained a problem involving Phragmen-Lindelöf. A student stated that this had not been covered and Henry, paused for a moment, and claimed 'I do not recall the exact statement, but this is how the proof goes' and then he proceeded to present a beautiful complete argument. Having been educated in the third-world version of Bourbaki, this was very illuminating. I thought that this was the way Mathematics should be presented.

After my graduation, Henry taught a course in Elliptic Functions. I commuted from Philadelphia every Monday to attend his lectures. It was simply a fascinating subject and Henry's presentation of it made it even better. I worked on the class notes, trying to fill all the gaps and kept Henry informed of what I was doing. It was a very exciting moment when Henry, waiting on line for coffee at the International Congress in Berkeley (1986), asked me if I wanted to convert this into a joint book. My state of complete happiness did not last very long: he thought for a moment and said to me 'It is not a good idea, you would have a book, and not tenure'. This was absolutely correct. In the coming years, while aspiring for tenure, I continue to work on the notes and when the moment came, I called him. My sabbatical at Courant was spent dealing with the finishing touches of the book. It was an incredible experience. Specially receiving his hand-written notes with comments on each chapter. I can actually read his hand-writing.

Through the years, I have had a chance to visit him periodically. The only change that I have noticed, is that his door is no longer open: it has to be closed to be able to smoke. My mathematical interests have moved in a different direction, but telling him about a result that interests me, is always rewarding. The mathematics that I learned form him is everywhere in what I have tried to pursue. One day, a conversation with a graduate student, developed into a map of rational functions that preserve the integral. The beauty of Landen, learned from Henry in his elliptic functions class, was the key idea to place our work in the right context.

My involvement in this project came late. I was aware that Alberto and Pierre were in charge of Henry's selected papers, but I knew no details. One day I decided to email Alberto to find out the status of the project. He told me that he had a large collection of files that needed editing. I offered my help and I was lucky that they accepted it. This effort is a small way of me saying: *Thanks Henry.*

[6] Curvature and the Eigenvalues of the Laplacian

[6] Curvature and the Eigenvalues of the Laplacian. *Jour. Diff. Geometry* **1** (1967), 43–69.

Curvature and the Eigenvalues of the Laplacian

6

H. P. McKean, Jr.[1,2] and I. M. Singer[3]

6.1. Introduction

A famous formula of H. Weyl [17] states that if D is a bounded region of R^d with a piecewise smooth boundary B, and if $0 > \gamma_1 \geq \gamma_2 \geq \gamma_3 \geq$ etc. $\downarrow -\infty$ is the spectrum of the problem

$$\Delta f = \left(\partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_d^2\right) f = \gamma f \quad \text{in } D, \tag{6.1.1a}$$

$$f \in C^2(D) \cap C(\overline{D}), \tag{6.1.1b}$$

$$f = 0 \quad \text{on } B, \tag{6.1.1c}$$

then

$$-\gamma_n \sim C(d)(n/\text{vol } D)^{2/d} \quad (n \uparrow \infty), \tag{6.1.2}$$

or, what is the same,

$$Z \equiv \text{sp}\, e^{t\Delta} = \sum_{n\geq 1} \exp\left(\gamma_n t\right) \sim (4\pi t)^{-d/2} \times \text{vol}\, D \quad (t \downarrow 0), \tag{6.1.3}$$

where $C(d) = 2\pi[(d/2)!]^{d/2}$.

A. Pleijel [13] and M. Kac [6] took up the matter of finding corrections to (6.1.3) for plane regions D with a finite number of holes. The problem is to find how the spectrum of Δ reflects the shape of D. Kac puts things in the following amusing language: thinking of D as a drum and $0 < -\gamma_1 < -\gamma_2 \leq$ etc. as its fundamental tones, *is it possible, just by listening with a perfect ear, to hear the shape of D*? Weyl's estimate (6.1.2) shows that you can hear *the area of D*. Kac proved that for D bounded by a broken line B,

$$\begin{aligned} Z = {} & \frac{\text{area}\, D}{4\pi t} - \frac{\text{length}\, B/4}{\sqrt{4\pi t}} \\ & + \text{the sum over the corners of } \frac{\pi^2 - \gamma^2}{24\pi\gamma} + \text{o}(1) \quad (t \downarrow 0), \end{aligned} \tag{6.1.4a}$$

[1]Rockefeller University, New York, NY, USA.

[2]Communicated April 6, 1967. The partial support of the National Science Foundation under NSF GP-4364 and NSF GP-6166 is gratefully acknowledged.

[3]Massachusetts Institute of Technology, Cambridge, MA, USA.

$0 < \gamma < 2\pi$ being the inside-facing angle at the corner,[4] *esp.*, you can hear *the perimeter* of such D. By making the broken line B approximate to a smooth curve, Kac was led to conjecture

$$Z = \frac{\text{area}}{4\pi t} - \frac{\text{length}/4}{\sqrt{4\pi t}} + \frac{1}{6}(1-h) + \mathrm{o}(1) \quad (t \downarrow 0) \tag{6.1.4b}$$

for regions D with smooth B and $h < \infty$ holes, and was able to prove the correctness of the first two terms. This jibes with an earlier conjecture of A. Pleijel and suggests that you can hear *the number of holes*. Equation (6.1.4b) will be proved below in a form applicable both to open manifolds with compact boundary and to closed manifolds.

Given a closed d-dimensional, smooth Riemannian manifold M with metric tensor $g = (g_{ij})$, let Δ be the associated Laplace-Beltrami operator:

$$\Delta = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x_i} g^{ij} \sqrt{\det g} \frac{\partial}{\partial x_j},$$

where $(g^{ij}) = g^{-1}$, and let $0 = \gamma_0 > \gamma_1 \geq \gamma_2 \geq$ etc. $\downarrow -\infty$ be its spectrum. Define also the scalar curvature K at a point of M (= the negative of the spur $\sum_{i<j} R^{ij}_{ij}$ of the Ricci tensor) and the partition function $Z \equiv \mathrm{sp}\, e^{t\Delta} = \sum \exp(\gamma_n t)$. Then, as will be proved in Sects. 6.4 and 6.7,

$$\begin{aligned}(4\pi t)^{d/2} Z = {} & \text{the (Riemannian) volume of } M \\ & + \frac{t}{3} \times \text{ the curvatura integra } \int_M K \\ & + \frac{t^2}{180} \int_M (10A - B + 2C) + \mathrm{o}\left(t^3\right),\end{aligned} \tag{6.1.5a}$$

where $\int_M$ stands for the integral relative to the Riemannian volume element $\sqrt{\det g}\, dx$, and A, B, C stand for a particular basis of the space of polynomials of degree 2 in the curvature tensor R which are invariant under the action of the orthogonal group (see (6.7.2)); $\mathrm{o}(t^3)$ cannot be improved. For $d = 2$, $10A - B + 2C = 12K^2$, and by an application of the classical Gauss-Bonnet formula for the Euler characteristic E of $M (2\pi E = \int_M K)$, (6.1.5a) simplifies to

$$Z = \frac{\text{area}}{4\pi t} + \frac{E}{6} + \frac{\pi t}{60} \int_M K^2 + \mathrm{o}\left(t^2\right), \tag{6.1.5b}$$

esp., the Euler characteristic of M is audible.

Consider now an open d-dimensional manifold D with compact $(d-1)$-dimensional boundary B, $\overline{D} = D \cup B$ being endowed with a smooth Riemannian geometry, and let $0 > \gamma_1^- \geq \gamma_2^- \geq$

[4] Kac [6] expresses the corner correction $(\pi^2 - \gamma^2)/24\pi\gamma$ as complicated integral. D. B. Ray [private communication] derived it by a simpler argument, beginning with the Green function G for $s - \Delta (s > 0)$ expressed as a Kantorovich-Lebedev transform

$$\begin{aligned}G(A, B) = x^{-2} \int_0^\infty & dx K_{\sqrt{-1}x}(\sqrt{s}a) K_{\sqrt{-1}x}(\sqrt{s}b) \\ & \times \left[\cosh\left(\pi - |\alpha - \beta|\right)x - \frac{\sinh \pi x}{\sinh \gamma x} \cosh(\gamma - \alpha - \beta)x + \frac{\sinh(\pi - \gamma)x}{\sinh \gamma x} \cosh(\alpha - \beta)x\right],\end{aligned}$$

in which $A = ae^{\sqrt{-1}\alpha}$, $B = be^{\sqrt{-1}\beta}$, and K is the usual modified Bessel function. The corner correction $(\pi^2 - \gamma^2)/24\pi\gamma$ follows easily, and this jibes with Kac's integral upon applying Parseval's formula to the latter.

etc. $\downarrow -\infty$ and $0 = \gamma_0^+ > \gamma_1^+ \geq \gamma_2^+ \geq$ etc. $\downarrow -\infty$ be the spectra of

$$\Delta^- = \Delta \mid C^\infty(\overline{D}) \cap (u : u = 0 \text{ on } B),$$
$$\Delta^+ = \Delta \mid C^\infty(\overline{D}) \cap (u : u^\bullet = 0 \text{ on } B),$$

where $\bullet$ stands for differentiation in the inward-pointing direction perpendicular to B.

Bring in also the mean curvature J at a point of B (= double the spur of the second fundamental form) and the partition function $Z^\pm \equiv \operatorname{sp} e^{t\Delta\pm} = \sum \exp(\gamma_n^\pm t)$. Then, as will be proved in Sect. 6.5,

$$\begin{aligned}(4\pi t)^{d/2} Z^\pm &= \text{the (Riemannian) volume of } D \\ &\pm \frac{1}{4}\sqrt{4\pi t} \times \text{the (Riemannian) surface area of } B \\ &+ \frac{t}{3} \times \text{the curvatura integra } \int_D K \\ &- \frac{t}{6} \times \text{the integrated mean curvature } \int_B J + \mathrm{o}\left(t^{3/2}\right),\end{aligned} \tag{6.1.6}$$

where $\int_B$ stands for the integral over B relative to the element of Riemannian surface area; $\mathrm{o}(t^{3/2})$ cannot be improved. Kac-Pleijel's conjecture (6.1.4b) for a plane region D with smooth boundary B and $h < \infty$ holes is obtained from (6.1.6) and the Gauss-Bonnet formula ($\int_M K + \int_B J = 2\pi \times$ the Euler characteristic) for the closed manifold $M =$ *the double of* D upon noting that the Euler characteristic of the handle-body M is just $2(1-h)$.

The estimates leading to (6.1.5) and (6.1.6) will be proved not just for Δ but for any smooth elliptic partial differential operator of degree 2 (Sects. 6.2, 6.3, 6.4 and 6.5), and some additional comments will be made about $Z = \operatorname{sp} e^{t\Delta}$ for Δ acting on exterior differential forms (Sect. 6.6). The basic idea, due to Kac, is to make a *pointwise* estimate of the pole of the elementary solution of $\partial u/\partial t = \Delta u$ and then to integrate over M to get an estimate of $Z = \operatorname{sp} e^{t\Delta}$. The *curvatura integra* coefficient in (6.1.5a) is computed directly in Sect. 6.4 and then re-computed (for Δ only) in Sect. 6.7 using more sophisticated algebraic ideas about differential invariants of the orthogonal group. A list of open problems is placed at the end of the paper (Sect. 6.8).

The new results of this paper are mainly for the case of manifolds with boundary. For a closed manifold, N. G. de Bruijn [private communication] obtained the *curvatura integra* coefficient independently as did V. Arnold [private communication from M. Berger]. Berger also kindly communicated his formula for the next coefficient, which suggested the approach in Sect. 6.7. Berger's results for closed manifolds can be found in [1]. His method is different from ours, but we arrive at the same formula for the coefficient of t^2; his norms τ^2, $|\rho|^2$, and $|R|^2$ are equal to our $4A$, B, and $2C$ respectively.

It is a pleasant duty to thank M. Kac for suggesting this problem and for a number of stimulating conversations about it. Thanks are also due to T. Kotake for help with the Levi sums of Sect. 6.3.

6.2. Manifolds and Elliptic Operators

Consider a closed, d-dimensional, smooth manifold M and let $Q : C^\infty(M) \to C^\infty(M)$ be an elliptic partial differential operator of degree 2, with $Q(1) = 0$. On a patch $U \subset M$, Q can be expressed as

$$Q = a^{ij}\partial^2/\partial x_i \partial x_j + b^i \partial/\partial x_i \equiv a\partial^2 + b\partial$$

with coefficients $a = (a^{ij})$ and $b = (b^i)$ from $C^\infty(U)$. By changing the sign of Q if necessary, we can take the quadratic form based upon a as positive ($\sum a^{ij} y_i y_j > 0,\ y \neq 0$), and under a change of local coordinates $x \to \overline{x}$ with Jacobian c, a transforms according to the rule $\overline{a} = cac^*$, so $g = a^{-1}$ transforms like a Riemannian metric tensor. M is now endowed with this Riemannian geometry, and Q is re-expressed as the sum of the associated Laplace-Beltrami operator Δ plus a part of degree 1:

$$Q = \Delta + h\partial, \quad h\partial = h^i \frac{\partial}{\partial x_i}, \ \Delta = \frac{1}{\sqrt{\det g}} \frac{\partial}{\partial x_i} g^{ij} \sqrt{\det g} \frac{\partial}{\partial x_j}.$$

Because Δ does not depend upon the choice of local coordinates, $h\partial$ is a vector field.

Δ is symmetric ($\int u\Delta v = \int v\Delta u$) and non-positive ($\int u\Delta u \le 0$) relative to the Riemannian volume element $\sqrt{\det g}\,dx$, where $\int f = \int_M f$ always means $\int_M f\sqrt{\det g}\,dx$. Q enjoys the same properties relative to *some* volume element $e^w\sqrt{\det g}\,dx$ if and only if the vector field $h\partial$ is conservative; this is the same as to say that the exterior differential 1-form dual to this field is an exact differential ($= dw$), as is plain from the fact that, for a patch U and compact u and $v \in C^\infty(U)$,

$$\int_U (uQv - vQu)e^w = \int_U (u \operatorname{grad} v - v \operatorname{grad} u)\big(h - g^{-1} \operatorname{grad} w\big)$$

cannot vanish unless $h = g^{-1} \operatorname{grad} w$ (Nelson [12]). Here, grad $= (\partial/\partial x_1, \dots, \partial/\partial x_d)$.

Consider, next, the elementary solution $e = e(t,x,y)$ of $\partial u/\partial t = Qu$ computed relative to the volume element $\sqrt{\det g}\,dx$ and recall the following facts:

(6.2.1a) $$0 < e \in C^\infty\big[(0,\infty) \times M^2\big],$$

(6.2.1b) $$\partial e/\partial t = Q_x e = Q_y^* e,$$

(6.2.1c) $$\int e\sqrt{\det g}\,dy = 1,$$

(6.2.1d) $$\lim_{t\downarrow 0} t^{-1} \lg e = -\frac{1}{4}[xy]^2,$$

where Q^* is the dual of Q relative to $\sqrt{\det g}\,dx$, and $[xy]$ is the Riemannian distance between x and y; see [15] for (6.2.1d) and [10] for the rest.

Now if Q is symmetric relative to the volume element $e^w\sqrt{\det g}\,dx$, then $e(t,x,y)\exp[-w(y)]$ is symmetric in x and y, and since its spur $Z = \int e(t,x,x)$ converges, $e^{tQ} : f \to \int ef$ is a compact mapping of the (real) Hilbert space $H = L^2[M, e^w\sqrt{\det g}\,dx]$. This implies that Q has a discrete spectrum

(6.2.2) $$0 = \gamma_0 > \gamma_1 \ge \gamma_2 \ge \text{etc.} \quad \downarrow -\infty$$

with corresponding eigenfunctions $f_n \in C^\infty(M)$ forming a unit perpendicular basis of H; in addition,

$$e = \sum_{n\ge 0} \exp\big(\gamma_n t\big) f_n \otimes f_n$$

with uniform convergence on compact figures of $(0,\infty)\times M^2$, and the spur Z is easily evaluated as (see for example [10])

(6.2.3) $$Z = \int \sum_{n\ge 0} \exp\big(\gamma_n t\big) f_n^2 e^w = \sum_{n\ge 0} \exp\big(\gamma_n t\big).$$

Kac's method for the proof (6.1.4a) is now imitated to obtain (6.1.5a): one estimates the pole $e(t,x,x)$ locally and then integrates over M. This is done in Sects. 6.3 and 6.4 using a method of E. E. Levi; the actual estimate is just as easy for the general Q, so the condition that the vector field $h\partial$ be conservative is not insisted upon.

Now let $Q = \Delta + h\partial$ be defined on a smooth open, d-dimensional manifold D with smooth, compact, $(d-1)$ -dimensional boundary B, suppose that $g = a^{-1}$ is positive and smooth on the whole of $\overline{D}$ so that it induces a nice Riemannian geometry on $\overline{D}$ and let the vector field $h\partial$ be smooth on $\overline{D}$ too. Both $Q^- = Q \mid C^\infty(\overline{D}) \cap (u : u = 0 \text{ on } B)$ and $Q^+ = Q \mid C^\infty(D) \cap (u : u^\bullet = 0 \text{ on } B)$, $\bullet$ standing for differentiation in the inward-pointing direction perpendicular to B, have nice elementary solutions $e = e^\pm$ subject to

(6.2.4a) $$0 \le e \in C^\infty\big[(0,\infty) \times \overline{D}^2\big],$$

(6.2.4b) $$\frac{\partial e}{\partial t} = Q_x e = Q_y^* e,$$

Q^* being the dual of Q relative to $\sqrt{\det g}\, dx$,

(6.2.4c−) $$\int_D e^- \uparrow 1 \quad (t \downarrow 0),$$

(6.2.4c+) $$\int_D e^+ = 1,$$

(6.2.4d) $$\overline{\lim_{t\downarrow 0}}\, t^{-1} \lg e \le -\frac{1}{4}[xy]^2,$$

(6.2.4e−) $$e^- = 0 \text{ on } B \times D,$$

(6.2.4e+) $$e^{+\bullet} = 0 \text{ on } B \times D.$$

For Q symmetric relative to some volume element, the spectra are as before except at the upper end:

(6.2.5a) $$0 > \gamma_1^- \ge \gamma_2^- \ge \text{etc.} \quad \downarrow -\infty,$$

(6.2.5b) $$0 = \gamma_0^+ > \gamma_1^+ \ge \gamma_2^+ \text{etc.} \quad \downarrow -\infty,$$

and the formula for the partition function still holds:

(6.2.6) $$Z^\pm = \int_D e^\pm(t,x,x) = \sum \exp\left(\gamma_n^\pm t\right),$$

so that (6.2.6) can likewise be derived by estimating the pole $e^\pm(t,x,x)$.

6.3. Levi's Sum for the Elementary Solution

Given closed M and $Q = \Delta + h\partial$ as above, one can express the elementary solution $e = e(t,x,y)$ of $\partial u/\partial t = Qu$ by means of a sum due to E. E. Levi; this computation has been carried out in a very careful manner by S. Minakshisundaram [10], but it will be helpful to indicate the idea in a form suited to the present use.

Consider a little closed patch U of M with smooth $(d-1)$-dimensional boundary B, view U as part of R^d, extend $Q' = Q \mid U$ to the whole of R^d in such a way that the coefficients of the extension belong to $C^\infty(R^d)$ and $Q' = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_d^2$ near ∞, let e' be the elementary solution of $\partial u/\partial t = Q'u$, and let us prove that *inside* $U \times U$,

(6.3.1) $$|e' - e| \le \exp(-\text{constant}/t) \quad (t \downarrow 0)$$

with a positive constant depending only upon the distance to B.

Proof. Bring in the elementary solution e'' of $\partial u/\partial t = Qu$ subject to $u = 0$ on B. Given a compact function $v \in C^\infty(U)$, $u = \int (e'' - e)v$ solves $\partial u/\partial t = Qu$ on $(0,\infty) \times U$ and tends to

0 uniformly on $\overline{U}$ as $t \downarrow 0$. But this means that in the figure $[0,t] \times \overline{U}$, $|u|$ peaks on $[0,t] \times B$, so that by an application of the estimate of Varadhan [(6.2.1d), (6.2.4d)],

$$|u| \le \max_{[0,t]\times B} \left| \int (e'' - e)v \right| \le \exp\big(-R^2/5t\big)\|v\|_1,$$

R being the shortest (Riemannian) distance from $(v \neq 0) \subset U$ to B. The rest of the proof is self-evident.

Because of (6.3.1), it is permissible, for the estimation of the pole $e(t,x,x)$ up to an exponentially small error, to replace M by R^d and to suppose that $Q = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_d^2$ far out; this modification of the problem is now adopted.

Define now Q^0 to be Q with its coefficients *frozen* at $y \in R^d$, and let $e^0(t,x,y)$ be the elementary solution of $\partial u/\partial t = Q^0 u$ evaluated at $t > 0$, $x \in R^d$, and the *same* point $y \in R^d$ at which the coefficients of Q^0 are computed:

$$e^0(t,x,y) = (4\pi t)^{-d/2} \exp\Big(-\big|a^{0\frac{1}{2}}\big(y - x - b^0 t\big)\big|^2/4t\Big) \tag{6.3.2}$$

with an obvious notation. Because of (6.2.1b), (6.2.1c) and (6.2.1d),

$$\begin{aligned} e(t,x,y) - e^0(t,x,y) &= \int_0^t ds \frac{\partial}{\partial s} \int_{R^d} e(s,x,\cdot)e^0(t-s,\cdot,y) \\ &= \int_0^t ds \int_{R^d} \big(e^0 Q^* e - e Q^0 e^0\big) \\ &= \int_0^t ds \int_{R^d} e(s,x,\cdot)\big(Q - Q^0\big)e^0(t-s,\cdot,y), \end{aligned} \tag{6.3.3a}$$

in short,

$$e = e^0 + e\sharp f, \tag{6.3.3b}$$

with $\sharp$ denoting the composition on the final line of (6.3.3a) and

$$f = \big(Q - Q^0\big)e^0(t-s,x,y).$$

Upon iteration, this identity produces the (formal) sum for e:

$$e = e^0 + \sum_{n\ge1} e^0\sharp f\sharp\cdots\sharp f \quad (n\text{-fold}). \tag{6.3.4}$$

Actually this formal sum converges to e uniformly on compact figures of $(0,\infty) \times R^{2d}$; the main point is that since

$$\begin{aligned} &Q = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_d^2 \quad \text{near } \infty, \\ &|f| \le c_1 \left(\frac{|x-y|^3}{t^2} + \frac{|x-y|}{t} + 1 \right) t^{-d/2} \exp\big(-c_2|x-y|^2/t\big) \\ &\quad \le c_3 t^{-(d+1)/2} \exp\big(-c_4|x-y|^2 t\big), \end{aligned} \tag{6.3.5a}$$

$c_1, \ldots, c_4$ standing for positive constants, as can easily be verified by a direct computation, and this leads easily to the bound

$$\big|e^0\sharp f\sharp\cdots\sharp f\big| \le c_5^n \big[(n/2)!\big]^{-1} t^{(n-d)/2} \exp\big(-c_6|x-y|^2/t\big), \tag{6.3.5b}$$

Accordingly, the formal sum (6.3.4) converges rapidly to a nice function e of magnitude

$$|e| \le \sum_{n\ge0} \frac{\big(c_5\sqrt{t}\big)^n}{(n/2)!} t^{-d/2} \exp\big(-c_6|x-y|^2/t\big), \tag{6.3.6}$$

which satisfies (6.3.3b). A moment's reflection shows that e is an elementary solution of $\partial u/\partial t = Qu$. But $\partial u/\partial t = Qu$ has only 1 elementary solution subject to (6.3.6), so $e =$ (6.3.4) is it. This is proved by noticing that any elementary solution subject to (6.3.6) is also a solution of (6.3.3b), and then proving that (6.3.3b) + (6.3.6) has just 1 solution.

6.4. Estimation of the Pole

Levi's sum (6.3.4) can now be used to estimate the pole $e(t,x,x)$ for $t \downarrow 0$, up to terms of magnitude $t^{1-d/2}$:

$$(4\pi t)^{d/2} e(t,x,x) = 1 + \frac{t}{3}K - \frac{t}{2}\operatorname{div} h - \frac{t}{4}|h|^2 + \mathrm{o}\left(t^2\right), \tag{6.4.1}$$

in which K is the scalar curvature (= the negative spur $\sum_{i<j} R^{ij}_{ij}$ of the Ricci-tensor), $\operatorname{div} h$ is the (Riemannian) divergence $[= (\det g)^{-\frac{1}{2}}\partial h^i (\det g)^{\frac{1}{2}}/\partial x_i]$, and $|h|$ is the (Riemannian) length $(= g_{ij}h^i h^j)$. Equation (6.4.1) can be integrated over M to get an estimate of $Z = \int e(t,x,x)$ (since $\int \operatorname{div} h = 0$):

$$(4\pi t)^{d/2} Z = \int 1 + \frac{t}{3}\int K - \frac{t}{4}\int |h|^2 + \mathrm{o}\left(t^2\right), \tag{6.4.2}$$

esp., if $Q = \Delta$, then $h = 0$ and (6.4.2) = (6.1.5a). A little extra attention to the proof, which is left to the industrious reader, shows the existence of an expansion

$$(4\pi t)^{d/2} e(t,x,x) = 1 + k_1 t + k_2 t^2 + \cdots + k_n t^n + \mathrm{o}\left(t^{n+1}\right). \tag{6.4.3}$$

This was proved by S. Minakshisundaram [10] for $Q = \Delta$; the only novel point is the evaluation $k_1 = K/3 - (\operatorname{div} h)/2 - |h|^2/4$. k_2 is computed in Sect. 6.7, using a more sophisticated method.

Proof of (6.4.1)**.** e can be replaced by the sum (6.3.4), and the terms of index $n \geq 4$ can be neglected in view of (6.3.5b). Put $x = 0$ for simplicity and bring in new coordinates on R^d coinciding with the old near ∞ and such that

$$g_{ij}(x) = \delta_{ij} + \frac{1}{3}R_{ikjl}x_k x_l + \mathrm{o}\left(|x|^3\right) \quad \text{near } 0, \tag{6.4.4}$$

R being the curvature tensor associated with g; this is accomplished by applying the exponential map to the tangent space at 0 to obtain coordinates on a patch and then fixing things up outside [3, Chapter 10]. An estimate of $f = (Q - Q^0)e^0(t-s,x,y)$ finer than (6.3.5a) is now possible:

$$\left|f(t,x,y)\right| \leq c_1\left(\frac{|x|\,|y-x|^3}{t^2} + \frac{|x|\,|y-x|}{t} + 1\right)\exp\left(-c_2|x-y|^2/t\right), \tag{6.4.5}$$

where c_1, c_2, etc. stand for positive constants. This is used to prove

$$\begin{aligned} \left|e^0 \sharp f \sharp f\right| \leq & \int_0^t ds_1 \int_0^t ds_2 \int_{R^{2d}} \frac{c_3 e^{-c_4|x|^2/(t-s_1)}}{\left(t-s_1\right)^{d/2}} \\ & \times \left(\frac{|x|\,|y-x|^3}{t^2} + \frac{|x|\,|y-x|}{t} + 1\right)\frac{e^{-c_4|x-y|^2/(s_1-s_2)}}{\left(s_1-s_2\right)^{d/2}} \\ & \times \left(\frac{|y|^4}{t^2} + \frac{|y|^2}{t} + 1\right)\frac{e^{-c_4|y|^2/s_2}}{s_2^{d/2}} \\ \leq & \; c_5 t^{-d/2}\int_0^t ds_1 \int_0^{s_1} ds_2 \sqrt{\frac{t-s_1}{s_1-s_2}} = c_6 t^{2-d/2}, \end{aligned} \tag{6.4.6a}$$

and the similar but easier bound

$$\left|e^0\sharp f\sharp f\sharp f\right| \le c_7 t^{2-d/2}, \tag{6.4.6b}$$

which shows that, up to terms of magnitude $\le$ constant $\times t^{2-d/2}$, one is left with

$$\begin{aligned} e(t,0,0) &= e^0(t,0,0) \\ &\quad + \int_0^t ds \int_{R^d} e^0(t-s,0,x)\big(Q-Q^0\big)e^0(s,x,0)\sqrt{\det g}\,dx. \end{aligned} \tag{6.4.7}$$

A moment's reflection will convince the reader that, up to the desired precision, the integrand $e^0(t-s,0,x)(Q-Q^0)e^0(s,x,0)\sqrt{\det g}$ can be replaced by the product of a factor $1+$ a linear function f of $x + \mathrm{o}(t) + \mathrm{o}(|x|^2)$ and the expression

$$\begin{aligned} &\frac{e^{-|x|^2/4(t-s)}}{[4\pi(t-s)]^{d/2}}\left[\frac{1}{2}\frac{\partial^2 g^{ij}}{\partial x_k \partial x_l}(0)x_k x_l \frac{\partial^2}{\partial x_i \partial x_j}\right. \\ &\qquad + \left(\frac{\partial}{\partial x_k}\frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x_i} g^{ij}\sqrt{\det g}\right)(0)x_k\frac{\partial}{\partial x_j} \\ &\qquad \left. + \frac{\partial h^i}{\partial x_k}(0)x_k\frac{\partial}{\partial x_i}\right]\frac{e^{-|x|^2/4s}}{(4\pi s)^{d/2}} \\ &= (4\pi t)^{-d/2}\frac{e^{-|x|^2/4r}}{(4\pi r)^{d/2}}\left[\frac{1}{2}\frac{\partial^2 g^{ij}}{\partial x_k\partial x_l}(0)x_k x_l\left(\frac{x_i x_j}{4s^2} - \frac{\delta_{ij}}{2s}\right)\right. \\ &\qquad \left. - \left(\frac{\partial}{\partial x_k}\frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x_i}g^{ij}\sqrt{\det g}\right)(0)\frac{x_k x_j}{2s} - \frac{\partial h^i}{\partial x_k}(0)\frac{x_i x_k}{2s}\right], \end{aligned} \tag{6.4.8}$$

where $r = s(t-s)/t$. Now the factor alluded to above (6.4.8) can be replaced by 1, since $f \times$ (6.4.8) integrates to 0 while the last 2 terms contribute $\le c_8 t^{2-d/2}$. Consequently, up to the desired precision,

$$\begin{aligned} t^{-1}(4\pi t)^{d/2}e^0\sharp f &= t^{-1}\int_0^t ds \int_{R^d} (6.4.8)\,dx \\ &= \frac{1}{2}\frac{\partial^2 g^{ij}}{\partial x_k \partial x_l}(0) \times (0, 1/3,\ \text{or } 1 \text{ according as } ijkl \text{ comprises} \\ &\qquad\qquad \le 1 \text{ pair, } 2 \text{ unequal pairs, or } 2 \text{ equal pairs}) \\ &\quad - \frac{1}{4}\frac{\partial^2 g^{ij}}{\partial x_k\partial x_l}(0)\delta_{kl}\delta_{ij} \\ &\quad - \frac{1}{2}\left(\frac{\partial}{\partial x_k}\frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x_i}g^{ij}\sqrt{\det g}\right)(0)\delta_{kj} - \frac{1}{2}\frac{\partial h^i}{\partial x_k}(0)\delta_{ki} \\ &= \frac{1}{6}\frac{\partial^2 g^{ii}}{\partial x_j^2} + \frac{1}{3}\frac{\partial^2 g^{ij}}{\partial x_i \partial x_j}(\text{summed for } i \ne j) + \frac{1}{2}\frac{\partial^2 g^{ii}}{\partial x_i^2} \\ &\quad - \frac{1}{4}\frac{\partial^2 g^{ii}}{\partial x_j^2} - \frac{1}{2}\frac{\partial}{\partial x_j}\frac{1}{\sqrt{\det g}}\frac{\partial}{\partial x_i}g^{ij}\sqrt{\det g} - \frac{1}{2}\frac{\partial h^i}{\partial x_i}, \\ &\quad \text{all evaluated at } x = 0. \end{aligned} \tag{6.4.9a}$$

Cartan's formula (6.4.4), combined with the skew symmetry of the curvature tensor R, permits an additional simplification of (6.4.9a) to

$$(6.4.9\text{b})\qquad \begin{aligned} &-\frac{1}{6}\frac{\partial^2 g^{ij}}{\partial x_i \partial x_j} - \frac{1}{2}\frac{\partial^2 \sqrt{\det g}}{\partial x_i^2} - \frac{1}{12}\frac{\partial^2 g^{ii}}{\partial x_j^2} - \frac{1}{2}\operatorname{div} h \\ &= -\frac{1}{18}R_{ijij} - \frac{1}{6}R_{ijij} + \frac{1}{18}R_{ijij} - \frac{1}{2}\operatorname{div} h \\ &= -\frac{1}{3}\sum_{i<j} R_{ijij} - \frac{1}{2}\operatorname{div} h = \frac{K}{3} - \frac{1}{2}\operatorname{div} h, \end{aligned}$$

and (6.4.1) follows upon noting that

$$(6.4.10)\qquad (4\pi t)^{d/2} e^0(t,0,0) = e^{-|h(0)t|^2/4t} = 1 - \frac{t}{4}|h|^2 + \mathrm{o}\left(t^2\right).$$

6.5. Manifolds with Boundary

Now let D be an open manifold with compact boundary B as at the end of Sect. 6.2, $M = D \cup B \cup D^*$ = the (closed) double of D, and Q the double to M of a smooth elliptic operator of degree 2 on D, and, as in Sect. 6.2, define $Q^-(Q^+)$ to be $Q \mid C^\infty(\overline{D})$ subject to $u = 0$ ($u^\bullet = 0$) on B. The coefficients $(\det g)^{-\frac{1}{2}}\partial g^{ij}(\det g)^{\frac{1}{2}}/\partial x_i$ occurring in Q jump as x crosses B, but $\partial u/\partial t = Qu$ still has a nice elementary solution e of class $C^\infty[(0,\infty)\times(M-B)^2] \cap C^1(M^2)$, approximable even on B by Levi's sum, and the elementary solutions $e^\pm$ of $\partial u/\partial t = Q^\pm u$ can be expressed on $(0,\infty)\times D^2$ as

$$(6.5.1)\qquad e^\pm(t,x,y) = e(t,x,y) \pm e\big(t,x,\overset{*}{y}\big),$$

$\overset{*}{y} \in D^*$ being the double of $y \in D$. By use of this formula, $Z^\pm = \int_D e^\pm(t,x,x)$ can be estimated as follows:

$$(6.5.2)\qquad \begin{aligned} (4\pi t)^{d/2} Z^\pm \;=\; &\text{the (Riemannian) volume } \int_D 1 \\ &\pm \frac{1}{4}\sqrt{4\pi t}\times \text{ the (Riemannian) surface area } \int_B 1 \\ &\pm \frac{t}{2}\int_B \text{flux } h + \frac{t}{3} \times \text{the curvatura integra } \int_D K \\ &-\frac{t}{6}\times \text{ the integrated mean curvature } \int_B J \\ &-\frac{t}{2}\int_D \operatorname{div} h - \frac{t}{4}\int_D |h|^2 + \mathrm{o}\left(t^{3/2}\right). \end{aligned}$$

To explain the new terms involved in this formula, pick a self-double patch U of M covering a patch $U \cap B$ of B endowed as in the diagram (Fig. 6.1) with local coordinates x such that (a) $1 > x_1 > 0$ in $U \cap D$, (b) $x_1 = 0$ on $U \cap B$, (c) $x_1(x^*) = -x_1(x)$, and (d) the positive x_1-direction is perpendicular to B. This has the effect that

$$(6.5.3\text{a})\qquad \begin{aligned} g_{ij}(x^*) &= -g_{ij}(x) \quad \text{for } i = 1 < j \text{ or } i > j = 1 \\ &= +g_{ij}(x) \quad \text{for } i = j = 1 \text{ or } i,j \geq 2, \end{aligned}$$

$$(6.5.3\text{b})\qquad g_{ij}(x) = 0 \quad \text{for } i = 1 < j \text{ or } i > j = 1 \text{ on } B,$$

$$(6.5.3\text{c})\qquad \sqrt{\det g/g_{11}}\,dx_2 \cdots dx_d = \text{the element of (Riemannian) surface area on } B.$$

Now $\int_B$ stands for integration relative to $\sqrt{\det g/g_{11}}\,dx_2\cdots dx_d$, flux h is the (outward-pointing) flux of h at a point of $B(= -\sqrt{g_{11}}h^1)$, and the mean curvature J at a point of

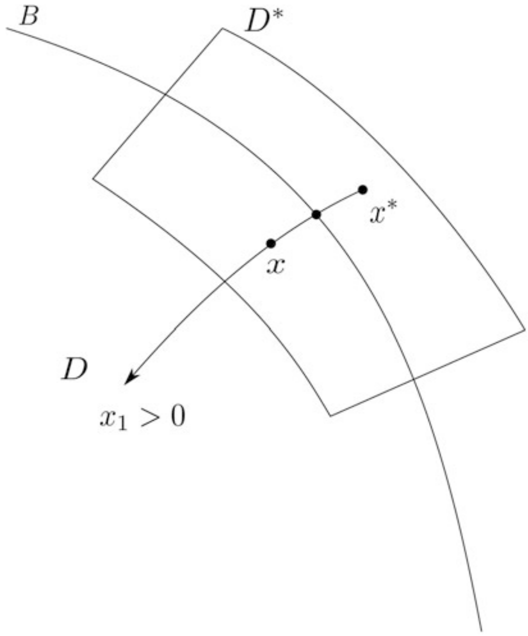

FIGURE 6.1.

B is (double) the spur of the second fundamental form $[= (g^{11}\det g)^{\bullet}\sqrt{g_{11}}/\det g]$,[5] representing (twice) the sum of inner curvatures of 2-dimensional sections perpendicular to B. Because of Green's formula ($\int_D \operatorname{div} h = \int_B \operatorname{flux} h$), a little cancellation occurs in (6.5.2) for Q^+. Equations (6.5.2) = (6.1.6) for $Q = \Delta$ ($h = 0$). The proof of (6.5.2) is broken up into a number of steps.

STEP 1. Consider a subregion $D' \subset D$ at a positive distance from B. Varadhan's bound (6.2.4d) implies that $\int_{D'} e(t, x, \overset{*}{x}) \le \exp(-c_1/t)$, so by (6.4.1),

$$(4\pi t)^{d/2}\int_{D'} e^{\pm}(t,x,x) = \int_{D'}\left[1 + \frac{t}{3}K - \frac{t}{2}\operatorname{div} h - \frac{t}{4}|h|^2\right] \tag{6.5.4a}$$
$$+ \text{ an exponentially small error,}$$

esp., it is enough to estimate $\int_{U\cap D} e^{\pm}(t,x,x)$ for such a patch U as described above. A close look at Levi's sum will convince the reader that $(4\pi t)^{d/2}\int_{U\cap D} e^{\pm}(t,x,x)$ can be developed in

[5]As before $\bullet$ stands for the one-sided partial in the positive 1-direction perpendicular to B. To prove that $(g^{11}\det g)^{\bullet}\sqrt{g_{11}}/\det g$ is (double) the spur of the second fundamental form of B, it is preferable to further specialize the local coordinates on U so as to make

$$g = \begin{pmatrix} g_{11} & 0 \\ 0 & h \end{pmatrix} \text{ on } U \quad \text{and} \quad g_{11} = 1 \text{ on } U \cap B.$$

The second fundamental form f is the (Riemannian) gradient along B of the inward-pointing unit normal field n:

$$f_{ij} = \frac{\partial n_i}{\partial x_j} + \left\{ {i \atop jk} \right\} n_k = \left\{ {i \atop 1j} \right\} = \text{the Christoffel bracket} \quad (i, j \ge 2).$$

Computing this for the special g adopted above gives $\frac{1}{2}h^{-1}h^{\bullet}$, so that double the spur is

$$\operatorname{sp} h^{-1}h^{\bullet} = (\lg\det h)^{\bullet} = \left(\lg g^{11}\det g\right)^{\bullet} = \left(g^{11}\det g\right)^{\bullet}/\det g,$$

as desired ($g^{11} = g_{11} = 1$ on B).

powers of $\sqrt{t}$. B can be covered by a finite number of patches U of *small total volume*, so terms like $t \times \operatorname{vol} U$ can be neglected: they can only influence the coefficient of $t^3/2$. As a simple application of this fact, the first term $e^0(t,x,x)$ of the expansion of $e^{\pm}(t,x,x)$ contributes

$$(4\pi t)^{d/2} \int_{U\cap D} e^0(t,x,x) = \int_{U\cap D} 1 \tag{6.5.4b}$$
$$+ \text{ an error of magnitude} \le \text{a constant multiple of } t \times \operatorname{vol} U,$$

so that, in view of (6.5.4a) and the fact that (6.3.5b) still holds, it suffices for the proof of (6.5.2) to check that

$$(4\pi t)^{d/2} \int_{U\cap D} e^0\big(t,x,\overset{*}{x}\big) = \frac{1}{4}\sqrt{4\pi t}\int_{U\cap D} 1 + \frac{t}{2}\int_{U\cap D} \text{flux } h + \mathrm{o}(t \times \operatorname{vol} U), \tag{6.5.5a}$$

$$(4\pi t)^{d/2} \int_{U\cap D} e^0\sharp f(t,x,x) = -\frac{t}{6}\int_{U\cap D} \frac{\big(g^{11}\det g\big)^{\bullet}}{\det g}\sqrt{g_{11}} + \mathrm{o}(t \times \operatorname{vol} U), \tag{6.5.5b}$$

$$(4\pi t)^{d/2} \int_{U\cap D} e^0\sharp f\big(t,x,\overset{*}{x}\big) = \mathrm{o}(t \times \operatorname{vol} U). \tag{6.5.5c}$$

STEP 2 (proof of (6.5.5a)).

$$\begin{aligned}(4\pi t)^{d/2} &\int_{U\cap D} e^0\big(t,x,\overset{*}{x}\big)\\ &= \int_{U\cap B} dx_2\cdots dx_d \int_0^1 \sqrt{\det g}\exp\Big\{-g\big(\overset{*}{x}\big)^{1/2}\big[\overset{*}{x} - x - b\big(\overset{*}{x}\big)t\big]^2/4t\Big\}dx_1\\ &= \int_{U\cap B} dx_2\cdots dx_d \int_0^1 \sqrt{\det g}\exp\big(-g_{11}x_1^2/t - fx_1 - |b|^2t/4\big)dx_1\end{aligned} \tag{6.5.6}$$

where $Q = \Delta + h\partial = a\partial^2 + b\partial$ and $f = g_{1k}b^k$; the following simplifications can be made by ignoring negligible terms:

(a) $\sqrt{\det g}$ can be replaced by $\sqrt{\det g^0} + (\sqrt{\det g})^{\bullet}x_1$, where 0 stands for evaluation at $x^0 = (0, x_2, \dots, x_d) \in B$, since

$$\int_0^1 x_1^2 dx_1 e^{-c_1x_1^2/t} \le c_2t^{3/2}.$$

(b) $\exp(-g_{11}x_1^2/t - fx_1 - |b|^2t/4)$ can be replaced by $e^{-g_{11}^0x_1^2/t}(1 - g_{11}^{\bullet}x_1^3/t - f^0x_1)$ for the same reason. $(0 \le e^{-x} - 1 + x \le x^2/2$ for $x \ge 0.)$

(c) $\int_0^1$ can be replaced by $\int_0^\infty$, since $\int_1^\infty e^{-c_1x_1^2/t} \le \exp(-c_2/t)$.

After these simplifications, (6.5.6) becomes

$$\int_{U\cap B} dx_2\cdots dx_d\sqrt{\det g^0}\int_0^\infty e^{-g^0_{11}x_1^2/t}dx_1\left[1+\frac{\left(\sqrt{\det g}\right)^\bullet}{\sqrt{\det g^0}}-x_1-g^\bullet_{11}\frac{x_1^3}{t}-f^0x_1\right] \tag{6.5.7}$$

up to a negligible error, and performing the inside integral gives

$$\begin{aligned}&\frac{1}{4}\sqrt{4\pi t}\int_{U\cap B}\frac{\sqrt{\det g^0}}{\sqrt{g^0_{11}}}dx_2\cdots dx_d\\ &\quad+\frac{t}{2}\int_{U\cap B}\frac{\sqrt{\det g^0}}{\sqrt{g^0_{11}}}dx_2\cdots dx_d\frac{1}{\sqrt{g^0_{11}}}\left[\frac{\left(\sqrt{\det g}\right)^\bullet}{\sqrt{\det g^0}}-\frac{g^\bullet_{11}}{g^0_{11}}-f^0\right].\end{aligned} \tag{6.5.8}$$

f^0 is now computed with the aid of (6.5.3):

$$f^0=g^0_{11}b^0_1=g^0_{11}\frac{\left(g^{11}\sqrt{\det g}\right)^\bullet}{\sqrt{\det g^0}}+g^0_{11}h^1=\frac{\left(\sqrt{\det g}\right)^\bullet}{\sqrt{\det g^0}}-\frac{g^\bullet_{11}}{g^0_{11}}-\sqrt{g^0_{11}}\text{flux } h,$$

and (6.5.5a) follows.

STEP 3 (proof of (6.5.5b)). This is not so cheap.

$$\begin{aligned}&(4\pi t)^{d/2}e^0\sharp f(t,x,x)\\ &=(4\pi t)^{d/2}\int_0^t ds\int_{R^d}\frac{\exp\Big\{-\big|g^{1/2}(y)\big[y-x-b(y)(t-s)\big]\big|^2/4(t-s)\Big\}}{\big[4\pi(t-s)\big]^{d/2}}\\ &\quad\times\frac{\exp\Big\{-\big|g^{1/2}(y)\big[x-y-b(x)s\big]\big|^2/4s\Big\}}{(4\pi s)^{d/2}}\sqrt{\det g(y)}dy\\ &\quad\times\Bigg\{\big[g^{ij}(y)-g^{ij}(x)\big]\\ &\qquad\times\left[\frac{1}{4s^2}g_{ik}(x)\big(y_k-x_k-b_k(x)s\big)g_{jl}(x)\big(y_l-x_l-b_l(x)s\big)-\frac{g_{ij}(x)}{2s}\right]\\ &\qquad-\big[b^i(y)-b^i(x)\big]\frac{g_{ik}(x)\big[y_k-x_k-b_k(x)s\big]}{2s}\Bigg\}.\end{aligned} \tag{6.5.9a}$$

Equation (6.5.9a) can actually be replaced by

$$\begin{aligned}&\int_0^t ds\int_{R^d}\frac{e^{-|g^{1/2}(x)(y-x)|^2/4r}}{(4\pi r)^{d/2}}\sqrt{\det g(y)}dy\\ &\quad\times\Bigg\{\big[g^{ij}(y)-g^{ij}(x)\big]\left[\frac{g_{ik}(x)\big(y_k-x_k\big)g_{jl}(x)\big(y_l-x_l\big)}{4s^2}-\frac{g_{ij}(x)}{4s}\right]\\ &\qquad-\big[f^i(y)-f^i(x)\big]\frac{g_{ik}(x)\big(y_k-x_k\big)}{2s}\Bigg\}\end{aligned} \tag{6.5.9b}$$

up to the desired degree of precision, where

$$r=s(t-s)/t,\quad f^j=(\det g)^{-\frac{1}{2}}\partial g^{ij}(\det g)^{\frac{1}{2}}/\partial x_i\quad(j\le d).$$

For example, to replace the first exponential in (6.5.9a) by

$$\exp\Big[-\big|g^{1/2}(x)(y-x)\big|^2/4(t-s)\Big],$$

it suffices to note the following points:

(a) The integration over R^d can be restricted to the figure $|y-x| < (t-s)^{2/5}$ since, for $t \downarrow 0$, the remainder makes a contribution of magnitude smaller than

$$
\begin{aligned}
c_1 t^{d/2} \int_0^t ds \int_{|y-x|\geq (t-s)^{2/5}} dy \frac{e^{-c_2|y-x|^2/(t-s)}}{(t-s)^{d/2}} \frac{e^{-c_2|y-x|^2/s}}{s^{d/2}} \\
\times \big(\text{terms like } s^{-2}|y-x|^3,\ s^{-1}|y-x|, \text{ etc., replaceable} \\
\text{by } c_3 s^{-1/2} \text{ after reducing } c_2 \text{ to } c_4 < c_2\big) \\
\leq c_5 \int_0^t \frac{ds}{\sqrt{s}} \int_{|w|>(t-s)^{2/5}} dw \frac{e^{-c_4|w|^2/r}}{r^{d/2}} \\
\leq c_6 \int_0^t \frac{ds}{\sqrt{s}} e^{-c_7(t-s)^{4/5}/r} < c_7 e^{-c_8/t},
\end{aligned}
$$

which is negligible.

(b) Performing the integral just over $|y-x| < (t-s)^{2/5}$ and using $e^{-A} - e^{-B} \leq (B-A)e^{-A}$ $(0 \leq A \leq B)$ to estimate the difference between the 2 integrands, one finds that the indicated replacement produces an error of magnitude smaller than

$$
\begin{aligned}
c_9 t^{d/2} \int_0^t ds \int_{|y-x|<(t-s)^{2/5}} dy \frac{e^{-c_{10}|y-x|^2/(t-s)}}{(t-s)^{d/2}} \\
\times \left[\frac{|y-x|^3}{t-s} + |y-x|^2 + t - s\right] \frac{e^{-c_{10}|y-x|^2/s}}{s^{d/2}} \\
\times \big(\text{terms like } s^{-2}|y-x|^3,\ s^{-1}|y-x|, \text{ etc.}\big) \\
\leq c_{11} \int_0^t \sqrt{\frac{t-s}{s}} = c_{12} t,
\end{aligned}
$$

which is also negligible after integrating over $U \cap D$.

(c) Finally, one makes use of the fact that for the new exponential, the integral over $|y-x| > (t-s)^{2/5}$ is likewise negligible.

Equation (6.5.9b) is also to be integrated over $U \cap D$; for this purpose, similar estimates permit us to replace it by

$$
\begin{aligned}
&\int_0^t ds \int_{R^d} \frac{e^{-|g^{0\frac{1}{2}}(y-x)|^2/4r}}{(4\pi r)^{d/2}} \sqrt{\det g^0} dy \\
&\times \left\{ \left[\widehat{g}^{ij}(y_1, x^0) - \widehat{g}^{ij}(x)\right] \left[\frac{g^0_{ik}(y_k - x_k) g^0_{jl}(y_l - x_l)}{4s^2} - \frac{g^0_{ij}}{2s}\right] \right. \\
&\left. - \left[\widehat{f}^i(y, x^0) - \widehat{f}^i(x)\right] \frac{g^0_{ik}(y_k - x_k)}{2s} \right\}. \qquad (6.5.9c)
\end{aligned}
$$

$\widehat{\ }$ has the following meaning: for fixed $x^0 = (0, x_2, \ldots, x_d) \in B$, $\widehat{g}$ is a broken line with the same corner as g at $x_1 = 0$ (and no other corners), while $\widehat{f}$ is a step function with a single jump at $x_1 = 0$ agreeing with f at $x_1 = 0^{\pm}$.

Do the integration $\int_{R^{d-1}} dy_2 \cdots dy_d$ and use the special form of g^0 (6.5.3b). This gives

$$
\begin{aligned}
&\int_0^t ds \int_{-\infty}^{+\infty} \frac{e^{-g^0_{11}(y_1-x_1)^2/4r}}{\sqrt{4\pi r/g^0_{11}}} dy_1 \\
&\quad \times \Bigg\{ \left[\widehat{g}^{ij}(y_1,x^0) - \widehat{g}^{ij}(x)\right] \left[\frac{g^0_{i1} g^0_{j1} (y_1-x_1)^2}{4r^2} + \sum_{k,l\ge 2} g^0_{ik} g^0_{jl} g^0_{kl} \frac{2r}{4s^2} - \frac{g^0_{ij}}{2s} \right] \\
&\qquad - \left[\widehat{f}^i(y_1,x^0) - \widehat{f}^i(x)\right] \frac{g^0_{i1}(y_1-x_1)}{2s} \Bigg\} \\
&= \int_0^t ds \int_{R^1} \frac{e^{-g^0_{11}(y_1-x_1)^2/4r}}{\sqrt{4\pi r/g^0_{11}}} dy_1 \\
&\quad \times \Bigg\{ \left[\widehat{g}^{11}(y_1,x^0) - \widehat{g}^{11}(x)\right] \left[\frac{(g^0_{11})^2 (y_1-x_1)^2}{4s^2} - \frac{g^0_{01}}{2s} \right] \\
&\qquad - \sum_{i,j\ne 1} \left[\widehat{g}^{ij}(y_1,x^0) - \widehat{g}^{ij}(x)\right] \frac{g^0_{ij}}{2t} - \left[\widehat{f}^1(y_1,x^0) - \widehat{f}^1(x)\right] \frac{g^0_{11}}{2s} (y_1-x_1) \Bigg\}.
\end{aligned}
\tag{6.5.10a}
$$

Equation (6.5.3a) implies that for $i=j=1$ or $i,j\ge 2$, $\widehat{g}^{ij}(y_1,x^0) - \widehat{g}^{ij}(x) = (y_1-x_1)g^{ij\bullet}$ or $-(y_1+x_1)g^{ij\bullet}$ according as $y_1>0$ or $y_1<0$, $g^{11}\sqrt{\det g}$ being even across B, so (6.5.10a) simplifies to

$$
\begin{aligned}
&\int_0^t ds \int_{-\infty}^{-x_1} \frac{e^{-g^0_{11} w_1^2/4r}}{\sqrt{4\pi r/g^0_{11}}} dw_1 \\
&\quad \times \Bigg\{ -2(x_1+w_1) g^{11\cdot} \left[\frac{(g^0_{11})^2 w_1^2}{4s^2} - \frac{g^0_{11}}{2s} \right] \\
&\qquad + 2(x_1+w_1) \sum_{i,j\ne 1} g^{ij\cdot} \frac{g^0_{ij}}{2t} + \frac{\left(g^{11}\sqrt{\det g}\right)^{\cdot}}{\sqrt{\det g^0}} \frac{g^0_{11}}{s} w_1 \Bigg\}.
\end{aligned}
\tag{6.5.10b}
$$

Do next the integral $\int_0^1$ (6.5.10b)$\sqrt{\det g}\,dx_1$, replacing g by g^0, extending the integration from 1 to $+\infty$, and changing $\int_0^\infty dx_1 \int_{-\infty}^{-x_1} dw_1$ into $\int_{-\infty}^0 dw_1 \int_0^{-w_1} dx_1$:

$$
\begin{aligned}
&\sqrt{\det g^0} \int_0^t ds \int_{-\infty}^0 \frac{e^{-g^0_{11} w_1^2/4r}}{\sqrt{4\pi r/g^0_{11}}} dw_1 \\
&\quad \times \Bigg[\frac{w_1^2}{2} g^{11\bullet} \left(\frac{(g^0_{11})^2 w_1^2}{4s^2} - \frac{g^0_{11}}{2s} \right) \\
&\qquad - \sum_{i,j\ne 1} g^{ij\bullet} g^0_{ij} \frac{w_1^2}{4t} - \frac{\left(g^{11}\sqrt{\det g}\right)^{\bullet}}{\sqrt{\det g^0}} g^0_{11} \frac{w_1^2}{s} \Bigg] \\
&= \sqrt{\det g^0} \int_0^t ds \Bigg\{ \frac{g^{11\bullet}}{2} \left[3\frac{(t-s)^2}{t^2} - \frac{t-s}{t} \right] \\
&\qquad - \frac{1}{2} \sum_{i,j\ne 1} g^{ij\bullet} g^0_{ij}/g^0_{11} \frac{s(t-s)}{t^2} - \frac{\left(g^{11}\sqrt{\det g}\right)^{\bullet}}{\sqrt{\det g^0}} \frac{t-s}{t} \Bigg\}
\end{aligned}
$$

$$= t\sqrt{\det g^0} \times \left[\frac{g^{11\bullet}}{4} - \frac{1}{12} \sum_{i,j \neq 1} g^{ij\bullet} g^0_{ij} / g^0_{11} - \frac{1}{2} \frac{\left(g^{11}\sqrt{\det g}\right)^\bullet}{\sqrt{\det g^0}} \right]$$

$$= -\frac{t}{6} \sqrt{\det g^0} \frac{\left(g^{11} \det g\right)^\bullet}{\det g^0}, \tag{6.5.11}$$

since $g^{ij\bullet} g_{ij} = -(\det g)\sqrt{\det g}$. An integration $\int_{U \cap B}$ (6.5.11)$dx_2 \cdots dx_d$ now gives the desired formula (6.5.5b).

STEP 4. The proof of (6.5.5c) is practically the same, so it is left to the industrious reader.

6.6. Δ on Differential Forms

Given a closed manifold M, let Δ act on the space Λ^p of smooth exterior differential p-forms ($p \le d$). Λ^p is a pre-Hilbert space relative to the inner product $(f_1, f_2) = \int \langle f_1, f_2 \rangle$, $\langle f_1, f_2 \rangle$ being the Riemannian inner product of p-forms at a point of M, and Δ can be expressed as $-(dd^* + d^*d)$, $d : \Lambda^{p-1} \to \Lambda^p$ ($1 \le p \le d$) being the exterior differential and $d^* : \Lambda^{p+1} \to \Lambda^p$ ($0 \le p < d$) its dual relative to the above inner product. Δ acting on Λ^p is symmetric with a discrete spectrum:

$$0 \ge \gamma_0 \ge \gamma_1 \ge \gamma_2 \ge \text{etc.} \downarrow -\infty,$$

the corresponding eigenforms f form a unit perpendicular basis of Λ^p, the sum

$$e^p = \sum_{n \ge 0} \exp\left(\gamma_n t\right) f_n \otimes f_n$$

converges uniformly on compact figures of $(0, \infty) \times M^2$ to the elementary solution of $\partial u / \partial t = \Delta u$ for p-forms, and the spur $Z^p = \sum \exp(\gamma_n t)$ of $e^{t\Delta}$ on Λ^p can be expressed as the integral over the manifold of the pole $\operatorname{sp} e^p = \sum \exp(\gamma_n t) \langle f_n, f_n \rangle : Z^p = \int \operatorname{sp} e^p$.

Define Z to be the alternating sum of Z^p ($p \le d$): $Z = Z^0 - Z^1 + \cdots \pm Z^d$. Then

$$Z = \text{the Euler Characteristic } E \text{ of } M, \tag{6.6.1}$$

as will be proved below. Poincaré duality makes this trivial for odd dimensions ($Z^p = Z^{d-p}$); also, in 2 dimensions $Z^0 = Z^2$ for the same reason, so from (6.1.5b) and (6.6.1) it follows that for $d = 2$,

$$Z^1 = 2Z^0 - E = \frac{\text{area}}{2\pi t} - \frac{2}{3} E + \frac{\pi t}{30} \int K^2 + o\left(t^2\right). \tag{6.6.2}$$

Given a number $\gamma \le 0$, define 3^p to be the eigenspace of p-forms f such that $\Delta f = \gamma f$. By de Rham's theorem [5],

$$\dim 3^0 - \dim 3^1 + \cdots \pm \dim 3^d = E \quad \text{for } \gamma = 0, \tag{6.6.3a}$$

so (6.6.1) is the same as

$$\dim 3^0 - \dim 3^1 + \cdots \pm \dim 3^d = 0 \quad \text{for } \gamma < 0. \tag{6.6.3b}$$

Chern [4] discovered a beautiful extension of the classical Gauss-Bonnet formula to manifolds of even dimension $d > 2$. Chern's formula states that $\int C = E$. C is a (complicated) homogeneous polynomial of degree $d/2$ in the entries of the curvature tensor, reducing to the classical integrand $K/2\pi = -R^{12}_{12}/2\pi$ for $d = 2$. Because of the complete cancellation of the time-dependent part of the alternating sum Z, it is natural to hope that some fantastic cancellation will also take place *in the small*, i.e., in the alternating pole sum:

$$\operatorname{sp} e^0 - \operatorname{sp} e^1 + \cdots \pm \operatorname{sp} e^d = \begin{cases} 0 \\ C \end{cases} + \mathrm{o}(1) \quad \text{for } d \begin{cases} \text{odd} \\ \text{even} \end{cases} . \tag{6.6.4}$$

Poincaré duality does it for odd d with $\mathrm{o}(1) = 0$, but the even-dimensional proof eludes us, except for $d = 2$; in which case

$$\operatorname{sp} e^0 - \operatorname{sp} e^1 + \operatorname{sp} e^2 = C + \frac{t}{6}\Delta C + \mathrm{o}\left(t^2\right) \tag{6.6.5}$$

(see Sect. 6.8 for additional information for $d = 4$). The proof of (6.6.5) is postponed until after the

Proof of (6.6.1) = (6.6.3b)**.** Choose $\gamma < 0$, let 3^p $(p \le d)$ be the corresponding eigenspaces, and make the convention that $3^{-1} = 3^{d+1} = 0$. $\Delta = -(d^*d + dd^*)$ commutes with d and d^*, so $d3^{p-1} + d^*3^{p+1} \subset 3^p$. Because $d^2 = 0$, $(d3^{p-1}, d^*3^{p+1}) = (d^2 3^{p-1}, 3^{p+1}) = 0$, so the sum is direct, and it fills up the whole of $3^p (= d3^{p-1} \bigoplus d^*3^{p+1})$ since, for $f \in 3^p$,

$$\left(f, d3^{p-1}\right) = \left(d^*f, 3^{p-1}\right) = 0 \quad \text{and} \quad \left(f, d^*3^{p+1}\right) = \left(df, 3^{p+1}\right) = 0$$

make $d^*f = df = 0$, so that $\gamma f = \Delta f = 0$ and $f = 0$ $(\gamma \ne 0)$; esp.,

$$\dim 3^p = \dim d3^{p-1} + \dim d^*3^{p+1} \quad (p \le d),$$

and so

$$\sum_{p \le d} (-1)^p \dim 3^p = \sum \left(- \dim d^*3^{2p} + \dim 3^{2p} - \dim d3^{2p}\right). \tag{6.6.6}$$

But $3^{2p} = d3^{2p-1} \bigoplus d^*3^{2p+1}$, so that

$$\begin{aligned} &\dim 3^{2p} - \dim d^*3^{2p} - \dim d3^{2p} \\ &\quad = \dim d3^{2p-1} + \dim d^*3^{2p+1} - \dim d^*d3^{2p-1} - \dim dd^*3^{2p+1} \ge 0, \end{aligned}$$

and also

$$\begin{aligned} &\dim 3^{2p} - \dim d^*3^{2p} - \dim d3^{2p} \\ &\quad = \dim \Delta 3^{2p} - \dim d^*3^{2p} - \dim d3^{2p} \\ &\quad \le \dim dd^*3^{2p} + \dim d^*d3^{2p} - \dim d^*3^{2p} - \dim d3^{2p} \le 0; \end{aligned}$$

in brief, $\dim 3^{2p} = \dim d3^{2p} + \dim d^*3^{2p}$, and the whole of the alternating dimension sum (6.6.6) collapses to 0.

Proof of (6.6.5) $(d = 2)$**.** $3^1 = d3^0 \bigoplus d^*3^2$ for $\gamma < 0$, and for $f \in 3^0$,

$$\|df\|^2 \equiv (df, df) = -\left(d^*df, f\right) = -(\Delta f, f) = -\gamma \|f\|^2$$

with a similar result $(\|d^*f\|^2 = -\gamma\|f\|^2)$ for $f \in 3^2$. Because of this,

$$\sum \exp\left(\gamma_n^0 t\right)\langle df_n^0, df_n^0\rangle + \sum \exp\left(\gamma_n^2 t\right)\langle d^*f_n^2, d^*f_n^2\rangle = -\sum \gamma_n^1 \exp\left(\gamma_n^1 t\right)\langle f_n^1, f_n^1\rangle$$

with a self-evident notation. But, for $f \in \Lambda^0$,

$$\langle df, df\rangle = g^{ij}\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{1}{2}\Delta f^2 - f\Delta f,$$

so, by the Poincaré duality between 3^0 and 3^2,

$$-\frac{\partial}{\partial t}\operatorname{sp} e^1 = -\sum \gamma_n^1 \exp\left(\gamma_n^1 t\right)\langle f_n^1, f_n^1\rangle = 2\sum \exp\left(\gamma_n^0 t\right)\langle df_n^0, df_n^0\rangle$$
$$= \sum \exp\left(\gamma_n^0 t\right)\left[\Delta\left(f_n^0\right)^2 - 2f_n^0\Delta f_n^0\right] = \Delta \operatorname{sp} e^0 - 2\frac{\partial}{\partial t}\operatorname{sp} e^0,$$

or, what is the same,

$$\frac{\partial}{\partial t}\left(\operatorname{sp} e^0 - \operatorname{sp} e^1 + \operatorname{sp} e^2\right) = \Delta \operatorname{sp} e^0.$$

$\operatorname{sp} e^0$ has an expansion beginning with a multiple of t^{-1} and proceeding by ascending powers of t as stated in Sect. 6.4, and a little extra attention to the proof shows that the formal application of Δ to this expansion gives the expansion for $\Delta \operatorname{sp} e^0$. Consequently, (6.4.1) implies

$$\operatorname{sp} e^0 - \operatorname{sp} e^1 + \operatorname{sp} e^2 = B + \frac{t}{6}\Delta C + \mathrm{o}\left(t^2\right)$$

with $C =$ the Gauss-Bonnet integrand $K/2\pi$. To complete the proof of (6.6.5), it remains to check that $B = C$. Pick local coordinates so that Cartan's formula (6.4.4) holds. A moment's reflection shows that B can be expressed as a (universal) combination of second partials of g_{ij} $(i, j \le 2)$; as such, it is a (universal) constant multiple of the one nonzero component R_{1212} of the Riemann tensor, and the constant can be identified as $-1/2\pi$ by using the Gauss-Bonnet formula in the special case of the Riemann sphere:

$$2 = E = \int \left(\operatorname{sp} e^0 - \operatorname{sp} e^1 + \operatorname{sp} e^2\right) = B$$
$$= \text{constant} \times \int R_{1212} = -\text{constant} \times 4\pi.$$

6.7. Algebraic Computation of k_1 and k_2

The style of proof just used to finish the verification of (6.2.5) will now be exploited to compute the third coefficient of the Minakshisundaram expansion (6.4.3) for $Q = \Delta$:

$$k_2 = \frac{1}{180}(10A - B + 2C) + \text{constant} \times \Delta K, \tag{6.7.1}$$

with

$$A = \left(\sum_{i<j} R_{ijij}\right)^2 = K^2, \tag{6.7.2a}$$

$$B = \sum_{j,k}\left(\sum_i R_{ijik}\right)^2, \tag{6.7.2b}$$

$$C = \sum_{i,j,k,l}\left(R_{ijkl}\right)^2. \tag{6.7.2c}$$

The constant multiplier of ΔK in (6.7.1) is not known, but $\int_M \Delta K = 0$, so

$$\int_M k_2 = \frac{1}{180}\int_M (10A - B + 2C), \tag{6.7.3}$$

as needed for (6.1.5a); in any case, this constant is *universal*, i.e., it is the same for all manifolds M. The method will also provide us with a new derivation of the formula $k_1 = K/3$. A short table of special expansions will be helpful for the proof; in this table Z is computed up to

Table 6.1

M	K	A	B	C	$Z/(4\pi t)^{d/2} \times \operatorname{vol} M$
S^2	1	1	2	2	$\frac{e^{t/4}}{\sqrt{\pi t}} \int_0^1 \frac{e^{-x/t}}{\sin \sqrt{x}} dx = 1 + \frac{t}{3} + \frac{t^2}{15} + \cdots$
S^3	3	9	12	6	$e^t = 1 + t + \frac{1}{2}t^2 + \cdots$
D^2	-1	1	2	2	$\frac{e^{-t/4}}{\sqrt{\pi t}} \int_0^1 \frac{e^{-x/t}}{\sinh \sqrt{x}} dx = 1 - \frac{t}{3} + \frac{t^2}{15} + \cdots$
D^3	-3	9	12	6	$e^{-t} = 1 - t + \frac{1}{2}t^2 + \cdots$

an exponentially small error for several standard manifolds. $D^2(D^3)$ is the 2(3)-dimensional Lobachevsky space modulo a discontinuous group of motions.

Pick exponential coordinates on a patch about a point $o \in M$ as for (6.4.4). The coefficients of the power series expansion of g about o will be polynomials in the curvature tensor R and its covariant derivatives [3, Chapter 10, §4], and it follows from this and from Levi's sum for the pole of e that the coefficients of (6.4.3) are expressible as polynomials of the same kind. A scaling argument now gives the degree of these polynomials. Change g into $C^2 g$ ($C^2 > 0$). Then Δ is changed into $C^{-2}\Delta$, and the pole of the elementary solution becomes $e(t/C^2, o, o)C^{-d}$, so that k_n is simply multiplied by C^{2n}. But also, an l-fold covariant derivative of $R(C^2 g)$ is a multiple of C^{2+l}. Consequently, $k_n = k_n(g)$ is a "homogeneous polynomial" of degree $2n$ in R and its covariant derivatives, if to an l-fold covariant derivative is ascribed the degree $2 + l$, *esp.*, k_1 is a form of degree 1 in R, while k_2 is a form of degree 2 in R plus a form of degree 1 in second covariant derivatives of R. Clearly, the coefficients of these forms depend upon M *only via the dimension*.

The next step is to exploit the fact that an orthogonal transformation of the tangent space changes one exponential coordinate system x into another. Because the pole of e depends on x only via $\sqrt{\det g}$, which is an orthogonal invariant, the coefficients of its expansion are likewise orthogonal invariants, *esp.*, k_1 is an invariant form of degree 1 in R, and as such, it is a constant multiple of $K = -\sum_{i<j} R_{ijij}$ [18, Chapter 5]. This constant depends upon the dimension of M only, so to complete the evaluation of k_1, it suffices to check that the constant is dimension-free and to compute it for $M = S^2$, say (see Table 6.1). To settle the first point, look at a product manifold, $M = M_1 \times M_2$. $\Delta(M) = \Delta(M_1) \bigotimes 1 \bigoplus 1 \bigotimes \Delta(M_2)$, so $e(M) = e(M_1) \bigotimes e(M_2)$, and it follows from (6.4.3) that $k_1(M) = k_1(M_1) + k_1(M_2)$. But also $R(M) = R(M_1) \bigoplus R(M_2)$, so that $K(M) = K(M_1) + K(M_2)$, and varying the dimension of M_2 leads at once to the proof.

k_2 is not so simple.

STEP 1 is to notice that the forms of degrees 2 and 1 into which k_2 is split are *separately* invariant under the action of the orthogonal group. As stated before, the coefficients of these forms depend upon dimension only.

STEP 2. For $d > 3$, the space of curvature tensors at a point of M, viewed as a representation space of the orthogonal group $\mathrm{o}(d)$, splits into 3 irreducible pieces. One piece is the kernel of the contraction map $R_{ijkl} \to R_{ijil}$. The orthogonal complement can be viewed as the space of symmetric matrices with $\mathrm{o}(d)$ acting by similarity ($x \to o^* x o$), and this piece splits into the scalars plus symmetric matrices with spur 0 [18, Chapter 5]. Consequently, the space of invariant polynomials of degree 2 is 3-dimensional, the 3 polynomials A, B, C exhibited in (6.7.2) provide us with a nice basis, and the corresponding part of k_2 is simply $c_0 A + c_1 B + c_2 C$ with coefficients depending (perhaps) on the dimension. The same still holds for dimensions 2 and 3, except that

(6.7.4a) $$B = C = 2A \quad (d = 2),$$

(6.7.4b) $$B = A + C/2 \quad (d = 3),$$

which make the splitting simpler.

STEP 3. The part of k_2 which is an invariant form of degree 1 in second covariant derivatives of R can only be obtained by a three-fold contraction [18, Chapter 5], and only 2 candidates present themselves: $R_{ijij;kk} = -2\Delta K$ and $R_{ikjk;ij}$. But, by the Bianchi identities,

$$R_{ikjk;ij} + R_{ikki;jj} + R_{ikij;kj} = 0,$$

so the second candidate is half the first, and

(6.7.5) $$k_2 = c_0 A + c_1 B + c_2 C + c_3 \Delta K$$

with coefficients depending upon dimension only.

STEP 4. is to prove that the coefficients are dimension-free. This is done, as in the proof of $k_1 = K/3$, by looking at a product $M = M_1 \times M_2$, $R(M) = R(M_1) \bigoplus R(M_2)$, so

(6.7.6a) $$A(M) = A(M_1) + A(M_2) + 2K(M_1)K(M_2),$$

(6.7.6b) $$B(M) = B(M_1) + B(M_2),$$

(6.7.6c) $$C(M) = C(M_1) + C(M_2);$$

also

(6.7.6d) $$e(M) = e(M_1) \otimes e(M_2),$$

and a comparison of the expansion

(6.7.7a) $$\begin{aligned} 1 + tk_1(M) + t^2 k_2(M) + \mathrm{o}\left(t^3\right) = 1 + \frac{t}{3}\left[K(M_1) + K(M_2)\right] \\ + t^2 \times \Big\{ c_0(d)\left[A(M_1) + A(M_2) + 2K(M_1)K(M_2)\right] \\ + c_1(d)\left[B(M_1) + B(M_2)\right] + c_2(d)\left[C(M_1) + C(M_2)\right] \\ + c_3(d)\left[\Delta K(M_1) + \Delta K(M_2)\right] \Big\} + \mathrm{o}\left(t^3\right), \end{aligned}$$

d being $\dim M$, with the expansion

$$
\begin{aligned}
&\left[1+tk_1(M_1)+t^2k_2(M_1)\right]\left[1+tk_1(M_2)+t^2k_2(M_2)\right]+\mathrm{o}\left(t^3\right)\\
&\quad=1+\frac{t}{3}\left[K(M_1)+K(M_2)\right]\\
&\qquad+t^2\times\Big[\frac{1}{9}K(M_1)K(M_2)+c_0(d_1)A(M_1)+c_1(d_1)B(M_1)\\
&\qquad\qquad+c_2(d_1)C(M_1)+c_3(d_1)\Delta K(M_1)+c_0(d_2)A(M_2)\\
&\qquad\qquad+c_1(d_2)B(M_2)+c_2(d_2)C(M_2)+c_3(d_2)\Delta K(M_2)\Big]\\
&\qquad+\mathrm{o}\left(t^3\right)
\end{aligned}
\tag{6.7.7b}
$$

in case M_1 is a flat torus $[R(M_1)=0]$ shows that the expression

$$c_0(d)A(M_2)+c_1(d)B(M_2)+c_2(d)C(M_2)+c_3(d)\Delta K(M_2) \tag{6.7.8}$$

is independent of $d \geq d_2$. The fact that the coefficients are dimension-free for $d \geq 4$ is immediate from this. For $d \leq 3$, the coefficients can be chosen to be the *same* as for higher dimensions.

STEP 5 is to compute the actual values of the coefficients. Comparison of the terms involving $K(M_1)K(M_2)$ in (6.7.7a) and (6.7.7b) gives

$$c_0 = 1/18, \tag{6.7.9a}$$

and, from Table 6.1 placed at the beginning of this section,

$$c_1 = -1/180, \tag{6.7.9b}$$

$$c_2 = 1/90, \tag{6.7.9c}$$

so that only c_3 is still unknown. This completes the proof.

For $d=4$, the integrand for Chern's extension of the Gauss-Bonnet formula [4] is easily evaluated as $(8\pi^2)^{-1}(A-B+C/2)$. The formula states that this integrates to the Euler characteristic E of M, whence, for $d=4$,

$$\int k_2 = \frac{2\pi^2}{45}E+\frac{1}{180}\int(9A+3C/2) \geq \frac{2\pi_2}{45}E, \tag{6.7.10a}$$

$$M \text{ is a flat space} \quad \text{if} \int k_2 = 0 \text{ and } E \geq 0, \tag{6.7.10b}$$

$$\int k_2 \neq 0 \quad \text{if } M \text{ is simply connected}, \tag{6.7.10c}$$

and if the sectional curvatures of M do not change sign, then

$$\int k_2 = 0 \quad \text{only for a flat space}, \tag{6.7.10d}$$

while, for $d \leq 3$,

$$\int k_2 \geq 0 \quad \text{and} \quad \int k_2 = 0 \quad \text{only for a flat space.} \tag{6.7.10e}$$

Proof. Equation (6.7.10a) is immediate from Chern's formula and (6.7.10b) follows, since $\int C = 0$ makes M flat. $E \geq 0$ if M is simply connected. But a flat compact space is not simply connected, so (6.7.10c) is proved. Equation (6.7.10d) is proved in the same way using the fact that $E \geq 0$ if the sectional curvatures of M do not change sign [4]. The proof of (6.7.10e) is immediate from (6.7.1) and (6.7.4).

The computation of k_3, k_4, etc. is a problem of classical invariant theory; see for instance [16]. It looks pretty hopeless.

6.8. Open Problems

1°. For $Q = \Delta$, compute *all* the coefficients of Minakshisundaran's expansion (6.4.3) and explain the geometrical significance of each. It is an open problem to find the corresponding corrections to Weyl's formula (6.1.2). But notice that even for $M = S^2$, $-\gamma_n$ does *not* behave like $c_{-1}n + c_0 + c_1 n^{-1} + \text{etc.}$

2°. Prove or disprove (6.6.4) for even $d \geq 4$; see (6.7.10a) for partial information in case $d = 4$.

3°. J. Milnor [8] proved that the spectrum of Δ acting on the differential forms of a closed manifold M is not sensitive enough to discriminate between the possible Riemannian geometries on M. Milnor's example depends upon an example of E. Witt of 2 self-dual 16-dimensional lattices Γ, dissimilar under the action of o(16), but with $\sharp(R) = \sharp(\omega \in \Gamma : |\omega| \leq R)$ the same for both. Because the lattices are dissimilar, the tori $M = R^{16}/\Gamma$ are not isometric. But the spectrum of Δ on functions is just the numbers $4\pi^2|\omega|^2$ with ω from Γ. Because $\Delta(f dx_{i_1} \wedge \cdots \wedge dx_{i_p}) = (\Delta f) dx_{i_1} \wedge \cdots \wedge dx_{i_p}$, the spectrum of Δ on p-forms is the same, but just repeated $16!/p!(16-p)!$ times, so that the 2 tori are identical from the spectral point of view. Despite this example, it may be possible to "hear" the geometry of M for small dimensions ($d = 2$, for instance) or for a special class of manifolds (topological spheres, for instance). Kac [6] has asked if the spectra of both $\Delta^{\pm}$ for a flat plane region D suffice to determine D up to a rigid motion; his conjecture is *no*. If that is so then probably the complete geometry of a closed manifold cannot be heard even for $d = 2$ and M a topological sphere. But it should be noted that for $D = (0, 1)$, $0 < f \in C[0, 1]$, and $Qu = fu''$, f can be recovered from the spectra of $Q^{\pm}$ [2].

4°. Jacobi's transformation of the theta-function shows that for Δ acting on functions on a flat torus $M = R^d/\Gamma$,

$$Z = \sum e^{-4\pi^2|\omega|^2 t} = \frac{\operatorname{vol} M}{(4\pi t)^{d/2}} \sum_{\omega \in \Gamma} e^{-|\omega|^2/4t}$$

$$= \frac{\operatorname{vol} M}{(4\pi t)^{d/2}} + \text{an exponentially small error,}$$

where Γ^* is the dual lattice of Γ. Does there exist a Jacobi like transformation of Z for any other manifolds? To our knowledge the only similar thing is the so-called Kramers-Wannier duality for the 2-dimensional Ising model of statistical mechanics. Both Kramers-Wannier and Jacobi's transformation are instances of Poisson's summation formula [7]. Perhaps Selberg's trace formula could be helpful in this. A simple case to look at would be a compact symmetric space $M = G/K$ of rank 1, since the pole $\operatorname{sp} e^0$ is constant on M and can be computed using just the radial part $A^{-1}\frac{\partial}{\partial R} A \frac{\partial}{\partial R}$ of Δ (A = the area of the spherical surface of radius R about the north pole). A second interesting case would be that of a closed Riemann surface of genus ≥ 2, viewed as the open unit disc modulo a discontinuous group. One may conjecture that the breaking off the expansion of Z at the first (volume) term happens for flat spaces only (see (6.7.10) for the proof in case $d \leq 3$ and for partial information in case $d = 4$).

5°. A Jacobi transformation for Z goes over into a Riemann like identity for the zeta-like function $\sum_{n\geq 1}|\gamma_n|^{-s}$ via the transformation

$$Z \longrightarrow \Gamma(s)^{-1}\int_0^\infty t^{s-1}(Z-1)dt.$$

Minakshisundaram [9] used (6.4.3) to prove that this zeta-function is meromorphic in the whole s-plane; see [11] for additional information. Expanding Z as $c_0 t^{-d/2} + c_1 t^{-d/2+1} +$ etc., one finds that the zeta-function has simple poles with residues c_n at the places $d/2-n$ $(n \geq 0)$ *if d is odd*, $(0 \leq n < d/2)$ *if d is even*. For even d, the value of the zeta-function at $s = 0$ is $c_{d/2} = \int k_{d/2}$, so that contact is made with R. Seeley's computation of this number [14] and with 2°.

Bibliography

[1] M. Berger. Sur le spectre d'une variètè riemanniene. *C. R. Acad. Sci. Paris Sèr. A-B*, 263:A13–A16, 1966.

[2] G. Borg. Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe Bestimmung der Differentialgleihung durch die Eigenwerte. *Acta Math.*, 78:1–96, 1946.

[3] È Cartan. *Leçons sur la Gèomètrie des Espaces de Riemann.* Gauthier-Villars, 1928.

[4] S. S. Chern. A simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds. *Ann. of Math.*, 2:747–752, 1944.

[5] G. de Rham. *Variètès Diffèrentiables.* Hermann, Paris, 1960.

[6] M. Kac. Can you hear the shape of a drum? *Amer. Math. Monthly*, 73:1–23, 1966.

[7] H. P. McKean. Kramers-Wannier duality for the 2-dimensional Ising model as an instance of Poisson's summation formula. *J. Math. Phys.*, 5:775–776, 1964.

[8] J. Milnor. Eigenvalues of the Laplace operator on certain manifolds. *Proc. Nat. Acad. Sci.*, 51:542, 1964.

[9] S. Minakshisundaram. A generalization of the Epstein zeta function. *Canadian J. Math.*, 1:320–327, 1949.

[10] S. Minakshisundaram. Eigenfunctions on Riemannian manifolds. *J. Indian Math. Soc. (N. S.)*, 17:159–163, 1953.

[11] S. Minakshisundaram and A. Pleijel. Some properties of the eigenfunctions of the Laplace operator on Riemannian manifolds. *Canadian J. Math.*, 1:242–256, 1949.

[12] E. Nelson. The adjoint Markoff process. *Duke Math. J.*, 25:671–690, 1958.

[13] A. Pleijel. A study of certain Green's functions with applications in the theory of vibrating membranes. *Ark. Mat.*, 2:553–569, 1954.

[14] R. Seeley. *The power A^s of an elliptic operator A.* The University of Michigan, 1966.

[15] S. R. S. Varadhan. Diffusion processes in small time intervals. *Comm. Pure Appl. Math.*, 20:659–685, 1967.

[16] R. Weitzenböck. *Invariantentheorie.* P. Noordhoff, Groningen, 1923.

[17] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenschwingungen eines beliebig gestalteten elastischen Körpers. *Rend. Circ. Mat. Palermo*, 39:1–50, 1915.

[18] H. Weyl. *The Classical Groups.* Princeton University Press, New Jersey, 1946.

[7] Hill's Operator and Hyperelliptic Function Theory in the Presence of Infinitely Many Branch Points

[7] Hill's Operator and Hyperelliptic Function Theory in the Presence of Infinitely Many Branch Points. *Comm. Pure Appl. Math.* **29** (1976), 143–226.

7 Hill's Operator and Hyperelliptic Function Theory in the Presence of Infinitely Many Branch Points

Henry P. McKean Jr. and E. Trubowitz[1]

7.1. Hill's Operator: Preliminaries

C_1^k, $k \leq \infty$, is the class of k times continuously differentiable real-valued functions of period 1. Q denotes the Hill's operator $-d^2/dx^2 + q(x)$ with a fixed q of class C_1^∞. The function $y_1(x,\lambda)$, respectively $y_2(x,\lambda)$, is the solution of $Qy = \lambda y$ with $y_1(0,\lambda) = 1$, $y_1'(0,\lambda) = 0$, respectively $y_2(0,\lambda) = 0$, $y_2'(0,\lambda) = 1$. The topics to be discussed are (1) the behavior of $y_1(x,\lambda)$ and $y_2(x,\lambda)$ as $\lambda \to \pm\infty$, (2) the periodic spectrum of Q arising from the eigenfunctions of period 1 or 2, (3) the discriminant $\Delta(\lambda) = y_1(1,\lambda) + y_2/(1,\lambda)$, (4) the tied spectrum comprising the roots of $y_2(1,\mu) = 0$, and (5) the reflecting spectrum comprising of the roots of $y_1'(1,\nu) = 0$. Most of the material can be found in Levitan-Sargsjan [25] and Magnus-Winkler [26]. Readers familiar with such matters can skip at once to Section 2.

Appraisal of y_1 and y_2. The proofs require a number of simple estimates of $y_1(x,\lambda)$ and $y_2(x,\lambda)$. These may be derived by elementary means from

$$\begin{aligned} y_1(\xi,\lambda) &= \cos\sqrt{\lambda}\xi + \int_0^\xi \frac{\sin\sqrt{\lambda}(\xi-\eta)}{\sqrt{\lambda}} q(\eta) y_1(\eta,\lambda)\, d\eta, \\ y_2(\xi,\lambda) &= \frac{\sin\sqrt{\lambda}\xi}{\sqrt{\lambda}} + \int_0^\xi \frac{\sin\sqrt{\lambda}(\xi-\eta)}{\sqrt{\lambda}} q(\eta) y_2(\eta,\lambda)\, d\eta. \end{aligned}$$

The necessary results are

$$\begin{aligned} y_1(\xi,\lambda) &= \cos\sqrt{\lambda}\xi + \frac{\sin\sqrt{\lambda}\xi}{2\sqrt{\lambda}} \int_0^\xi q(\eta)\, d\eta + O(\lambda^{-1}), \\ y_2(\xi,\lambda) &= \frac{\sin\sqrt{\lambda}\xi}{\sqrt{\lambda}} - \frac{\cos\sqrt{\lambda}\xi}{2\lambda} \int_0^\xi q(\eta)\, d\eta + O(\lambda^{-3/2}), \end{aligned}$$

for $\lambda \uparrow \infty$ and

$$\begin{aligned} y_1(\xi,\lambda) &= \cos\sqrt{\lambda}\xi \left[1 + o(1)\right], \\ y_2(\xi,\lambda) &= \frac{\sin\sqrt{\lambda}\xi}{\sqrt{\lambda}} \left[1 + o(1)\right], \end{aligned}$$

for $\lambda \downarrow -\infty$. These estimates may be differentiated with respect to ξ and/or λ and they are uniform on $0 \leq \xi \leq 1$ and in any ball $\max_{0\leq\eta\leq 1} q(\eta) \leq$ constant. They will be employed below without additional comment.

[1] Courant Insitute, New York, NY, USA

Periodic spectrum. This is the spectrum of Q acting on the class of twice differentiable functions of period 2; it is a sequence of simple or double eigenvalues tending to infinity:

$$\lambda_0 < \lambda_1 \le \lambda_2 < \lambda_3 \le \lambda_4 < \lambda_5 \le \lambda_6 < \cdots \uparrow \infty.$$

The lowest eigenvalue λ_0 is simple, the eigenfunction f_0 being root-free and of period 1. Then come pairs of eigenvalues $\lambda_{2i-1} \le \lambda_{2i}$, $i = 1, 2, \cdots$, equality signifying that the eigenspace is of dimension 2. Both the eigenfunctions f_{2i-1} and f_{2i} have i roots in a period $0 \le x < 1$ and are themselves of period 1 or 2 according to the parity of i, i.e. being of period 1 if $i = 2, 4, 6, \cdots$ and of period 2 if $i = 1, 3, 5, \cdots$. The eigenfunctions are normalized by

$$\int_0^1 f_i^2(x)\,dx = 1, \qquad i = 0, 1, 2, \cdots.$$

The eigenvalues obey the estimate

$$\lambda_{2i-1}, \lambda_{2i} = i^2\pi^2 + \int_0^1 q(x)\,dx + O(i^{-2}) \quad \text{as } i \uparrow \infty.$$

The periodic spectrum falls into two parts: the double spectrum of pairs $\lambda_{2i-1} = \lambda_{2i}$ and the simple spectrum comprising $\lambda_0^o = \lambda_0$ and $n \le \infty$ pairs $\lambda_{2i-1}^o < \lambda_{2i}^o$, $i = 1, 2, \cdots, n$, listed in natural order as

$$\lambda_0^o < \lambda_1^o < \lambda_2^o < \cdots < \lambda_{2n-1}^o < \lambda_{2n}^o.$$

The interval $(\lambda_{2i-2}, \lambda_{2i-1})$ is an interval of stability; the nomenclature is explained by the fact that every solution of $Qy = \lambda y$ is bounded on the line if $\lambda_{2i-2} < \lambda < \lambda_{2i-1}$. The complementary intervals of instability $(-\infty, \lambda_0]$, $[\lambda_{2i-1}, \lambda_{2i}]$, $i = 1, 2, \cdots$, have the opposite property: no solution of $Qy = \lambda y$ is bounded for $\lambda < \lambda_0$ or for $\lambda_{2i-1} < \lambda < \lambda_{2i}$. The simple eigenfunctions f_{2i-1}^o, f_{2i}^o both have m_i roots in a period; naturally, $m_1 < m_2 < \cdots < m_n$. The number $m_i - m_{i-1}$ is one more than the number of pairs of double eigenvalues between λ_{2i-2}^o and λ_{2i-1}^o.

Discriminant. The periodic spectrum may also be described by means of the discriminant $\Delta(\lambda) = y_1(1,\lambda) + y_2'(1,\lambda)$; in fact, $\Delta(\lambda_i) = \pm 2$ with signature $+1$ or -1 according to whether $i \equiv 0, 3 \bmod 4$ or $i \equiv 1, 2 \bmod 4$, the periodic spectrum being just the roots of $\Delta^2(\lambda) - 4 = 0$. The discriminant is an integral function of order $\frac{1}{2}$ and type 1, so $\Delta(\lambda) - 2$, respectively $\Delta(\lambda) + 2$, may be expressed as a constant multiple of the canonical product formed from the roots $\lambda_0, \lambda_3, \lambda_4, \cdots$, respectively $\lambda_1, \lambda_2, \lambda_5, \lambda_6, \cdots$, and the multiplier may be determined from the known estimate

$$y_1(x,\lambda) \sim y_2'(1,\lambda) \sim \cos\sqrt{\lambda} \quad \text{as } \lambda \downarrow -\infty.$$

The moral is that the periodic spectrum λ_i, $i = 0, 1, 2, \cdots$, and the discriminant $\Delta(\lambda)$ are equivalent pieces of information. The general shape of $\Delta(\lambda)$ is seen in Fig. 7.1 below; in particular, the roots of [2] $\Delta^\bullet(\lambda) = 0$ interlace the periodic spectrum: a so-called trivial root appears at a double eigenvalue $[\lambda_{2i-1} = \lambda_{2i}]$, while the remaining non-trivial roots $\lambda_i^\bullet$, $i = 1, \cdots, n$, fall inside the non-degenerate intervals of instability $[\lambda_{2i-1}^o < \lambda_i^\bullet < \lambda_{2i}^o]$.

Tied spectrum. The roots μ_i, $i = 1, 2, \cdots$, of $y_2(1,\mu) = 0$ form the spectrum of Q acting on the class of functions $f \in C^2[0,1]$ with $f(0) = f(1) = 0$. These interlace the periodic spectrum

$$\lambda_1 \le \mu_1 \le \lambda_2 < \lambda_3 \le \mu_2 \le \lambda_4 < \lambda_5 \le \mu_3 \le \lambda_6 < \cdots$$

and fall into two classes: the trivial roots at the double periodic eigenvalues and the remaining non-trivial roots in the non-degenerate intervals of instability $[\lambda_{2i-1}^o, \lambda_{2i}^o]$, $i = 1, \cdots, n$: the

[2] $\bullet$ means $d/d\lambda$.

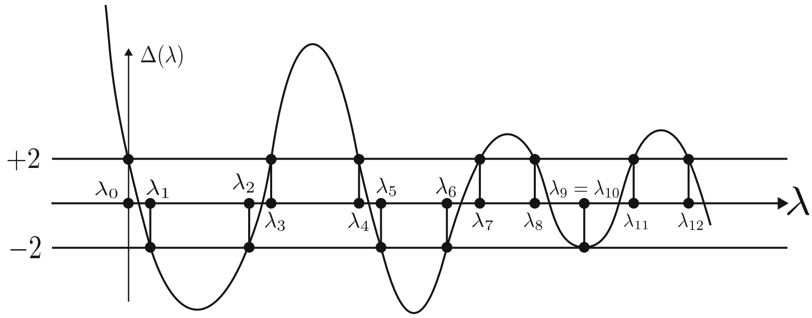

FIGURE 7.1.

latter are denoted by μ_i^o, $i = 1, \cdots, n$. The i-th eigenfunction is proportional to $y_2(x, \mu_i)$ with norming constant[3]

$$\int_0^1 y_2^2(x, \mu_i)\, dx = y_2'(1, \mu_i) y_2^{\bullet}(1, \mu_i).$$

The roots μ_i, $i = 1, 2, \cdots$, determine $y_2(1, \mu)$, the latter being an integral function of order $\frac{1}{2}$, type 1, and of growth $(1/\sqrt{\mu}) \sin \sqrt{\mu}$ for $\mu \downarrow -\infty$.

Reflecting spectrum. A second useful spectrum is formed by the roots $\nu_i, i = 0, 1, 2, \cdots$, of $y_1'(1, \nu) = 0$ forming the spectrum of Q acting on the class of functions $f \in C^2[0, 1]$ with $f'(0) = f'(1) = 0$. These are disposed in the same way as the roots of $y_2(1, \mu) = 0$ except that there is an extra root ν_0 in the interval of instability $(-\infty, \lambda_0]$. The non-trivial roots are denoted by $\nu_0^o = \nu_0$ and $\nu_i^o \in [\lambda_{2i-1}^o, \lambda_{2i}^o]$, $i = 1, \cdots, n$. The i-th eigenfunction is proportional to $y_1(x, \nu_i)$ with norming constant

$$\int_0^1 y_1^2(x, \nu_i)\, dx = -y_1(1, \nu_i) y_1'^{\bullet}(1, \nu_i).$$

The roots $\nu_i, i = 0, 1, 2, \cdots$, suffice to determine $y_1'(1, \nu)$ with the aid of the estimate $y_1'(1, \nu) \sim -\sqrt{\nu} \sin \sqrt{\nu}$ for $\nu \downarrow -\infty$.

7.2. General Introduction

Let M be the class of functions $q \in C_1^\infty$ giving rise to a fixed periodic spectrum $\lambda_0 < \lambda_1 \leq \lambda_2 < \lambda_3 \leq \lambda_4 < \cdots$. The purpose of this paper is to study the geometry of M, with special emphasis upon the connection to hyperelliptic function theory in the presence of infinitely many branch points and with a view to applications to a class of non-linear partial differential equations arising naturally as Hamiltonian flows on M, the Korteweg-de Vries equation $\partial q/\partial t = 3q(\partial q/\partial x) - \frac{1}{2}(\partial^3 q/\partial x^3)$ being the most notable example. Let $2n+1$ be the number of simple periodic eigenvalues. The case of $n < \infty$ was studied by Novikov [32, 33], Dubrovin and Novikov [6], Its and Matveev [16] and McKean and Moerbeke [28]. The case $n = 1$ was already investigated by Hochstadt [13] and partial results were obtained by Flaschka [9, 10], Goldberg [11], and Lax [22]. The investigations of Flaschka [9, 10], Kac-Moerbeke [17, 18], and Moser [29] are closely related. Lax [22] and McKean-Moerbeke [28] may be consulted for additional literature. A general review of the contents of these papers is presented below together with a sketch of the principal results obtained in the present paper.

[3]$'$ means $\partial/\partial x$, $^\bullet$ means $\partial/\partial \mu$.

Local Hamiltonian flows: $n \le \infty$. The partition function

$$Z(t) = \sum_{i=0}^{\infty} e^{-\lambda_i t}$$

possesses an expansion[4]

$$Z(t) \sim \frac{1}{\sqrt{\pi t}} \sum_{j=0}^{\infty} \frac{(-t)^j H_{j-1}}{(2j-3)\cdots 3\cdot 1} \qquad \text{as } t \downarrow 0$$

in which $H_{-1} \equiv 1$, the other coefficients H_j, $j \ge 0$, being integrals over a period $0 \le x < 1$ of universal polynomials in $q(x)$, $q'(x)$, etc., without constant term. They may be computed by the rule[5]

$$V_j q = L\partial H_{j-1}/\partial q, \quad j = 1, 2, \cdots,$$

in which

$$V_j : q \to D\partial H_j/\partial q$$

and L is the skew symmetric operator[6] $qD + Dq - \frac{1}{2}D^3$. The results obtained for $j = 0, 1, 2$ are displayed in the table. Now[7]

j	0	1	2
H	$\int_0^1 q$	$\frac{1}{2}\int_0^1 q^2$	$\frac{1}{2}\int_0^1 q^3 + \frac{1}{4}\int_0^1 q'^2$
$\partial H/\partial q$	1	q	$\frac{3}{2}q^2 - \frac{1}{2}q''$
Vq	0	q'	$3qq' - \frac{1}{2}q'''$

$V = V_j$ induces a local Hamiltonian flow $exp(tV)$ in C_1^∞. This means that $\partial q/\partial t = Vq$ can be solved in C_1^∞ for all time $-\infty < t < \infty$; the nomenclature refers first to the fact that $q \to Vq$ is a local (differential) operator, and second, to the Hamiltonian form of Vq : the gradient of H followed by the skew symmetric operator D. These flows preserve M. For instance, if λ is a simple periodic eigenvalue of Q and if f is the associated eigenfunction, then $\partial\lambda/\partial q = f^2$ and $Lf^2 = 2\lambda Df^2$ by elementary computation, so that under the flow $\partial q/\partial t = V_i q$

$$\begin{aligned} 2\lambda V_i\lambda &= 2\lambda \int_0^1 \frac{\partial\lambda}{\partial q} V_i q = \int_0^1 f^2 2\lambda D \frac{\partial H_i}{\partial q} \\ &= -\int_0^1 \frac{\partial H_i}{\partial q} 2\lambda D f^2 \\ &= -\int_0^1 \frac{\partial H_i}{\partial q} L f^2 \\ &= \int_0^1 f^2 L \frac{\partial H_i}{\partial q} \\ &= \int_0^1 f^2 V_{i+1} q \\ &= V_{i+1}\lambda. \end{aligned}$$

$V_i\lambda = 0$, $i = 1, 2, \cdots$, follows inductively from the self-evident fact that translation of q does not change λ, i.e., $V_1\lambda = 0$.

---

[4] $2j-3$ is to be interpreted as 1 if $j - 0$ or 1.

[5] $\partial/\partial q$ stands for the gradient defined by $H^\bullet = \int_0^1 [\partial H/\partial q(x)]\, q^\bullet(x)\, dx$ for any smooth functional H and any reasonable infinitesimal variation $q^\bullet$ of q, precisely as in advanced calculus.

[6] D means d/dx.

[7] Sjöberg [35] proved the existence of the flow $exp\{tV\}$ in C_1^∞; this fact will be confirmed for all the flows $exp(tV_j)$ in a different way in Section 13, below.

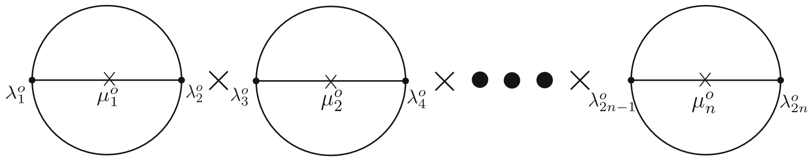

FIGURE 7.2.

M as a torus: $n < \infty$. Now for a bit of geometry. The vectors $V_i q$, $i = 1, \cdots, n$, span the tangent space of M at each point; in particular, $V_{n+1}q$ must be dependent upon $V_i q$, $i = 1, \cdots, n$, i.e., you must have an identity $c_1 V_1 q + \cdots + c_n V_n q + V_{n+1} q = 0$ with coefficients c_i, $i = 1, \cdots, n$, depending upon M but not upon q. This is a differential equation for q; for $n = 0$, it states that $V_1 q = q' = 0$, proving a theorem of Borg [2] that the periodic spectrum is all double except for λ_0 if and only if q is constant, while, for $n = 1$, it states that $c_1 V_1 q + V_2 q = c_1 q' + 3qq' - \frac{1}{2}q''' = 0$, from which you may derive the theorem of Hochstadt [12] that $n = 1$ if and only if $q = a\wp + b$ with $a = 2$ and a Weierstrassian elliptic function $\wp$ with real period 1 and real elliptic invariants $e_1 = -\lambda_2^o + c$, $e_2 = -\lambda_1 + c$, $e_3 = -\lambda_0^o + c$, $c = \frac{1}{3}(\lambda_0^o + \lambda_1^o + \lambda_2^o)$, looked at on $(0 \le x < 1) + \frac{1}{2}\times$ the imaginary period, but this is to digress. The geometrical fact that the $V_i q$, $i = 1, \cdots, n$, span the tangent space of M permits the introduction of a natural coordinate system: a point $t \in \mathbb{R}^n$ is associated with a point on M by letting $exp(t_1 V_1 + \cdots + t_n V_n)$ act on a fixed origin of M, and this association is 1:1 modulo the lattice L of periods $\omega \in \mathbb{R}^n$ such that $exp(\omega_1 V_1 + \cdots + \omega_n V_n)$ acts as the identity on M. The upshot is that M may be identified as the factor space $\mathbb{R}^n / L$.

M can be visualized as a torus in a second way. The periodic spectrum is fixed, so that the discriminant $\Delta(\lambda)$ is known. Let the non-trivial tied spectrum μ_i^o, $i = 1, \cdots, n$, be fixed, also. Then the trivial tied spectrum, alias the double periodic eigenvalues, is known, and the norming constants

$$\int_0^1 y_2^2(x, \mu_i)\, dx = y_2'(1, \mu_i) y_2^\bullet(1, \mu_i), \quad i = 1, 2, \cdots,$$

are almost known. In fact, $y_2(1, \mu)$ is determined from the full tied spectrum; thus $y_2'(1, \mu_i)$ is known. Besides, $y_1(1, \mu_i) y_2'(1, \mu_i) = 1$ in view of $y_1 y_2' - y_1' y_2 = 1$ and $y_2(1, \mu_i) = 0$. Therefore, you may solve $\Delta(\mu_i) = y_1(1, \mu_i) + y_2'(1, \mu_i)$ for

$$y_2'(1, \mu_i) = \tfrac{1}{2}\left[\Delta(\mu_i) \pm \sqrt{\Delta^2 - 4}\right],$$

up to an ambiguous signature in front of the radical. This presents a bona fide ambiguity only if $\Delta^2(\mu_i) \neq 4$. i.e., only for the non-trivial roots $\mu_i^o \neq \lambda_{2i-1}^o$, λ_{2i}^o. Borg [2] proved that the roots μ_i, $i = 1, 2, \cdots$, together with the norming constants determine q uniquely; see Levinson [24] for a simple proof. This means that you have a 1:1 map *into* the n-dimensional torus of Fig. 7.2, this being the product of n circles, one for each non-trivial interval of instability $[\lambda_{2i-1}^o, \lambda_{2i}^o]$, $i = 1, \cdots, n$, the upper (lower) semi-circle corresponding to the positive (negative) determination of the radical $\sqrt{\Delta^2(\mu_i^o) - 4}$. The fact that the map is *onto* is not hard to confirm.

Geometry of M. Let the periodic spectrum λ_i, $i = 0, 1, 2, \cdots$, be fixed. Then M is the variety comprising the common roots q in[8] L_1^2 of the relations

$$2 \times (-1)^i = \Delta(\lambda_{2i}) \quad \text{when } \lambda_{2i} \text{ is simple},$$

[8] L_1^2 denotes the space of real-valued measurable functions of period 1 with $\int_0^1 |f(x)|^2\, dx < \infty$.

and

$$(-1)^i = y_1(1, \lambda_{2i}) \quad \text{when } \lambda_{2i} \text{ is double.}$$
$$0 = y_1'(1, \lambda_{2i})$$

These relations are analytic as regards q, so that M is an analytic variety. The proof appears in Section 8.

This way of presenting M suggests a candidate for its normal space. By analogy with finite-dimensional geometry, the gradients

$$\frac{\partial \Delta}{\partial q}(\lambda_{2i}) = -\Delta'(\lambda_{2i}) f_{2i}^2(x) \qquad \text{when } \lambda_{2i} \text{ is simple,}$$

and

$$\left.\begin{aligned} \tfrac{\partial y_1}{\partial q}(1, \lambda_{2i}) &= -y_1(1, \lambda_{2i}) y_1(x, \lambda_{2i}) y_2(x, \lambda_{2i}) \\ \tfrac{\partial y_1'}{\partial q}(1, \lambda_{2i}) &= y_2'(1, \lambda_{2i}) y_1^2(x, \lambda_{2i}) \end{aligned}\right\} \text{when } \lambda_{2i} \text{ is double}$$

ought to be normal fields spanning out the normal space at every point of M. Write N^o for the span of the gradients $\dfrac{\partial \Delta}{\partial q}(\lambda_{2i}^o)$, $i = 0, 1, \cdots, n$, N_x for the span of $\partial y_1(1, \lambda)/\partial q$ and $\partial y_1'(1, \lambda)/\partial q$ for such double eigenvalues as may exist, and N for the sum of these two. The latter is the normal space.

By the same analogy, it is natural to ask if M is non-singular in the sense that the gradients are independent at each point. This is indeed the case: no single gradient lies in the span[9] of the others. Now in a finite number of dimensions, it is a consequence of the implicit function theorem that a non-singular variety is both smooth and smoothly imbedded, but here this is not at all apparent. First of all, if M is infinite-dimensional, it cannot have a local structure modeled by an infinite-dimensional topological vector space, because M is compact while such a vector space is never locally compact, so it is not even clear what smooth means. Another problem is that the gradients may be independent but the "proportion" of the ambient space spanned by them may very well change from point to point, and this is an intuitively unacceptable characteristic of an imbedding. These difficulties may be overcome. The normal spaces can be compared with one another in a very satisfactory way; for example, if M is finite-dimensional, then the infinite-dimensional normal spaces all have the same finite codimension, and in general it will develop that M "sits smoothly" in L_1^2.

The next topic is the tangent space. In Section 3, it is shown that the flows

$$\frac{\partial q}{\partial t} = D \frac{\partial \Delta}{\partial q}(\lambda_{2i}^o) \equiv X_i q, \quad i = 0, 1, \cdots, n,$$

preserve M, and so it makes good sense to regard the $X's$ as tangent fields defined on M, acting on functions F of q by the rule

$$X_i \text{ applied to } F \text{ at } q = \int_0^1 \frac{\partial F}{\partial q} X_i q \, dx.$$

Write T for the closed span of the $X_i q$, $i = 1, \cdots, n$. The latter is the tangent space. Note that $X_0(q)$ is omitted, the reason being that X_0 is dependent upon the other fields.

This picture of how M is situated in L_1^2 is justified in Sections 8, 9, and 13. N and T are seen to be perpendicular subspaces adding up to L_1^2. The perpendicularity confirms the idea that N is normal and T is tangent, and the fact that they span the whole of L_1^2 shows that every direction of the ambient space has been accounted for. Moreover, it is shown that the normal and tangent spaces, as framed respectively by the gradients of the defining relations and $Xq's$, "look alike" from point to point in a very satisfactory sense, elaborated in Section 8.

[9] Span means the closed linear span in L_1^2.

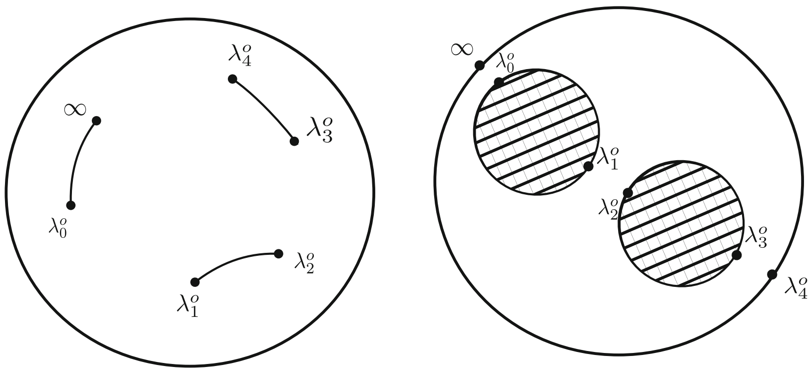

FIGURE 7.3.

The implication that M is smoothly imbedded in L_1^2 is confirmed in Section 13. where it is shown that M is the smooth imbedding in L_1^2 of the product of n circles.

It is an important geometrical fact that M has a symplectic structure, or something like one. D maps N^o onto T. This provides M with a symplectic structure with Poisson bracket

$$\int_0^1 \frac{\partial F}{\partial q}\, D\, \frac{\partial G}{\partial q}\, dx;$$

moreover, the system is completely integrable. The only discrepancy between this and the classical case is that the dimension of the normal space is now 1 more than the dimension of the tangent space, reflecting the non-classical one-dimensional degeneracy of the Poisson bracket. Relative to the bracket, the flows induced by the vector fields X_i, $i = 1, \cdots, n$, are in Hamiltonian form, i.e., $X_i q$ is the gradient $\partial/\partial q$ of the Hamiltonian $\Delta(\lambda_{2i}^o)$ followed by the skew-symmetric operator D. Moreover, they commute. The skew-symmetric operator $L = qD + Dq - \frac{1}{2}D^3$ also maps N^o onto T owing to the identity $Lf_{2i}^{o2} = 2\lambda_{2i}^o D f_{2i}^{o2}$. The interaction between D and L is very important for the whole subject; it would be valuable to understand its real significance from the standpoint of symplectic geometry.

Connection with Hyperelliptic function theory: $n < \infty$. Now comes the main point. The Riemann surface S of the hyperelliptic irrationality

$$\ell(\lambda) = \sqrt{-(\lambda - \lambda_0^o)(\lambda - \lambda_1^o)(\lambda - \lambda_2^o)\cdots(\lambda - \lambda_{2n}^o)}$$

is a sphere with n handles, obtained by cutting 2 copies of the number sphere along the intervals of instability $[-\infty, \lambda_0^o]$, $[\lambda_{2i-1}^o, \lambda_{2i}^o]$, $i = 1, 2, \cdots, n$, and pasting them together as in Fig. 7.3. Let $\mathbf{o}_i$ be the point λ_{2i-1}^o of S and let $\mathbf{p}_i = (\mu_i^o, \sqrt{\Delta^2(\mu_i^o) - 4})$ be a variable point of S situated on the circle arising from $[\lambda_{2i-1}^o, \lambda_{2i}^o]$. Then $(\mathbf{p}_1, \cdots, \mathbf{p}_n)$ is a point of the torus of Fig. 7.2 and is related to the point $t \in \mathbb{R}^n$ by the Jacobi map

$$\sum_{i=1}^{n} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \frac{\mu^{j-1}\, d\mu}{\ell(\mu)} = t_j, \quad j = 1, \cdots, n,$$

the point being that M is, so to say, the real part of the Jacobi variety of the Riemann surface S.[10] The Jacobi map, or better the addition theorem for abelian integrals, may be regarded as an isomorphism between the two different operations of addition with which M is equipped:

[10]To be quite honest, this is not correct. The present coordinate t agrees with the previous one modulo the period lattice L only after an affine transformation, but that is of no importance now.

the natural addition on the torus of Fig. 7.2 and the addition provided by the presentation of M as $\mathbb{R}^n/L$.[11]

Connection with Hyperelliptic function theory: $n = \infty$. The general case $n = \infty$ leads to a similar theory, though new obstacles arise; it is the purpose of this paper to overcome them. Most importantly, it is necessary to develop to some extent the hyperelliptic function theory in the presence of infinitely many branch points. The only previous attempt known to us is that of Myrberg [30, 31]. The methods employed below are quite different from his. M is now an infinite-dimensional torus identifiable (with tears) as the real part of the Jacobi variety of the transcendental hyperelliptic irrationality $\sqrt{\Delta^2(\lambda) - 4}$. *The new aspect of this identification is best explained for a purely simple spectrum.* M is pictured as in Fig. 7.2 with $n = \infty$, and the Jacobi map takes the form

$$\sum_{i=1}^{\infty} 2 \int_{\mathbf{o}_i}^{\mathbf{p}_i} \frac{\phi(\mu)\, d\mu}{\sqrt{\Delta^2(\lambda) - 4}} = \sum_{j=1}^{\infty} \phi(\lambda_{2j}) x_j,$$

in which $\mathbf{o}_i = \lambda_{2i-1}$ and $\mathbf{p}_i = (\mu_i, \sqrt{\Delta^2(\mu_i) - 4})$, $i = 1,\, 2, \cdots,$ ϕ belongs to the class $I^{3/2}$ of integral function of order $\frac{1}{2}$ and type at most 1 with

$$\int_0^{\infty} |\phi(\mu)|^2 \mu^{3/2}\, d\mu < \infty,$$

x is a point of the Hilbert space $I^{3/2\dagger}$ for the quadratic form $\sum |x_i|\, i^{-4}$, and the path of integration from $\mathbf{o}_i$ to $\mathbf{p}_i$ is permitted to wind about the circle n_i times provided $\sum n_i^2 i^{-2} < \infty$. The map requires some explanation. The left-hand side defines and element x of the space $I^{3/2\dagger}$ dual to $I^{3/2}$. Now any function $\phi \in I^{3/2}$ can be interpolated off $\lambda_{2i}, i = 1,\, 2, \cdots,$ the quadratic form $\sum \phi(\lambda_{2i})^2 i^{-4}$ being comparable to

$$\int_0^{\infty} |\phi(\mu)|^2 \mu^{3/2}\, d\mu.$$

Therefore, the sum at the left-hand side in the Jacobi map may be expressed by the kind of sum at the right, $x \in I^{3/2\dagger}$ being identified with the point $(x_1, x_2, \cdots)$, from the Hilbert space of the dual quadratic form $\sum |x_i|^2 i^{-4}$. This explains the Jacobi map $\mathbf{p}_i$, $i = 1,\, 2, \cdots, \to x \in I^{3/2\dagger}$. The paths of integration are not specified, so the map is well-defined if and only if you factor out the lattice $L^{3/2}$ of periods $\omega \in I^{3/2\dagger}$ arising from closed paths of integration, permitting the identification of M as the factor space $I^{3/2\dagger}/L^{3/2}$, alias the real part of the Jacobi variety of S. The primitive periods ω_i, $i = 1,\, 2, \cdots,$ are determined by the rule

$$4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \frac{\phi(\mu)\, d\mu}{\sqrt{\Delta^2(\mu) - 4}} = \sum_{j=1}^{\infty} \phi(\lambda_{2j}) \omega_{ij} \equiv \omega_i(\phi)$$

and every period $\omega \in L^{3/2}$ can be expressed as $\sum n_i \omega_i$ with $\sum i^{-2} n_i^2 < \infty$. The chief new point in all this is that the differential of the first kind on the Riemann surface S employed classically if $n < \infty$, namely $\mu^{j-1} \ell^{-1}(\mu)\, d\mu$, $j = 1,\, 2, \cdots, n$, must now be replaced by the more recondite differentials $\phi(\mu)[\Delta^2(\mu) - 4]^{-1/2}\, d\mu$ with $\phi \in I^{3/2}$. This is confirmed by the fact that the map $\phi \to \omega_i(\phi)$, $i = 1, 2, \cdots,$ is 1:1 from $I^{3/2}$ onto the Hilbert space for the quadratic form $\sum |\omega_i(\phi)|^2 i^2$; in particular, $\omega_i(\phi) = 0, i = 1,\, 2, \cdots,$ only of $\phi = 0$, this being an extension of the classical theorem that if the periods of a differential of the first kind vanish on the cycles of S arising from $[\lambda_{2i-1}^o, \lambda_{2i}^o]$, $i = 1, \cdots, n$, then the differential itself must vanish identically. This kind of result is based upon a general interpolation theorem

[11] Siegel [34] is the best source of information about the classical function theory involved here.

to the effect that $\phi \in I^{3/2}$ can be interpolated off any tied spectrum μ_i, $i = 1, 2, \cdots$, the quadratic form $\sum |\phi(\mu_i)|^2 i^4$ being comparable to

$$\int_0^\infty |\phi(\mu_i)|^2 \mu^{3/2}\, d\mu;$$

but more of this below. The case of partly simple spectrum is more complicated; see Section 15 for full explanations. The theory of the complex part of the Jacobi variety for $n = \infty$ is not yet fully understood. Now, the map

$$q \to \mathbf{p}_i,\ i = 1, 2, \cdots, \to x \in I^{3/2\dagger}$$

is defined modulo periods. It is inverted by the exponential map of $I^{3/2\dagger}$ *onto* M, namely,

$$x \to X = \sum_{i=1}^{\infty} x_i X_i \to q = e^X \text{ applied to the origin.}$$[12]

This permits the identification of $L^{3/2}$ with the lattice of points $x \in I^{3/2\dagger}$ for which $\exp(\sum_{i=1}^\infty x_i X_i)$ acts as the identity on M. It is clear that the exponential map *cannot be* locally 1:1 because M is compact while the unit ball of $I^{3/2\dagger}$ is not. Indeed, every ball in $I^{3/2\dagger}$ contains periods from $L^{3/2}$.

Off hand, it seems that $I^{3/2\dagger}$ ought to be called a tangent space of M even though the space of tangent vectors $Xq = \sum_{i=1}^\infty x_i X_i q$ with $x \in I^{3/2\dagger}$ is much smaller than T: unfortunately, $I^{3/2\dagger}$ does not contain *any* of the local Hamiltonian vector fields such as the infinitessimal translation $V_1 q = q'$ or the Korteweg-de Vries field $V_2 q = 3qq' - \frac{1}{2}q'''$. This defect is remedied by the introduction of a hierarchy of tangent spaces $I^{j/2\dagger}, j$ odd,

$$I^{3/2\dagger} \subset I^{5/2\dagger} \subset \cdots \subset I^{\infty/2\dagger} \subset T$$

with $V_i \in I^{(4i+1)/2\dagger}$, $i = 1, 2, \cdots$, as well as corresponding period lattices $L^{j/2} \subset I^{j/2\dagger}$ introduced in such a way that M may be identified with $I^{j/2\dagger}/L^{j/2}$ by means of a generalization of the Jacobi map. $I^{j/2\dagger}$ is the Hilbert space for the quadratic form $\sum_{i=1}^\infty x_i^2 i^{-j-1}$, and $I^{j/2\dagger} \subset T$ in the sense that $Xq \in T$ if $x \in I^{j/2\dagger}$. $I^{\infty/2\dagger}$ is not a Hilbert space, being simply the union of the $I^{j/2\dagger}s$; in spite of this $I^{\infty/2\dagger}$ is a very nice tangent space, accounting for all the local flows induced by V_1, V_2, etc.; see Section 14 for more information.

Applications to partial differential equations. The relationship between M, S and the pair $I^{j/2\dagger}, L^{j/2} (j$ odd) provided by the sequence of maps

$$M \xrightarrow[\text{tied spectrum and norming constants}]{} S \xrightarrow[\text{Jacobi map}]{} I^{j/2^\dagger}/L^{j/2} \xrightarrow[\text{exponential map}]{} M$$

for $j = 4i + 1$ can be used to study the local Hamiltonian flow $\partial q/\partial t = V_i q$. The first application is a proof of the existence and smoothness of all the local flows. Sjöberg [35] proved global existence for $\exp\{tV_2\}$ in a different way. Another application is to prove that $\exp\{tV_i\}q$, $i = 1, 2, \cdots$, is almost periodic in time for any $q \in C_1^\infty$ as conjectured by Lax [22] and numerically confirmed by M. Hyman in the appendix to [22]; see Section 16.

The machinery can also be used to study the periodicity of the local flows. For example, the Korteweg-de Vries flow induced by $V_2 q = 3qq' - \frac{1}{2}q'''$ is of period T if and only if

$$2\sum_{i=1}^\infty n_i \int_{\lambda_{2i-1}}^{\lambda_{2i}} \frac{y_2(1,\mu)}{\mu^j} \frac{d\,\mu}{\sqrt{\Delta^2(\mu) - 4}} = \begin{cases} cT & \text{for} \quad j = 1, \\ 2T & \text{for} \quad j = 2, \\ 0 & \text{for} \quad j \geq 3, \end{cases}$$

[12]The origin is now fixed as the point on M with $\mu_i = \lambda_{2i-1}$, $i = 1, 2, \cdots$.

with $c = \lambda_0 - \sum_{i=1}^{\infty}(\lambda_{2i} - \lambda_{2i-1})$ and integers n_i, $i = 1, 2, \cdots$, satisfying $\sum n_i^2 i^{-8} < \infty$. The periodicity of the flows induced by $V_i \in I^{(4i+1)/2\dagger}$ can be discussed in a similar manner.

Acknowledgement. The authors would like to thank P. Lax, P. van Moerbeke, J. Moser, and L. Nirenberg for helpful conversations and friendly encouragement.

Note Added in Proof. We have learned from F. J. Dyson that O. Steinmann has observed M to be a product of circles and has connected the $df^2/dx's$ with isospectral deformations of q; see his Zurich Diplomarbeit (1957) and his paper *Äquivalente periodische Potentiale*, Helv. Phys. Acta 30, 1957, p. 515 – 520. Dyson also told us that W. Kohn has studied the analytic nature of the periodic spectrum; see his paper *Analytic properties of the Bloch waves and Wannier functions*, Phys. Rev. 115, 1959, pp. 802–821. We would like to thank F. J. Dyson for this information and for providing us with a copy of his 1965 La Jolla Seminar, *Determination of a Periodic Potential from its Band Structure.*

7.3. A New Class of Hamiltonian Flows

The study of the geometry of M is based upon the fact that the periodic spectrum, and so M itself, is preserved by the flow with Hamiltonian $H = \Delta(\lambda)$, λ fixed. The flow is defined by solving $\partial q/\partial t = Xq$ with $Xq = D\partial H/\partial q$; its existence is proven below. The present flows are closely allied to those of Lax [22]. They represent an improvement over the local Hamiltonian flows of Section 2 in that it seems the latter need not be transitive on M; see Section 10 below. They are also easier to deal with from a technical standpoint. A number of additional facts about $\partial\Delta/\partial q$ are prepared in this section. L stands for the skew symmetric operator $qD + Dq - \frac{1}{2}D^3$.

LEMMA 1.

$$\begin{aligned}\frac{\partial y_1(\xi,\lambda)}{\partial q(\eta)} &= y_2(\eta,\lambda)y_1^2(\eta,\lambda) - y_1(\xi,\lambda)y_1(\eta,\lambda)y_2(\eta,\lambda),\\ \frac{\partial y_2(\xi,\lambda)}{\partial q(\eta)} &= y_2(\eta,\lambda)y_1(\eta,\lambda)y_2(\eta,\lambda) - y_1(\xi,\lambda)y_2^2(\eta,\lambda),\end{aligned}$$

for $0 \le \eta \le \xi \le 1$, *the gradients vanishing otherwise. Both evaluations may be differentiated with regard to* ξ.

Proof: Let $q^\bullet$ be an infinitessimal variation of q. Then $(\lambda - Q)y_1^\bullet = q^\bullet y_1$ can be solved for $y_1^\bullet(\xi,\lambda)$:

$$y_1^\bullet(\xi,\lambda) = \int_0^\xi [y_2(\xi,\lambda)y_1(\eta,\lambda) - y_1(\xi,\lambda)y_2(\eta,\lambda)]q^\bullet(\eta)y_1(\eta,\lambda)\,d\eta,$$

and $\partial y_1/\partial q$ can be read off. The computation of $\partial y_2/\partial q$ is similar.

LEMMA 2.

$$\frac{\partial\Delta(\lambda)}{\partial q(x)} = [y_2'(1,\lambda)y_1(1,\lambda)]y_1(x,\lambda)y_2(x,\lambda) - y_1'(x,\lambda)y_2^2(x,\lambda) + y_2(1,\lambda)y_1^2(x,\lambda).$$

Proof: See Lemma 1.

LEMMA 3. *The product* z *of any* 2 *solutions of* $Qy = \lambda y$ *satisfies* $Lz = 2\lambda z'$; *in particular,* y_1^2, y_1y_2, y_2^2 *is a base for the null space of* $L - 2\lambda D$ *since the three functions are independent and the dimension of the null space is* $\deg L = 3$.

Proof: Let $z = y_-y_+$. Then

$$\begin{aligned}Lz &= (qD + Dq)z - \tfrac{1}{2}D(y_-''y_+y + y_-y_+'') - (y_-''y_+' + y_-'y_+'')\\ &= (qD + Dq)z - D(q-\lambda)z - (q-\lambda)Dz\\ &= 2\lambda z'.\end{aligned}$$

The rest is plain.

THEOREM 1. *Let* λ *be fixed and let* $\Delta(\lambda) = H$. *Then* $X : q \to D\partial H/\partial q$ *induces a smooth flow* $\exp\{tX\}$ *in* C_1^∞ *by solving* $\partial q/\partial t = Xq$; *this flow preserves* $\Delta(\mu)$ *for any value of* μ *and so also the manifold* M.

AMPLIFICATION 1. The theorem may be interpreted as saying that X is a tangent field to M.

AMPLIFICATION 2. The flows commute. The point is that $\partial q/\partial t = Xq$ is in Hamiltonian form: Xq is the gradient of H followed by the skew symmetric operator D. Now it is a general fact of Hamiltonian mechanics that if two flows preserve each other's Hamiltonian, then they commute, and so it is here: the λ-flow preserves the Hamiltonian $\Delta(\mu)$ of the μ-flow; compare [28], pp. 228–229, for the kind of computation involved.

Proof of Theorem 1: It is required to solve $\partial q/\partial t = Xq$ in C_1^∞. We have

$$Xq = \frac{d}{dx}[(y_2'(1,\lambda) - y_1(1,\lambda))y_1y_2 + y_2(1,\lambda)y_1^2 - y_1'(1,\lambda)y_2^2]$$

by Lemma 2. Introduce the conventional Sobolev norms

$$\|f\|_i = \sqrt{\sum_{j\le i}\int_0^1 |D^i f|^2\,dx}, \quad i = 0, 1, 2, \cdots,$$

and the dual (negative) norms $\|f\|_i$, $i = -1, -2, \cdots$. Then it is easy to see from

$$\begin{aligned} y_1(\xi,\lambda) &= \cos\sqrt{\lambda}\xi + \int_0^\xi \frac{\sin\sqrt{\lambda}(\xi-\eta)}{\sqrt{\lambda}} q(\eta)y_1(\eta,\lambda)\,d\eta, \\ y_2(\xi,\lambda) &= \frac{\sin\sqrt{\lambda}\xi}{\sqrt{\lambda}} + \int_0^\xi \frac{\sin\sqrt{\lambda}(\xi-\eta)}{\sqrt{\lambda}} q(\eta)y_2(\eta,\lambda)\,d\eta \end{aligned}$$

that $\|y_1\|_i$ and $\|y_2\|_i$ are controlled by $\|q\|_{i-2}$. Now use Lemma 1 to prove that[13] $\|\partial(Xq)/\partial q(\eta)\|_i$ is bounded in any ball $B : \|q\|_i \le b$ of the i-th Sobolev space of functions of period 1. This permits you to solve $\partial q/\partial t = Xq$ in B until such time as q hits ∂B. In [22] and [28] it was proved that $\|q\|_i$ is controlled by H_j, $j \le i+1$, with the local Hamiltonian of Section 2. The flow $\partial q/\partial t = Xq$ preserves these Hamiltonians; in fact, it preserves $\Delta(\mu)$ for any fixed μ and hence the periodic spectrum from which the Hamiltonians are derived. The proof is easy: $\partial\Delta/\partial q = f$ satisfies $Lf = 2\lambda f'$, so

$$\begin{aligned} 2\lambda X\Delta(\mu) &= \int_0^1 \frac{\partial\Delta(\mu)}{\partial q} 2\lambda Xq\,dx \\ &= \int_0^1 \frac{\partial\Delta(\mu)}{\partial q} 2\lambda \frac{d}{dx}\frac{\partial\Delta(\lambda)}{\partial q}\,dx \\ &= \int_0^1 \frac{\partial\Delta(\mu)}{\partial q} L \frac{\partial\Delta(\lambda)}{\partial q}\,dx, \end{aligned}$$

which is antisymmetric in λ and μ in view of the skew-symmetry of L, i.e.,

$$2\lambda X\Delta(\mu) = -2\mu X\Delta(\mu) = -\mu\int_0^1 \frac{\partial\Delta(\mu)}{\partial q}\frac{d}{dx}\frac{\partial\Delta(\mu)}{\partial q}\,dx = 0,$$

by periodicity of $\partial\Delta/\partial q$. Now fix a region A in the i-th Sobolev space of functions of period 1 by limiting the values of H_j, $j = 0, 1, \cdots, i+1$, and enclose it in a ball $B : |q|_i \le b$. Then a solution of $\partial q/\partial t = Xq$ starting in A exists until such time as it meets ∂B, and as that cannot happen, the solution exists for all time, the point is that the solution cannot exit from A since the Hamiltonians are preserved. A flow in C_1^∞ is now obtained by noting that, for

[13] $\|\partial(Xq)(\xi)/\partial(\eta)\|_i$ is the Sobolev norm in 2 variables $0 \le \xi, \eta \le 1$.

initial data from C_1^∞, the solution of $\partial q/\partial t = Xq$ in the i-th periodic Sobolev space does not depend upon i and so belongs to their intersection C_1^∞. The smoothness of the flow is proved as follows. Let q_0 be the initial value of q. Formally, $\partial q/\partial t = Xq$ implies

$$\frac{d}{dt}\frac{\partial q(\xi)}{\partial q_0(\eta)} = \frac{\partial Xq(\xi)}{\partial q_0(\eta)} = \int_0^1 \frac{\partial Xq(\xi)}{\partial q(x)}\frac{\partial q(x)}{\partial q_0(\eta)}\,dx,$$

which is to say that the operator $J(\xi,\eta) = \partial q(\xi)/\partial q_0(\eta)$ out to satisfy $\partial J/\partial t = KJ$ with $K(\xi,\eta) = \partial Xq(\xi)/\partial q(\eta)$. Now $K(\xi,\eta)$ is smooth on $[0,1]^2 \times C_1^\infty$ by Lemma 1, and $\partial J/\partial t = KJ$ can be solved for J, starting from the unit operator $J_0 = 1$. It is required to prove that, with this determination of J,

$$e^{tX}(q_0 + \varepsilon q_1) = e^{tX}q_0 + \varepsilon q_1 + o(\varepsilon).$$

This is easy: $\varepsilon^{-1}[e^{tX}(q_0+\varepsilon q_1) - e^{tX}q_0] = p$ solves

$$\partial p/\partial t = \varepsilon^{-1}]X(q+\varepsilon p) - Xq] = Kp + O(\varepsilon),$$

so $p = Jq_1 + O(\varepsilon)$, as required. The existence of the higher derivatives of $q(\xi)$ with regard to $q_0(\eta)$ is proved in a similar way.

A number of additional facts about $\partial\Delta/\partial q$ are collected below.

LEMMA 4. $\partial\Delta(\lambda)/\partial q(x)$ *is the only function of period 1 in the null space of* $L - 2\lambda D$ *unless* $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 = 2$.

Proof: $Lz = 2\lambda z'$ only if $f = ay_1^2 + by_1y_2 + cy_2^2$, by Lemma 3. The condition of periodicity is that z, z', and z'' match at $x = 0$ and at $x = 1$, i.e., that (a,b,c) be annihilated by

$$K = \begin{bmatrix} [y_1(1,\lambda)]^2 - 1 & y_1(1,\lambda)y_2(1,\lambda) & [y_2(1,\lambda)]^2 \\ y_1(1,\lambda)y_1'(1,\lambda) & y_1'(1,\lambda)y_2(1,\lambda) & y_2(1,\lambda)y_2'(1,\lambda) \\ [y_1'(1,\lambda)]^2 & y_1'(1,\lambda)y_2'(1,\lambda) & [y_2'(1,\lambda)]^2 - 1 \end{bmatrix}.$$

The condition for two independent solutions of period 1 is that $\det(K - k) = 0$ have a root of multiplicity at least 2 at $k = 0$, that is to say $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 = 2$, by a tiresome computation. The proof is finished by noticing that $\partial\Delta/\partial q = z$ solves $Lz = 2\lambda z'$, is of period 1, and cannot vanish if $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 \neq 2$ because you cannot have (a) $y_2'(1,\lambda) = y_1(1,\lambda)$ and (b) $y_1'(1,\lambda) = y_2(1,\lambda) = 0$; in fact, (b) places you at a double eigenvalue $[\Delta(\lambda) = \pm 2]$, so (a) implies $y_1(1,\lambda) = y_2'(1,\lambda) = \pm 1$ and $[y_1(1,\lambda)]^2 + [y_2(1,\lambda)]^2 = 2$.

LEMMA 5. *Let* $0 \le x \le 1$ *be fixed. Then* $\partial\Delta(\lambda)/\partial q(x)$ *coincides with the function* $y_2(1,\lambda)$ *computed for* q *translated by* x.

Proof: The cited function $y_2(1,\lambda)$ is expressed in terms of the original functions y_1 and y_2 as

$$\begin{aligned} &y_2(1+x,\lambda)y_1(x,\lambda) - y_1(1+x,\lambda)y_2(x,\lambda) \\ &\qquad = [y_2(1,\lambda)y_1(x,\lambda) + y_2'(1,\lambda)y_2(x,\lambda)]y_1(x) \\ &\qquad\quad - [y_1(1,\lambda)y_1(x,\lambda) + y_1'(1,\lambda)y_2(x,\lambda)]y_2(x,\lambda) \\ &\qquad = [y_2'(1,\lambda) - y_1(1,\lambda)]y_1(x,\lambda)y_2(x,\lambda) \\ &\qquad\quad - y_1'(1,\lambda)y_2^2(x,\lambda) + y_2(1,\lambda)y_1^2(x,\lambda), \end{aligned}$$

in agreement with the formula for $\partial\Delta/\partial q$ of Lemma 2.

LEMMA 6.

$$\int_0^1 [\partial\Delta(\lambda)/\partial q(x)]\,dx = -\Delta^\bullet(\lambda).$$

Proof: Under the flow $\partial q/\partial t$, $\Delta(\lambda)$ is changed into $\Delta(\lambda - t)$, thus

$$-\Delta^\bullet(\lambda) = \int_0^1 (\partial\Delta/\partial q)q^\bullet = \int_0^1 \partial\Delta/\partial q,$$

as stated.

LEMMA 7. $\partial\Delta(\lambda)/\partial q(x) = -\Delta^\bullet(\lambda)f^2(x)$ *at any simple periodic eigenvalue* $\lambda = \lambda_i^o$, $i = 0, \cdots, 2n$, *with eigenfunction* $f = f_i^o$. *The same holds at a double periodic eigenvalue* $\lambda = \lambda_{2i-1} = \lambda_{2i}$, *except that* $\Delta^\bullet(\lambda) = 0$ *there so that* $\partial\Delta(\lambda)/\partial q(x) = 0$.

Proof: At a double eigenvalue $\lambda_{2i-1} = \lambda_{2i}$, $y_1'(1,\lambda) = y_2(1,\lambda) = 0$ and $y_1(1,\lambda) = y_2'(1,\lambda)$ since $y_1(1,\lambda)y_2'(1,\lambda) = 1$ and $\Delta^2 = [y_1(1,\lambda) + y_2'(1,\lambda)]^2 = 4 = 4y_1(1,\lambda)y_2'(1,\lambda)$. This implies that $\partial\Delta(\lambda)/\partial q(x) = 0$, by Lemma 2. Contrariwise, at a simple eigenvalue λ_i^o, $\partial\Delta(\lambda)/\partial q(x)$ is a non-trivial member of the null-space of $L - 2\lambda D$ of period 1, as is $[f_i^o(x)]^2$, by Lemma 3; hence $\partial\Delta/\partial q$ is proportional to $(f_i^o)^2$ if $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 \neq 2$, by Lemma 4. Besides, $\Delta(\lambda) = y_1(1,\lambda) + y_2'(1,\lambda) = \pm 2$, and you cannot have $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 = 2$ unless $y_1(1,\lambda) = y_2'(1,\lambda) = \pm 1, y_1'(1,\lambda)y_2(1,\lambda) = 0$, and either $y_1'(1,\lambda) = 0$ or $y_2(1,\lambda) = 0$, so that the proportionality of $\partial\Delta/\partial q$ and $(f_i^o)^2$ takes place anyhow. The proof is finished by use of Lemma 6:

$$\int_0^1 (f_i^o)^2 = 1,$$

while

$$\int_0^1 \partial\Delta/\partial q = -\Delta^\bullet,$$

so $\partial\Delta/\partial q = -\Delta^\bullet(f_i)^2$.

LEMMA 8. *Let* $\lambda \neq \lambda_i$, $i = 0, 1, 2, \cdots$, *be fixed and let* $G(\xi,\eta)$ *be the Green function for* $\lambda - Q$ *acting on functions of period* 2. *Then*

$$2G(x,x) = \sum_{i=0}^{\infty} (\lambda - \lambda_i)^{-1} f_i^2(x) = \frac{2\Delta(\lambda)}{4 - \Delta^2(\lambda)} \frac{\partial\Delta(\lambda)}{\partial q(x)}.$$

Proof: The normalized eigenfunctions for Q acting on functions of period 2 are $2^{-1/2}f_i$, $i = 0, 1, 2, \cdots$. This explains the factor 2 in the identity between $2G(x,x)$ and the sum. Now $G(x,x)$ is of period 1, and you have only to rehearse mentally the standard computation of the $G(\xi,\eta)$ to realize that $G(x,x)$ is a sum of products $y_1^2(x,\lambda)$, $y_1(x,\lambda)y_2(x,\lambda)$, $y_2^2(x,\lambda)$ and, as such, a solution of $Lf = 2\lambda f'$ of period 1. $G(x,x)$ is seen to be proportional to $\partial\Delta/\partial q$ by Lemma 4, if $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 \neq 2$ and the constant of proportionality is fixed by Lemma 6 and the self-evident identity

$$2\int_0^1 G(x,x)\,dx = \sum_{i=0}^{\infty} (\lambda - \lambda_i)^{-1} = [\log(\Delta^2(\lambda) - 4)]^\bullet.$$

Clearly, the exceptional values of $\lambda \neq \lambda_i$, $i = 0, 1, 2, \cdots$, at which $[y_1(1,\lambda)]^2 + [y_2'(1,\lambda)]^2 = 2$ may be dealt with by continuity.

AMPLIFICATION 3. The formula $[-\Delta^\bullet(\lambda_i^o)]^{-1}\partial\Delta(\lambda_i^o)/\partial q(x) = [f_i^o(x)]^2$ of Lemma 7 has a counterpart for any double periodic eigenvalue $\lambda_{2i-1} = \lambda_{2i}$: at such eigenvalue

$$2 \times [-\Delta^\bullet(\lambda)]^{-1} \frac{\partial\Delta(\lambda)}{\partial q(x)} = f_{2i-1}^2(x) + f_{2i}^2(x) > 0,$$

as can be seen from Lemma 8, and the functions

$$\begin{matrix} f_-(\xi) \\ f_+(\xi) \end{matrix} = \sqrt{\frac{2\partial\Delta(\lambda)/\partial q(\xi)}{-\Delta^\bullet(\lambda)}} \begin{matrix} \sin \\ \cos \end{matrix} \left[\frac{\sqrt{2}}{\sqrt{|\Delta^{\bullet\bullet}(\lambda)|}} \int_0^\xi \frac{-\Delta^\bullet(\lambda)\,d\eta}{\partial\Delta(\lambda)/\partial q(\eta)} \right]$$

form a unit perpendicular base of the eigenspace. The proof employs the identity

$$\left(\frac{\partial \Delta'}{\partial q}\right)^2 - 2\frac{\partial \Delta}{\partial q}\frac{\partial \Delta''}{\partial q} + 4\left(\frac{\partial \Delta}{\partial q}\right)^2 (q-\lambda) = \Delta^2 - 4.$$

The latter is verified by differentiation of the left-hand side to prove its constancy. The value is now computed at $x = 0$; compare [28], p. 20, for the case $n < \infty$.

AMPLIFICATION 4. The formula of Lemma 2 may be used to prove that the case of purely simple spectrum is typical of $q \in C_1^\infty$. The proof is easy. The part of C_1^∞ in which $\lambda_{2i-1} = \lambda_{2i}$ is closed; it is to be proved that is cannot contain an open region. In such a region, $\Delta(\lambda_{12}) \equiv \pm 2$ for $\lambda_{12} = \frac{1}{2}\lambda_{2i-1} + \frac{1}{2}\lambda_{2i}$. But $\Delta(\lambda)$ is an analytic function of $q \in C_1^\infty$ for fixed $-\infty < \lambda < \infty$, as is apparent from Lemma 3.2 and the integral equations for $y_1(x,\lambda)$ and $y_2(x,\lambda)$ cited in Section 1. The same is true of λ_{12}; this follows from the preceding remark and from the formula

$$\lambda_{12} = \frac{1}{4\pi\sqrt{-1}} \oint \frac{\Delta^\bullet(\mu)}{\Delta(\mu) \pm 2}\, \mu\, d\mu$$

in which the integral is extended about a circle enclosing λ_{2i-1} and λ_{2i} but excluding λ_j, $j \neq 2i-1$, $2i$. Therefore, the composite function $\Delta(\lambda_{12})$ is also analytic in q. Thus $\Delta(\lambda_{12}) \equiv \pm 2$ in the whole of C_1^∞, and that is not the case: namely, if $q = \sin 2\pi x$, then $\lambda_{2i-1} < \lambda_{2i}$ and $\Delta(\lambda_{12}) \neq \pm 2$; see Magnus-Winkler [26], p. 90.

AMPLIFICATION 5. The prevalence of purely simple spectra may be confirmed in another way. The fact is that if q has only simple periodic spectrum, then so does cq for all but a countable number of exceptional values of $-\infty < c < \infty$ with no finite limit point. The case $q(x) = c \sin 2\pi x$ with purely simple spectrum if $c \neq 0$ was mentioned above. A second example is due to Ince [14, 15]: Let $q(x) = (2kK)^2 \mathrm{sn}^2(2kKx, k)$, $\mathrm{sn}(x,k)$ being the customary Jacobi elliptic function and K the complete elliptic integral of the first kind, both with modulus $0 < k < 1$. Then cq has $2n+1$ simple periodic eigenvalues if $c = n(n+1)$, n integral, and only simple periodic eigenvalues in every other case. Now let q be general and let $\Delta(\lambda, c)$ be the discriminant of cq. This function is analytic in the pair (λ, c), permitting you to view the periodic spectrum $\lambda_i(c)$, $i = 1, 2, \cdots$, as the branches of one or more multiple-valued functions of c produced by solving $\Delta^2(\lambda, c) - 4 = 0$. For $c = 1$, the spectrum is simple, i.e., the sheets of the Riemann surface are distinct, and as c ranges over the line, these sheets may meet, but only in pairs, and then only at a countable number of exceptional values of c with no finite limit point. The proof is finished.

7.4. Tied Spectrum: M as a Torus

Figure 7.2 of Section 2, depicting M as an n-dimensional torus, is also valid if $n = \infty$. Recall what is involved. The periodic spectrum is fixed so $\Delta(\lambda)$ is known as well as the trivial part of the tied spectrum, alias the double eigenvalues $\lambda_{2i-1} = \lambda_{2i}$. Let the non-trivial tied spectrum μ_i^o, $i = 1, \cdots, n$, be fixed, too. Then $y_2^\bullet(1, \mu_i^o)$ is known and so is the norming constant

$$\int_0^1 y_2^2(x, \mu_i^o)\, dx = y_2^\bullet(1, \mu_i^o)\, \tfrac{1}{2}[\Delta(\mu_i^o) \pm \sqrt{\Delta^2(\mu_i^o) - 4}],$$

up to the signature of the radical; the latter presents a bona fide ambiguity only if $\lambda_i^o < \mu_i < \lambda_{2i}^o$. Now the theorem of Borg [2][14] cited in this connection in Section 2 state that the map

$$q \to \mathbf{p}_i = (\mu_i^o, \sqrt{\Delta^2(\mu_i^o) - 4}), \quad i = 1, 2, \cdots, n,$$

of M into the n-dimensional torus of Fig. 7.4 is 1:1.

THEOREM 1. *M is topologically equivalent to the n-dimensional torus of Figure 7.4.*

[14] See [24].

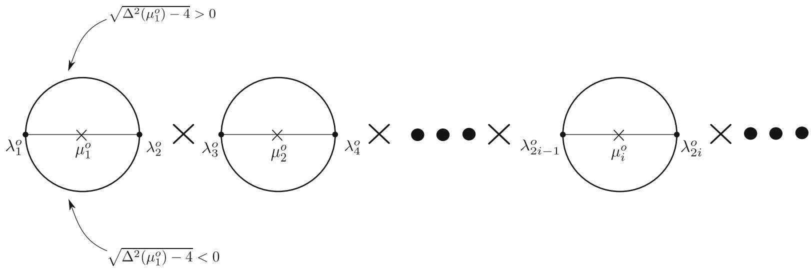

FIGURE 7.4. $n=\infty$

Proof: It is required to prove that the map is onto, i.e., that the eigenvalue μ_i^o can be placed arbitrarily in the interval of instability $[\lambda_{2i-1}^o, \lambda_{2i}^o]$ with an arbitrary signature to the radical $\sqrt{\Delta^2(\mu_i^o)-4}$, simultaneously for $i=1,\cdots,n$.

LEMMA 1. *Let $i=1,\cdots,n$ be fixed. Then the vector field*

$$X: q \to D\,\partial\Delta(\lambda)/\partial q \qquad \textit{evaluated at } \lambda=\mu_i^o$$

induces a smooth flow on M under which

$$X\mu_j^o = \begin{cases} \frac{1}{2}\sqrt{\Delta^2(\mu_j^o)-4} & \textit{for } j=i, \\ 0 & \textit{for } j\neq i. \end{cases}$$

Warning. The present flows commute among themselves but not with the flows induced by $X: q \to D\partial\Delta(\lambda)/\partial q$ with λ fixed. Besides, the present flows are not Hamiltonian: $D[\partial\Delta(\mu_i^o)/\partial q] \neq D\partial\Delta(\lambda)/\partial q$ evaluated at $\lambda=\mu_i^o$.

Proof: The proof of the existence of the flow is routine, making use of Theorem 3.1 and the fact that $\partial\mu_i^o/\partial q = f^2$, $f(x)$ being the normalized eigenfunction $[y_2^\bullet(1,\mu_i^o)y_2'(1,\mu_i^o)]^{-1/2}y_2(x,\mu_i^o)$. Now let f be the normalized eigenfunction for $\mu=\mu_i^o$. Then

$$X\mu_i^o = \int_0^1 \frac{\partial\mu_j^o}{\partial q} Xq\,dx = \int_0^1 f^2(x) D\frac{\partial\Delta(\lambda)}{\partial q(x)}\,dx \text{ evaluated at } \lambda=\mu_i^o,$$

and for general λ,

$$\begin{aligned} 2(\lambda-\mu)\int_0^1 f^2 D\frac{\partial\Delta(\lambda)}{\partial q}\,dx &= \int_0^1 f^2 L\frac{\partial\Delta(\lambda)}{\partial q}\,dx + \int_0^1 \frac{\partial\Delta(\lambda)}{\partial q} Lf^2\,dx \\ &= -(f')^2\frac{\partial\Delta}{\partial q}\Big|_0^1 \\ &= \frac{1-[y_2'(1,\mu)]^2}{y_2^\bullet(1,\mu)y_2'(1,\mu)} y_2(1,\mu). \end{aligned}$$

Therefore, $X\mu_j^o=0$ if $j\neq i$, while for $j=i$,

$$\begin{aligned} X\mu_j^o &= \tfrac{1}{2}[y_2'(1,\mu_j^o)]^{-1} - y_2'(1,\mu_j^o) \\ &= \tfrac{1}{2}\sqrt{\Delta^2(\mu_j^o)-4}, \end{aligned}$$

as advertised.

The proof of the theorem is now before you. The flow induced by X fixes μ_j^o, $j\neq i$, and moves $\mathbf{p}_i = (\mu_i, \sqrt{\Delta^2(\mu_i^o)-4})$ about the circle of Fig. 7.5 according to $X\mu_i^o = \frac{1}{2}\sqrt{\Delta^2(\mu_i^o)-4}$. This motion does not pause at the critical points $\mu_i^o = \lambda_{2i-1}^o$, λ_{2i}^o even, though $\sqrt{\Delta^2(\mu_i^o)-4}$

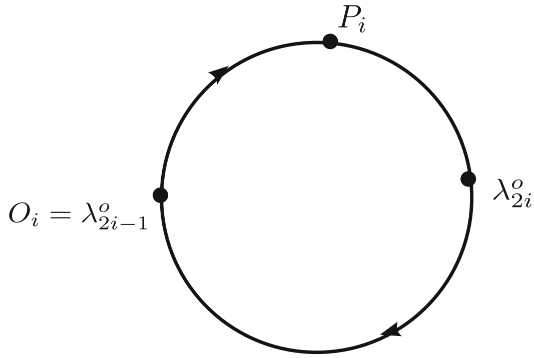

FIGURE 7.5.

vanishes there. The point is that if μ^o_i sticks at λ^o_{2i-1} on λ^o_{2i}, then nothing moves, not even q by Borg's theorem, while

$$Xq = \frac{d}{dx}\{[y_2'(1,\mu^o_i) - y_1(1,\mu^o_i)]y_1(x,\mu^o_i) + y_1'(1,\mu^o_i)y_2^2(x,\mu^o_i)\}$$

cannot vanish unless $y_1'(1,\mu^o_i) = 0$, and that is never the case, μ^o_i being at a simple periodic eigenvalue. These observations permit you to move the $\mathbf{p}_i$, $i = 1, \cdots, m$, to any desired positions for any $m < \infty$ if $n = \infty$ and for $m = n$ if $n < \infty$. The proof that the map is onto is finished for $n < \infty$, while for $n = \infty$, it suffices to note that M is *compact*: M is closed in C^∞_1, the discriminant $\Delta(\lambda)$ being a nice function of q; also, the Hamiltonians H_i, $i = 0, 1, 2, \cdots$, are constant upon M and[15] for any i the i-th Sobolev norm of q is controlled by $H_0, \cdots, H_{i+1}$. The proof that the map is topological is self-evident.

AMPLIFICATION 1. The fields $X : q \to D\partial\Delta(\lambda)/\partial q$, λ fixed, act upon μ^o_j, $j = 1, 2, \cdots$, by the rule

$$X\mu^o_j = \frac{\sqrt{\Delta^2(\mu^o_j) - 4}}{2y_2^\bullet(1,\mu^o_j)}\,\frac{y_2(1,\lambda)}{\lambda - \mu^o_j};$$

in particular, for $X_i = X$ evaluated at $\lambda = \lambda_{2i}$, $\det X_i\mu^o_j$, $1 \le i, j \le m$, does not vanish if $\mu^o_i \ne \lambda^o_{2i-1}, \lambda^o_{2i}$, $i = 1, \cdots, n$. This fact will be helpful in Section 9.

AMPLIFICATION 2. The origin of M is now fixed as that point at which $\mu^o_i = \lambda^o_{2i-1}$, $i = 1, \cdots, n$. In this connection, it is noted that $\mu^o_i = \lambda_{2i-1}$ *or* λ_{2i}, $i = 1, \cdots, n$, *if and only if* $q(x) = q(-x)$.

Proof: If $q(x) = q(-x)$, then both $y_2(1 + x, \mu^o_i)$ and $-y_2(1 - x, \mu^o_i)$ solve $Qy = \mu^o_i y$ and they agree by computation of value and slope at $x = 0$. Now compute the value and slope at $x = -1$ to obtain $y_2(0) = y_2(0) = 0$ and $y_2'(0) = y_2'(2) = 1$. The conclusion is that y_2 is of period 2 and $\mu^o_i = \lambda_{2i-1}$ or λ_{2i}. As to the converse, the change of $q(x)$ into $\bar{q}(x) = q(-x) = q(1 - x)$ does not change either the periodic or tied spectrum; you are merely flipping over the interval $0 \le x \le 1$. The proof is finished by observing that the tied eigenfunctions $-y_2(-x, \mu_i)$ for $\bar{q}$ have the same norming constants as the eigenfunctions $y_2(x, \mu_i)$ for q if $\mu^o_i = \lambda^o_{2i}$ or λ^o_{2i-1}, $i = 1, \cdots, n$. The point is that $y_2^2(x, \mu_i)$ is of period 1 for $i = 1, 2, \cdots$.

[15] See [21], p. 147; [28], p.226.

The points of M at which $\mu_i^o = \lambda_{2i-1}^o$ or λ_{2i}^o for $m \le n$ values of $i = 1, \cdots, n$ are points of ramification of M over the cell

$$C = \bigtimes_{i=1}^{n} [\lambda_{2i-1}^o, \lambda_{2i}^o]$$

in which the nontrivial tied spectrum lies. M is a 2^n-sheeted covering of C. The covering group $G = \mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2$ (n-fold) is the group of changes of signature generated by

$$\begin{aligned}
\sigma_i \mathbf{p}_j &= \mathbf{p}_j, \qquad j \neq i, \\
\sigma_i \mathbf{p}_i &= (\mu_i^o, -\sqrt{\Delta^2(\mu_i^o - 4)} \qquad i = 1, \cdots, n, \\
\sigma_i q(\xi) &= q(\xi) + 2\frac{d^2}{d\xi^2} \log \left[1 + \frac{\sqrt{\Delta^2(\mu_i^o) - 4}}{y_2^\bullet(1, \mu_i^o)} \int_0^\xi y_2^2(\eta, \mu_i^o)\, d\eta \right];
\end{aligned}$$

see [28], p. 221. The map $\sigma = \sigma_1 \sigma_2 \cdots \sigma_n$ acts on q as $\sigma\, q(x) = q(-x)$. As noted above, $\bar{q}(x) = q(-x)$ and $q(x)$ have the same tied spectrum, the tied eigenfunctions of $\bar{q}$ being $-y_2(-x, \mu_i)$, $i = 1, \cdots, n$. To identify $\bar{q}$ with σq, it is enough to check that the norming constants

$$\int_0^1 y_2^2(-x, \mu_i^o)\, dx = y_2'(-1, \mu_i^o)[-y_2^\bullet(-1, \mu_i^o)]$$

are the same as those for q

$$\int_0^1 y_2^2(x, \mu_i^o)\, dx = \tfrac{1}{2}[\Delta(\mu_i^o) + \sqrt{\Delta^2(\mu_i^o) - 4}] y_2^\bullet(1, \mu_i^o)$$

with the signatures of the radicals flipped. The $y_2^\bullet(1, \mu_i^o)$, $i = 1, \cdots, n$, are the same by the coincidence of the tied spectra, while

$$y_2'(-1, \mu_i^o) = \tfrac{1}{2}[\Delta(\mu_i^o) - \sqrt{\Delta^2(\mu_i^o) - 4}] = [y_2(1, \mu_i^o)]^{-1}$$

because $y_2(-1, \mu_i^o) = 0$ and

$$\begin{aligned}
y_2')(-1, \mu_i^o) y_2'(1, \mu_i^o) &= y_2(-1, \mu_i^o) y_1'(\mu_i^o) + y_2'(-1, \mu_i^o) y_2'(1, \mu_i^o) \\
&= y_2'(0, \mu_i) \\
&= 1.
\end{aligned}$$

AMPLIFICATION 3. $\mu_i^o = \lambda_{2i-1}^o$, say, if and only if $f_{2i-1}^o(0) = 0$. Now, under translation of q, f_{2i-1} is also translated. Therefore, under translation thorough a full period $0 \le x \le 1$, μ_i^o hits λ_{2i-1}^o as many times as $f_{2i-1}^o(x) = 0$ has roots, namely m_i times; clearly, μ_i^o hits λ_{2i}^o the same number of times. This may be confirmed by Lemmas 5 and 7 which yield the formulas

$$\begin{aligned}
-\Delta^\bullet(\lambda_{2i-1}^o)[f_{2i-1}^o(x)]^2 &= (\mu_i^o - \lambda_{2i-1}^o) \times \text{ a nonvanishing factor}, \\
-\Delta^\bullet(\lambda_{2i}^o)[f_{2i}^o(x)]^2 &= (\lambda_{2i}^o - \mu_i^o) \times \text{ a nonvanishing factor}.
\end{aligned}$$

The picture is clarified by the computation of $V_1 \mu_i^o$, $i = 1, \cdots, n$, V_1 being the infinitessimal translation $q \to q'$. Let $f(x)$ be the normalized eigenfunction

$[y_2^\bullet(1,\mu_i^o)y_2'(1,\mu_i^o)]^{-1/2}y_2(x,\mu_i^o)$; then

$$\begin{aligned} V_1\mu_i^o &= \int_0^1 f^2q'\,dx \\ &= \int_0^1 f^2L\,1\,dx \\ &= 2qf^2\Big|_0^1 - \tfrac{1}{2}(f^2)''\Big|_0^1 - \int_0^1 Lf^2\,dx \\ &= -(f')^2\Big|_0^1 - 2\mu_i^o\int_0^1 (f^2)'\,dx \\ &= \frac{1-[y_2'(1,\mu_i^o)]^2}{y_2^\bullet(1,\mu_i^o)\,y_2'(1,\mu_i^o)} \\ &= \pm\frac{\sqrt{\Delta^2(\mu_i^o)-4}}{y_2^\bullet(1,\mu_i^o)}. \end{aligned}$$

Now $y_2'(1,\mu_i^o)$ is of one signature since it cannot vanish, and it is clear that when $\mathbf{p}_i = (\mu_i^o, \sqrt{\Delta^2(\mu_i^o)-4})$ hits λ_{2i-1}^o or λ_{2i}^o it passes thought with change of signature. The upshot is that $\mathbf{p}_i$ moves steadily about its little circle m_i times, returning to its starting place at $x = 1$.

7.5. Interpolation Theorems

The purpose of this section is to prepare some interpolations theorems for integral functions of order $\frac{1}{2}$ and type at most 1. They will be in continual use below. Let ν_i, $i = 0, 1, 2, \cdots$, be the reflecting spectrum of Q comprising the roots of $y_1'(1,\nu) = 0$, one from each of the intervals of instability $(-\infty, \lambda_0]$, $[\lambda_{2i-1}, \lambda_{2i}]$, $i = 1, 2, \cdots$, and let $\nu_0 = 0$ for ease of writing.[16] Then

$$e(\omega) = y_1(1,\omega^2) + \frac{\sqrt{-1}}{\omega}y_1'(1,\omega^2)$$

is an integral function of order 1 and type 1 with no roots on the line.

LEMMA 1. $|e(\omega)| > |e(\omega^*)|$ *in the open upper half-plane.*

Proof: Fix ω in the open upper half plane. Then

$$\begin{aligned} \tfrac{1}{4}|e(\omega)|^2 - \tfrac{1}{4}|e(\omega^*)|^2 &= \text{imag}\,\frac{1}{\omega^*}y_1(1,\omega^2)y_1'^*(1,\omega^2) \\ &= \text{imag}\,\frac{1}{\omega^*}\int_0^1 [y_1(x,\omega^2)y_1'^*(x,\omega^2)]'\,dx \\ &= \text{imag}\,\frac{1}{\omega^*}\int_0^1 [|y_1'(x,\omega^2)|^2 + q(x)|y_1(x,\omega^2)|^2]\,dx \\ &\qquad -\text{imag}\,\omega^*\int_0^1 |y_1(x,\omega^2)|^2\,dx \\ &> |\omega|^{-2}\text{imag}\,\omega\times\inf\int_0^1 f^*Qf\,dx. \end{aligned}$$

The infimum is computed for $f\in C^2[0,1]$, say, with $f'(0) = f'(1) = 0$ and

$$\int_0^1 |f|^2\,dx = \int_0^1 |y_1(x,\omega^2)|^2\,dx,$$

[16] If $\nu_0 \neq 0$, replace $y_1(1,\omega^2)$ by $y_1(1,\omega^2+\nu_0)$ and $y_1'(1,\omega^2)$ by $y_1'(1,\omega^2+\nu_0)$.

the value of the infimum being

$$0 = \int_0^1 |f|^2 dx \times \text{ the lowest reflecting eigenvalue } (\nu_0 = 0).$$

To spell it out, the final term

$$-\text{imag } \omega^* \int_0^1 |y_1|^2 \, dx$$

is dropped causing a strict inequality. The remaining integral I can be approximated as closely as you please by

$$\int_0^1 (|f'|^2 + q|f|^2) \, dx = \int_0^1 f^* Q f \, dx$$

with f as stated. This is done by choosing f' in $C^1[0,1]$ vanishing near $x = 0$ and $x = 1$ with

$$\int_0^1 |f' - y_1'|^2 \, dx$$

small. Take

$$f(x) = 1 + \int_0^x f'(y) \, dy.$$

Then I is well-approximated by

$$\int_0^1 f^* Q f \, dx$$

and is therefore larger than

$$\inf \int_0^1 f^* Q f \, dx.$$

The proof is finished.

The function e is now use to define a class B of integral functions f with

$$(a) \qquad B[f,f] = \frac{1}{\pi} \int_{-\infty}^{\infty} |f/e|^2 \, d\omega < \infty,$$

$$(b) \qquad |f(\omega)|^2 \leq \text{ a constant multiple of } |\text{ imag } \omega|^{-1} \times |e(\omega)|^2,$$

and you prove that

$$\mathbf{1}_\alpha(\beta) = \frac{e^*(\alpha)e(\beta) - e(\alpha^*)e^*(\beta^*)}{-2\sqrt{-1}(\beta - \alpha^*)}$$

belongs to B for each complex number α and that[17] and to verify that B is a Hilbert space; see De Branges [4] for such matters.

Now let A be the Hilbert space of functions f defined on $\cdots -\sqrt{\nu_1},\ 0 = \sqrt{\nu_0},\ \sqrt{\nu_1},\ \sqrt{\nu_2}, \cdots$ with

$$\begin{aligned} A[f,f] &= \tfrac{1}{2} \sum_{i=1}^{\infty} \ell_i^{-2} |f(-\sqrt{\nu_i})|^2 + \ell_0^{-2}|f(0)|^2 + \tfrac{1}{2} \sum_{i=1}^{\infty} \ell_i^{-2} |f(\sqrt{\nu_i})|^2 < \infty, \\ \ell_i &= \int_0^1 y_1^2(x,\nu_i) \, dx = -y_1(1,\nu_i) y_1'^{\bullet}(1,\nu_i),\ i = 0,\ 1,\ 2, \cdots. \end{aligned}$$

LEMMA 2. *The restriction map*

$$f \to \cdots, f(-\sqrt{\nu_1}),\ f(0),\ f(\sqrt{\nu_1}),\ f(\sqrt{\nu_2}), \cdots$$

is an isomorphism of B upon A, inner products and all, inverted by[18] $f(\omega) = A[f, \mathbf{1}_\omega]$.

---

[17] $B[f, \mathbf{1}_\omega]$ is the inner product $(1/\pi) \int f \, \mathbf{1}_\omega^* |e|^{-1}$.

[18] $A[f, \mathbf{1}_\omega]$ is the inner product of A.

Proof: The functions $\mathbf{1}_\omega$ span B because $f(\omega) = B[f, \mathbf{1}_\omega]$ cannot vanish identically if $f \neq 0$. Besides, the restrictions of these functions also span A: for $\alpha = \pm\sqrt{\nu_i}$ and $\beta = \pm\sqrt{\nu_i}$, $\mathbf{1}_\alpha(\beta) = -(\beta-\alpha)^{-1} y_1(1,\alpha^2) y_1'(1,\beta^2)$ vanishes for $\beta \neq \alpha$, while for $\beta = \alpha$, it assumes the value ℓ_0^2 if $i = 0$ and the value $2\ell_i^2$ if $i \neq 0$. It remains to prove that $A[\mathbf{1}_\alpha, \mathbf{1}_\beta] = B[\mathbf{1}_\alpha, \mathbf{1}_\beta]$. To begin with,

$$\begin{aligned} A[\mathbf{1}_\alpha, \mathbf{1}_\beta] &= \ell_0^{-2}\mathbf{1}_\alpha(0)\mathbf{1}_\beta^*(0) + \tfrac{1}{2}\sum_{i=1}^{\infty} \ell_i^{-2}[\mathbf{1}_\alpha(-\sqrt{\nu_i})\mathbf{1}_\beta^*(-\sqrt{\nu_i}) + \mathbf{1}_\alpha(\sqrt{\nu_i})\mathbf{1}_\beta^*(\sqrt{\nu_i})] \\ &= \frac{y_1'^*(1,\alpha^2)y_1'(1,\beta^2)}{\alpha^*\beta} \sum_{i=0}^{\infty} \ell_i^{-2} y_1^2(1,\nu_i) \frac{\alpha^*\beta + \nu_i}{(\alpha^{*2}-\nu_i)(\beta^2-\nu_i)} \end{aligned}$$

by substitution of

$$\mathbf{1}_\omega(\pm\sqrt{\nu_i}) = -\frac{y_1(1,\nu_i)y_1'^*(1,\omega^2)}{\omega^*(\omega^* \pm \sqrt{\nu_i})}.$$

The sum is now computed by means of the identity

$$\sum_{i=0}^{\infty} \ell_i^{-2}(\omega^2-\nu_i)^{-1} y_1(\xi,\nu_i) y_1(\eta,\nu_i) = y_1(\xi,\omega^2)\frac{y_2'(1,\omega^2)y_1(\eta,\omega^2) - y_1'(1,\omega^2)y_2(\eta,\omega^2)}{-y_1'(1,\omega^2)}$$

valid for $0 \leq \xi, \eta \leq 1$, which you recognize to be the expansion in terms of the unit perpendicular eigenfunctions $\ell_i^{-1} y_1(x,\nu_i)$, $i = 0, 1, 2, \cdots$, of the Green function of $\omega^2 - Q$ acting on the class of functions $f \in C^2[0,1]$ with $f'(0) = f'(1) = 0$. The identity is employed for $\xi = \eta = 1$, only: it reads

$$\sum_{i=0}^{\infty} \ell_i^{-2}(\omega^2-\nu_i) y_1^2(1,\nu_i) = -\frac{y_1(1,\omega^2)}{y_1'(1,\omega^2)},$$

the upshot being that

$$\begin{aligned} A[\mathbf{1}_\alpha, \mathbf{1}_\beta] &= \frac{y_1'^*(1,\alpha^2)y_1'(1,\beta^2)}{\alpha^*\beta}\left[\frac{-\beta}{\alpha^*-\beta}\frac{y_1^*(1,\alpha^2)}{y_1'(1,\alpha^2)} + \frac{\alpha^*}{\alpha^*-\beta}\frac{y_1(1,\beta^2)}{y_1'(1,\beta^2)}\right] \\ &= \frac{1}{\alpha^*-\beta}\left[y_1^*(1,\alpha^2)\frac{y_1'(1,\beta^2)}{\beta} - \frac{y_1'(1,\alpha^2)}{\alpha^*} y_1(1,\beta^2)\right] \\ &= \frac{e^*(\alpha)e(\beta) - e(\alpha^*)e^*(\beta^*)}{2\sqrt{-1}(\beta-\alpha^*)} \\ &= \mathbf{1}_\alpha(\beta) \\ &= B[\mathbf{1}_\alpha, \mathbf{1}_\beta], \end{aligned}$$

as required.

LEMMA 3. *A may be identified as the space of functions f defined on $\pm\sqrt{\nu_i}$, $i = 0, 1, 2, \cdots$, with $\sum |f(\pm\sqrt{\nu_i})|^2 < \infty$, the sum being comparable*[19] to $A[f,f]$. B may be identified as the space of integral functions f of order 1 and type at most 1 with

$$\int_{-\infty}^{\infty} |f|^2 \, d\omega < \infty,$$

the integral being comparable to $B[f,f]$.

Proof: The statement about A is self-evident from the estimate

$$\ell_i^2 = \int_0^1 y_1^2(x,\nu_i)\,dx = \int_0^1 \cos^2\sqrt{\nu_i}x\,dx + o(1) = \tfrac{1}{2} + o(1).$$

[19]The adjective *comparable* signifies that $\sum |f(\pm\sqrt{\nu_i})|^2$ is bounded above and below by positive multiples of $A[f,f]$, independent of $f \in A$. The meaning is similar for B.

As to B, the fact that $\int |f|^2$ is comparable to $B[f,f] = \int |f/e|^2$ is immediate from the estimate

$$|e(\omega)|^2 = y_1^2(1,\omega^2) + \omega^{-2}[y_1'(1,\omega^2)]^2 = \cos^2\omega + \sin^2\omega + o(1) = 1 + o(1)$$

for large real ω. Let I be the class of integral functions of order 1 and type at most 1 with $\int |f|^2 < \infty$. Then $\mathbf{1}_\omega \in I$ by inspection and $B \subset I$ follows by comparability of $\int |f|^2$ and $B[f,f]$ and by the fact that I is closed in $L^2(\mathbb{R}^1)$. The opposite inclusion is only a little less transparent. To begin with, $f \in I$ can be expressed as

$$f(\omega) = (2\pi)^{-1/2} \int_{-1}^{1} e^{-\sqrt{-1}\omega x} \hat{f}(x)\, dx$$

by the Paley-Wiener theorem; hence

$$|f(\omega)|^2 \le \frac{1}{2\pi} \int_{-1}^{1} |\hat{f}(x)|^2\, dx \int_{-1}^{1} e^{2bx}\, dx = O(b^{-1} \sinh 2b)$$

for $b = \operatorname{imag}\omega$. The proof that $f \in B$ is now achieved by confirming (a) that $\int |f/e|^2 < \infty$ and (b) that $|f(\omega)|^2 \le$ a constant multiple of $|b|^{-1}|e(\omega)|$ for $b > 0$, say. (a) is trivial, while (b) follows from the appraisal $|e(\omega)|^{-1} \le O(e^b)$ for $\omega \to \infty$ in the upper half-plane. To begin with, $|e(\omega)| > |e(\omega^*)|$ in the upper half-plane implies $|e(\omega)|$ is increasing on vertical lines $\omega = a + \sqrt{-1}b$, $0 \le b < \infty$. Therefore, $|e(\omega)|^{-1}$ is bounded in the closed upper half-plane. Besides,

$$|e(\sqrt{-1}b)| = y_1(1,-b^2) + b^{-1}y_1'(1,-b^2) \sim e^b \text{ as } b \uparrow \infty.$$

The proof is finished by a self-evident application of the Phragmén-Lindelöf principle to the function $\exp\{-\sqrt{-1}\omega\}[e(\omega)]^{-1}$ in the sector $[0, \frac{1}{2}\pi]$ and $[\frac{1}{2}\pi, \pi]$.

The results obtained to date are now applied to the class $I^{-1/2}$ of integral functions ϕ of order $\frac{1}{2}$ and type at most 1 with

$$\|\phi\|^2_{-1/2} = \int_0^\infty |\phi(\nu)|^2 \nu^{-1/2}\, d\nu < \infty.$$

Let $I^{-1/2*}$ be the space of functions ϕ defined for $\nu = \nu_i$, $i = 0, 1, 2, \cdots$, with

$$\|\phi\|^2_{-1/2*} = \sum_{i=0}^\infty |\phi(\nu_i)|^2 < \infty.$$

THEOREM 1. *The restriction $\phi \to \phi(\nu_i)$, $i = -0, 1, 2, \cdots$, is a 1:1 map of $I^{-1/2}$ onto $I^{-1/2*}$, the sum $\|\phi\|^2_{-1/2*}$ being comparable to the integral $\|\phi\|^2_{-1/2}$. Moreover, ϕ may be recovered from its restriction by means of the interpolation formula*[20]

$$\phi(\nu) = \sum_{i=0}^\infty \frac{\phi(\nu_i) y_1'(1,\nu)}{y_1'^{\bullet}(1,\nu_i)(\nu - \nu_i)} = \sum_{i=0}^\infty \phi(\nu_i) \prod_{j \ne i} \frac{1 - \nu/\nu_j}{1 - \nu_i/\nu_j}.$$

Proof: The functions $f(\omega) = \phi(\omega^2)$ with $\phi \in I^{-1/2}$ fill up the even part of B, by Lemma 3. The rest is immediate from Lemma 2; especially, the interpolation formula is just a spelling

---

[20] $\nu_i \ne 0$, $i = 0, 1, 2, \cdots$, is now assumed for ease of writing.

out of $f(\omega) = A[f, \mathbf{1}_\omega]$ for $\omega^2 = \nu$:

$$\begin{aligned} \phi(\nu) &= A[f, \mathbf{1}_\omega] \\ &= \sum_{i=0}^{\infty} \ell_i^{-2} \phi(\nu_i) \tfrac{1}{2} [\mathbf{1}_\omega^*(-\sqrt{\nu_i}) + \mathbf{1}_\omega^*(\sqrt{\nu_i})] \\ &= \sum_{i=0}^{\infty} \frac{\phi(\nu_i) y_1'(1,\nu)}{y_1'^{\bullet}(1,\nu_i)(\nu - \nu_i)} \\ &= \sum_{i=0}^{\infty} \phi(\nu_i) \prod_{j \neq i} \frac{1 - \nu/\nu_j}{1 - \nu_i/\nu_j}. \end{aligned}$$

The last line employs the canonical product for $y_1'(1,\nu)$ in terms of its roots ν_i, $i = 0, 1, 2, \cdots$.

A couple of similar results concerning interpolation off the tied spectrum μ_i, $i = 1, 2, \cdots$, are developed for future use. Let $I^{1/2}$ be the class of integral functions of order $\frac{1}{2}$ and type at most 1 with

$$\|\phi\|_{1/2}^2 = \int_0^\infty |\phi(\mu)|^2 \mu^{1/2} \, d\mu < \infty,$$

and let $I^{1/2*}$ be the class of functions ϕ defined for $\mu = \mu_i$, $i = 1, 2, \cdots$, with

$$\|\phi\|_{1/2*}^2 = \sum_{i=1}^{\infty} |\phi(\mu_i)|^2 i^2 < \infty.$$

THEOREM 2. *The restriction $\phi \to \phi(\mu_i)$, $i = 1, 2, \cdots$, is a 1:1 map of $I^{1/2}$ onto $I^{1/2*}$, the sum $\|\phi\|_{1/2*}^2$ being comparable to the integral $\|\phi\|_{1/2}^2$. Morever, ϕ may be recovered from the restriction by means of the interpolation formula*[21]

$$\phi(\mu) = \sum_{i=1}^{\infty} \frac{\phi(\mu_i) y_2(1,\mu)}{y_2^{\bullet}(1,\mu_i)(\mu - \mu_i)} = \sum_{i=1}^{\infty} \phi(\mu_i) \prod_{j \neq i} \frac{1 - \mu/\mu_j}{1 - \mu_i/\mu_j}.$$

Proof: The proof is similar to that of Theorem 1, employing

$$e(\omega) = y_2'(1,\omega^2) - \sqrt{-1}\omega \, y_2(1,\omega^2)$$

in place of the previous function of that name. The details may be left to the reader.

THEOREM 3. *The same interpolation formula applies to the class $I^{3/2}$ of integral functions ϕ of order $\frac{1}{2}$ and type at most 1 with*

$$\|\phi\|_{3/2}^2 = \int_0^\infty |\phi(\mu)|^2 \mu^{3/2} \, d\mu < \infty.$$

The map $\phi \to \phi(\mu_i)$, $i = 1, 2, \cdots$, is now a 1:1 application of $I^{3/2}$ onto the class $I^{3/2*}$ of restricted functions with

$$\|\phi\|_{3/2*}^2 = \sum_{i=1}^{\infty} \phi(\mu_i)| i^4 < \infty,$$

the latter being comparable to the integral $\|\phi\|_{3/2}^2$.

Proof: The functions $f(\omega) = \omega \, \phi(\omega^2)$ with $\phi \in I^{3/2}$ fill up the odd part of the space B associated with the current function e.

[21] $\mu_i \neq 0$, $i = 1, 2, \cdots$, for ease of writing.

7.6. A Remarkable Identity

A number of consequences of the application of Theorem 5.1 to the gradient of Δ are derived below, both for their own interest and for future use. The chief result is the remarkable identity embodied in Proposition 2; it will have numerous consequences below.

USAGE. $\nu_i = \lambda_{2i}$, $i = 0, 1, 2, \cdots$, at the origin of M; in fact, $q(x) = q(-x)$ at the origin (Amplification 4.2), so $y_1(x, \nu_i)$ is even, $y_1(-1, \nu_i) = y_1(1, \nu_i)$, $y_1'(-1, \nu_i) = 0$, and consequently

$$\begin{aligned} y_1(2, \nu_i) &= y_1(-1, \nu_i) y_1(1, \nu_i) + y_1'(-1, \nu_i) y_2(1, \nu_i) = y_1(0, \nu_i) = 1, \\ y_1'(2, \nu_i) &= y_1(1, \nu_i) y_1'(1, \nu_i) + y_1'(1, \nu_i) y_2'(1, \nu_i) = 0. \end{aligned}$$

Therefore, $y_1(x, \nu_i)$ is of period 2, which is to say $\nu_i = \lambda_0$ if $i = 0$ and $\nu_i = \lambda_{2i-1}$ or λ_{2i} of $i \geq 1$. The final step is to note that μ_i^o and ν_i^o cannot sit at the same end of their interval of instability, λ_{2i-1}^o and λ_{2i}^o being *simple* periodic eigenvalues. The moral is that Theorem 4.1 may be used with $\nu_i = \lambda_{2i}$, $i = 0, 1, 2, \cdots$. The function $y_1'(1, \lambda)$ employed below is computed for that circumstance, regardless of the actual position of $q \in M$; such abuses of notation will be noted specifically every time they occur.

PROPOSITION 1.

$$\frac{\partial \Delta(\lambda)}{\partial q(x)} = \sum_{i=0}^{n} \frac{y_1'(1, \lambda)}{y_1'^{\bullet}(1, \lambda_{2i}^o)\,(\lambda - \lambda_{2i}^o)} \times -\Delta^{\bullet}(\lambda_{2i}^o)[f_{2i}^o(x)]^2.$$

Proof: Fix $0 \leq x < 1$. Then $\partial \Delta(\lambda)/\partial q(x)$ is of class $I^{-1/2}$ by Lemma 3.5 and the estimate $y_2(1, \lambda) = O(\lambda^{-1/2})$ for $\lambda \uparrow \infty$. The formula follows by interpolation off λ_{2i}, $i = 0, 1, 2, \cdots$, and by the evaluation $\partial \Delta(\lambda_{2i})/\partial q = -\Delta^{\bullet}(\lambda_{2i}) f_{2i}^2$ of Lemma 3.7.

LEMMA 1.

$$\epsilon_i \equiv \frac{\Delta^{\bullet}(\lambda_{2i}^o)}{y_1'^{\bullet}(1, \lambda_{2i}^o)} = o(\lambda_{2i}^o - \lambda_{2i-1}^o) \quad as \quad i \uparrow \infty.$$

AMPLIFICATION 1. Hochstadt [12] proved that $\lambda_{2i} - \lambda_{2i-1}$ is of rapid decrease[22] as $i \uparrow \infty$ if $q \in C_1^{\infty}$, and thus the same is true of ϵ_i; see Amplification 3 below for the converse of Hochstadt's therem.

Proof: Let $y_2(1, \lambda)$ be computed for the point of M with the non-trivial tied spectrum $\mu_i^o = \lambda_i^{\bullet}$, $i = 1, \cdots, n$, so that $\Delta^{\bullet}(\lambda)$ is a constant multiple of $y_2(1, \lambda)$. Then ϵ_i is a constant multiple of

$$\frac{y_2(1, \lambda_{2i}^o)}{-y_1'^{\bullet}(1, \lambda_{2i}^o)} = \frac{(\lambda_{2i}^o - \lambda_1^{\bullet}) y_2^{\bullet}(1, \lambda) y_1(1, \lambda_{2i}^o)}{-y_1(1, \lambda_{2i}^o) y_1'^{\bullet}(1, \lambda_{2i}^o)}$$

for some intermediate $\lambda \in [\lambda_i^{\bullet}, \lambda_{2i}^o]$. Now $-y_1(1, \lambda_{2i}^o) y_1'^{\bullet}(1, \lambda_{2i}^o)$ is a norming constant of the form

$$\int_0^1 y_1^2(x, \lambda_{2i}^o)\, dx = \int_0^1 \cos^2 \sqrt{\lambda_{2i}^o} x\, dx + o(1) = \tfrac{1}{2} + o(1),$$

while $y_1(1, \lambda_{2i}^o) = \pm 1$ according to whether $\Delta(\lambda_{2i}^o) = -2$ or $+2$ and $y_2^{\bullet}(1, \lambda) = O(1/\lambda)$. The whole quantity is therefore $o(\lambda_{2i}^o - \lambda_i^{\bullet})$. The proof is finished.

PROPOSITION 2. $1 = \sum_{i=0}^{\infty} \epsilon_i [f_{2i}^o(x)]^2$*; in particular,* $1 = \sum_{i=0}^{n} \epsilon_i$.

AMPLIFICATION 2. The statement is that for any $0 \leq x < 1$, the point $[f_0^o(x), f_2^o(x), \cdots, f_{2n}^o(x)] \in \mathbb{R}^{n+1}$ lies on the ellipsoid with semi-axes $\epsilon_i^{-1/2}$, $i = 1, \cdots, n$. The geometrical meaning of this fact is unclear to us. The derived identity $0 = \sum_{i=1}^{n} \epsilon_i D(f_{2i}^o)^2$ may be viewed as a dependency upon M between the vector fields $X_i : q \to D\partial\Delta(\lambda_{2i}^o)/\partial q$,

[22]Rapid decrease means that $\lambda_{2i} - \lambda_{2i-1} = O(i^{-p})$ as $i \uparrow \infty$ for every $p = 1, 2, \cdots$.

$i = 0, \cdots, n$, and it is the only such dependency, as will appear in Section 8; for $n < \infty$, it reduces to a differential equation for q, as in [11, 22], or [27].

Proof: The first identity follows from Proposition 1 upon letting $\lambda \downarrow -\infty$, indeed, $\partial\Delta(\lambda)/\partial q \sim \lambda^{-1/2} \sin\sqrt{\lambda}$ by Lemma 3.5 and $y_1'(1,\lambda) \sim -\sqrt{\lambda}\sin\sqrt{\lambda}$; hence

$$1 \sim -\lambda\frac{\partial\Delta/\partial q}{y_1'(1,\lambda)} = \sum_{i=0}^{n} \frac{\lambda}{\lambda - \lambda_{2i}^o}\epsilon_i(f_{2i}^o)^2 \sim \sum_{i=0}^{n} \epsilon_i(f_{2i})^2,$$

making use of Lemma 1 and the elementary estimates

$$\begin{aligned} f_{2i}^o(x) &= a_i y_1(x, \lambda_{2i}^o) + b_i\sqrt{\lambda_{2i}^o}\, y_2(x, \lambda_{2i}^o) \\ &= a_i \cos\sqrt{\lambda_{2i}^o}x + b_i \sin\sqrt{\lambda_{2i}^o} + o(1), \\ 1 &= \int_0^1 |f_{2i}^o|^2\, dx = \tfrac{1}{2}(a_i^2 + b_i^2)[1 + o(1)] + o(\sqrt{a_i^2 + b_i^2}) + o(1), \\ |f_{2i}^o(x)| &\le \sqrt{a_i^2 + b_i^2} + o(1) = O(1) \end{aligned}$$

in order to deal with the sum.

PROPOSITION 3. *Fix* $m = 0, 1, 2, \cdots$. *Then*

$$\frac{\partial H_m}{\partial q} = \sum_{i=0}^{n} \epsilon_i p_n(\lambda_{2i}^o)[f_{2i}^o(x)]^2$$

with a polynomial p_n *of precise degree* n *depending upon* M *but not upon* q.

Proof: The case $n = 0$ is covered by Proposition 2. For $\lambda = \lambda_{2i}^o$ and $f = f_{2i}^o$,

$$f(\xi) = f(0)\cos\sqrt{\lambda}\xi + f'(0)\frac{\sin\sqrt{\lambda}\xi}{\sqrt{\lambda}} + \int_0^\xi \frac{\sin\sqrt{\lambda}(\xi - \eta)}{\sqrt{\lambda}} q(\eta) f(\eta)\, d\eta,$$

from which you may derive $|D^m f(x)| = O(\lambda^{m/2})$, uniformly in $0 \le x < 1$, for $m = 1, 2, \cdots$, starting from the case $m = 0$ established in the proof of Proposition 2. This permits you to apply $L = qD + Dq - \frac{1}{2}D^3$ to the identity $1 = \sum \epsilon_i(f_{2i}^o)^2$ of proposition 2:

$$Dq = L1 = \sum_{i=0}^{n} \epsilon_i L(f_{2i}^o)^2 = \sum_{i=0}^{n} \epsilon_i 2\lambda_{2i}^o D(f_{2i}^o)^2,$$

and this may be integrated with respect to x to obtain

$$q = \sum_{i=0}^{n} \epsilon_i 2\lambda_{2i}^o (f_{2i}^o)^2 + \text{ a constant of integration.}$$

The stated identity for $m = 1$ is now before you: the left-hand side is just $\partial H_1/\partial q$ and the constant of integration can be incorporated into the sum by means of $1 = \sum \epsilon_i(f_{2i}^o)^2$. The procedure can be repeated. L may be applied to the identity for $m = 1$:

$$D\frac{\partial H_2}{\partial q} = L\frac{\partial H_1}{\partial q} = \sum_{i=0}^{n} \epsilon_i p_1(\lambda_{2i}^o) 2\lambda_{2i}^o D(f_{2i}^o)^2,$$

and you can integrate back and incorporate the constant of integration to get the identity for $m = 2$. The proof is finished by induction, using the recursion $L\partial H_{m-1}/\partial q = D\partial H_m/\partial q$ at the m-th stage.

AMPLIFICATION 3. Hochstadt [12] proved the rapid decrease of $\lambda_{2i} - \lambda_{2i-1}$ for $q \in C_1^\infty$ as noted in Amplification 1. Now the formula of Proposition 2 and its derivation are valid if $q \in L_1^2$ and $\lambda_{2i} - \lambda_{2i-1}$ is rapidly decreasing. This remark may be used to prove the converse of Hochstadt's theorem. *If* $q \in L_1^2$ *and if* $\lambda_{2i} - \lambda_{2i-1}$ *is rapidly decreasing, then* $q \in C_1^\infty$. The first step is to deduce from $1 = \sum \epsilon_i (f^o_{2i})^2$ that q can be expressed as $q = \sum \epsilon_i 2\lambda^o_{2i}(f^o_{2i})^2 + c$, as in Proposition 3. Fix $\psi \in C^\infty_1$. Then

$$\int_0^1 2\lambda^o_{2i}(f^o_{2i})^2 D\psi\, dx = \int_0^1 (f^o_{2i})^2(qD - \tfrac{1}{2}D^3)\psi\, dx - \int_0^1 q\psi D(f^o_{2i})^2\, dx$$

by approximation of $q \in L^2_1$ by functions from C^∞_1. Now multiply by ϵ_i and sum to obtain

$$\begin{aligned}\int_0^1 \sum \epsilon_i 2\lambda^o_{2i}(f^o_{2i})^2 D\psi\, dx &= \int_0^1 \sum \epsilon_i (f^o_{2i})^2(qD - \tfrac{1}{2}D^3)\psi\, dx \\ &\quad - \int_0^1 q\psi \sum \epsilon_i D(f^o_{2i})^2\, dx \\ &= \int_0^1 (qD - \tfrac{1}{2}D^3)\psi\, dx \\ &= \int_0^1 qD\psi\, dx.\end{aligned}$$

The interchange of sums and integrals is justified by the elementary estimates $(f^o_{2i})^2 = O(1)$, $D(f^o_{2i})^2 = O(i)$ and by the rapid decrease of ϵ_i. The disappearance of the sum $\sum \epsilon_i D(f^o_{2i})^2$ is a consequence of $1 = \sum \epsilon_i (f^o_{2i})^2$. Therefore $q = \sum \epsilon_i 2\lambda^o_{2i}(f_{2i})^2 + c$, and now you may proceed inductively. The eigenfunctions f^o_{2i} are of class C^1_2 or better with $D(f^o_{2i})^2 = O(i)$ uniformly in $0 \le x < 1$, so $q \in C^1_2$, and if $q \in C^{m-2}_1$ for $m \ge 3$, then $f^o_{2i} \in C^m_2$ and $D^m(f^o_{2i})^2 = O(i^m)$ uniformly in $0 \le x < 1$ as in the proof of Proposition 2. This permits the sum for q to be differentiated twice more, so $q \in C^m_2$, and the result follows.

7.7. Reflecting Spectrum

The interpolation theorems of Section 5 permits a full discussion of the reflecting spectrum ν_i, $i = 0, 1, 2, \cdots$, comprising the roots of $y'_1(1, \nu) = 0$, parallel to the discussion of the tied spectrum in Section 4. Borg [2] proved that ν_i, $i = 0, 1, 2, \cdots$, together with the norming constants

$$\int_0^1 y^2_1(x, \nu_i)\, dx = -y_1(1, \nu_i) y'^{\bullet}_1(1, \nu_i)$$

determining q uniquely. Let the non-trivial eigenvalues ν^o_i, $i = 0, 1, \cdots, n$, be fixed. The rest of the reflecting spectrum is already known from the double periodic spectrum; hence $y'^{\bullet}_1(1, \nu_i)$ is known, and the remaining part of the i-th norming constant can be determined by solving $\Delta(\nu_i) = y_1(1, \nu_i) + [y_1(1, \nu_i)]^{-1}$ for

$$y_1(1, \nu_i) = \tfrac{1}{2}[\Delta(\nu_i) \pm \sqrt{\Delta^2(\nu_i) - 4}],$$

up to signature of the radical. The latter represents a bonafide ambiguity only if $\lambda_{2i-1} < \nu_i < \lambda_{2i}$ if $i \ge 1$ or $\nu_0 < \lambda_0$ if $i = 0$, suggesting a picture of M as a product of a line and a circle, as in Fig. 7.6 drawn for $n < \infty$. Figure 7.6 appears to contradict Figs. 7.2 and 7.4; in fact, the line arising from $(-\infty, \lambda_0]$ is superfluous. The suspicion that ν_0 is anomalous introduces itself upon noting that $\nu_0 \ge \min_{0 \le x < 1} q(x)$.

LEMMA 1. *ν_0 and the signature of $\sqrt{\Delta^2(\nu_0) - 4}$ are already determined by ν_i and the signature of $\sqrt{\Delta^2(\nu_i) - 4}$, $i = 1, 2, \cdots$.*

Proof: The method of Propositions 6.1 and 6.2 can be used to prove that

$$1 = \sum_{i=0}^{\infty} \Delta^{\bullet}(\nu_i)[y'^{\bullet}_1(1, \nu_i)]^{-1},$$

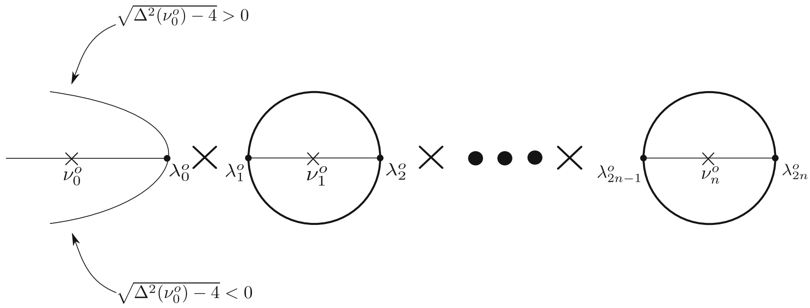

FIGURE 7.6.

in which ν_i, $i = 0, 1, 2, \cdots$, and $y_1'(1,\nu)$ are computed for the actual point $q \in M$ in hand; you have only to perform the interpolation of $\partial\Delta/\partial q$ off ν_i, $i = 0, 1, 2, \cdots$, instead of λ_{2i}, $i = 0, 1, 2, \cdots$. This is a relation between ν_0 and ν_i, $i = 1, 2, \cdots$. To spell it out,

$$1 = \frac{\Delta^\bullet(\nu_0)}{\prod_{j\neq 0}(1-\nu_0/\nu_j)(-1/\nu_0)} + \sum_{i=1}^{\infty} \frac{\Delta^\bullet(\nu_i)}{\prod_{j\neq i,0}(1-\nu_i/\nu_j)(-1/\nu_i)(1-\nu_i/\nu_0)},$$

the right-hand side being a known function $\phi(\nu_0)$ of ν_0 if the ν_i, $i = 1, 2, \cdots$, are known. The function $\phi(\lambda)$ may be computed by means of Theorem 7.5.1: $\Delta^\bullet(\lambda)$ is of the form $y_2(1,\lambda) \in I^{-1/2}$, the roots of $\Delta^\bullet(\lambda) = 0$ being the tied spectrum of some point of M, so $\Delta^\bullet(\lambda)$ may be interpolated off ν_i, $i = 0, 1, 2, \cdots$, and you have

$$\begin{aligned}
\sum_{i=1}^{\infty} \Delta^\bullet(\nu_i)\left[\prod_{j\neq i,0}\left(1-\frac{\nu_i}{\nu_j}\right)\left(-\frac{1}{\nu_i}\right)\left(1-\frac{\nu_i}{\lambda}\right)\right]^{-1} &= \sum_{n=1}^{\infty} \Delta^\bullet(\nu_i)\frac{1-(\nu_i/\nu_0)}{y_1'^\bullet(1,\nu_i)}\frac{\lambda}{\lambda-\nu_i} \\
&= \frac{\lambda}{\nu_0}\sum_{i=0}^{\infty}\frac{\Delta^\bullet(\nu_i)}{y_1'^\bullet(1,\nu_i)}\left[1-\frac{\lambda-\nu_0}{\lambda-\nu_i}\right] \\
&= \frac{\lambda}{\nu_0}\left[1-(\lambda-\nu_0)\frac{\Delta^\bullet(\lambda)}{y_1'(1,\lambda)}\right]
\end{aligned}$$

from which you deduce

$$\phi(\lambda) = \frac{\lambda}{\nu_0} + \frac{1-(\lambda/\nu_0)}{1/\lambda}\frac{\Delta^\bullet(\lambda)}{y_1'(1,\lambda)} + \frac{1-(\lambda/\nu_0)}{-1/\lambda}\frac{\Delta^\bullet(\lambda)}{y_1'(1,\lambda)} = \frac{\lambda}{\nu_0}.$$

Thus, $\phi(\lambda) = 1$ only if $\lambda = \nu_0$, which is thereby determined from ν_i, $i = 1, 2, \cdots$. It remains to prove that the signature of $\sqrt{\Delta^2(\nu_0)-4}$ is also determined by ν_i, $\sqrt{\Delta^2(\nu_i)-4}$, $i = 1, 2, \cdots$. Let the latter be fixed and let $y_1^-(1,\lambda)$ and $y_1^+(1,\lambda)$ be the two determinations of $y_1(1,\lambda)$ corresponding to the two possible signatures of $\sqrt{\Delta^2(\nu_0)-4}$. Then $\phi(\lambda) = y_1^+(1,\lambda) - y_1^-(1,\lambda)$ is of class $I^{1/2}$ being $O(1/\lambda)$ for $\lambda \uparrow \infty$, it vanishes at ν_i, $i = 1, 2, \cdots$, because $y_1^+(1,\nu_i) = y_1^-(1,\nu_i) = \frac{1}{2}[\Delta(\nu_i) \pm \sqrt{\Delta^2(\nu_i)-4}]$ is determined, and the conclusion is that $\phi(\lambda) = 0$ by Theorem 5.2, ν_i, $i = 1, 2, \cdots$, being the tied spectrum of a point of M. In particular, $y_1^-(1,\nu_0) = y_1^+(1,\nu_0)$ which is to say that the change of signature in $\sqrt{\Delta^2(\nu_0)-4}$ was illusory. The proof is finished.

THEOREM 1. *M is topologically equivalent to the product of the $n \leq \infty$ circles depicted in Fig. 7.6.*

Proof: The present state of affairs is that you have a 1:1 map of M *into* the torus formed by the n circles of Fig. 7.6; it is required ti prove that the map is *onto*. Pick a point

$\mathbf{p}_i = (\nu_i^o, \sqrt{\Delta^2(\nu_i^o - 4)})$, $i = 1, 2, \cdots, n$, of the image. Then if $\nu_i^o = \lambda_{2i-1}^o$, say, you will have $(f_{2i-1}^o)'(0) = 0$, and a small translation of q will break all such equalities, keeping q on M. The moral is that every point of the image is close to image points in general position $[\lambda_{2i-1}^o < \nu_i^o < \lambda_{2i}^o, i = 1, 2, \cdots, n]$. Now let $X_i : q \to D\partial\Delta(\lambda)/\partial q$ evaluated at $\lambda = \mu_i^o$, $i = 1, 2, \cdots, n$, as in Section 4. Then

$$X_i \nu_j^o = \frac{1}{2} \times \frac{\sqrt{\Delta^2(\nu_i^o) - 4}}{-y_1'^{\bullet}(1, \nu_j^o)} \frac{y_1'(1, \mu_i^o)}{\mu_i - \nu_j^o}, \qquad i, j = 1, 2, \cdots,$$

much as in Lemma 4.1, so $\det X_i \nu_j^o$, $1 \le i, j \le m$, cannot vanish while ν_i^o, $i = 1, \cdots, m$, is in general position for any $m \le n$ if $n < \infty$, respectively $m < \infty$ if $n = \infty$, and it follows from the compactness of M that its image is a sum of closed sheets of the torus, picturing the latter as a 2^n-sheeted covering in the cell $C = \bigtimes_{i=1}^n [\lambda_{2i-1}^o, \lambda_{2i}^o]$ with sheets distinguished by the signatures of the radicals $\sqrt{\Delta^2(\nu_i^o) - 4}$, $i = 1, 2, \cdots, n$; compare Amplification 4.2. The final step is to prove that every sheet appears in the image, i.e., that the covering map σ_i of M over C which fixes the signature of $\sqrt{\Delta^2(\nu_j^o) - 4}$, $j \ne i$, and reverses the signature of $\sqrt{\Delta^2(\nu_i^o) - 4}$ is a bonafide automorphism of M. This may be done in the manner of Faddeev [8]. The procedure will be outlined; compare Amplification 4.2. To begin with, σ_i would have to flip the signature of $\sqrt{\Delta^2(\nu_0) - 4}$; in fact, the difference $\phi(\lambda) \in I^{1/2}$ of the functions $y_1(1, \lambda)$ for q and for $\sigma_i q$ could be interpolated off ν_j, $j = 1, 2, \cdots$, producing a relation between $\phi(\nu_i^o)$ and $\phi(\nu_0)$ which forces the signature of $\sqrt{\Delta^2(\nu_0) - 4}$ to flip with the signature of $\sqrt{\Delta^2(\nu_i^o) - 4}$. Let A_j, respectively B_j, stand for the reciprocal of $\int_0^1 y_1^2(x, \nu_j)\, dx$ evaluated at $\sigma_i q$, respectively q, Introduce the kernel

$$\begin{aligned} L(\xi, \eta) &= \sum_{j=0}^{\infty} y_1(\xi, \nu_j) y_1(\eta, \nu_j) \times (A_j - B_j) \\ &= y_1(\xi, \nu_0) y_1(\eta, \nu_0) \times \frac{\sqrt{\Delta^2(\nu_0) - 4}}{-y_1'^{\bullet}(1, \nu_0)} \\ &\quad + y_1(\xi, \nu_i) y_1(\eta, \nu_i^o) \times \frac{\sqrt{\Delta^2(\nu_i^o) - 4}}{-y_1'^{\bullet}(1, \nu_i^o)}. \end{aligned}$$

Now solve $K + KL + L\,(\eta \le \xi)$ for the triangular kernel $K(\xi, \eta)$ vanishing for $\xi < \eta$ and expressed for $\xi \ge \eta$ by

$$K(\xi, \eta) = \begin{bmatrix} \frac{\sqrt{\Delta^2(\nu_i^o) - 4}}{y_1'^{\bullet}(1, \nu_i^o)} y_1(\xi, \nu_i^o) \\ \frac{\sqrt{\Delta^2(\nu_0) - 4}}{y_1'^{\bullet}(1, \nu_0)} y_1(\xi, \nu_0) \end{bmatrix} \frac{\begin{bmatrix} k_{11} & k_{21} \\ k_{12} & k_{22} \end{bmatrix}}{(k_{11}k_{22} - k_{12}k_{21})} \begin{bmatrix} y_1(\eta, \nu_i^o) \\ y_1(\eta, \nu_0) \end{bmatrix}$$

with

$$\begin{aligned} k_{11} &= 1 - \frac{\sqrt{\Delta^2(\nu_0) - 4}}{y_1'^{\bullet}(1, \nu_0)} \int_0^{\xi} y_1^2(x, \nu_0)\, dx, \\ k_{12} &= \frac{\sqrt{\Delta^2(\nu_i^o) - 4}}{y_1'^{\bullet}(1, \nu_i^o)} \int_0^{\xi} y_1(x, \nu_0) y_1(x, \nu_i^o)\, dx, \\ k_{21} &= \frac{\sqrt{\Delta^2(\nu_0) - 4}}{y_1'^{\bullet}(1, \nu_0)} \int_0^{\xi} y_1(x, \nu_0) y_1(x, \nu_i^o)\, dx, \\ k_{22} &= 1 - \frac{\sqrt{\Delta^2(\nu_i^o) - 4}}{y_1'^{\bullet}(1, \nu_i^o)} \int_0^{\xi} y_1^2(x, \nu_i^o)\, dx. \end{aligned}$$

Then you can check (with tears) that the function

$$\sigma_i q(x) = q(x) - 2\frac{d}{dx}K(x,x) = q(x) + \frac{d^2}{dx^2}\log(k_{11}k_{22} - k_{12}k_{21})$$

belongs to M and has reflecting spectrum ν_i, $i = 0, 1, 2, \cdots$, with the same signatures for the radicals $\sqrt{\Delta^2(\nu_j^o) - 4}$, $j = 1, \cdots, n$, as for q except that the signature of $\sqrt{\Delta^2(\nu_i^o) - 4}$ is flipped. The computations are too lengthy to reproduce here, but do not offer any difficulty. M is now seen to map *onto* the torus. The fact that the map is topological is self-evident.

AMPLIFICATION 1. The reflecting spectrum is in special position $\nu_0 = \lambda_0$, $\nu_i = \lambda_{2i-1}$ or λ_{2i}, $i = 1, 2, \cdots,$, if and only if $q(x) = q(-x)$, just as for the tied spectrum; see Amplification 4.2. The proof is similar. Notice that if ν_i^o sits at one end of $[\lambda_{2i-1}^o, \lambda_{2i}^o]$, $i = 1, \cdots, n$, then μ_i^o sits at the other end, λ_{2i-1}^o and λ_{2i}^o being *simple* eigenvalues. The map $\sigma : q(x) \to q(-x)$ is the product of the sheet maps σ_i, $i = 1, 2, \cdots, n$, as before.

7.8. Normal and Tangent Spaces: Formal Discussion

So far nothing has been made of the fact that M sits inside the Euclidean space L_1^2. It turns out that the geometry of M is clarified by this point of view. For example, the normal and tangent spaces can be calculated. A formal account is given here, the proofs being postponed until Sections 9 and 13.

As a start, it is not hard to see that M is an $(n \le \infty)$-dimensional analytic variety in L_1^2. For if $q_0 \in C_1^\infty$ is fixed and $\lambda_i(q_0)$, $i = 0, 1, 2, \cdots$, is its periodic spectrum, then *the space M in which q_0 lies is the variety comprising the common roots q in L_1^2 of the analytic relations*

$$2 \times (-1)^i = \Delta(\lambda_{2i}(q_0)), \quad i = 0, 1, 2, \cdots.$$

Proof: The analyticity of $\Delta(\lambda_{2i}(q_0))$ *qua* function of q is clear, since the numbers $\lambda_{2i}(q_0)$, $i = 0, 1, 2, \cdots$, are fixed; see Amplification 3.4. Now let $\Delta(\lambda)$ be computed for $q \in L_1^2$ and let it satisfy the stated relations. Then $[\lambda - \lambda_0(q)]^{-1}[\Delta(\lambda) - 2]$ is in $I^{1/2}$ and so can be interpolated off its known values $[\lambda_{2i}(q_0) - \lambda_0(q_0)]^{-1} \times (-4 \text{ or } 0)$ at $\lambda = \lambda_{2i}(q_0)$, $i = 1, 2, \cdots$. The same is true of the discriminant of q_0, so the two coincide. Therefore q and q_0 have the same periodic spectrum. Finally, q is in C_1^∞ by the result of Amplification 6.3.

Now a finite-dimensional variety defined by smooth relations $r_i(x) = 0$, $i = 1, \cdots, c$, the gradients $\partial r_i/\partial x$ are normal vectors, and provided they are everywhere independent, the variety is smooth. By analogy, the gradients $\partial\Delta(\lambda_{2i})/\partial q = -\Delta^\bullet(\lambda_{2i})f_{2i}^2$, $i = 0, 1, \cdots$, ought to be normal vector fields on M. So far so good. But the gradient $\partial\Delta(\lambda_{2i})/\partial q$ vanishes identically when λ_{2i} is double, so the relations may not determine the normal space even though they do determine M. Therefore we seek a new set of relations with nowhere vanishing gradients. These will account for all normal directions since now there cannot be more normals than defining relations. Naturally, we would like the new relations and their gradients to be independent.

Now the eigenvalue λ_{2i} is double only if

$$(a) \qquad (-1)^i = y_1(1, \lambda_{2i}), \qquad 0 = y_1'(1, \lambda_{2i}),$$

$$(b) \qquad 0 = y_2(1, \lambda_{2i}), \qquad (-1)^i = y_2'(1, \lambda_{2i}).$$

In fact, $y_1 y_2' = 1$, and $y_1 = (-1)^i = y_2'$ follows from $y_1 + y_2' = 2 \times (-1)^i$. On the other hand, either (a) or (b) implies $\Delta(\lambda_{2i}) = 2 \times (-1)^i$. *Therefore, M can be presented as the variety of common roots of the analytic relations*

$$2 \times (-1)^i = \Delta(\lambda_{2i}(q_0)) \textit{ when } \lambda_{2i}(q_0) \textit{ is simple},$$

supplemented by

$$\left.\begin{array}{ll}(-1)^i & = y_1(1,\lambda_{2i}(q_0)) \\ 0 & = y_1'(1,\lambda_{2i}(q_0))\end{array}\right\} \text{ when } \lambda_{2i}(q_0) \text{ is double}$$

The choice of the supplementary relations (a) is a matter of taste. The relations (b) would do just as well. Clearly, none of the relations may be omitted since the interpolation of $[\lambda - \lambda_0(q)]^{-1}[\Delta(\lambda) - 2]$ requires the use of every λ_{2i}, $i = 1, 2, \cdots$, and both $y_1(1, \lambda_{2i}) = (-1)^i$, $y_1'(1, \lambda_{2i}) = 0$ are necessary to recover $\Delta(\lambda_{2i}) = 2 \cdot (-1)^i$. Therefore, the relations ought to be independent and their non-vanishing gradients:

$$\frac{\partial \Delta}{\partial q}(\lambda_{2i}) = -\Delta^\bullet(\lambda_{2i}) f_{2i}^2 \text{ when } \lambda_{2i} \text{ is simple,}$$

$$\left.\begin{array}{ll}\frac{\partial y_1}{\partial q}(1,\lambda_{2i}) & = -y_1(1,\lambda_{2i}) y_1(x,\lambda_{2i}) y_2(x,\lambda_{2i}) \\ \frac{\partial y_1'}{\partial q}(1,\lambda_{2i}) & = y_2'(1,\lambda_{2i}) y_1^2(x,\lambda_{2i})\end{array}\right\} \text{ when } \lambda_{2i} \text{ is double.}$$

ought to be independent normal vectors. Now for any double eigenvalue $\lambda_{2i} = \lambda_{2i-1}$, the gradients of the relations (b):

$$\begin{aligned}\frac{\partial y_2}{\partial q}(1,\lambda_{2i}) &= -y_1(1,\lambda_{2i}) y_2^2(x,\lambda_{2i}), \\ \frac{\partial y_2'}{\partial q}(1,\lambda_{2i}) &= y_2'(1,\lambda_{2i}) y_1(x,\lambda_{2i}) y_2(x,\lambda_{2i})\end{aligned}$$

should also be normal, so that the 3-dimensional null-space of $L - 2\lambda_{2i}D$, spanned by f_{2i-1}^2, $f_{2i-1}f_{2i}$ and f_{2i}^2, is normal to M. But three normal vectors is one too many; indeed, $f_{2i-1}^2 + f_{2i}^2$ already lies in the span of $(f_{2i}^o)^2$, $i = 0, 1, \cdots, n$, in view of Proposition 6.1 and the evaluation

$$f_{2i}^2 + f_{2i-1}^2 = \lim_{\lambda \to \lambda_{2i}} (\lambda - \lambda_{2i}) 2\Delta(\lambda)[4 - \Delta^2(\lambda)]^{-1} \frac{\partial \Delta(q)}{\partial q},$$

which follows from Lemma 3.8. Thus, we ought to keep the product $f_{2i-1}f_{2i}$ and only one of the pairs f_{2i-1}^2, f_{2i}^2. The discussion of the normal space is finished for now.

The next topic is the tangent space. It was shown in Section 3 that the flows

$$\frac{\partial q}{\partial t} = D\frac{\partial \Delta}{\partial q}(\lambda_{2i}^o) = -\Delta^\bullet(\lambda_{2i}^o) D(f_{2i}^o)^2 \equiv X_i q, \quad i = 0, 1, \cdots, n,$$

preserve M. Thus the X's should be tangent fields. By Amplification 6.2,

$$0 = \sum_{i=0}^n \epsilon_i D(f_{2i})^2;$$

X_0 may be omitted. A formal dimension count now shows that you have as many normal as tangent directions as the dimension of the ambient space L_1^2. The X's span a tangent space of dimension n while the dimension of the proposed normal space is $n + 1 + (2 \times$ the number of double eigenvalues) for a total count of

$$(2n + 1 = \text{ the number of simple eigenvalues}) \\ + (2 \times \text{ the number of double eigenvalues}).$$

The precise theorem may now be stated.

THEOREM 1. *Let T be the span in L_1^2 of $D(f_{2i})^2$, $i = 1, \cdots, n$, and let N be the span of $(f_{2i}^o)^2$, $i = 0, 1, 2, \cdots, n$, supplemented by the span of*[23] *$(f_{2i-1}^\times)^2$ and $f_{2i-1}^\times f_{2i}^\times$ for such double*

[23] $f_{2i-1}^\times, f_{2i}^\times$ is the i-th pair of double eigenfunctions.

eigenvalues as may exist. Then (i) N and T are perpendicular, (ii) $N \oplus T = L^2_1$, (iii) the unit function 1 belongs to N, while the functions

$$F : \sqrt{2}[f^2_{2i} - 1], \quad \sqrt{2}[(f^{\times}_{2i})^2 - 1], \quad -2^{3/2} f^{\times}_{2i-1} f^{\times}_{2i}, \quad -\sqrt{2}(\pi i)^{-1} f^{o}_{2i}(f^{o}_{2i})'$$

form an oblique base to the annihilator 1° of the unit function, meaning that any function in 1° can be uniquely written as $f = \sum c_i F_i$, the length

$$\sqrt{\int_0^1 |f(x)|^2 \, dx}$$

being comparable to $\sqrt{\sum c^2_j}$.

AMPLIFICATION 1. The geometry is most transparent if the periodic spectrum is purely simple. The normal space is the span of the gradients $\partial\Delta(\lambda_{2i})/\partial q = -\Delta^\bullet(\lambda_{2i}) f^2_{2i}$, $i = 0, 1, \cdots$, of the defining relations $2 \times (-1)^i = \Delta(\lambda_{2i})$, $i = 0, 1, \cdots$, and the skew-symmetric operator D carries it onto the tangent space $T = \text{span } f_{2i} f^{'}_{2i}$, $i = 1, 2, \cdots$, annihilating the extra function 1. This state of affairs may be thought as a symplectic structure as explained in Section 2. The fact that $\dim M = \infty$ while $\dim N = \infty + 1$ reflects the one-dimensional degeneracy of the Poisson bracket

$$\int_0^1 \frac{\partial F}{\partial q} D \frac{\partial G}{\partial q} \, dx;$$

in classical symplectic geometry the Poisson bracket is always non-degenerate.

AMPLIFICATION 2. Item (iii) of the theorem may be regarded as stating that M sits in L^2_1 without creases, *to wit, the local splitting $L^2_1 = N \oplus T$ looks the same at every point of M.* In a finite-dimensional setting, you would say *that M is nicely embedded in the ambient space.* This line of thought will be developed further in Section 13. The connection of T with flows on M will be explained in Amplification 14.1 in which *it is proved for simple spectrum that (a) every vector $Xq \in T$ can be uniquely written as $\sum_{j=1}^{\infty} x_j X_j q$ with $X_j q = -\Delta^\bullet(\lambda_{2j}) D f^2_{2j}$, $j = 1, 2, \cdots$, and*

$$\sum i^2 |\Delta^\bullet(\lambda_{2i})|^2 x^2_i$$

comparable to

$$\int_0^1 |Xq|^2 \, dx,$$

and (b) that for fixed x, the vector field X defines a continuous flow on M; the same is true for mixed spectrum, only it is a little more complicated to state, see Section 15 for more information. More refined tangent spaces are introduced in Sections 11–14 in connection with the Jacobi variety.

EXAMPLE 1. The simplest case of all is when q is constant. The lowest eigenfunction $f_0 \equiv 1$ is the only simple one, the tangent space evaporates, and the double eigenspaces $[\sin \pi i x, \cos \pi i x]$, $i = 1, 2, 3, \cdots$, produce all the normal vectors perpendicular to the unit function via the trigonometrical identities

$$\begin{aligned} 2\sin^2 \pi i x &= \cos 2\pi i x + 1, \\ 2 \sin \pi i x \times \cos \pi i x &= \sin 2\pi i x, \\ 2\cos^2 \pi i x &= \cos 2\pi i x + 1. \end{aligned}$$

EXAMPLE 2. The next simplest case ($n = 1$) is already quite complicated. Hochstadt [13] proved that $q = 2\wp + c$, in which $3c = \lambda^o_0 + \lambda^o_1 + \lambda^o_2$ and $\wp$ is the Weierstrassian elliptic functions with invariants $e_1 = -\lambda_2^o + c$, $e_2 = -\lambda_1^o + c$, $e_3 = -\lambda_0^o + c$, primitive real period

$$\omega = \frac{1}{m} = \int_{\lambda_1^o}^{\lambda_2^o} \frac{d\mu}{\sqrt{(\mu-\lambda_0^o)(\mu-\lambda_1^o)(\lambda_2^o-\mu)}},$$

and primitive imaginary period

$$\sqrt{-1}\omega' = \sqrt{-1}\int_{\lambda_0^o}^{\lambda_1^o} \frac{d\mu}{\sqrt{(\mu-\lambda_0^o)(\lambda_1^o-\mu)(\lambda_2^o-\mu)}},$$

$\wp$ being evaluated at $x + \sqrt{-1}\frac{1}{2}\omega'$, $0 \le x < 1$. The simple eigenfunctions f_0^o, f_1^o, f_2^o are constant multiples of the 3 Jacobi functions $\sqrt{\wp - e_3}$, $\sqrt{\wp - e_2}$, $\sqrt{\wp - e_1}$. The tangent space is one-dimensional ($c\wp'$, $-\infty < c < \infty$), the simple part of the normal space is two-dimensional ($a + b\wp$, $-\infty < a$, $b < \infty$), and the rest of the normal space comes from the double spectrum. The latter is determined by the formula of Hochstadt [13]

$$\Delta(\lambda) = 2\cos\left[\frac{1}{2}\int_{\lambda_0}^{\lambda} \frac{(\mu-\lambda_1^\bullet)\,d\mu}{\sqrt{(\mu-\lambda_0^o)(\mu-\lambda_1^o)(\mu-\lambda_2^o)}}\right]$$

with

$$\lambda_1^\bullet = m\int_{\lambda_1^o}^{\lambda_2^o} \frac{\mu\,d\mu}{\sqrt{(\mu-\lambda_0^o)(\mu-\lambda_1^o)(\lambda_2^o-\mu)}}.$$

The perpendicular double eigenfunctions are

$$\sqrt{2\frac{\lambda_{2i}^\times + \wp(\xi) - \frac{1}{3}c}{\lambda_{2i}^\times - \lambda_1^\bullet}} \times \begin{matrix}\sin\\ \cos\end{matrix}\left[\sqrt{(\lambda_{2i}^\times-\lambda_0^o)(\lambda_{2i}^\times-\lambda_1^o)(\lambda_{2i}^\times-\lambda_2^o)}\int_0^\xi \frac{d\eta}{\lambda_{2i}^\times + \wp(\eta) - \frac{1}{2}c}\right],$$

$\lambda_{2i}^\times$ being the i-th double eigenvalue; see Amplification 3.3 and, for more information, [28], p.238.

7.9. Normal and Tangent Spaces: Proofs

The informal discussion of Section 8 was made precise in Theorem 1. Here, for simplicity, we prove the theorem for a purely imaginary spectrum; the proof in the presence of double eigenvalues will be outlined. The method of proof goes back, in part, to Borg [2].

THEOREM 1. *Let $q \in C_1^\infty$ have a purely simple spectrum, let N be the span of f_{2j}^2, $j = 0, 1, 2, \cdots$, and T be the span of Df_{2j}^2, $j = 1, 2, \cdots$. Then (i) N and T are perpendicular, (ii) $N \oplus T = L_1^2$, and (iii) the unit function 1 belongs to N, while the functions*

$$F:\ \sqrt{2}[f_{2j}^2 - 1], \qquad -\sqrt{2}(\pi j)^{-1} f_{2j} f_{2j}', \quad j = 1, 2, \cdots,$$

form an oblique base for the space perpendicular to the unit function.

Proof: For any $i = 0, 1, 2, \cdots$, $f_i^2(x)$ is of period 1; in fact, at a simple eigenvalue, $f_i(x+1) = \pm f_i(x)$, the former being a solution of $Qf = \lambda_i f$ and so proportional to the latter. Item (i) now follows from the skew-symmetry of the operator $L = qD + Dq - \frac{1}{2}D^3$ and the fact that $Lf_i^2 = 2\lambda_i Df_i^2$. In detail,

$$\begin{aligned} 2\lambda_{2i}\int_0^1 f_{2i}^2 Df_{2j}^2\,dx &= \int_0^1 f_{2i}^2 Lf_{2j}^2\,dx = -\int_0^1 f_{2j}^2 Lf_{2i}^2\,dx \\ &= -2\lambda_{2i}\int_0^1 f_{2j}^2 Df_{2i}^2\,dx = 2\lambda_{2i}\int_0^1 f_{2i}^2 Df_{2j}^2\,dx \end{aligned}$$

and $\lambda_{2i} \ne \lambda_{2j}$, so the integral must vanish.

LEMMA 1. *As* $j \uparrow \infty$,

$$\begin{aligned} f_{2j}(x) &= \sqrt{2}\cos \pi j(x+\phi_j) + O(1/j), \\ f'_{2j}(x) &= -\sqrt{2}\pi j \sin \pi j(x+\phi_j) + O(1), \end{aligned}$$

with suitable phases $0 \le \phi_j < 2\pi$, *uniformly in* $0 \le x < 2$ *and in any ball* $\|q\|_0 \le C$.

Proof: Recall the estimate $\lambda_{2j} = \pi^2 j^2 + O(1)$; see Section 1. Now

$$\begin{aligned} \int_0^1 [f_{2j}^2 + \lambda_{2j}^{-1} f_{2j}'^2]\, dx &= 1 - \lambda_{2j}^{-1} \int_0^1 f_{2j} f''_{2j}\, dx \\ &= 1 + \lambda_{2j}^{-1} \int_0^1 f_{2j}^2 (\lambda_{2j} - q)\, dx \\ &= 2 + O(1/j^2); \end{aligned}$$

thus

$$f_{2j}^2(x_j) + \lambda_{2j}^{-1} f_{2j}'^2(x_j) = 2 + O(1/j^2)$$

for some $0 \le x_j < 1$, making use of the fact that $f_{2j}(x+1) = \pm f_{2j}(x)$. Let $y_1(x,\lambda)$ and $y_2(x,\lambda)$ be normalized in the usual way at $x = x_j$ instead of at $x = 0$, i.e., $y_1(x_j,\lambda) = 1$, etc. Then with $f_{2j}(x_j) = a$, $(\lambda_{2j})^{-1/2} f'_{2j}(x_j) = b$, and $\lambda = \lambda_{2j} = \pi^2 j^2 + O(1)$, you have

$$\begin{aligned} f_{2j}(x) &= a y_1(x,\lambda) + b\sqrt{\lambda} y_2(x,\lambda) \\ &= a\cos\sqrt{\lambda}(x - x_j) + b \sin\sqrt{\lambda}(x-x_j) + O(\lambda^{-1/2}) \\ &= \sqrt{a^2+b^2}\cos\sqrt{\lambda}(x+\phi_j) + O(\lambda^{-1/2}) \\ &= \sqrt{2}\cos \pi j(x+\phi_j) + O(1/j) \end{aligned}$$

with a suitable phase $0 \le \phi_j < 2\pi$. This confirms the first estimate. The proof of the second is similar.

COROLLARY 1. As $j \uparrow \infty$,

$$\begin{aligned} f_{2j}^2 - 1 &= \cos 2\pi j(x+\phi_j) + O(1/j), \\ -(\pi j)^{-1} f_{2j} f'_{2j} &= \sin 2\pi j(x+\phi_j) + O(1/j). \end{aligned}$$

Let

$$\begin{aligned} F_j &= \sqrt{2}(f_{2j}^2 - 1), \qquad & F_j^o &= \sqrt{2}\cos 2\pi j(x+\phi_j), \\ F_{-j} &= -\sqrt{2}(\pi j)^{-1} f_{2j} f'_{2j}, \qquad & F_{-j}^o &= \sqrt{2}\sin \pi j(x+\phi_j), \end{aligned}$$

for $j = 1, 2, \cdots$. Items (ii) and (iii) are proven together. The plan is as follows. F_j^o, $j \ne 0$, is a unit perpendicular base for the annihilator 1° of the unit function. $F_j^{\o}$, $j \ne 0$, also lies in 1° and may be compared to F_j^o, $j \ne 0$, by means of the matrix J with entries

$$J_{ij} = \int_0^1 F_i F_j^o\, dx, \qquad i,\, j \ne 0.$$

J acts on the space l^2 of vectors $c = (\cdots, c_{-1}, c_1, c_2, \cdots)$ with $\sum c_i^2 < \infty$ by the rule $(Jc)_i = \sum_{j\ne 0} J_{ij} c_j$. It is easy to prove that J is of the form *identity + compact*. If it were invertible, F_j, $j \ne 0$, would span 1°, and the identity $1 = \sum_{i=0}^{\infty} \epsilon_i f_{2i}^2$ of Proposition 6.3 would finish the proof of item (ii). Item (iii) would also follow from the existence of J^{-1}. Assume that J is *not* invertible. Then you can find (with some tears) numbers c_i, $i = 1, 2, \cdots$, with fairly rapid decrease at infinity so as to make $\sum_{i=1}^{\infty} c_i f_{2i} f'_{2i} = 0$, with excellent convergence of the sum. Introduce the vector field

$$Y : p \to \sum_{i=1}^{\infty} c_i f_{2i} f'_{2i}$$

for general $p \in M$. Then $Yp = 0$ at $p = q$ from which it follows that $Yp = 0$ on M. One can see this fact in the following way. Let X be a finite linear combination of the Hamiltonian vector fields $p \to D\partial\Delta(\lambda)/\partial p(x)$ of Section 3. Then the induced mappings $\exp\{X\}$ of M commute and distribute a fixed point, such as q, densely over M. Now Y is an *infinite* linear combination of Hamiltonian fields (see Lemma 3.1). Technical difficulties aside, you ought to have $e^{tY}e^X = e^X e^{tY}$ on M, and you should be able to differentiate this identity with respect to t at $t = 0$ to obtain the vanishing of Y at the general point $p = e^X q$ of M:

$$Yp = \frac{d}{dt} e^{tY} e^X q = \frac{d}{dt} e^X e^{tY} q = \int_0^1 \frac{\partial e^X q}{\partial q} Y q \, dx = 0.$$

A contradiction is then obtained by direct computation of $Yp \neq 0$ at the point p on M with tied spectrum $\mu_i = \lambda_{2i}$, $i = 1, 2, \cdots$. The plan will be implemented by a series of simple lemmas.

LEMMA 2. *J is of the form: identity + compact.*

Proof: Let $K_{ij} = J_{ij}$, $i \neq j$, and let $K_{ii} = J_{ii} - 1$. We know that[24]

$$\begin{aligned}
\sum_{j\neq 0} K_{ij}^2 &= \sum_{j\neq i,0} (F_i, F_j^o)^2 + [(F_i, F_i^o) - 1]^2 \\
&= \sum_{j\neq i,0} (F_i - F_i^o, F_j^o)^2 + (F_i - F_i^o, F_i^o)^2 \\
&= \|F_i - F_i^o\|^2 \\
&= O(i^{-2}),
\end{aligned}$$

by Corollary 1 and the fact that F_j, $j \neq 0$, is a unit perpendicular base of 1°. Therefore, $\sum K_{ij}^2 < \infty$ and K is compact.

LEMMA 3. *J is invertible and F_j, $j \neq 0$, spans 1° if and only if $\sum_{j\neq 0} c_j F_j \neq 0$ for every $c \neq 0$ in l^2.*

Proof: The sum converges in L_1^2 by Corollary 1. $F \in 1^\circ$ is perpendicular to F_i, $i \neq 0$, if and only if

$$0 = \sum_{j\neq 0} (F_i, F_j^o)(F, F_j^o), \qquad i \neq 0,$$

which is to say that J annihilates the non-trivial vector $c \in l^2$ with $c_j = (F, F_j^o)$, $j \neq 0$. Now $J =$ *identity + compact*; thus the dimension of the null space of J is the same as that of $J^\dagger$. The proof is finished by observing that $J^\dagger c = 0$ if and only if $F = \sum c_j F_j$ is perpendicular to F_i^o, $i \neq 0$.

A refinement of Lemma 3 is needed. Let l^{2+} be the space of vectors $c = (c_1, c_2, \cdots)$ with $\sum_{i>0} c_i^2 < \infty$.

LEMMA 4. *The condition of Lemma 3 may be replaced by $\sum_{j>0} c_j F_j \neq 0$ for every $c \neq 0$ from l^{2+}.*

Proof: The content of the lemma is that $\sum_{j\neq 0} c_j F_j = 0$ with $c \neq 0$ from l^2 implies $\sum_{j>0} c'_j F_j = 0$ for some $c' \neq 0$ from l^{2+}. The vanishing of $\sum_{j\neq 0} c_j F_j$ implies the vanishing

[24] $(F, G) = \int_0^1 F(x) G(x)\, dx$; $\|F\|^2 = \int_0^1 F^2\, dx$.

of the sums for $j < 0$ and for $j > 0$, separately, by perpendicularity, and *either* $c_j \not\equiv 0$, $j > 0$, in which case the proof is finished, *or* $c_j \not\equiv 0$, $j < 0$, in which case

$$0 = \sum_{j<0} c_j F_j = -\sum_{j>0} c_{-j} \frac{\sqrt{2}}{\pi j} f_{2j} f_{2j}',$$

and you may integrate with regard to x to obtain a sum of the required form:

$$0 = \sum_{j>0} c_{-j} (\pi j)^{-1} (f_{2j}^2 - 1) = \sum_{j>0} c_j' F_j,$$

the constant of integration being absorbed in the -1's.

LEMMA 5. *Let* $\sum_{j>0} a_j F_j = 0$ *for some* $a \neq 0$ *from* l^{2+}. *Then it is possible to find* $b \neq 0$ *in* l^{2+} *so that* $\sum_{j>0} b_j f_{2j} f_{2j}'$, *qua function of* $0 \leq x < 1$ *and a general point* $p \in M$, *is continuous in the pair* (x, p) *and vanishes at* $p = q$.

Proof: Fix $\lambda \neq \lambda_i$, $i = 0, 1, 2, \cdots$. The estimates $f_{2j} = O(1)$ and $f_{2j}' = O(j)$ of Lemma 1 are uniform on $[0, 2) \times M$. In view of the first estimate, $\sum_{j>0} a_j (\lambda - \lambda_{2j})^{-1} f_{2j}^2 = \psi$ converges uniformly for $0 \leq x < 1$ in view of $\lambda_{2j} \sim \pi^2 j^2$ as $j \uparrow \infty$. For any $\psi \in C_1^\infty$,

$$(f_{2j}^2, (L - 2\lambda D)\psi) = -((L - 2\lambda D) f_{2j}^2, \psi) = 2(\lambda_{2j} - \lambda)(f_{2j}^2, D\psi);$$

consequently,

$$\begin{aligned}
(\psi, (L - 2\lambda D)\psi) &= \sum a_j (\lambda - \lambda_{2j})^{-1} (f_{2j}^2, (L - 2\lambda D)\psi) \\
&= -2 \sum a_j (f_{2j}^2, D\psi) \\
&= -\sqrt{2} \sum a_j (F_j, \psi) \\
&= 0.
\end{aligned}$$

Therefore, ψ is a weak solution of $L\psi - 2\lambda D\psi = 0$, and as such, a bonafide solution of class C_1^∞. Thus, by Lemma 3.4 and choice of λ (almost any one will do), $\psi = c \times \partial\Delta/\partial q$ with a constant depending upon λ. By Section 6,

$$\psi = -c \sum_{j=0}^{\infty} \frac{y_1'(1, \lambda)}{y_1'^{\bullet}(1, \lambda_{2j})} \frac{1}{\lambda - \lambda_{2j}} \Delta^{\bullet}(\lambda_{2j}) f_{2j}^2$$

with rapidly decreasing $\Delta^{\bullet}(\lambda_{2j})/y_1'^{\bullet}(1, \lambda_{2j}) = \epsilon_j$. Now you have two expansions for ψ in terms of f_{2j}^2, $j = 0, 1, 2, \cdots$, and subtracting one from the other produces the identity

$$\begin{aligned}
0 &= \sum_{j=1}^{\infty} \frac{a_j}{\lambda - \lambda_{2j}} f_{2j}^2 + c \sum_{j=0}^{\infty} \frac{y_1'(1, \lambda)}{y_1'^{\bullet}(1, \lambda_{2j})} \frac{\Delta^{\bullet}(\lambda_{2j}}{\lambda - \lambda_{2j}} f_{2j}^2 \\
&= \sum_{j=1}^{\infty} \left[\frac{a_j}{\lambda - \lambda_{2j}} + \frac{c y_1'(1, \lambda)}{\lambda - \lambda_{2j}} \epsilon_j \right] (f_{2j}^2 - 1) + \frac{c y_1'(1, \lambda)}{\lambda - \lambda_0} \epsilon_0 (f_0^2 - 1) \\
&= \sum_{j=1}^{\infty} \left[\frac{a_j}{\lambda - \lambda_{2j}} + \frac{c y_1'(1, \lambda)}{\lambda - \lambda_{2j}} \epsilon_j - \frac{c y_1'(1, \lambda)}{\lambda - \lambda_0} \epsilon_j \right] (f_{2j}^2 - 1) \\
&= \sum_{j=1}^{\infty} b_j (f_{2j}^2 - 1).
\end{aligned}$$

The insertion of the $-1's$ in line 2 is justified by the fact that

$$\int_0^1 (f_{2i}^2 - 1)\, dx = 0;$$

the identity $0 = \sum_{i=0}^{\infty} \epsilon_i(f_{2i}^2 - 1)$ is used in line 3. The proof is finished by observing that $\sum_{j=1}^{\infty} b_j f_{2j} f'_{2j}$ converges uniformly on $[0,1) \times M$ and that the coefficients b_j do not all vanish. The first point is settled by the estimates $f_{2j}f'_{2j} = O(j)$, $\lambda_{2j} \sim \pi^2 j^2$, $b_j = a_j \times O(j^{-2}) + O(j^{-4})$, say, and $\sum a_j^{23} < \infty$. The second is dealt with as follows: If $b_j = 0$, $j = 1, 2, \cdots$, then $a_j = k\epsilon_j(\lambda_{2j} - \lambda_0)$, $j = 1, 2, \cdots$, with $k \neq 0$, so

$$0 = \sum_{j>0} a_j F_j = \sum_{j=0}^{\infty} \epsilon_j 2(\lambda_{2j} - \lambda_0)(f_{2j}^2 - 1) = q,$$

by Proposition 6.4 and the identity $\sum_{j=0}^{\infty} \epsilon_j(f_{2j}^2 - 1) = 0$. This is a contradiction. The proof is finished.

We are now in a position to prove item (ii). Let J be non-invertible, pick $b \in l^{2+}$ as in Lemma 5, and let Y be the vector field

$$Y : p \to \sum_{j=1}^{\infty} b_j f_{2j} f'_{2j}$$

defined for general $p \in M$; Yp is continuous on $[0,1) \times M$ and $Yq = 0$, by Lemma 5.

LEMMA 6. *Let F_j and $G_j = -2^{-1/2} D[y_1(x, \lambda_{2j}) y_2(x, \lambda_{2j})]$ be computed for $j = 1, 2, \cdots$ at the point of M with tied spectrum $\mu_i = \lambda_{2i}$, $i = 1, 2, \cdots$. Then $(F_i, G_j) = 1$ or 0 according to whether $i = j$ or not.*

Proof: If $\mu_j = \lambda_{2j}$, then $f_{2j}(x) =$ constant $\times\, y_2(x, \lambda_{2j})$ vanishes at $x = 0$ and at $x = 1$. Now

$$\sqrt{2} \int_0^1 F_i G_j \, dx = -\int_0^1 f_{2i}^2(x) D[y_1(x, \lambda_{2j}) y_2(x, \lambda_{2j})] \, dx,$$

and the stated perpendicularity for $i \neq j$ follows from $\lambda_{2i} \neq \lambda_{2j}$ by a familiar partial integration:

$$\begin{aligned}
2\lambda_{2j} \int_0^1 f_{2i}^2 D[y_1 y_2] \, dx &= \int_0^1 f_{2i}^2 L y_1 y_2 \, dx \\
&= 2q f_{2i}^2 y_1 y_2 - \tfrac{1}{2} f_{2i}^2 (y_1 y_2)'' + \tfrac{1}{2}(f_{2i}^2)'(y_1 y_2)' - \tfrac{1}{2}(f_{2i}^2)'' y_1 y_2 \Big|_0^1 \\
&\quad - \int_0^1 y_1 y_2 L f_{2i}^2 \, dx \\
&= -2\lambda_{2i} \int_0^1 y_1 y_2 D f_{2i}^2 \, dx \\
&= 2\lambda_{2i} \int_0^1 f_{2i}^2 D[y_1 y_2] \, dx.
\end{aligned}$$

Finally, for $i = j$,

$$\begin{aligned}
\int_0^1 F_i G_i \, dx &= -\int_0^1 f_{2i}^2 (y_1 y_2)' \, dx \\
&= \int_0^1 y_1 y_2 (f_{2i}^2)' \, dx \\
&= \int_0^1 f_{2i}^2 (y_1 y_2' - y_1 y_2') \, dx \\
&= \int_0^1 f_{2i}^2 \, dx \\
&= 1,
\end{aligned}$$

by use of $f_{2j}(x) =$ constant $\times\, y_2(x, \lambda_{2j})$ to pass from line 2 to line 3. The proof is finished.

COROLLARY 2. *Y does not vanish identically on M.*

Proof: Let p be the point distinguished in Lemma 6, at which $\mu_i = \lambda_{2i}$, $i = 1, 2, \cdots$. Then $Yp = \sum b_j f_{2j} f'_{2j} = 0$ would imply $\psi = \sqrt{2} \sum b_j (f_{2j}^2 - 1) = \sum b_j F_j = 0$ also, and hence $b_j = (\psi, G_j) = 0$, $j = 1, 2, \cdots$, contradicting the choice of the b's.

LEMMA 7. *Let X be any finite linear combination of the vector fields*

$$X_j : p \to D\partial\Delta(\lambda_{2j})/\partial p(x) = -\Delta^\bullet(\lambda_{2j}) D f_{2j}^2.$$

Then Y vanishes at $p = e^X q$.

Proof: Let

$$Y_m : p \to \sum_{j \le m} b_j f_{2j} f'_{2j}.$$

Then e^{tY} commutes with e^X, and, by differentiation at $t = 0$,

$$Y_m e^X q = \frac{d}{dt} e^{tY_m} e^X q = \frac{d}{dt} e^X e^{tY_m} q = \int_0^1 \frac{\partial e^X q}{\partial q} Y_m q \, dx.$$

The proof is finished by the uniform convergence of $Y_m p$ to Y_p; see Lemma 5.

The proof of item (ii) is now before you. To obtain a contradiction, you have only to pick $e^X q$ as in Lemma 7, tending to the point p of Lemma 6, and to use the fact that Y is continuous on M. This is achieved as follows. Besides the maps e^X, it is also permissible to use the flow of translation. If $\mu_i = \lambda_{2i}$, say, then $f_{2i}(0) = 0$ and an appropriate small translation will break that for every i, simultaneously; in short, you can assume that q is in general position: $\lambda_{2i-1} < \mu_i < \lambda_{2i}$, $i = 1, 2, \cdots$. The fact that $\det X_j \mu_i$, $1 \le i, j \le m$, does not vanish for $m = 1, 2, \cdots$ and q in general position (see Amplification 4.1) permits you to move q as close as you please to p by a suitable map e^X. The proof of item (ii) is finished, and item (iii) is easily disposed of. Every function $f \in 1^\circ$ can be uniquely expressed as $\sum_{j \ne 0} c_j F_j$ and also as $\sum_{j \ne 0} c_j^o F_j^o$; the coefficients being related by $\sum_{i \ne 0} c_i^o J_{ij} = c_j$, $j \ne 0$. The rest is clear from the fact that J^{-1} must be bounded, being everywhere defined: $c^o = (J^*)^{-1} c$, and so

$$\|J\|^2 \sum_{i \ne 0} c_i^2 \le \sum_{i \ne 0} (c_i^o)^2 = \int_0^1 |f|^2 \, dx \le \|J^{-1}\|^2 \sum_{i \ne 0} c_i^2,$$

i.e., the functions F_j, $j \ne 0$, form an oblique base for 1°.

The proof in the presence of double eigenvalues. The proof for a mixed spectrum is much the same. We have $F_{-j} = -\sqrt{2}(\lambda_{2j}^o)^{-1} f_{2j}^o f_{2j}^o$, $j = 1, \cdots, n$, while the functions[25]

$$\sqrt{2} f_{2i-1}^o, \quad \sqrt{2}(f_{2i}^\times - 1), \quad -2^{3/2} f_{2i-1}^\times f_{2i}^\times$$

are listed in any convenient order as F_j, $j = 1, 2, \cdots$. The only substantial change is in Lemma 4 in which the condition $\sum_{j=1}^\infty c_j F_j$ is replaced by $\sum_{j=1}^n c_j (f_{2j}^{o^2} - 1)$, and the point at issue is whether

$$\sum_{j=1}^n c_j (f_{2j}^{o^2} - 1) + \sum a_j (f_{2i}^{\times^2} - 1) + \sum b_i f_{2i-1}^\times n f_{2i}^\times = 0,$$

implies $a_i = b_j = 0$, $i = 1, 2, \cdots$.

[25] $f_{2i-1}^\times, f_{2i}^\times$ is the i-th pair of double eigenfunctions.

Proof: $(f^{\times}_{2i-1}f^{\times}_{2i})'$, respectively $(f^{\times^2}_{2i})'$, is perpendicular to every summand except perhaps $f^{\times^2}_{2i}-1$, respectively $f^{\times}_{2i-1}f^{\times}_{2i}$, so it is enough to check

$$\int_0^1 f^{\times^2}_{2i}(f^{\times}_{2i-1}f^{\times}_{2i})'\,dx = -\int_0^1 f^{\times}_{2i-1}f^{\times}_{2i}(f^{\times^2}_{2i})'\,dx \neq 0.$$

Now $f^{\times}_j(x+1) = \pm f^{\times}_j(x)$, $j = 2i-1, 2i$, so $f^{\times^2}_{2i}$ and $(f^{\times}_{2i-1}f^{\times}_{2i})'$ are of period 1 and the integration may be performed over *any* period. Thus, you may assume $f^{\times}_{2i-1}(x) = ay_1(x,\lambda)+by_2(x,\lambda)$ and $f^{\times}_{2i}(x) = cy_2(x,\lambda)$ with $a \neq 0$, λ being the eigenvalue. Then the inner product of $f^{\times}_{2i-1}f^{\times}_{2i}$ and $(f^{\times}_{2i})'$ is a non-vanishing multiple of

$$\begin{aligned}\int_0^1 y_1(x,\lambda)y_2(x,\lambda)[y_2^2(x,\lambda)]'\,dx &= -y_1(1,\lambda)y_2^{\bullet}(1,\lambda)[y_2'(1,\lambda)]^2\\ &= -y_2^{\bullet}(1,\lambda)y_2'(1,1,\lambda)\\ &= -\frac{1}{2}\Delta(\lambda)\\ &= \pm 1.\end{aligned}$$

The first line is proved by partial integration of

$$\int_0^1 y_1(x,\mu)y_2(x,\mu)Dy_2^2(x,\lambda)\,dx \qquad \text{for } \lambda \neq \mu,$$

using Lemma 3.3 and the skew-symmetry of D and $L = qD+Dq-\frac{1}{2}D^3$. The proof is finished.

7.10. Local Hamiltonian Tangent Vectors

Let V_i, $i = 1, 2, \cdots$, be the local Hamiltonian fields $V_1 : q \to q'$, $V_2 : q \to 3qq' - \frac{1}{2}q'''$, etc. Then $V_iq \in T$, $i = 1, 2, \cdots$; in fact, by Proposition 6.2,

$$V_iq = \sum_{j=0}^{\infty} \epsilon_j p_i(\lambda^o_{2j})D(f^o_{2j})^2$$

with rapidly deceasing coefficients $\epsilon_j = \Delta^{\bullet}(\lambda^o_{2j})/y_1'^{\bullet}(1,\lambda^o_{2j})$ and a polynomial p_i of precise degree i.

Warning. The function $y_1'(1,\lambda)$ is computed for the origin of M, so that $y_1'(1,\lambda_{2i}) = 0$, $i = 0, 1, 2, \cdots$.

The vectors V_iq, $i = 1, \cdots, n$, span T if $n < \infty$; it is believed that this can fail if $n = \infty$. The next theorem presents a criterion; see Amplification 1 below for more information.

THEOREM 1. *The vectors V_iq, $i = 1, \cdots, n$, span T if and only if polynomials span the space of functions $\phi(\lambda^o_{2j})$, $j = 1, \cdots, n$, with*

$$\sum_{j=1}^{n}(\lambda^o_{2j}-\lambda^o_{2j-1})^2|\phi(\lambda^o_{2j})|^2 < \infty;$$

naturally, this is automatic if $n < \infty$.

Proof: The proof starts from the expansion of V_iq, $i = 1, 2, \cdots$, in terms of $D(f^o_{2j})^2$, $j = 0, 1, 2, \cdots$, revised by use of the identity $V_0q = 0 = \sum_{j=0}^{\infty}\epsilon_j D(f^o_{2j})^2$ to make the sum commence at $j = 1$ instead of $j = 0$. The condition that $D(f^o_{2i})^2$ belong to the span of V_iq, $i = 1, 2, \cdots$, is

$$\inf\left\|\sum_{j=1}^{n}\epsilon_j p(\lambda^o_{2j})D(f^o_{2j})^2 - D(f^o_i)^2\right\|_2 = 0,$$

the infimum being reckoned over all polynomials p with $p(\lambda_0) = 0$. Now $\sqrt{2}(\lambda_{2j}^o)^{-1/2} D(f_{2j}^o)^2$, $j = 1, \cdots, n$, is an (oblique) base of T as explained in Theorem 8, so the condition that $D(f_{2i})^2$, $i \neq 0$, belongs to the span can be reexpressed as

$$\inf \sum_{j=1}^{n} \lambda_{2j}^o \epsilon_j^2 |p(\lambda_{2j}^o) - \chi_i(\lambda_{2j}^o)|^2,$$

in which $\chi_i(\lambda)$ is the indicator of $\lambda = \lambda_{2i}^o$ and the infimum is reckoned as before, The rest of the proof consists of a sharpening of Lemma 6.1.

LEMMA 1. *ϵ_i is comparable to $(\lambda_{2i}^o)^{-1} \times (\lambda_{2i}^o - \lambda_{2i-1}^o)$ as $i \uparrow \infty$.*

Proof: Let $y_2(1, \lambda)$ be computed for a point of M with $\mu_i^o = \lambda_i^\bullet$, $i = 1, \cdots, n$, so that $\Delta^\bullet(\lambda)$ is proportional to $y_2(1, \lambda)$, as in the proof of Lemma 6.1, and follow that proof so far as to see that ϵ_i is comparable to $(\lambda_{2i}^o - \lambda_i^\bullet) y_2^\bullet(1, \lambda)$ with an intermediate $\lambda \in [\lambda_i^\bullet, \lambda_{2i}^o]$. For any $\lambda_{2i-1}^o \leq \lambda \lambda_{2i}^o$,

$$\begin{aligned} y_2^\bullet(1,\lambda) y_2'(1,\lambda) - y_2(1,\lambda) y_2'^\bullet(1,\lambda) &= \int_0^1 y_2^2(x,\lambda)\, d\lambda \\ &= \frac{1}{\lambda} \int_0^1 \sin^2 \sqrt{\lambda} x \, dx \, [1 + o(1)] \\ &\sim \frac{1}{2\lambda}, \end{aligned}$$

while $y_2(1, \lambda) = o(\lambda - \lambda^\bullet) = o(\lambda_{2i}^o - \lambda_{2i-1}^o)$, and $y_2'(1, \lambda) = O(1)$, where $|2\lambda y_2^\bullet(1, \lambda)|$ is bounded away from 0 for $\lambda_{2i-1}^o \leq \lambda \leq \lambda_{2i}^o$, independently of $i = 1, 2, \cdots$. Besides, $y_2^\bullet(1, \lambda) = O(1/\lambda)$ as $\lambda \uparrow \infty$ anyhow, and the upshot is that $y_2^\bullet(1, \lambda)$ is comparable to $1/\lambda_{2i}^o$, in the whole of $[\lambda_{2i-1}^o, \lambda_{2i}^o]$ independently of i and that ϵ_i is comparable to $(\lambda_{2i}^o)^{-1} \times (\lambda_{2i}^o - \lambda_i^\bullet)$. It remains to prove that $\lambda_{2i}^o - \lambda_i^\bullet$ is comparable to $\lambda_{2i}^o - \lambda_{2i-1}^o$. $\Delta^\bullet(\lambda) =$ constant $\times\, y_2(1, \lambda)$, so from the estimate of $y_2^\bullet(1, \lambda)$, you have

$$\begin{aligned} c_1(\lambda_i^\bullet - \lambda) &\leq \pm \lambda_{2i}^o \Delta^\bullet(\lambda) \leq c_2(\lambda_i^\bullet - \lambda), & \lambda_{2i-1}^o \leq \lambda \leq \lambda_i^\bullet, \\ c_1(\lambda - \lambda_i^\bullet) &\leq \mp \lambda_{2i}^o \Delta^\bullet(\lambda) \leq c_2(\lambda - \lambda_i^\bullet), & \lambda_i^\bullet \leq \lambda \leq \lambda_{2i}^o, \end{aligned}$$

with positive constants c_1 and c_2 not depending upon i and the same signature as $\Delta(\lambda_{2i}^o) = \pm 2$. But

$$\int_{\lambda_{2i-1}^o}^{\lambda_i^\bullet} \Delta^\bullet(\lambda)\, d\lambda = -\int_{\lambda_i^\bullet}^{\lambda_{2i}^o} \Delta^\bullet(\lambda)\, d\lambda,$$

so

$$\frac{c_1}{c_2}(\lambda_i^\bullet - \lambda_{2i-1}^o)^2 \leq (\lambda_{2i}^o - \lambda_i^\bullet)^2 \leq \frac{c_2}{c_1}(\lambda_i^\bullet - \lambda_{2i-1}^o)^2,$$

by integration of the inequalities for $\lambda_{2i}^o \Delta^\bullet(\lambda)$. The proof of the lemma is finished.

The proof of the theorem is now before you: The condition of spanning is expressed as[26]

$$\inf \sum_{j=1}^{n} (\lambda_{2j}^o)^{-1} (\lambda_{2j}^o - \lambda_{2j-1}^o)^2 |(\lambda_{2j}^o - \lambda_0) p(\lambda_{2j}^o) - \chi_i(\lambda_{2j}^o)|^2 = 0.$$

The proof is finished by an elementary remark: the success or non-success of polynomial approximation is unaffected by removal of the factor λ_{2j}^o in front of $(\lambda_{2j}^o - \lambda_{2j-1}^o)^2$.

AMPLIFICATION 1. Magnus-Winkler [26], p. 65, cite the fact that $\lambda_{2i} - \lambda_{2i-1} \leq a e^{-bi}$ with a, $b > 0$ if $q(x) = \sin(2\pi x)$, so the spanning takes place; presumably, the same is true for any real analytic q. Let $\ell_p^2 = \sum (\lambda_{2i}^o)^p (\lambda_{2i}^o - \lambda_{2i-1}^o)^2$, $p = 0, 1, 2, \cdots$. Then the spanning

[26] $\lambda_0 = 0$ is now assumed so that $\lambda_{2j}^o > 0$, $j \neq 0$, for ease of writing.

takes place if $\sum(\ell_p)^{-1/p} = \infty$ by a criterion of Carleman [3]; see Akhiezer [1], p. 45. Let $\lambda_{2i-1} < \lambda_{2i}$, $i = 1, 2, \cdots$, so that $\lambda^o_{2i} = \lambda_{2i} \sim i^2\pi^2$ as $i \uparrow \infty$. Then the spanning cannot take place unless $\sum i^{-2} \log(\lambda_{2i} - \lambda_{2i-1}) = -\infty$; see Koosis [19]. The result of Erdélyi [7] should be mentioned at this point: if $k = \int_0^1 q^2\,dx$ is uniformly small, then $\lambda_{2j} - \lambda_{2j-1}$ is approximately $k \times \left|\int_0^1 e^{2\pi\sqrt{-1}jx} q(x)\,dx\right|$ for each j, separately. McKean-Moerbeke [28] p.251, proved that the Hamiltonian series H_j, $j = 0, 1, 2, \cdots$, determines M if $n < \infty$; it is believed that this happens for $n = \infty$ if and only if the spanning takes place.

7.11. M as a Jacobi Variety: Purely Simple Spectrum

Let $\lambda_{2i-1} < \lambda_{2i}$, $i = 1, 2, \cdots$, so that the whole periodic spectrum is simple. The purpose of this section is to prove that M can be visualized as the real part of the Jacobi variety of the Riemann surface S of $\sqrt{\Delta^2(\lambda) - 4}$. This surface is a sphere with an infinite number of handles obtained by cutting two copies of the number sphere along the intervals of instability $[-\infty, \lambda_0]$, $[\lambda_{2i-1}, \lambda_{2i}]$, $i = 1, 2, \cdots$, and pasting them together; compare Fig. 7.3. The present development represents an extension of the classical function theory of the hyperelliptic irrationality $\sqrt{-(\lambda - \lambda^o_0)(\lambda - \lambda^o_1)\cdots(\lambda - \lambda^o_{2n})}$ to the case of infinitely many branch points. To the best of our knowledge, the papers of Myrberg [30, 31] represent the only previous work on this subject.

LEMMA 1. $\left|\int_0^1 [\Delta^2(\mu) - 4]^{-1/2}\,d\mu\right|$ *is comparable to* i *as* $i \uparrow \infty$.

Proof: Let $y_1'(1,\lambda)$ and $y_2(1,\lambda)$ be computed at the origin[27] of M so that[28] $y_1'(1,\lambda_{2i}) = 0$, $i = 0, 1, 2, \cdots$, and $y_2(1,\lambda_{2i-1}) = 0$, $i = 1, 2, \cdots$. Then $\Delta^2(\lambda) - 4 = y_1'(1,\lambda)y_2(1,\lambda)$, up to a multiplicative constant which is ignored for simplicity. For $\lambda_{2i-1} \le \mu \le \lambda_{2i}$, you have

$$\frac{\Delta^2(\mu) - 4}{(\lambda_{2i} - \mu)(\mu - \lambda_{2i-1})} = -y_2^\bullet(1,\mu^*)y_1'^\bullet(1,\mu^{**})$$

for some intermediate points $\lambda_{2i-1} \le \mu^*, \mu^{**} \le \lambda_{2i}$, But $y_2^\bullet(1,\mu)$ is comparable to $1/\lambda_{2i}$ in the whole interval $\lambda_{2i-1} \le \mu \le \lambda_{2i}$, as in the proof of Lemma 7.10.1, while if $y_1'^\bullet(1,\mu)$ is comparable to 1 by a similar proof. Therefore,

$$\int_{\lambda_{2i-1}}^{\lambda_{2i}} \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}}$$

is comparable to

$$(\lambda_{2i})^{1/2} \int_{\lambda_{2i-1}}^{\lambda_{2i}} \frac{d\mu}{\sqrt{(\lambda_{2i} - \mu)(\mu - \lambda_{2i-1})}} \sim \pi^2 i \quad \text{as } i \uparrow \infty,$$

as required.

USAGE. The functions $y_1'(1,\lambda)$ and $y_2(1,\lambda)$ are computed at the origin of M until further notice. The notation

$$\mathbf{1}_j(\mu) = \prod_{1 \le i \ne j} \frac{1 - \mu/\lambda_{2i}}{1 - \lambda_{2j}/\lambda_{2i}} = \frac{y_1'(1,\mu)}{\mu - \lambda_0} \frac{\lambda_{2j} - \lambda_0}{y_1'^\bullet(1,\lambda_{2j})} \frac{1}{\mu - \lambda_{2j}}, \quad j = 1, 2, \cdots,$$

is employed is connection with interpolation of $I^{3/2}$ off λ_{2j}, $j = 1, 2, \cdots$, by means of Theorem 7.5.3.

[27] $\mu_i = \lambda_{2i-1}$, $i = 1, 2, \cdots$.
[28] See Amplification 7.1.

Now define a map of the point $\mathbf{p}$ with coordinates $\mathbf{p}_i = (\mu_i, \sqrt{\Delta^2(\mu_i) - 4})$, $i = 1, 2, \cdots$, of the ∞-dimensional torus of Fig. 7.4 to a point $x = (x_1, x_2, \cdots)$ of the ∞-dimensional real number space by the rule

$$2\sum_{i=1}^{\infty}\int_{\mathbf{o}_i}^{\mathbf{p}_i} \mathbf{1}_j(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}, \quad j = 1, 2, \cdots,$$

in which $\mathbf{o}_i = (\lambda_{2i-1}, 0)$, $i = 1, 2, \cdots$, and the integral is extended clockwise about the i-th circle of Fig. 7.4 *less than once around.*

THEOREM 1. $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \cdots)$ *maps to* x *if and only if*

$$2\sum_{i=1}^{\infty}\int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = \sum_{j=1}^{\infty}\phi(\lambda_{2j})x_j$$

for every integral function ϕ *of class* $I^{3/2}$*; moreover,* x *belongs to the (real) dual space* $I^{3/2\dagger}$ *of* $I^{3/2*}$*, i.e.* $\sum j^{-4}x_j^2 < \infty$.

Proof: The sum $2\sum\int_{\mathbf{o}_i}^{\mathbf{p}_i}\phi(\Delta^2-4)^{-1/2}$ defines a bounded functional on $I^{3/2}$. To see this, let ϕ be the real or imaginary on the line, every function of class $I^{3/2}$ being expressible as the sum of such. Then, by the mean value theorem

$$\int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = \phi(\bar{\mu}_i)\int_{\mathbf{o}_i}^{\mathbf{p}_i}\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}$$

for some intermediate $\bar{\mu}_i \in [\lambda_{2i-1}, \lambda_{2i}]$, and

$$\begin{aligned}
\sum_{i=1}^{\infty}\left|\int_{\mathbf{o}_i}^{\mathbf{p}_i}\phi(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}\right| &\le \sum_{i=1}^{\infty}|\phi(\bar{\mu}_i)| \times 2\left|\int_{\lambda_{2i-1}}^{\lambda_{2i}}\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}\right| \\
&\le \text{constant} \times \sqrt{\sum_{i=1}^{\infty} i^4|\phi(\bar{\mu}_i)|^2} \times \sqrt{\sum_{i=1}^{\infty} i^{-2}} \\
&\le \text{constant} \times \|\phi\|_{3/2},
\end{aligned}$$

in view of Lemma 1 and the fact that $\bar{\mu}_i$, $i = 1, 2, \cdots$, is the tied spectrum of a point of M so that $\sum i^4|\phi(\bar{\mu}_i)|^2$ is comparable to $\|\phi\|_{3/2}^2 = \int_0^{\infty}|\phi(\mu)|^2\mu^{3/2}\,d\mu$, by Theorem 5.3. Therefore,

$$2\sum_{i=1}^{\infty}\int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = \sum_{j=1}^{\infty}\phi(\lambda_{2j})x_j$$

with $x \in I^{3/2\dagger}$, by a second application of Theorem 5.3 with λ_{2j}, $j = 1, 2, \cdots$, as tied spectrum, and x identified as the image of $\mathbf{p}$ by choice of $\phi = \mathbf{1}_j \in I^{3/2}$ for $j = 1, 2, \cdots$, in turn.

THEOREM 2. *The map* $\mathbf{p} \to x$ *is 1:1.*

Proof: If the map is not 1:1, you can find points $\mathbf{p}'$ and $\mathbf{p}''$ of the torus of Fig. 7.4 such that

$$\sum_{i=1}^{\infty}\int_{\mathbf{p}_i'}^{\mathbf{p}_i''}\phi(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = 0$$

for every $\phi \in I^{3/2}$. Let $\mathbf{p}_i' \neq \mathbf{p}_i''$, $i = 1, 2, \cdots$, for simplicity and fix $m < \infty$. Then

$$\det \int_{\mathbf{p}_i'}^{\mathbf{p}_i''} \mathbf{1}_i(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}}$$
$$= \int_{\mathbf{p}_1'}^{\mathbf{p}_1''} \frac{d\mu_1}{\sqrt{\Delta^2(\mu_1) - 4}} \cdots \int_{\mathbf{p}_m'}^{\mathbf{p}_m''} \frac{d\mu_1}{\sqrt{\Delta^2(\mu_m) - 4}} \det \mathbf{1}_j(\mu_i) \neq 0, \quad 1 \leq i,\, j \leq m,$$

since

$$\begin{aligned} \det \mathbf{1}_j(\mu_i) &= \det \frac{1}{\mu_i - \lambda_{2j}} \times \text{ a non-vanishing factor} \\ &= \frac{\prod_{i>j}(\mu_i - \mu_j) \prod_{i>j}(\lambda_{2i} - \lambda_{2j})}{\prod_{i,j}(\mu_i - \lambda_{2j})} \times \text{ the same factor} \end{aligned}$$

is of one signature for $\mu_i \in [\lambda_{2i-1}, \lambda_{2i}]$, $i = 1, \cdots, m$. This permits you to pick $\phi(\lambda) = \sum_{i=1}^m c_i \ell_i(\lambda) \in I^{3/2}$ with real coefficients, so as to make

$$\int_{\mathbf{p}_i'}^{\mathbf{p}_i''} \phi(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} = \begin{cases} \int_{\mathbf{p}_1'}^{\mathbf{p}_1''} \frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} & \text{for } i = 1, \\ 0 & \text{for } 2 \leq i \leq m. \end{cases}$$

But now $\phi(\bar{\mu}_1) = 1$, $\phi(\bar{\mu}_i) = 0$, $i = 2, 3, \cdots, m$, for some intermediate $\bar{\mu}_i \in [\lambda_{2i-1}, \lambda_{2i}]$, $i = 1, 2, \cdots, m$, while $\phi(\lambda_{2i}) = 0$, $i > m$, so that

$$1 = \sum_{i=1}^{m} i^4 |\phi(\bar{\mu_i})|^2 + \sum_{i>m} i^4 |\phi(\lambda_{2i})|^2$$

and the comparable $\|\phi\|_{3/2}^2$ is bounded independently of m, $\bar{\mu_i}$, $i = 1, \cdots, m$, augmented by λ_{2i}, $i > m$, being the tied spectrum of a point of M. The proof is finished by letting $m \uparrow \infty$ in such a way that the above $\phi = \phi_m$ tends weakly in $I^{3/2}$ to a function ϕ_∞ with

$$\int_{\mathbf{p}_i'}^{\mathbf{p}_i''} \phi_\infty(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} = \begin{cases} \int_{\mathbf{p}_1'}^{\mathbf{p}_1''} \frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} & \text{for } i = 1, \\ 0 & \text{for } i \neq 1, \end{cases}$$

contradicting

$$\sum_{i=1}^{\infty} \int_{\mathbf{p}_i'}^{\mathbf{p}_i''} \phi_\infty(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} = 0.$$

LEMMA 2. *Let X_i be the Hamiltonian vector field $q \to D\partial\Delta(\lambda_{2j})/\partial q$, $j = 1, 2, \cdots$. Then $x \cdot X = \sum x_j X_j$ induces a smooth flow* $\exp\{tx \cdot X\}$ *on M for any choice of $x \in I^{3/2\dagger}$.*

Proof: Let ξ and η be any two tame points of $I^{3/2\dagger}$ and write e^X in place of of $\exp\{X\}$. Then for any $q \in M$ and $0 \leq x < 1$,

$$e^{\xi \cdot X} q(x) - e^{\eta \cdot X} q(x) = \int_0^1 (\xi - \eta) \cdot X e^{\zeta \cdot X} q(x)\, dc$$

with $\zeta = \eta + c(\xi - \eta)$. Let $\|\ \|_i$ be any Sobolev norm. Then

$$\|e^{\xi \cdot X} q - e^{\eta \cdot X} q\|_i \leq \int_0^1 dc \sum_{j=1}^{\infty} |\xi_i - \eta_i| |\Delta^\bullet(\lambda_{2j})| \|Df_{2j}^2\|_i$$

with f_{2j} evaluated at the point $e^{\zeta \cdot X} q$ of M. But now $j^{-1-i}\|Df_{2j}^2\|_i$ is controlled by $\|e^{\zeta \cdot X} q\|_{i-1}$ and hence[29] by $H_0, \cdots, H_i$, independently of ζ and q, while $\Delta^\bullet(\lambda_{2j})$ is rapidly decreasing as

[29] See [22], p. 147 and [28], p. 226.

$j \uparrow \infty$, so

$$\|e^{\xi \cdot X} q - e^{\eta \cdot X} q\|_i \le c_i(M) \|\xi - \eta\|_{3/2\dagger}$$

with a constant depending only upon i and M.

Now approximate $x \in I^{3/2\dagger}$ by tame $y \in I^{3/2\dagger}$. Then by the above estimates, $p = \exp\{ty \cdot Y\}q$ converges *in the topology of infinitely differentiable functions from* C_1^∞ *to* C_1^∞ to a point $p_\infty \in C_1^\infty$. Moreover,

$$p_\infty - q = \lim \int_0^t y \cdot Xp \, dt' = \int_0^t x \cdot Xp_\infty \, dt' :$$

thus p_∞ is the unique smooth solution of $\partial p_\infty / \partial t = x \cdot X p_\infty$ reducing to q at $t = 0$. This proves that $x \cdot X$ induces a flow in C_1^∞ which is once differentiable in t.

The proof is finished.

THEOREM 3. *Extend the map* $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \cdots) \to x \in I^{3/2\dagger}$ *by permitting such paths of integration* $\mathbf{o}_i$, $\mathbf{p}_i$, $i = 1, 2, \cdots$, *as make*

$$\sum \int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi \, (\Delta^2 - 4)^{-1/2}$$

convergent in the dual space of $I_{3/2}$. *Then the extended map is onto* $I^{3/2\dagger}$ *and is inverted by application of* $\exp\{x \cdot X\}$ *to the origin of* M, $x \cdot X = \sum x_j X_j$ *being formed with the image* $x \in I^{3/2\dagger}$ *of* $\mathbf{p}$.

Proof: Fix $x \in I^{3/2\dagger}$. Then under the flow $\partial q / \partial t = x \cdot Xq$, the tied spectrum moves according to the rule

$$\frac{2\mu_i^\bullet}{\sqrt{\Delta^2(\mu_i) - 4}} = \sum_{j=1}^{\infty} x_j \prod_{k \ne i} \frac{1 - \lambda_{2j}/\mu_k}{1 - \mu_i/\mu_k};$$

see Amplification 4.1. Let the flow carry $\mathbf{o} = (\mathbf{o}_1, \mathbf{o}_2, \cdots)$ to $\mathbf{p}$ in unit time and fix $\phi \in I^{3/2}$. Then

$$2 \sum_{i=1}^{m} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} = \sum_{j=1}^{\infty} x_j \int_0^1 \left[\sum_{i=1}^{m} \phi(\mu_i) \prod_{k \ne i} \frac{1 - \lambda_{2i}/\mu_k}{1 - \mu_i/\mu_k} \right] dt,$$

in which μ_i, $i = 1, 2, \cdots$, depends upon $0 \le t \le 1$. The inner sum tends to $\phi(\lambda_{2i})$ as $m \uparrow \infty$ in $I^{3/2*}$, uniformly in $0 \le t \le 1$; hence the functional

$$2 \sum_{i=1}^{m} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi (\Delta^2 - 4)^{-1/2}$$

converges in the dual space $I^{3/2\dagger}$ to

$$\begin{aligned} 2 \sum_{i=1}^{\infty} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 2}} &= \sum_{j=1}^{\infty} \phi(\lambda_{2j}) \, dt \\ &= \sum_{j=1}^{\infty} \phi(\lambda_{2j}) x_j. \end{aligned}$$

The proof is finished.

THEOREM 4. *Let the primitive periods* $\omega_i \in I^{3/2\dagger}$ *be defined by the rule*

$$\omega_{ij} = 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_j(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}}, \quad j = 1, 2, \cdots .$$

Then ω_i, $i = 1, 2, \cdots$, *is a basis of* $I^{3/2\dagger}$*: every point* $x \in I^{3/2\dagger}$ *can be uniquely written as* $\sum y_i \omega_i$ *with* $\sum i^{-2} y_i^2 < \infty$, *the sum being comparable to* $\sum i^4 x_i^2 = \|x\|_{3/2\dagger}^2$.

Proof: Let $\phi \in I^{3/2}$ take real values on the line and pick intermediate $\bar{\mu_i} \in [\lambda_{2i-1}, \lambda_{2i}]$ so that

$$\int_{\lambda_{2i-1}}^{\lambda_{2i}} \phi(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = \phi(\bar{\mu_i}) \int_{\lambda_{2i-1}}^{\lambda_{2i}} \frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}, \quad i = 1, 2, \cdots .$$

Then $\bar{\mu_i}$, $i = 1, 2, \cdots$, is the tied spectrum of a point of M, so $\sum i^4 |\phi_(\bar{\mu_i})|^2$ is comparable to

$$\|\phi\|_{3/2}^2 = \int_0^\infty \|\phi(\lambda)|^2 \lambda^{3/2} \, d\lambda.$$

Now think of $x \in I^{3/2\dagger}$ as an element of the dual space of $I^{3/2}$ $(x(\phi) = \sum \phi(\lambda_{2j}) x_{2j})$. Then $x = \sum y_i \omega_i$ is convergent in $I^{3/2\dagger}$ if and only if

$$\begin{aligned} \mathfrak{x}(\phi) = \sum y_i \omega_i(\phi) &= \sum y_i 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \phi(\Delta^2-4)^{-1/2} \\ &= \sum y_i 4\phi(\bar{\mu_i}) \int_{\lambda_{2i-1}}^{\lambda_{2i}} (\Delta^2-4)^{-1/2} \end{aligned}$$

is convergent for every $\phi \in I^{3/2}$, i.e., if and only if $\sum i^{-2} y_i^2 < \infty$, the point being that

$$\int_{\lambda_{2i-1}}^{\lambda_{2i}} (\Delta^2-4)^{-1/2}$$

is comparable to i and $\|x\|^{3/2\dagger}$ is comparable to $\sum i^{-2} y_i^2$. The proof of the uniqueness of the expansion $\sum y_i \omega_i$ proceeds as in Theorem 2: if $\sum y_i \omega_i = 0$, then you can pick $\phi \in I^{3/2}$ so that $\omega_i(\phi) = 0$, $i \neq j$, and $\omega_j(\phi) = 1$, whence $y_j = 0$, $j = 1, 2, \cdots$. The proof is finished by observing that ω_i, $i = 1, 2, \cdots$, spans $I^{3/2\dagger}$: if not, you could find non-trivial $\phi \in I^{3/2}$ and $\bar{\mu_i} \in [\lambda_{2i-1}, \lambda_{2i}]$, $i = 1, 2, \cdots$, as above so that

$$0 = \omega_i(\phi) = 4\phi(\bar{\mu_i}) \int_{\lambda_{2i}} (\Delta^2-4)^{-1/2}, \quad i = 1, 2, \cdots,$$

contradicting the fact that $\sum i^4 |\phi(\bar{\mu_i})|^2$ is comparable to $\|\phi\|_{3/2}^2 > 0$.

Now let the extended map $\mathbf{p} \to x$ of Theorem 3 be expressed as $\mathbf{p} \to \sum y_i \omega_i = \sum (e_i + n_i)\omega_i$ with $0 \le e_i < 1$ and integral n_i, $i = 1, 2, \cdots$, and let n_i' be the number of times the path of integration $\mathbf{o}_i$, $\mathbf{p}_i$ winds clockwise about the i-th circle. Then

$$\begin{aligned} x(\phi) &= \sum_{i=1}^\infty 2 \int_{\mathbf{o}_i}^{\mathbf{p}_i'} \phi(\Delta^2-4)^{-1/2} + \sum_{i=1}^\infty n_i' \omega_i(\phi) \\ &= \sum_{i=1}^\infty e_i \omega_i(\phi) + \sum_{i=1}^\infty n_i \omega_i(\phi), \end{aligned}$$

in which the path $\mathbf{o}_i$, $\mathbf{p}_i'$ extends clockwise less than once around the i-th circle. Now

$$2 \int_{\mathbf{o}_i}^{\mathbf{p}_i'} \phi(\Delta^2-4)^{-1/2} = \sum_{j=1}^\infty \epsilon_{ij}' \omega_j(\phi),$$

and it is easy to prove, as in Theorem 2, that $e_{ij}' = 0$, $i \neq j$. Therefore ,

$$2 \int_{\mathbf{o}_i}^{\mathbf{p}_i'} \phi(\Delta^2-4)^{-1/2} = e_i' \omega_i(\phi)$$

for every $\phi \in I^{3/2}$, and by choice of $\phi(\lambda) = (\lambda - \lambda_1)^{-1} y_2(1, \lambda)$ which is one signature in $[\lambda_{2i-1}, \lambda_{2i}]$, you see that $0 \le e'_i < 1$. Therefore,

$$x(\phi) = \sum_{i=1}^{\infty} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\Delta^2 - 4)^{-1/2} = \sum_{i=1}^{\infty} (e'_i + n_i)\omega_i,$$

and you identify $e'_i = e_i$ and $n'_i = n_i$, $i = 1, 2, \cdots$. This proves almost all of

THEOREM 5. *Let L be the lattice of periods $\omega \in I^{3/2\dagger}$ such that* $\exp\{\omega \cdot X\}$ *acts trivially on M. Then $\omega \in L$ if and only if $\omega = \sum n_i \omega_i$ with integral n_i, $i = 1, 2, \cdots$, subject to $\sum i^{-2} n_i^2 < \infty$. The extended map $\mathbf{p} \to x$ is well-defined precisely for such winding numbers n_i, $i = 1, 2, \cdots$, effecting a 1:1 topological map of M onto the factor space $I^{3/2\dagger}/L$. The latter is compact and may be identified with the fundamental cell $[0, 1)^\infty$ with edge identifications by means of the expansion of $x \in I^{3/2\dagger}$ as $x = \sum(e_i + n_i)\omega_i$ with $0 \le e_i < 1$ and integral n_i. The map $\mathbf{p} \to x$ is ambiguous owing to the indeterminacy of the winding numbers n_i, $i = 1, 2, \cdots$, but the ambiguity is removed by reduction to e.*

Proof: $\omega \in I^{3/2\dagger}$ belongs to the period lattice if and only if all the paths $\mathbf{o}_i$, $\mathbf{p}_i$ are trivial, i.e., if and only if they consist of an integral number of full revolutions. Then $\omega = \sum n_i \omega_i$, the sum being convergent in $I^{3/2\dagger}$ and $\sum i^{-2} n_i^2 < \infty$, and conversely. The rest of the proof is just as easy. The only point requiring comment is the compactness of the factor space $I^{3/2\dagger}/L$. This follows easily from $\sum i^2 |\omega_i(\phi)|^2 \le$ constant $\times \|\phi\|_{3/2}^2$, which permits you to bound the tail of the sum $\sum e_i \omega_i$ independently of $0 \le e_i < 1$, $i = 1, 2, \cdots$. The fact that the map $M \to I^{3/2\dagger}/L$ is topological is plain.

AMPLIFICATION 1. The fact that ω_i, $i = 1, 2, \cdots$, spans $I^{3/2\dagger}$ is the same as to say that a differential of the first kind $\phi(\Delta^2 - 4)^{-1/2}\, d\mu$ with $\phi \in I^{3/2}$ is completely determined by its periods

$$\omega_i(\phi) = 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \phi(\Delta^2 - 1)^{-1/2}, \quad i = 1, 2, \cdots.$$

The analogy with classical function theory is perfect.

AMPLIFICATION 2. The exponential map

$$x \to \exp\{x \cdot X\} \text{ applied to the origin of } M$$

carries $I^{3/2\dagger}$ *onto* M, permitting you to view $I^{3/2\dagger}$ as a common tangent space to $I^{3/2\dagger}/L$ and M. The map is *not* 1:1 in any ball of $I^{3/2\dagger}$ centered at $x = 0$, unlike the exponential map of finite-dimensional differential geometry. The reason is that every ball contains infinitely many primitive periods in view of $\sum i^2 |\omega_i(\phi)|^2 \le$ constant $\times \|\phi\|_{3/2}^2$. Actually, this is fortunate: if the map had been 1:1, it would have effected a topological map of a *non-compact* closed ball in $I^{3/2\dagger}$ into a *compact* piece of M!

AMPLIFICATION 3. x is now regarded as a global coordinate on M and the relation of $\mathbf{p}$ to x is interpreted as saying that $M = I^{3/2\dagger}/L$ is the real part of the Jacobi variety of the Riemann surface S of $\sqrt{\Delta^2(\lambda) - 4}$. The same state of affairs is epitomized by the vector field identity

$$2 \sum_{i=1}^{\infty} \left(D \frac{\partial \Delta}{\partial q} \text{ evaluated at } \lambda = \mu_i \right) \times \frac{d\mu_i}{\sqrt{\Delta^2(\mu_i) - 4}} = \sum_{j=1}^{\infty} \left(D \frac{\partial \Delta}{\partial q} \text{ evaluated at } \lambda_{2j} \right) \times dx_j.$$

The relation may also be expressed somewhat more vaguely by means of the addition theorem

$$\int_{\mathbf{o}}^{\mathbf{p}'} \phi(\Delta^2 - 4)^{-1/2} + \int_{\mathbf{o}}^{\mathbf{p}''} \phi(\Delta^2 - 4)^{-1/2} = \int_{\mathbf{o}}^{\mathbf{p}} \phi(\Delta^2 - 4)^{-1/2},$$

in which $\phi \in I^{3/2}$,

$$\int_{\mathbf{o}}^{\mathbf{p}} = \sum \int_{\mathbf{o}_i}^{\mathbf{p}_i},$$

and the theorem asserts the existence of $\mathbf{p}$ and appropriate paths of integration on the right-hand side for every choice of $\mathbf{p}', \mathbf{p}''$, and permissible paths of integration on the left.

The next few sections contain additional information about these matters.

7.12. Volume elements

The line element identity

$$\frac{2y_2^\bullet(1, \mu_i)}{\sqrt{\Delta^2(\mu_i) - 4}}\, d\mu_i = \sum_{j=1}^{\infty} \frac{y_2(1, \lambda_{2j})}{\lambda_{2j} - \mu_i}\, dx_i$$

is easily verified by putting $\phi(\lambda) = (\lambda - \mu_i)^{-1} y_2(1, \lambda)$ in the differential form of the Jacobi map, $y_2(1, \lambda)$ being computed for the actual point q of M in hand. This leads to entertaining formulas of volume elements. The plan is to compute the formal wedge product of the differentials

$$\frac{2y_2^\bullet(1, \mu_i)}{\sqrt{\Delta^2(\mu_i)}}\, d\mu_i, \quad i = 1, 2, \cdots,$$

by means of the line element identity and the evaluation

$$\det \frac{1}{a_i - b_j} = \frac{\prod_{i>j}(a_i - a_j) \prod_{i>j}(b_i - b_j)}{\prod_{i,\,j}(a_i - b_j)}.$$

The latter is employed freely even though i and j run from 1 to ∞. The computation is elementary: the left-hand side produces

$$\prod_i \frac{2y_2^\bullet(1, \mu_i)}{\sqrt{\Delta^2(\mu_i) - 4}}\, d\mu_i = \prod_i \frac{2\, d\mu_i}{\sqrt{\Delta^2(\mu_i) - 4}} \prod_{j \neq i} \left(1 - \frac{\mu_i}{\mu_j}\right) \left(-\frac{1}{\mu_i}\right),$$

while the right-hand side produces

$$\prod_j y_2(1, \lambda_{2j}) dx_j \times \det \frac{1}{\lambda_{2j} - \mu_i}$$
$$= \prod_j dx_j \prod_i \left(1 - \frac{\lambda_{2j}}{\mu_i}\right) \times \frac{\prod_{i>j}(\lambda_{2i} - \lambda_{2j}) \prod_{i>j}(\mu_i - \mu_j)}{\prod_{i,j}(\lambda_{2j} - \mu_i)},$$

and if you carry out all the (formal) cancellations, you will find

$$\prod_{i>j} \frac{\mu_i - \mu_j}{\lambda_{2i} - \lambda_{2j}} \times \prod_i \frac{2\, d\mu_i}{\sqrt{\Delta^2(\mu_i) - 4}} = \prod_j dx_j.$$

Notice that the left-hand side is one of one signature and that the "density"

$$f(\mu_1, \mu_2, \cdots) = \prod_{i>j} \frac{\mu_i - \mu_j}{\lambda_{2i} - \lambda_{2j}}$$

is bounded from 0 and ∞ because

$$\begin{aligned}\sum_{i>j}\left|1-\frac{\mu_i-\mu_j}{\lambda_{2i}-\lambda_{2j}}\right| &\le \sum_{i>j}\frac{\lambda_{2i}-\lambda_{2i-1}+\lambda_{2j}-\lambda_{2j-1}}{\lambda_{2i}-\lambda_{2j}}\\ &\le \text{constant}\times\sum_{i=1}^{\infty}(\lambda_{2i}-\lambda_{2i-1})\sum_{i\ne j}\frac{1}{|i^2-j^2|}\end{aligned}$$

is bounded independently of μ_i, $i = 1, 2, \cdots$.

The volume of a finite-dimensional torus is the determinant of the primitive periods defining the associate lattice. The primitive periods ω_i, $i = 1, 2, \cdots$, of the Jacobi variety are expressed by

$$\omega_{ij} = 4\int_{\lambda_{2i-1}}^{\lambda_{2i}}\frac{y_1'(1,\mu)}{y_1'^{\bullet}(1,\lambda_{2j})}\frac{1}{\mu-\lambda_{2j}}\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}$$

with $y_1'(1,\lambda)$ computed for the origin of M, and $\det\omega_{ij}$ may be computed in the same formal way:

$$\begin{aligned}\det\omega_{ij} &= \prod_i 4\int_{\lambda_{2i-1}}^{\lambda_{2i}}\frac{y_1'(1,\mu_i)}{\sqrt{\Delta^2(\mu_i)-4}}\,d\mu_i\times\prod_j\frac{1}{y_1'^{\bullet}(1,\lambda_{2j})}\times\det\frac{1}{\mu_i-\lambda_{2j}}\\ &= \prod_i 4\int_{\lambda_{2i-1}}^{\lambda_{2i}}\prod_j\left(1-\frac{\mu_i}{\lambda_{2j}}\right)\frac{d\mu_i}{\sqrt{\Delta^2(\mu_i)-4}}\times\prod_j\prod_{i\ne j}\frac{1}{(1-\lambda_{2j}/\lambda_{2i})(-1/\lambda_{2i})}\\ &\quad\times\frac{\prod_{i>j}(\mu_i-\mu_j)\prod_{i>j}(\lambda_{2i}-\lambda_{2j})}{\prod_{i,j}(\mu_i-\lambda_{2j})}\\ &= \int_{\lambda_1}^{\lambda_2}\int_{\lambda_3}^{\lambda_4}\int_{\lambda_5}^{\lambda_6}\cdots f(\mu_1,\mu_2,\cdots)\prod_i\frac{4\,d\mu_i}{\sqrt{\Delta^2(\mu_i)-4}}.\end{aligned}$$

This is precisely the normalization constant needed in the preceding formula, the upshot being that

$$\frac{f(\mu_1,\mu_2,\cdots)\prod_i\frac{2\,d\mu_i}{\sqrt{\Delta^2(\mu_i)-4}}}{\int_{\lambda_1}^{\lambda_2}\int_{\lambda_3}^{\lambda_4}\cdots f(\mu_1,\mu_2\cdots)\prod_i\frac{4d\mu_i}{\sqrt{\Delta^2(\mu_i)-4}}} = \frac{\prod_j dx_j}{\det(\omega_{ij})}$$

is the normalized invariant volume element of M viewed as $I^{3/2\dagger}/L$. Notice that the right-hand side is formally the same as the normalized volume element $de = \prod_i de_i$ in view of $dx_j = \sum_i de_i\omega_{ij}$.

Naturally, these results have only a formal status until verified in a more convincing manner, but as they are not needed below it is unnecessary to do so.

7.13. An Imbedding

Consider the map

$$e \to x = \sum e_i\omega_i \to X = \sum x_jX_j \to q = e^X(\text{origin})$$

of the standard torus, $[0,1)^\infty$ with edge identifications, onto $M \subset L_1^2$. The purpose of this section is to prove that *this map is an imbedding of the torus into L_1^2 in the following sense: (i) the map is topological, (ii) $q(x)$, $0 \le x < 1$, has indefinitely many tame partials with regard to e_i, $i = 1, 2, \cdots$, (iii) the differential map from the natural tangent vectors $\partial/\partial e_i$, $i = 1, 2, \cdots$, of the torus to the tangent space T is non-singular at every point.* (iii) needs to be spelled out. The differential map carries $\partial/\partial e_i$ to $\partial q/\partial e_i \in T$; it is to be proved that these vectors comprise an oblique base for T, meaning that every $Xq \in T$ can be uniquely written as

$$Xq = \sum_{i=1}^{\infty} c_i \ell_i^{-1} \frac{\partial q}{\partial e_i}, \quad \ell_i^2 = \int_0^1 \left| \frac{\partial q}{\partial e_i} \right|^2 dx,$$

and that

$$\int_0^1 |Xq|^2 \, dx$$

is comparable to $\sum c_i^2$. (i) is known from Section 11, so the next step is the proof *that* $e \to q$ *is differentiable*. Pick $x \in I^{3/2\dagger}$ and expand it as $\sum e_i \omega_i$. Then $\partial x/\partial e_i = \omega_i$, $\partial X/\partial e_i = \sum \omega_{ij} X_j$, and

$$\begin{aligned} \frac{\partial q}{\partial e_i} &= \frac{\partial}{\partial e_i} e^X (\text{ origin }) \\ &= \frac{\partial X}{\partial e_i} q \\ &= \sum_{j=1}^{\infty} \omega_{ij} X_j q \\ &= \sum_{j=1}^{\infty} 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_j (\Delta^2 - 4)^{-1/2} \, d\mu \, X_j q. \end{aligned}$$

This formal computation may be used to confirm the existence of $\partial q/\partial e_i$, but for the discussion of the higher partials it is better to bring the final sum into a more compact form. One would like to pull the sum inside the integral and interpolate to obtain

$$\begin{aligned} \frac{\partial q}{\partial e_i} &= 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \left[\sum_{j=1}^{\infty} \mathbf{1}_j D \frac{\partial \Delta(\lambda_{2j})}{\partial q} \right] \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} \\ &= 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} D \frac{\partial \Delta}{\partial q} \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} \end{aligned}$$

but that is not legitimate since $D(\partial \Delta/\partial q)$, *qua* function of μ, does not belong to $I^{1/2}$. However, it *does* belong to $I^{-1/2}$, so

$$\begin{aligned} \sum_{j=1}^{\infty} \omega_{ij} \frac{\partial \Delta}{\partial q}(\lambda_{2j}) &= \sum_{j=1}^{\infty} 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_j (\Delta^2 - 4)^{-1/2} d\mu \frac{\partial \Delta}{\partial q}(\lambda_{2j}) \\ &= 4 \sum_{j=1}^{\infty} \int_{\lambda_{2i-1}}^{\lambda_{2i}} \left[\frac{1}{\mu - \lambda_{2j}} - \frac{1}{\mu - \lambda_0} \right] \frac{y_1'(1,\mu)}{y_1'^{\bullet}(1,\lambda_{2j})} \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} \frac{\partial \Delta}{\partial q}(\lambda_{2j}) \\ &= 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \left[\frac{\partial \Delta}{\partial q} + \frac{y_1'(1,\mu)}{\mu - \lambda_0} \right] \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}}. \end{aligned}$$

The rapid decrease of the factor $\Delta^{\bullet}(\lambda_{2j})$ in $\partial \Delta(\lambda_{2j})/\partial q = -\Delta^{\bullet}(\lambda_{2j}) f_{2j}^2$ permits term wise differentiation of the sum with regard to $0 \le x < 1$, and the stated formula for $\partial q/\partial e_i$ drops out. The rest of the proof will be plain.

Proof *that the differential map is non-singular.* $j^{-1}(f_{2j}^2)'$, $j = 1, 2, \cdots$, is an oblique base for T, by Theorem 8. Now

$$\frac{\partial q}{\partial e_i} = \sum_{j=1}^{\infty} \omega_{ij} X_j q = -\sum_{j=1}^{\infty} \omega_{ij} j \Delta^{\bullet}(\lambda_{2j}) [j^{-1}(f_{2j}^2)'];$$

thus

$$\ell_i^2 = \int_0^1 \left| \frac{\partial q}{\partial e_i} \right|^2 dx$$

is comparable to

$$\sum_{j=1}^{\infty} \omega_{ij}^2 j^2 |\Delta^\bullet(\lambda_{2j})|^2,$$

by the same theorem, and it is required to prove that the matrix

$$\left[\frac{\omega_{ij} j \Delta^\bullet(\lambda_{2i})}{\sqrt{\sum_{j=1}^{\infty} \omega_{ij}^2 j^2 |\Delta^\bullet(\lambda_{2j})|^2}} \right], \quad i, j = 1, 2, \cdots,$$

regarded as an operator in the Hilbert space of the quadratic form $\sum c_i^2$ is bounded, 1:1, onto, and so boundedly invertible. The proof occupies the next three lemmas.

LEMMA 1.

$$\sum_{j=1}^{\infty} \omega_{ij}^2 j^2 |\Delta^\bullet(\lambda_{2j})|^2$$

is comparable to $\omega_{ii}^2 i^2 |\Delta^\bullet(\lambda_{2i})|^2$ *as* $i \to \infty$.

Proof: The lemma states that the (off-diagonal) sum for $j \neq i$ is majored by the (diagonal) summand for $j = i$. Now

$$|\omega_{ij}| = 4 \left| \int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_j (\Delta^2 - 4)^{-1/2} \right| \geq \text{constant}| \times \left| \int_{\lambda_{2i-1}}^{\lambda_{2i}} (\Delta^2 - 4)^{-1/2} \right|$$

is comparable to i, by Lemma 1.1, so that the diagonal summand is comparable to $|\lambda_{2i} - \lambda_{2i-1}|^2$, by Lemma 10.1. The estimate of the off-diagonal sum is just as easy: if $j \neq i$ and if $\lambda_{2i-1} \leq \mu \leq \lambda_{2i}$, then

$$|\mathbf{1}_j(\mu)| = \left| \frac{y_1'(1,\mu)}{\mu - \lambda_0} \frac{\lambda_{2j} - \lambda_0}{y_1'^\bullet(1, \lambda_{2j})} \frac{1}{\mu - \lambda_{2j}} \right| \leq \lambda_{2i} - \lambda_{2i-1} \times \frac{j^2}{i^2},$$

so

$$\begin{aligned} |\omega_{ij}| &\leq 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} \left| \mathbf{1}_j (\Delta^2 - 4)^{-1/2} \right| \\ &\leq \text{constant} \times \lambda_{2i} - \lambda_{2i-1} \times \frac{j^2}{i^2} \times i, \end{aligned}$$

and the sum is over-estimated by a constant multiple of

$$\sum_{j \neq i} \frac{(\lambda_{2i} - \lambda_{2i-1})^2}{i^2} \times j^6 \times \frac{(\lambda_{2j} - \lambda_{2j-1})^2}{j^4} \leq \text{constant} \times i^{-2} (\lambda_{2i} - \lambda_{2i-1})^2,$$

which is plenty.

The estimate of Lemma 1 implies that is suffices to deal with the modified matrix

$$(J_{ij}) = \left[\frac{\omega_{ij}}{\omega_{ii}} \frac{j}{i} \frac{\Delta^\bullet(\lambda_{2j})}{\Delta^\bullet(\lambda_{2i})} \right], \qquad i, j = 1, 2, \cdots.$$

LEMMA 2. *J is of the form: identity + compact.*

Proof: The lemma is proved by noting that $J-1$ is of Hilbert-Schmidt class: $J_{ii}=1$, and by the estimate of the off-diagonal sum in the proof of Lemma 1,

$$\begin{aligned}\sum_{i=1}^{\infty}\sum_{j\neq i}^{\infty} J_{ij}^2 &= \sum_{i=1}^{\infty}[\omega_{ii} i\Delta^\bullet(\lambda_{2i})]^{-2}\sum_{j\neq i}[\omega_{ij} j\Delta^\bullet(\lambda_{2j})]^2 \\ &\leq \text{constant}\times\sum_{i=1}^{\infty}\frac{i^{-2}(\lambda_{2i}-\lambda_{2i-1})^2}{i^4\times i^{-4}(\lambda_{2i}-\lambda_{2i-1})^2} \\ &= \text{constant}\times\sum_{i=1}^{\infty} i^{-2} \\ &< \infty.\end{aligned}$$

Lemma 2 implies that J is 1:1, onto, and boundedly invertible as soon as it is 1:1, so the proof is finished by proving this.

LEMMA 3. *J is 1:1.*

Proof: Pick c_j, $j=1,2,\cdots$, with $\sum c_j^2<\infty$ so as to make $\sum J_{ij}c_j=0$ for every $i=1,2,\cdots$. The condition states that

$$0=\sum_{j=1}^{\infty}\omega_{ij} j\Delta^\bullet(\lambda_{2j})c_j = 4\int_{\lambda_{2i-1}}^{\lambda_{2i}}\phi(\Delta^2-4)^{-1/2}, \quad j=1,2,\cdots,$$

with an integral function

$$\phi(\lambda)=\sum_{j=1}^{\infty}\mathbf{1}_j(\lambda) j\Delta^\bullet(\lambda_{2j})c_j\in I^{3/2},$$

as you know from Amplification 11.2 that such functions must have nontrivial periods or vanish identically. Therefore, $0=[j\Delta^\bullet(\lambda_{2j})]^{-1}\phi(\lambda_{2j})=c_j$, $j=1,2,\cdots$, as required.

7.14. The Hierarchy of Tangent Spaces for a Purely Simple Spectrum

The point $x\in I^{3/2\dagger}$ gives rise to the vector field $x\cdot X=\sum x_jX_j$ permitting the interpretation of $I^{3/2\dagger}$ as a tangent space to M, as in Amplification 11.2. Unfortunately, this space does not contain any of the local Hamiltonian fields $V_i=D\partial H_i/\partial q$, $i=1,2,\cdots$, the study of which the Jacobi variety is supposed to facilitate; in fact

$$V_i=\sum_{j=1}^{\infty}p_i(\lambda_{2j})[y_1'^\bullet(1,\lambda_{2j})]^{-1}X_j$$

with a polynomial p_i of precise degree i (see Section 10), and the corresponding point x with coordinates

$$x_j=p_i(\lambda_{2j})[y_1'^\bullet(1,\lambda_{2j})]^{-1}\sim\text{constant}\times j^{2i}$$

never belongs to $I^{3/2\dagger}$ for $i=1$ or more.

The remedy is not far off. Define $I^{m/2\dagger}$ as the Hilbert space for the quadratic form $\sum j^{-m-1}x_j^2<\infty$ for odd $m=3,5,7,\cdots$. Then V_i belongs to $I^{m/2\dagger}$ if and only if $m>4i$, and thus $V_1\in I^{5/2\dagger}$, $V_2\in I^{9/2}\dagger$,etc., with a self-explanatory abuse of notation. In this way you obtain a hierarchy of tangent spaces

$$I^{3/2\dagger}\subset I^{5/2\dagger}\subset I^{7/2\dagger}\cdots\subset I^{m/2\dagger}\subset T,$$

$I^{\infty/2\dagger}$ being the union of $I^{m/2\dagger}$ for $m=3,5,7,\cdots$. The exponential map $x\to\exp\{x\cdot X\}$ provides a topological map of $I^{m/2\dagger}$ onto M as before, and M can be identified with the factor space $I^{m/2\dagger}/L^{m/2}$ in which $L^{m/2}$ is the lattice of periods $\omega \in I^{m/2\dagger}$ such that $\exp\{\omega \cdot X\}$ acts as the identity on N. The proof is the same as for $m = 3$.

AMPLIFICATION 1. The inclusion $I^{\infty/2\dagger} \subset T$ requires some explanation. $T \subset L_1^2$ is the space of space of tangent vectors Xq to M at q expressible as $\sum_{i=1}^{\infty} c_i[-\sqrt{2}(\pi i)^{-1} f_{2i} f'_{2i}]$ with

$$\int_0^1 |Xq|^2\,dx$$

comparable to $\sum c_i^2$. Now you may write $-\sqrt{2}(\pi i)^{-1} f_{2i} f'_{2i}$ as $[\sqrt{2}\pi i \Delta^\bullet(\lambda_{2i})]^{-1} X_i q$ and so express Xq as $\sum x_i X_i q$ with $x_i = [\sqrt{2}\pi i \Delta^\bullet(\lambda_{2i})]^{-1} c_i$, $i = 1, 2, \cdots$. The upshot is that T may be visualized as the Hilbert space for the quadratic form $\sum i^2 |\Delta^\bullet(\lambda_{2i})|^2 x_i^2$, thus placing it on the same footing as $I^{m/2\dagger}$, $m = 3, 5, 7, \cdots$.

THEOREM 1. *$\omega \in L^{m/2}$ if and only if it can be written as $\sum n_i \omega_i$ with the primitive periods of Theorem 11.4 and integral n_i, $i = 1, 2, \cdots$, subject to $\sum i^{-m+1} n_i^2 < \infty$.*

Proof: The case $m = 3$ was treated before; thus you may take $m = 5$ or more. First, you prove that $\omega = \sum n_i \omega_i$ is convergent in $I^{m/2\dagger}$ if and only if and only if $\sum i^{-m+1} n_i^2 < \infty$, precisely as in Theorem 11.4, and you prove that $\exp\{\omega \cdot X\}$ acts trivially on M, the moral being that $\omega \in L^{m/2}$. Now let x be any point of $L^{m/2}$. Then, the flow $\exp\{tx \cdot X\}$ is of period 1, and during this motion

$$\frac{2\mu_i^\bullet}{\sqrt{\Delta^2(\mu_i) - 4}} = \sum_{j=1}^{\infty} x_j \prod_{k \neq i} \frac{1 - \lambda_{2j}/\mu_k}{1 - \mu_i/\mu_k},$$

as in the proof of Theorem 11.3. The result of the motion during a full period $0 \le t < 1$ is that $\mathbf{p}_i = (\mu_i, \sqrt{\Delta^2(\mu_i) - 4})$ makes an integral number n_i of full revolution about the i-th circle, so

$$n_i 4 \int_{\lambda_{2i-1}}^{\lambda_{2i}} (\Delta^2 - 4)^{-1/2} = \sum_{j=1}^{\infty} \int_0^1 \left[\prod_{k \neq i} \frac{1 - \lambda_{2j}/\mu_k}{1 - \mu_i/\mu_k} \right] dt \times x_j$$

for $i = 1, 2, \cdots$. The proof is finished by two simple lemmas.

LEMMA 1. $\sum i^{-m+1} n_i^2 < \infty$.

Proof: The product inside the right-hand integral is expressed as $y_2(1, \lambda_{2j})[(\lambda_{2j} - \mu_i) y_2^\bullet(1, \mu_i)]^{-1}$, the function $y_2(1, \lambda)$ being computed for the actual point of M at hand. Therefore,

$$\int_0^1 \left[\prod_{k \neq i} \frac{1 - \lambda_{2j}/\mu_k}{1 - \mu_i/\mu_k} \right] dt = \begin{cases} O(1) & \text{for } j = 1, \\ O(\lambda_{2j} - \lambda_{2j-1}) \times i^2/j^2 & \text{for } j \neq i, \end{cases}$$

by a self-evident appraisal; hence the modulus of

$$4 n_i \int_{\lambda_{2i-1}}^{\lambda_{2i}} (\Delta^2 - 1)^{-1/2}$$

cannot exceed a constant multiple of

$$|x_i| + \sum_{j \neq i} (\lambda_{2j} - \lambda_{2j-1}) \frac{i^2}{j^2} |x_j| = O(|x_i| + i^2),$$

and

$$\sum i^{-m+1} n_i^2 \le O(1) \sum i^{-m-1} (|x_i|^2 + i^4) < \infty,$$

m being 5 or more.

LEMMA 2. $x = \sum n_i \omega_i$.

Proof: Let $y_2(1,\lambda)$ be computed for the actual point of M in hand, as in the proof of Lemma 1. Then

$$\begin{aligned} n_i\omega_{ik} &= 4n_i\int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_k(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} \\ &= \sum_{i=1}^{\infty}\int_0^1\left[\frac{y_2(1,\lambda_{2j})}{(\lambda_{2j}-\mu_i)y_2^{\bullet}(1,\mu_i)}\mathbf{1}_k(\mu_i)\right] dt \times x_j. \end{aligned}$$

This is to be summed over i and the latter sum is to be interchanged with the sum over k and with the integral to obtain the desired result:

$$\begin{aligned} \sum_{i=1}^{\infty} n_i\omega_{ij} &= \sum_{j=1}^{\infty}\int_0^1\left[\sum_{i=1}^{\infty}\frac{y_2(1,\lambda_{2j})}{(\lambda_{2j}-\mu_i)y_2^{\bullet}(1,\mu_i)}\mathbf{1}_k(\mu_i)\right] dt \times x_j \\ &= \sum_{j=1}^{\infty}\mathbf{1}_k(\lambda_{2j})x_j \\ &= x_k. \end{aligned}$$

The only point at issue is the interchange, and that can be settled directly:

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\int_0^1\left|\frac{y_2(1,\lambda_{2j})}{(\lambda_{2j}-\mu_i)y_2^{\bullet}(1,\mu_i)}\mathbf{1}_k(\mu_i)\right| dt\,|x_j|$$

$$\le O(1)\times\sum_{i=1}^{\infty}\left[|x_i|+\sum_{j\neq i}(\lambda_{2j}-\lambda_{2j-1})\frac{i^2}{j^2}(\lambda_{2i}-\lambda_{2i-1})\frac{k^2}{i^2}|x_j|\right]<\infty.$$

The methods employed above lead easily to the principal result.

THEOREM 2. *Define a map* $\mathbf{p}$ *to a point* x *of the infinite-dimensional number space by the rule*

$$2\sum_{i=1}^{\infty}\int_{\mathbf{o}_i}^{\mathbf{p}_i}\mathbf{1}_j(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = x_j, \qquad j=1,\,2,\cdots,$$

permitting such paths of integration $\mathbf{o}_i$, $\mathbf{p}_i$ *as make* n_i *revolutions about the* i*-th circle with* $\sum i^{-m+1}n_i^2<\infty$. *Then the map is onto* $I^{m/2\dagger}$ *and can be rendered 1:1 by identification modulo the period lattice* $L^{m/2}$, *i.e.,* $\mathbf{p}\to x$ *can be viewed as a topological map of* M *onto the compact factor space* $I^{m/2\dagger}/L^{m/2}$. *The inverse map is implemented by the exponential map as for* $m=3$.

AMPLIFICATION 2. The existence and smoothness of the exponential map

$$x\in I^{m/2\dagger}\to X=\sum x_jX_j\to\exp\{X\}$$

is proven exactly as in Lemma 11.2 for each $m=3,\,5,\,7,\cdots$, separately. The local Hamiltonian flows $\exp\{tV_i\}$, $i=1,\,2,\cdots$, form a special case; *in particular, for* $i=2$, *you obtain an extraordinarily simple proof of the existence and smoothness of solutions of the Korteweg-de Vries equation* $\partial q/\partial t=3q\partial q/\partial x-\frac{1}{2}D^3q$; see Sjöberg [35] for a proof in another style. The same methods may be employed to prove the existence and continuity of $\exp\{tX\}$ for any vector field X with

$$\sum j^2|\Delta^{\bullet}(\lambda_{2j})|^2\,x_j^2<\infty, \qquad \text{for } X\in T\supset I^{\infty/2\dagger}.$$

The current restriction to a purely simple spectrum is for ease of writing, only. The facts and their proofs have a general validity, as will be plain in the next section.

AMPLIFICATION 3. $I^{m/2\dagger}$ is closely allied to the space $I^{m/2}$ of integral functions $\phi(\lambda)$ of order $\frac{1}{2}$ and type at most 1 with

$$\|\phi\|_{m/2}^2 = \int_0^\infty |\phi(\lambda)|^2 \lambda^{m/2}\, d\lambda < \infty.$$

Let $x \in I^{m/2\dagger}$ be the image of $\mathbf{p}$. Then

$$2\sum_{i=1}^\infty \int_{\mathbf{o}_i}^{\mathbf{p}_i} \phi(\Delta^2-4)^{-1/2} = \sum_{j=1}^\infty \phi(\lambda_{2j})x_j$$

for $\phi \in I^{m/2}$, much as for $m=3$. The key to the proof is the fact that $\sum i^{m+1}|\phi(\mu_i)|^2$ is comparable to $\|\phi\|_{m/2}^2$ for any choice of tied spectrum μ_i, $i = 1, 2, \cdots$. The formula embodies a map of $I^{m/2\dagger}$ into the dual space of $I^{m/2}$; indeed, the map is already onto, with this defect: *it ceases to be 1:1 for $m = 5, 7, \cdots$, the kernel being the span of V_i, $i = 1, \cdots, p$, for $m = 2p+3$.*

Proof for $m=5$: $y_1'(1,\lambda)\times[(\lambda-\lambda_0)(\lambda-\lambda_{2i})(\lambda-\lambda_{2j})]^{-1}$ is of class $I^{5/2}$ for any $0<i<j$, so $\sum \phi(\lambda_{2k})x_k = 0$ for $\phi \in I^{5/2}$ implies

$$0 = \frac{y_1'^{\bullet}(1,\lambda_{2i})}{(\lambda_{2i}-\lambda_0)(\lambda_{2i}-\lambda_{2j})}x_i + \frac{y_1'^{\bullet}(1,\lambda_{2j})}{(\lambda_{2i}-\lambda_0)(\lambda_{2i}-\lambda_{2j})}x_j,$$

and either $x=0$ or else x_j is proportional to $-2(\lambda_{2j}-\lambda_0)[y_1'^{\bullet}(1,\lambda_{2j})]^{-1}$; it is to be proved that the second choice of x annihilates $I^{5/2}$. Let $\phi \in I^{5/2}$. Then $(\lambda-\lambda_0)\phi(\lambda)$ is of class $I^{1/2}$, and

$$\sum_{j=1}^\infty \frac{y_1'(1,\lambda)}{(\lambda-\lambda_{2j})y_1'^{\bullet}(1,\lambda_{2j})}(\lambda_{2j}-\lambda_0)\phi(\lambda_{2j}) = (\lambda-\lambda_0)\phi(\lambda).$$

Now a simple application of the Paley-Wiener theorem confirms that

$$\phi(\lambda) = O\left(e^{\sqrt{-\lambda}}\lambda^{-7/4}\right)$$

for $\lambda \downarrow -\infty$; hence

$$\sum_{j=1}^\infty \frac{\lambda_{2j}-\lambda_0}{y_1'^{\bullet}(1,\lambda_{2j})}\phi(\lambda_{2j}) = \lim_{\lambda\downarrow-\infty}\frac{\lambda^2\phi(\lambda)}{y_1'(1,\lambda)} = \lim_{\lambda\downarrow-\infty} O(\lambda^{-1/4}) = 0.$$

The proof is finished by checking that V_i is, in fact, the vector field $x\cdot X$ corresponding to x:

$$V_1 = -\sum_{j=1}^\infty \frac{2(\lambda_{2j}-\lambda_0)}{y_1'^{\bullet}(1,\lambda_{2j})}X_j;$$

this is immediate from Proposition 7.6.3.

7.15. The Jacobi Variety in the Presence of Double Eigenvalues

It is necessary to modify the presentation of Sections 11–14 in the presence of double eigenvalues. Let $y_1(1,\lambda)$ be formed for the point with tied spectrum $\mu_i = \lambda_{2i-1}$, $i = 1, 2, \cdots$, and $y_1'(1,\lambda)$ for the point with reflecting spectrum $\nu_i = \lambda_{2i}$, $i = 0, 1, 2, \cdots$, as in Section 11. Then $y_2(1,\lambda)$ splits as $y_2^o(\lambda)y_2^\times(\lambda)$, in which the first, respectively second, factor accounts for the simple, respectively double, periodic eigenvalues; similarly, $y_1'(1,\lambda)$ splits as $y_1'^o(\lambda)y_2^\times(\lambda)$ and $\Delta^2(\lambda)-4 = 4y_1'(1,\lambda)y_2(1,\lambda)$ as $[\Delta^{o2}(\lambda)-4]\times[y_2^\times(\lambda)]^2$. The Jacobi map from the torus of Fig. 7.2 or 7.4 into the n-dimensional number space is now defined much as in Section 11:

the image $x = (x_1, \cdots, x_n)$ of a point $\mathbf{p}_i = (\mu_i^o, \sqrt{\Delta^{o2}(\mu_i^o) - 4})$, $i = 1, 2, \cdots, n$, of the torus is given by

$$x_j = 2\sum_{i=1}^{n} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \mathbf{1}_j^o(\mu) \frac{d\mu}{\sqrt{\Delta^{o2}(\mu) - 4}}, \quad j = 1, \cdots, n,$$

in which $\mathbf{o}_i = (\lambda_{2i-1}^o, 0)$, $i = 1, \cdots, n$, and $\mathbf{1}_j^o$ is formed for the simple eigenvalues in the obvious way. The rule may be spelled out as

$$x_j = 2\sum_{i=1}^{n} \int_{\mathbf{o}_i}^{\mathbf{p}_i} \frac{y_1'(1,\mu)}{\mu - \lambda_0} \frac{\lambda_{2j}^o - \lambda_0}{y_1'^{\bullet}(1, \lambda_{2j}^o)} \frac{1}{\mu - \lambda_{2j}^o} \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} \times y_2^{\times}(\lambda_{2j}^o),$$

and it follows from Section 11 that

$$\|x\|_{3/2\dagger}^2 = \sum_{j=1}^{n} (\lambda_{2j}^o)^{-2} |y_2^{\times}(\lambda_{2j}^o)|^{-2} x_j^2 < \infty$$

provided the winding numbers n_i, $i = 1, 2, \cdots$, of the paths of integration satisfy

$$\sum_{i=1}^{n} (\lambda_{2i}^o)^{-1} n_i^2 < \infty.$$

Now it is easy to prove much as before

(1) *that the map $q \to \mathbf{p} \to x$ is inverted by the exponential*

$$x \to X = \sum_{j=1}^{n} x_j D\partial\Delta^o(\lambda_{2j}^o)/\partial q \to q = e^X(\mathit{origin}),$$

in which $\partial\Delta^o/\partial q$ is the quotient of $\partial\Delta/\partial q$ by $y_2^{\times}$,

(2) *that the image if the full Hilbert space $I^{3/2\dagger}$ for the quadratic form $\|x\|_{3/2\dagger}^2 = \sum (\lambda_{2j}^o)^{-1} |y_2^{\times}(\lambda_{2j}^o)|^{-2} x_j^2$,*

(3) *that the map is 1:1 modulo the lattice $L^{3/2}$ of integral sums $\omega = \sum n_i \omega_i$ of the primitive periods*

$$\omega_{ij} = 4\int_{\lambda_{2i-1}^o}^{P} \lambda_{2i}^o \mathbf{1}_j^o(\mu) \frac{d\mu}{\sqrt{\Delta^{o2}(\mu) - 4}}, \quad j = 1, 2, \cdots,$$

with $\sum_{j=1}^{n} (\lambda_{2j}^o)^{-1} n_j^2 < \infty$, and

(4) *that M may be identified thereby as the factor space $I^{3/2\dagger}/L^{3/2}$, alias the real part of the Jacobi variety of $\sqrt{\Delta^{o2}(\lambda) - 4}$.*

Let $I^{3/2}$ be the class of integral functions $\phi(\lambda)$ of order $\frac{1}{2}$ and type at most $1 -$ (the type of $y_2^{\times}$) with

$$\|\phi\|_{3/2}^2 = \int_0^{\infty} |\phi(\lambda) y_2^{\times}(\lambda)|^2 \lambda^{3/2}\, d\lambda < \infty.$$

Then you have a perfect analogy with the results of Section 5:

(5) *the restriction $\phi \to \phi(\mu_i^o)$, $i = 1, \cdots, n$, is a 1:1 map of $I^{3/2}$ onto the Hilbert space of the quadratic form $\sum_{i=1}^{n} (\mu_i^o)^2 |\phi(\mu_i^o) y_2^{\times}(\mu_i^o)|^2$, the latter being comparable to $\|\phi\|_{3/2}^2$, and*

(6) *$I^{3/2}$ inherits from the former class of that name the interpolation formula*

$$\phi(\lambda) = \sum_{i=1}^{n} \phi(\mu_i^o) \prod_{j \neq i} \frac{1 - \lambda/\mu_j^o}{1 - \mu_i^o/\mu_j^o};$$

also,

(7) *the Jacobi map* $\mathbf{p} \to x$ *is equivalent to the relation*

$$2\sum_{i=1}^{n}\int_{\mathbf{o}_i}^{\mathbf{p}_i}\phi(\mu)\frac{d\mu}{\sqrt{\Delta^{o2}(\mu)-4}}=\sum_{j=1}^{n}\phi(\lambda_{2j}^{o})x_j$$

for $\phi \in I^{3/2}$, *and*

(8) *a differential of the first kind* $\phi(\mu)(\Delta^{o2}(\mu)-4)^{-1/2}\,d\mu$ *is uniquely determined by its periods*

$$4\int_{\lambda_{2i-1}^{o}}^{\lambda_{2i}^{o}}\phi(\mu)\frac{d\mu}{\sqrt{\Delta^{o2}(\mu)-4}}\qquad i=1,\cdots,n.$$

The higher tangent spaces $I^{m/2\dagger}$, $m = 3, 5, 7, \cdots, \infty$, and period lattices are introduced in a similar way, and the upshot is that everything goes through routinely provided you systematically factor out the function $y_2^{\times}(\lambda)$ containing the double eigenvalues.

AMPLIFICATION 1. Let n be finite. Then $\Delta^{o2}(\lambda)-4$ is proportional to $\ell^2(\lambda) = -(\lambda-\lambda_0^o)\cdots(\lambda-\lambda_{2n}^o)$, and $I^{3/2}$ is the class of polynomials of degree less than n. This may be seen from the fact that, for $\phi(\lambda)$ from that class,

$$\int_0^\infty |\phi y_2^{\times}|^2\,\lambda^{3/2}\,d\lambda=\int_0^\infty\frac{|\phi y_2|^2}{|y_2^o|^2}\lambda^{3/2}\,d\lambda$$

is comparable to

$$\int_0^\infty|\phi(\lambda)|^2|\lambda^{-1/2}\sin\sqrt{\lambda}+O(\lambda^{-3/2})|^2\lambda^{3/2}(1+\lambda^2)^{-n}\,d\lambda,$$

and so, by a standard sampling theorem for functions of order 1 and type at most 1, comparable to

$$\int_0^\infty|\phi(\lambda)|^2\lambda^{1/2}(1+\lambda^2)^{-n}\,d\lambda.$$

This is the case of classical hyperelliptic function theorem treated in [5, 16, 32], and [28].

AMPLIFICATION 2. In connection with the present development, it is comforting to observe that the *simple periodic spectrum uniquely determines the double periodic spectrum.*[30] To begin with,

$$\Delta(\lambda)=2\cos\left[\sqrt{-1}\epsilon\int_{\lambda_0}^{\lambda}\frac{\Delta^{\bullet}(\mu)}{\sqrt{\Delta^2(\mu)-4}}\,d\mu\right]$$

with a real constant ϵ so determined as to make $\Delta(\lambda)\sim 2\cos\sqrt{\lambda}$ for $\lambda\downarrow-\infty$. Now

$$\frac{\Delta^{\bullet}(\mu)}{\sqrt{\Delta^2(\mu)-4}}=\text{constant}\times\frac{\prod_{i=1}^{n}(1-\mu/\lambda_i^o)}{\sqrt{\prod_{i=0}^{2n}(1-\mu/\lambda_i^o)}},$$

from which it appears that the simple spectrum λ_i^o, $i=0,\cdots,2n$, determines the double if the non-trivial roots $\lambda_i^{\bullet}$, $i=1,\cdots,n$, of $\Delta^{\bullet}(\lambda)=0$ are known; *the fact is that the simple spectrum already determines them uniquely.*

Proof: The increment to the argument of the cosine across a non-trivial interval of instability $[\lambda_{2i-1}^o,\lambda_{2i}^o]$ is pure imaginary and must vanish if you are to have $\Delta(\lambda_{2i-1}^o-)=\Delta(\lambda_{2i}^o+)$. Threfore

$$\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o}\frac{\prod_{i=1}^{n}(1-\mu/\lambda_i^{\bullet})}{\sqrt{\prod_{i=0}^{2n}(1-\mu/\lambda_i^o)}}\,d\mu=0,\quad i=1,\cdots,n.$$

[30] Compare Hochstadt [13].

Now $\prod_{i=1}^{n}(1-\mu/\lambda_i^\bullet)y_2^\times(\mu)$ is proportional to $y_2(1,\mu)$ at some point of M and thus can be estimated as a constant multiple of $\lambda^{-1/2}\sin\sqrt{\lambda}+O(\lambda^{-3/2})$ for $\lambda\uparrow\infty$. Let λ_i' and λ_i'' be two different determinations of $\lambda_i^\bullet$, $i=1,\cdots,n$. Then the difference

$$\phi(\mu)=\prod_{i=1}^{n}\left(1-\frac{\mu}{\lambda_i'}\right)-c\prod_{i=1}^{n}\left(1-\frac{\mu}{\lambda_i''}\right)$$

belongs to $I^{3/2}$ for some choice of $c\neq 0$, determining a differential of the first kind with vanishing periods

$$\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o}\phi(\mu)\frac{d\mu}{\sqrt{\Delta^{o2}(\mu)-4}}=0,\quad i=1,\cdots,n.$$

But such a ϕ has to vanish identically by (8), and you must have $\lambda_i'=\lambda_i''$, $i=1,\cdots,n$. The proof is finished.

7.16. Periodicity, Almost Periodicity, and Metric Transitivity

A rectilinear motion on a compact torus such as $M=I^{3/2\dagger}/L^{3/2}$ is periodic, or metrically transitive, or else a mixture of the two. The purpose of the present section is to study this aspect of the flows $\exp\{tX\}$ arising from $I^{\infty/2\dagger}$. For ease writing, the material is presented for a purely simple spectrum; the modifications necessary for a mixed spectrum will be self-evident from Section 15.

PERIODICITY. The fact that translation $\exp\{tV_1\}$ is of period 1 on M is expressed in the language of the Jacobi variety as

$$2\sum_{i=1}^{\infty}i\int_{\lambda_{2i-1}}^{\lambda_{2i}}\frac{y_1'(1,\mu)}{\mu-\lambda_0}\frac{1}{\mu-\lambda_{2i}}\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}}=-1,\quad j=1,\,2,\cdots.$$

Proof: As t runs from 0 to 1, $\mathbf{p}_i$ moves through i full revolutions, and during the motion

$$\frac{2\mu_i^\bullet}{\sqrt{\Delta^2(\mu_i)-4}}=\sum_{j=1}^{\infty}\frac{-2(\lambda_{2j}-\lambda_0)}{y_1'^\bullet(1,\lambda_{2j})}\prod_{l\neq i}\frac{1-\lambda_{2i}/\mu_l}{1-\mu_i/\mu_l}$$

according to the expansion

$$V_1=-\sum_{j=1}^{\infty}2(\lambda_{2j}-\lambda_0)[y_1'^\bullet(1,\lambda_{2j})]^{-1}X_j$$

of Amplification 13.3. Therefore,

$$\begin{aligned}\sum_{i=1}^{\infty}4i\int_{\lambda_{2i-1}}^{\lambda_{2i}}\mathbf{1}_k(\mu)\frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} &= \sum_{j=1}^{\infty}\frac{-2(\lambda_{2j}-\lambda_0)}{y_1'^\bullet(1,\lambda_{2j})}\int_0^1\sum_{i=1}^{\infty}\left[\mathbf{1}_k(\mu_i)\prod_{l\neq i}\frac{1-\lambda_{2j}/\mu_l}{1-\mu_i/\mu_l}\right]dt\\ &= \sum_{j=1}^{\infty}\frac{-2(\lambda_{2j}-\lambda_0)}{y_1'^\bullet(1,\lambda_{2j})}\mathbf{1}_k(\lambda_{2j})\\ &= \frac{-2(\lambda_{2k}-\lambda_0)}{y_1'^\bullet(1,\lambda_{2k})}.\end{aligned}$$

The proof is finished by the substitution

$$\mathbf{1}_k(\mu)=\frac{y_1'(1,\mu)}{\mu-\lambda_0}\frac{\lambda_{2k}-\lambda_0}{y_1'^\bullet(1,\lambda_{2k})}\frac{1}{\mu-\lambda_{2k}}.$$

The periodicity of the Korteweg-de Vries flow $\exp\{tV_2\}$ may be discussed in the same way.

LEMMA 1.
$$V_2 = -\sum_{j=1}^{\infty} \frac{(2\lambda_{2j}+c)2(\lambda_{2j}-\lambda_0)}{y_1'^{\bullet}(1,\lambda_{2j})} X_j$$
with $c = \lambda_0 - \sum_{i=1}^{\infty}(\lambda_{2i}-\lambda_{2i-1})$.

Proof: Let the expansion for V_1 act on q and integrate back with regard to x to produce
$$q = -\sum_{j=1}^{\infty} \frac{2(\lambda_{2j}-\lambda_0)}{y_1'^{\bullet}(1,\lambda_{2j})} \frac{\partial\Delta(\lambda_{2j})}{\partial q} + c$$
with a constant of integration c. This number is now evaluated as $c = \lambda_0 - \sum(\lambda_{2i}-\lambda_{2i-1})$ by noting that it must be constant on M:
$$H_0 = \int_0^1 q\, dx = -\sum \frac{2(\lambda_{2j}-\lambda_0)}{y_1'^{\bullet}(1,\lambda_{2j})} \Delta^{\bullet}(\lambda_{2j}) + c$$
and then evaluating $q(0)$ at the special point of M with tied spectrum $\mu_i = \lambda_{2i}$, $i = 1, 2, \cdots$, with the help of the first trace formula
$$q(0) = \lambda_0 + \sum_{i=1}^{\infty}(\lambda_{2i-1}+\lambda_{2i}-2\mu_i)$$
in [28], p. 254, and the fact that $\partial\Delta(\lambda_{2j})/\partial q(0) = 0$, $j = 1, 2, \cdots$, at the point in hand. The expansion of V_2 is obtained by applying $L = qD + Dq - \frac{1}{2}D^3$ to the formula for q and then using the expansion of V_1.

Now $V_2 \in I^{9/2\dagger}$, and if the flow is of period T, then as time runs from $t = 0$ to $t = T$, $\mathbf{p}_i$ moves through n_i full revolutions with $\sum i^{-8} n_i^2 < \infty$; whence
$$4\sum_{i=1}^{\infty} n_i \int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_j(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = \frac{-(2\lambda_{2j}+c)2(\lambda_{2j}-\lambda_0)}{y_1'^{\bullet}(1,\lambda_{2j})} T, \quad j = 1, 2, \cdots,$$
and similarly in general: *if* $x \in I^{m/2\dagger}$ *and if* X *is the corresponding vector field, then* $\exp\{tX\}$ *is of period* T *on* M *if and only if there exists integers* n_i, $i = 1, 2, \cdots$, *with* $\sum i^{-m+1} n_i^2 < \infty$ *such that*
$$4\sum_{i=1}^{\infty} n_i \int_{\lambda_{2i-1}}^{\lambda_{2i}} \mathbf{1}_j(\mu) \frac{d\mu}{\sqrt{\Delta^2(\mu)-4}} = x_j T, \qquad j = 1, 2, \cdots,$$
i.e., if and only if $xT \in L^{m/2}$.

ALMOST PERIODICITY. Hyman[31] and Kruskal-Zabusky [20] demonstrated numerically that the Korteweg-de Vries flow is almost periodic in C_1^{∞}. Let $x \in I^{\infty/2\dagger}$ and let X be the associated vector field. *Then the flow* $\exp\{tX\}$ *is almost periodic on* M *in the following extraordinarily strict sense: Fix* $i \geq 1$. *Then to every* $\epsilon > 0$ *corresponds a number* $l(\epsilon) < \infty$ *such that any interval of length* $l(\epsilon)$ *or more contains an approximate period* T *for which*[32]
$$\left\| e^{(t+T)X} q - e^{tX} q \right\|_i < \epsilon,$$
independently of $-\infty < t < \infty$ *and of* $q \in M$. Lax conjectured this fact in a less precise form in [23].

[31] See the appendix to [22].

[32] $\|f\|_i$ is the i-th Sobolev norm $\sqrt{\sum_{j\leq i}\int_0^1 |D^i f|^2\, dx}$.

Proof: Let $x \in I^{m/2\dagger}$. The bound

$$\begin{aligned} \left\| e^{\xi \cdot X} q - e^{\eta \cdot X} q \right\|_i &\leq C_m(M) d(\xi, \eta) \\ d(\xi, \eta) &= \inf_{L^{m/2}} \|\xi - \eta - \omega\|_{m/2\dagger} \end{aligned}$$

is immediate from the proof of Lemma 11.2, revamped for general $m = 3, 5, 7, \cdots$. The proof of the almost periodicity of e^{tX} is now before you: The rectilinear flow $x_0 \to x_0 + tx$ is almost periodic on $I^{m/2\dagger}/L^{m/2}$ by the compactness of that space. But also,

$$\left\| e^{(t+T)X} q - e^{tX} q \right\|_i \leq c_m(M) d(Tx, x),$$

independently of $-\infty < t < \infty$ and of $q \in M$; thus every approximate period T of the rectilinear motion is an approximate period of $e^{tX} q$, independently of $q \in M$. The proof is finished.

METRIC TRANSITIVITY. M is a compact commutative group. The dual group M' comprises of the characters $e^{2\pi\sqrt{-1}k\cdot x}$, with k from the dual lattice $(L^{\infty/2})'$, i.e., the points k of the ∞-dimensional number space with $\sum i^m k_i^2 < \infty$ for every $m = 1, 2, \cdots$ and $k \cdot \omega$ integral for every $\omega \in L^{\infty/2}$. Notice that if ω_i, $i = 1, 2, \cdots$, are the primitive periods of M, then $n_i = k \cdot \omega_i$ is tame:

$$|n_i| \leq \text{constant} \times \|\omega_i\|_{3/2} = O(i^{-1}) \quad \text{as } i \uparrow \infty;$$

thus $n_i = 0$ from some $i = j < \infty$ on. The rest of the discussion follows Weyl [36]. The condition of metric transitivity for a rectilinear motion $x_0 \to x_0 + tx$ of $I^{\infty/2\dagger}/L^{\infty/2}$ is that

$$\lim_{T\uparrow\infty} \frac{1}{T} \int_0^T e^{2\pi\sqrt{-1}k\cdot(x_0+tx)}\, dt = 1 \text{ or } 0$$

according to whether $k = 0$ or not; hence the transitivity takes place if and only if $k \cdot x$ does not vanish for any $k \in (L^{\infty/2})'$. The criterion can be reformulated without reference to the dual lattice. Let $n_i = k \cdot \omega_i$, $i = 1, 2, \cdots$, be as above and recall that $n_i = 0$ from some $i = j$ on. Expand x as $\sum y_i \omega_i$. Then $k \cdot x = \sum_{i \leq j} y_i n_i \neq 0$ *for every k if and only if the numbers y_i, $i = 1, 2, \cdots$, are rationally independent.* For the Korteweg-de Vries flow $\exp\{tV_2\}$ the condition may be spelled out as follows: *the numbers y_i, $i = 1, 2, \cdots$, determined by*

$$-2 \sum_{i=1}^{\infty} y_i \int_{\lambda_{2i-1}}^{\lambda_{2i}} \frac{y_1'(1,\mu)}{\mu - \lambda_0} \frac{1}{\mu - \lambda_{2j}} \frac{d\mu}{\sqrt{\Delta^2(\mu) - 4}} = \lambda_{2j} + c, \quad j = 1, 2, \cdots,$$

must be rationally independent; it looks hopeless to check this.

Bibliography

[1] N. I. Akhiezer. *The Classical Moment Problem.* Hafner Publishing Co., New York, 1965.

[2] G. Borg. Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe. Bestimmung der Differentialgleihung durch die Eigenwerte. *Acta Math.*, 78:1–96, 1946.

[3] T. Carleman. *Les fonctions quasi-analytiques.* Gauthier-Villars, Paris, 1926.

[4] L. De Branges. *Hilbert Spaces of Entire Functions.* Prentice-Hall, Englewood Cliffs, New Jersey, 1968.

[5] B. Dubrovin and S. Novikov. Periodic and conditionally periodic analogues of multi-soliton solutions of the Kortweg-de Vries equation. *Dokl. Akad. Nauk SSSR*, 6:2131–2144, 1974.

[6] B. Dubrovin and S. Novikov. A periodicity problem for the Kortweg-de Vries and Sturm-Liouville equations. *Soviet Math. Dokl.*, 15:1597–1601, 1974.

[7] A. Erdélyi. Uber die freien Schwingungen in Kondersatorkreisen von veraendlicher Kapazitaet. *Ann. der Physik*, 19:585–622, 1934.

[8] L. D. Faddeev. The inverse problem in the quantum theory of scattering. *Uspehi Mat. Nauk.*, 14:57–119, 1959. Translation in *J. Mathematical Phys.* 4:72–104, 1963.

[9] H. Flaschka. Discrete and periodic illustrations of some aspects of the inverse method. In J. Moser, editor, *Dynamical Systems Theory and Applications*. Springer-Verlag, New York, 1975.

[10] H. Flaschka. On the inverse problem for Hill's operator. *Arch. Rational Mech. Anal.*, 54:293–309, 1975.

[11] W. Goldberg. Hill's equation for a finite number of instability intervals. *Jour. Math. Anal. Appl.*, 51:705–723, 1975.

[12] H. Hochstadt. Function-theoretic properties of the discriminant of Hill's equation. *Math. Zeit.*, 82:237–242, 1963.

[13] H. Hochstadt. On the determination of Hill's equation from its spectrum. *Arch. Rat. Mech. Anal.*, 19:353–362, 1965.

[14] E. L. Ince. Further investigations into the periodic Lamé functions. *Proc. Roy. Soc. Edin.*, 60:83–99, 1940.

[15] E. L. Ince. The periodic Lamé functions. *Proc. Roy. Soc. Edin.*, 60:47–63, 1940.

[16] A. R. Its and V. B. Matveev. On Hill's operators with a finite number of lacunae. *Funct. An. Appl.*, 9:65–66, 1975.

[17] M. Kac and P. van Moerbeke. A complete solution of the periodic Toda problem. *Proc. Nat. Acad. Sci. U.S.A.*, 72:2879–2880, 1975.

[18] M. Kac and P. van Moerbeke. On some periodic Toda lattices. *Proc. Nat. Acad. Sci. U.S.A.*, 72:1627–1629, 1975.

[19] P. Koosis. Weighted polynomial approximation on arithmetic progressions of intervals points. *Acta Math.*, 116:223–277, 1966.

[20] M. D. Kruskal and N. J. Zabusky. Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Letters*, 15:240–243, 1965.

[21] P. Lax. Periodic solutions of the KdV equations. In *Nonlinear Wave Motion (Proc. AMS-SIAM Summer Sem., Clarkson Coll. Tech., Postdam, N.Y. 1972)*, volume 15 of *Lectures in Appl. Math,*, pages 85–96. Amer. Math. Soc., Rhode Island, 1975.

[22] P. Lax. Periodic solutions of the Korteweg-de Vries equation. *Comm. Pure Appl. Math.*, 28:141–188, 1975.

[23] P. Lax. Almost periodic solutions of the KdV equation. *SIAM Review*, 18:351–375, 1976.

[24] N. Levinson. The inverse Sturm-Liouville problem. *Mat. Tidsskr. B*, 1949:25–30, 1949.

[25] B. M. Levitan and I. S. Sargsjan. *Introduction to Spectral Theory*, volume 39 of *Translations of the Mathematical Monographs*. Amer. Math.Soc., 1975.

[26] W. Magnus and W. Winkler. *Hill's equation*. Interscience Wiley, New York, 1st edition, 1966.

[27] H. P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piscounov. *Comm. Pure Appl. Math.*, 28:323–331, 1975.

[28] H. P. McKean and P. van Moerbeke. The spectrum of Hill's equation. *Inv. Math.*, 30:217–274, 1975.

[29] J. Moser. Finitely many mass points on the line under the influence of an exponential potential. *Springer Lecture Notes in Physics*, 38:467–497, 1975.

[30] P. J. Myrberg. Über transzendente hyperelliptische Integrale erster Gattung. *Ann. Acad. Sci. Fenn. Math.*, 14:1–32, 1943.

[31] P. J. Myrberg. Über analytische Funktionen auf transzendenten zweiblättrigen Riemannschen Flächen mit reellen Verzweigungspunkten. *Acta Math.*, 76:185–224, 1945.

[32] S. P. Novikov. The periodic problem for the Korteweg-de Vries equation. *Funct. Anal. Appl.*, 8:236–246, 1974.

[33] S. P. Novikov. A periodic problem for the Korteweg-de Vries equation. I. *Funkcional. Anal. i Priložen.*, 8(3):54–66, 1974.

[34] C. L. Siegel. *Topics in Complex Function Theory I. Elliptic Functions and Uniformization.* Wiley-Interscience, 1969.

[35] A. Sjöberg. On the Korteweg-de Vries equation: existence and uniqueness. *J. Math. Anal. Appl.*, 29:569–579, 1970.

[36] H. Weyl. Über die Gleichverteilung von Zahlen mod Eins. *Math. Ann.*, 77:111–147, 1916.

[8] Book Reviews: *Riemann Surfaces of Infinite Genus* by J. Feldman, H. Knörrer, and E. Trubowitz. CRM Monograph series. Vol. 20, Amer. Math. Soc.

[8] Book Reviews:*Riemann Surfaces of Infinite Genus* by J. Feldman, H. Knörrer, and E. Trubowitz. *Bull. Amer. Math. Soc.* **42** (2005), 79–87.

8 Book Reviews: *Riemann Surfaces of Infinite Genus* by J. Feldman, H. Knörrer, and E. Trubowitz. CRM Monograph series. Vol. 20, Amer. Math. Soc.

Henry P. McKean Jr.[1]

8.1. The Classical Story

One of the loveliest parts of mathematics is the subject of projective curves, developed, notably, by Jacobi, Abel, Riemann, and Poincaré over most of the nineteenth century. "Projective curve" means that you take an irreducible polynomial $p \in \mathbb{C}[x_1, x_2]$, bring each of its terms up to top degree by means of an auxiliary variable x_0, and think of the vanishing of the homogeneous polynomial $p(x_0, x_1, x_2)$ so produced as cutting out a locus or "curve" $\mathbb{X}$ in the projective space $\mathbb{CP}^2$ of triples $(x_0, x_1, x_2) \in \mathbb{C}^3 - 0$ with identification of (complex) lines $c(x_0, x_1, x_2)$: $c \neq 0$ to single points. $\mathbb{X}$ is then a compact, complex manifold of (complex) dimension 1, with possible singularities which may be resolved in a routine way, and conversely: any such manifold arises in this way. Alternatively, the solution x_2 of $p(x_1, x_2) = 0$ is a many-valued "algebraic" function of x_1 and $\mathbb{X}$ is its "Riemann surface".

$\mathbb{X}$ is an (oriented) sphere with $0 \leq g < \infty$ handles attached, this number being its genus. It carries a field $\mathbb{K}$ of functions of "rational character", imitating the common rational functions $\mathbb{K} = \mathbb{C}(x)$ on the sphere ($g = 0$), but more complicated, e.g. for the torus ($g = 1$), $\mathbb{K}$ is an elliptic function field of the form $\mathbb{K} = \mathbb{C}(x)[\sqrt{(x-e_1)(x-e_2)(x-e_3)}]$ with distinct numbers e. Here, it is already a remarkable fact that the algebra carries within it all the geometry: if the $\mathbb{K}$'s are isomorphic *as fields*, then the underlying curves $\mathbb{X}$ have a 1:1 onto "morphism" between them, rationally expressible both forward and back.

$\mathbb{X}$ also carries exactly g independent differentials of the first kind (DFK) of the form $\omega = f(z)dz$ with pole-free f on any little patch with "local parameter" z. These may be organized as follows. Let $a_1, \ldots, a_g$ and $b_1, \ldots, b_g$ be a standard homology basis of $\mathbb{X}$, the a's/b's passing around/through the holes. The standard basis of DFK is specified by requiring $[a_i(\omega_j) : 1 \leq i,\ j \leq g]$ = the identity.

Now form the "Abel sum" as follows: Take (1) a base point $\mathfrak{o} \in \mathbb{X}$, fixed, (2) a variable "divisor" $\mathfrak{P}$ comprising so and so many points $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ of $\mathbb{X}$, (3) paths from $\mathfrak{o}$ to each $\mathfrak{p}$, and write

$$\int_{\mathfrak{O}}^{\mathfrak{P}} \omega \equiv \sum_{i=1}^{n} \int_{\mathfrak{o}}^{\mathfrak{p}_i} (\omega_1, \ldots, \omega_g) = \mathfrak{x} \in \mathbb{C}^g.$$

Here, the paths of integration produce ambiguities, which may be removed by considering $\mathfrak{x}$ modulo "periods", i.e. modulo the lattice $\mathfrak{L} \subset \mathbb{C}^g$ produced by closed paths, from $\mathfrak{o}$ and

[1] Courant Institute of Mathematical Sciences, New York, NY, USA, mckean@cims.nyu.edu.

back again, in each summand. In this way, $\mathfrak{x}$ is reduced to an unambiguous point of $\mathbb{C}^g/\mathfrak{L} \equiv$ Jac = the Jacobi variety of $\mathbb{X}$, this being a compact torus because the several periods $a(\omega)$ and $b(\omega)$ span C^g over $\mathbb{R}$. The beautiful theorem of Abel is expressed in this language: $\mathfrak{P}/\mathfrak{Q}$ is the divisor of poles/roots of a function $f \in \mathbb{K}$ if and only if they comprise an equal number of points and have the same image in Jac. Riemann-Roch may also be mentioned here. It states that the class $F \subset \mathbb{K}$ of functions with poles $\mathfrak{P}$ or softer and the class $D \subset$ DFK of differentials of the first kind with these roots or harder are related by $\dim F =$ the number of points in $\mathfrak{P} + 1 - g + \dim D$, showing (in part) how $\mathbb{K}$ and DFK are intertwined.

The next item in the classical story is Riemann's theta function: with the prior normalization $A = [a_i(\omega_j)] =$ the identity, it turns out that $B = [b_i(\omega_j)]$ is symmetric, with positive-definite imaginary part, guaranteeing the rapid convergence of the sum

$$\vartheta(\mathfrak{x}) = \sum_{\mathfrak{n}\in\mathbb{Z}^g} e^{2\pi\sqrt{-1}\mathfrak{n}\bullet\mathfrak{x}+\pi\sqrt{-1}\mathfrak{n}\bullet B\mathfrak{n}} = \text{“theta”}$$

for $\mathfrak{x} \in \mathbb{C}^g$. This is almost a function on Jac itself: ϑ is unchanged by addition to $\mathfrak{x}$ of any “real” period $a(\omega)$, while addition of any “imaginary” period $b(\omega)$ multiplies it by a simple exponential factor. It follows that the vanishing of ϑ cuts out a sub-variety Θ of Jac, the so-called “theta divisor”. This may be described by Riemann's “vanishing theorem” to the effect that if $\mathfrak{P}$ is comprised of exactly g points $\mathfrak{p}_1, \ldots, \mathfrak{p}_g$ in general position and if $\mathfrak{x}$ is its image in Jac, then, with a suitable fixed “Riemann constant” K, $f(\mathfrak{p}) = \vartheta(x - \int_{\mathfrak{o}}^{\mathfrak{p}} + K)$ vanishes simply at $\mathfrak{p} = \mathfrak{p}_1, \ldots, \mathfrak{p}_g$ and no place else. This fact can now be used to express the general function $f \in \mathbb{K}$ as a ratio of products of translates of ϑ, comparable to the way an ordinary rational function is written as a ratio of products of translates of x. Here it must be mentioned that ϑ contains all the information about $\mathbb{X}$ in a deeper sense: if the imaginary period matrix $B = [b_i(\omega_j)]$ is known, then so is $\mathbb{X}$, up to a morphism. This is “Torelli's theorem”.

These are the bare bones of the classical story.

8.2. Bigger and Better Curves

Think about curves which are 2-sheeted covers of the plane, of which Jacobi's elliptic curve $y^2 = (1-x^2)(1-k^2x^2)$ with $k \neq 0, 1, \infty$ is the genus 1 prototype. But why just a polynomial in x? Why not $y^2 = \Pi_{n\neq 0}(1 - x^2/n^2)$? In short, why not a “transcendental” curve? Now the old (projective) compactness is lost and transcendental points of a new character appear where infinitely many handles pile up, as over $x = \infty$ in the present example.

Historically, a number of attempts were made to come to grips with such curves. Here are a few references: Nevanlinna [26], Myrberg [22, 23, 24], Ahlfors [2], Heins [11], Andreotti [4], and Accola [1]. But I think it is fair to say that these trials, gallant as they may have been, were none of them fully satisfactory, and this for three reasons: (1) It is necessary to deal with the piling up of handles at “transcendental” points of $\mathbb{X}$. (2) It is also pretty obvious that you must bring in transcendental functions to the function field. Already, H.F. Baker [6] understood this point, introducing the analogue of the exponential function which bears his name, together with that of Akhiezer [3], who reinvented it for another reason. (3) To extend the classical story to transcendental curves, it is, if not indispensable, then surely an enormous help to have in mind some concrete problem to which the machinery is to be applied as an aid to finding the best technical conditions: not so wide that the story comes to a stop, not so narrow that no application of any consequence can be accommodated. Baker [7] was almost successful in this respect, too: on p. 48, close inspection will reveal the solution of KdV expressed in the fashion of Its-Matveev [12] reported below. Unfortunately, KdV was unknown to him (Baker), so he did not understand what he had. But this is jumping ahead.

8.3. KdV and All That

A few connections between mechanics and projective curves were known in the nineteenth century. Jacobi in his *Vorlesungen über Dynamik* [13] had used Abel sums to separate variables in the Hamilton-Jacobi equation in connection with the geodesic flow on the surface of a 3-dimensional ellipsoid, etc. Turning things backwards, he even made a mechanical proof of the addition theorem for such integrals in the case of 2-sheeted curves. C. Neumann [25] employed the same trick to integrate a system of initially uncoupled harmonic oscillators, constrained to have their joint displacements move on a fixed sphere. Kovalevskaya's integration of her top by means of theta functions [15] is the most famous instance—this in the face of Picard's advice that theta functions are surely too simple to solve any interesting mechanical problem. These are all examples of "completely integrable" Hamiltonian mechanics, meaning that the system of (say) *2d* degrees of freedom has, in addition to its own Hamiltonian, d-1 more, independent, "commuting" constants of motion, where the adjective signifies that all the Hamiltonian flows produced by these commute, one with another. The number d is maximal in this regard: you cannot have more than d such flows in $2d$ dimensions. But if you *do* have so many, then (in principle) you can change coordinates to convert the original flow into straight-line motion at constant speed and map that back to solve your problem. The only trouble is that there is no effective way to be sure you have enough constants of motion and no effective recipe to find the correct change of coordinates. Jacobi complains of this in his *Vorlesungen*, where he says something like this: "We are supposed to be solving differential equations but of course we don't known how to do it. What we *can* do is look for some geometrically attractive substitution and, having found one that pleases us, seek out mechanical problems that could yield to this particular trick." The moral is that integrability, as defined above, is elusive. I much prefer Hermann Flaschka's practical definition, *to wit*: "You didn't think I could integrate that, but I can!" Take your choice. There are effectively no useful theorems, only a series of beautiful examples; indeed, the whole subject died a sad and early death when Poincaré showed that the 3-body problem is *not* integrable.

But then came startling news in Gardiner, Greene, Kruskal, and Miura [10] that the Korteweg-de Vries equation KdV: $\partial v/\partial t = 3v\partial v/\partial x - (1/2)\partial^3 v/\partial x^3$, descriptive of the leading edge of long waves in shallow water, had an apparently unlimited number of commuting constants of motion and could be integrated explicitly. Now the number of degrees of freedom is $2d = 2\infty$, and it is not enough to exhibit $d = \infty$ constants of motion. Somehow, you must have $d = \infty$ such *exactly*, not one more or less. Be that as it may, KdV was solved in a completely satisfactory way, using (and this is really odd) the quantum-mechanical apparatus of reflection and transmission coefficients ("scattering") for the allied "spectral problem" $-\psi'' + v\psi = \lambda\psi$. I have no space to explain this right now but will come back to it briefly at the end after describing what happens when v is periodic, of period 1, say.

KdV was first expressed in a clever way by P. Lax [16, 17]: with $L\psi = -\psi'' + v\psi$ and $P\psi = (3/2)(v'\psi + 2v\psi') - 2\psi'''$, KdV may be expressed in commutator language as $v^\bullet = L^\bullet = [P, L]$, in which I have obeyed Gelfand's rule that the right-hand side should read P(eter) L(ax). Here L is symmetric and P is skew, which leads at once to the easily verified surmise that the periodic/anti-periodic eigenvalues of L are constants of motion as v moves under KdV; in fact, this a complete list of such constants.

Now $L\psi = \lambda\psi$ is Hill's problem, studied by him in connection with the motion of the moon, and well-understood. There is a simple periodic ground state λ_0 followed by an infinite number of separated, alternately anti-/periodic pairs $\lambda_n^- \leq \lambda_n^+$, tending to $+\infty$ like $\pi^2 n^2$. The whole discussion now centers upon a variant of Hill's "discriminant" Δ in the form

$$\Delta^2(\lambda) - 1 = \big(\lambda_0 - \lambda\big)\Pi_{n=1}^{\infty}\big(\lambda_n^+ - \lambda\big)\big(\lambda_n^- - \lambda\big)/\pi^4 n^4.$$

For each value of $\lambda \in \mathbb{C}$, $L\psi = \lambda\psi$ has two "multiplicative" solutions ψ_- and ψ_+ with multipliers $m_-/m_+ = \Delta \pm \sqrt{\Delta^2 - 1}$, meaning that $\psi(x+1) = m\psi$. Here, the irrationality $\sqrt{\Delta^2 - 1}$ comes in, and it is natural to think in terms of the transcendental curve $\mu^2 = \Delta^2 - 1$, lying in two sheets over the complex plane where λ sits. I write $\mathfrak{p} = (\lambda, \sqrt{\Delta^2 - 1})$ for points of $\mathbb{X}$ and $e(x, \mathfrak{p}) = \psi_\pm(x)$, the signature of the radical $\sqrt{\Delta^2 - 1}$ in $\mathfrak{p}$ dictating which of the two functions ψ is meant. With the normalization $e(0, \mathfrak{p}) \equiv 1$, this is the Baker-Akhiezer function I spoke of before. Over $\lambda = \infty$, $e(x, \mathfrak{p})$ is of the form $\exp[kx + o(1)]$ with $k = \sqrt{-\lambda}$, $1/k$ serving as local parameter there. This is the only transcendental point of $\mathbb{X}$ where holes pile up, and even there things are pretty nice: the holes are spaced more and more widely and (if v is smooth) of rapidly vanishing diameter, so as to approximate common double points and to be, so to say, self-effacing. The "finite" part of $\mathbb{X}$ imitates a sphere with handles, one such to each open spectral "gap" $\lambda_n^- < \lambda_n^+$, and on the "real oval" a_n covering $[\lambda_n^-, \lambda_n^+]$ is found (a) a simple "immovable" pole of $e(x, \mathfrak{p})$, independent of $0 \le x < 1$, and (b) a simple moving root $\mathfrak{p}_n(x)$ starting at $\mathfrak{p}_n(0) =$ the pole. The projection of this root to the spectral plane is the nth eigenvalue of $Le = \mu e$ subject to $e(x, \mathfrak{p}) = 0$; that is why it moves with x.

Now it is fortunate that

(1) There are plenty of differentials of the first kind.
(2) The moving root divisor of $e(x, \mathfrak{p})$ is in "real position"; i.e. it has just one point in each real oval a.
(3) For such divisors, Abel's sum can be adjusted so as to make perfect sense, mapping the divisor to the (highly compact) real part of a nice Jacobi variety.
(4) The composite map from (a) the class of velocity profiles with common constants of motion, i.e. with fixed Hill's anti-/periodic spectrum, to (b) the root divisor, to (c) the Jacobi variety is $1:1$ and onto, and better still:
(5) v moves in a complicated way under KdV; likewise the motion of the root divisor is not simple, but (and this is quite a miracle) the corresponding point $\mathfrak{x} \in$ Jac moves in straight lines at constant speed, both under translation of v and under KdV.

It remains to undo this "substitution" to return from the simple motion of $\mathfrak{x}$ to the complicated motion of v. Riemann's theta function, adapted to infinite genus, comes in here. The commuting flows of v, by translation and by KdV, show up as infinitesimal directions $\mathfrak{n}_1$ and $\mathfrak{n}_2$ tangent to Jac, and if now $\mathfrak{x} \in$ Jac corresponds to the initial velocity profile, then the moving profile is

$$v(t,x) = -2\big(\partial^2/\partial x^2\big)\ell n \vartheta\big(\mathfrak{x} + x\mathfrak{n}_1 + t\mathfrak{n}_2\big).$$

This was worked out for $g < \infty$ open gaps, first by S.P. Novikoff [27] and later (independently) by McKean-van Moerbeke [18]. For the general transcendental case ($g = \infty$), see McKean-Trubowitz [19, 20]. The final recipe $v = -2(\ell n \vartheta)''$ is due to Its-Matveev [12]. This is the formula, cited before, that Baker [7] had, not knowing its hydrodynamical interpretation.

8.4. Other Examples

KdV is not the only problem that such machinery can solve. I have said that what we have is a body of examples, lacking much of a general theory, so all I offer here is a selection of these:

(1) Cubic-Schrödinger: $\sqrt{-1}\partial\psi/\partial t = \frac{\partial^2\psi}{\partial x^2} \pm |\psi^2|\psi$, coming from optics, waves in deep water, etc.;
(2) Sine-Gordon: $\partial^2\theta/\partial t^2 = \partial^2\theta/\partial x^2 + \sin\theta$, coming from nineteenth century geometry of surfaces of constant negative curvature, super-conductivity, fiber optics, etc.;
(3) Boussinesq: $\partial^2 v/\partial t^2 = \partial^2 v^2/\partial x^2 - (1/3)\partial^4 v/\partial x^4$, coming from long waves in shallow water once more;

(4) Camassa-Holm: $\partial v/\partial t + v\partial v/\partial x + \partial p/\partial x = 0$ with "pressure" $p(x) = \frac{1}{2}\int e^{-|x-y|}[v^2 + \frac{1}{2}(v')^2]dy$, much simpler than KdV and a lot closer to 3-dimensional Euler: $\partial v/\partial t + (v \bullet \mathrm{grad})v + \mathrm{grad}\, p = 0$ (Feynman's "dry water").

(5) Kadomtsev-Petviashvilli: $\frac{\partial}{\partial x_1}(\frac{\partial v}{\partial t} - 3v\frac{\partial v}{\partial x_2} - \frac{1}{2}\frac{\partial^3 v}{\partial x_2^3}) + \frac{3}{2}\frac{\partial^2 v}{\partial x_1^2} = 0$, a $(2+1)$-dimensional variant of KdV coming from lasers, etc.

For each of these systems, and more besides, the KdV story, a little modified, is repeated. The flow is expressed by a Lax-type pair $L^\bullet = [P, L]$, permitting the motion to be interpreted as an isospectral deformation of an associated spectral problem $L\psi = \lambda\psi$. The latter produces spectral invariants forming a complete list of commuting constants of motion; a (normally transcendental) multiplier curve comes in, comparable to the Riemann surface of Hill's $\sqrt{\Delta^2 - 1}$; and then the machine rolls: BA functions, root divisors, Abel sums, Jacobi variety and all: The multiplier curves of (1), (2), and (4) lie in two sheets over the plane having 1 or 2 transcendental points; that of (3) in three sheets, with two transcendental points; while (5) leads to effectively arbitrary curves with a limited number of transcendental points.

This sounds simpler than it really is: You have to *find* the Lax pair and its attendant spectral problem, in which connection I must mention the name of V. Zakharov and his indispensable role in this regard. The joke used to be: "You have a problem. Send it to Zakharov. If he sends you the Lax pair (waiting time 2 weeks), then it is integrable; if not, then it is not." This has been the best known way to hit upon Jacobi's "attractive substitution", in accord with Flaschka's "You didn't think I could integrate that, but I can!"

8.5. Aside on Complex Structure

What is it doing here? Nobody really knows, but here's an idle thought. Classical integrability with $2d$ degrees of freedom requires d independent, commuting constants of motion: $H_1, \ldots, H_d$. These functions have vanishing Poisson brackets

$$[H_i, H_j] = \frac{\partial H_i}{\partial P} \bullet \frac{\partial H_j}{\partial Q} - \frac{\partial H_i}{\partial Q} \bullet \frac{\partial H_j}{\partial P} = 0 \quad (i < j),$$

representing "d choose 2" partial differential constraints imposed upon are mere d functions, and if d is 100, say, then "d choose 2" is already 4950 of these. This over-kill reflects the delicacy of the situation (poke it and you break it) and may perhaps imply, what all experience confirms, that this over-kill entails some complex structure in the background.

Besides, it is not only in the periodic case that complex structure is present. I go back to KdV in the "scattering" case when $v(\pm\infty) = 0$. This was solved by GGKM [10] as reported before, put into a proper (integrable) Hamiltonian form by Faddeev-Zakharov [31], and reformulated by Dyson [8] in the following elegant manner, reminiscent of the Its-Matveev formula $v = -2(\ell n\vartheta)''$ of Sect. 8.3. Think of the velocity v as the potential in $\sqrt{-1}\partial\psi/\partial t = -\psi'' + v\psi$ and send in a wave $e^{-\sqrt{-1}kx}$ from $+\infty$: part will be reflected back off v in the form $s_-(k)e^{\sqrt{-1}kx}$ and part will be transmitted in the form $s_+(k)e^{-\sqrt{-1}kx}$; the (complex) numbers $s_\pm(k)$ are, *resp.* the "transmission" and "reflection" coefficients at (real) wave number k. Now turn on KdV: $\partial v/\partial t = 3v/(\partial v/\partial x) - (1/2)\partial^3 v/\partial x^3$. It turns out that the several numbers $s_+(k)$ are (commuting) constants of motion, while $s_-(k)$ is modified by a factor $e^{-\sqrt{-1}4k^3 t}$. Besides, in the absence of bound states, which I ignore for simplicity, the initial reflection coefficient determines the whole KdV flow, as in Dyson's version of the recipe: $v(t, x) = -2(\partial^2/\partial x^2)\ell n\vartheta$, where now ϑ is the Fredholm determinant $\det[I + w(\xi + \eta) : \xi, \eta \geq x]$ with

$$w(\xi + \eta) = \frac{1}{2\pi}\int e^{\sqrt{-1}k(\xi+\eta)} s_-(k) e^{-\sqrt{-1}4k^3 t} dk.$$

Now comparison with Its-Matveev is instructive: (1) $|s_-|^2 + |s+|^2 \equiv 1$, so it is only the phase of s_- that moves (in "straight lines" at constant speed), by addition of $2kx$ in response to translation by x, and by $-4k^3t$ in response to KdV, which I take to mean that the phase of s_- lives in some "Jacobi variety". (2) ϑ as a function of that phase must be Riemann's function. (3) There must be some kind of BA function and an Abel sum mapping its root divisor to phase s_-, and so forth. This speculation could be confirmed by a simple experiment: Take your favorite profile v, decaying rapidly at $\pm\infty$, periodize it as in $\overline{v}(x) = \sum_{\mathbb{Z}} v(x+np)$, apply Its-Matveev to that, and hope that Riemann's theta will pass over into Dyson's determinant as the period $p \uparrow \infty$. This was done most elegantly by Venakides [30] and in a more synthetic way by Ercolani-McKean [9]. In fact, Dyson's ϑ, properly complexified, imitates Riemann's ϑ in nearly every aspect, a nice instance being a nearly perfect analogue of Riemann's description of his theta divisor alluded to in Sect. 8.1, for which see Kempf [14]. But enough of such details.

What I want to suggest is that the language and the technology of projective curves, like everything stemming from complex structure, is extraordinarily robust: Pushed in favorable directions, it will enable you to recognize old projective friends going about their business in new ∞-dimensional settings—such friends as, formerly, you might have thought to have no algebraic meaning at all.

8.6. The Present Book

I come back to my theme: that to get a hold of transcendental curves, you need a specific problem you want to solve. It will tell you what to do: what technical conditions are favorable and, too, by what machinery to compute. Speaking for a moment classically, the introduction of moduli permits the most *efficient* description of a projective curve. But efficiency is in the eye of the beholder. It depends what you want to do. If you want to *compute* anything, efficient moduli are hopeless, since it is next to impossible ($g \leq 1$ excepted) to decode them into useful information. How about *inefficient* moduli? For example, a Hill's curve is determined by a single profile v with the right constants of motion. Horribly redundant to be sure, but what an advantage here with the whole machinery of Hill's equation behind you for help in computation.

The present book arises out of evidence obtained in this way, as to the "true" idea of a transcendental curve, at once effectively computable and flexible enough to cover all applications that have come to light so far. The class of curves described here is very wide. They are "small" deformations of a sphere with widely spaced double points (nodes), pasted together out of patches and handles under strict (but not too strict) rules of behavior, so that everything is true that should be true. I cannot enter now into the details, which are scrupulously explained, nothing too much or too little. It will be enough to say that, some few mysteries aside, $\mathbb{K}$, DGK, BA=Baker-Akhiezer, Abel's sum, Jacobi variety, Riemann's theta and its vanishing theorem, Riemann-Roch, and Torelli's theorem, too, all survive in a robust form. Here are included (1) very general 2-sheeted curves, as for KdV, but more; (2) "heat curves" connected to Kadomtsev-Petviashvilli [(5) of Sect. 8.4], imitating most any classical curve you could want; (3) "Fermi curves" which enter into a separate, very elaborate story about superconductivity, not explained here. These are the chief examples spelled out in detail.

The few complaints I have may be quickly told. (1) A little more chat would have been welcome, both mathematical and historical. The exposition is determinedly technical, and while it is lucid and done with much care, it could be discouraging to the immature reader. These may find Schmidt [28] helpful. (2) I would have liked to see more about Riemann-Roch, but for this there is the splendid article of Merkl [21]. (3) And last: the omission of an index makes it nearly impossible to browse.

Be that as it may. This is a big piece of work, brought to a very successful conclusion after some 10 laborious years, for which the authors have my warmest congratulations.

But isn't it all remarkable, this whole story? I mean, who would have guessed that projective curves could help us to understand long waves in shallow water, not to mention fiber optics, self-transparency, and so on and on. And who would ever have thought that physical problems of this type could be any guide to algebraic geometry. I will take just a moment to explain this last allusion. Riemann's theta function involves the imaginary periods $B = [b_i(\omega_j)]$. These Riemann matrices are special: They are symmetric with positive imaginary part, and something more, something hidden, owing to their origination from a curve. What is it? Well, there is an Its-Matveev-type formula expressing the solution of Kadomtsev-Petviashvilli in terms of Riemann's theta function, and this works *only if* the B appearing in the theta sum of Sect. 8.1 is a bona fide Riemann matrix. S.P. Novikoff conjectured this; the proof is due to Arbarello and de Concini [5] and to Shiota [29]. You never can tell.

Bibliography

[1] R. D. M. Accola. Some classical theorems on open Riemann surfaces. *Bull. Amer. Math. Soc.*, 73:13–26, 1967.

[2] L. V. Ahlfors. Normalintegrale auf offenen Riemannschen Flächen. *Ann. Acad. Sci. Fenn. Math.*, 1947(35):1–24, 1947.

[3] N. I. Akhiezer. Continuous analogues of orthogonal polynomials on a system of intervals. *Dokl. Akad. Nauk SSSR*, 141:263–266, 1961.

[4] A. Andreotti. On a theorem of Torelli. *Amer. J. Math.*, 80:801–828, 1958.

[5] E. Arbarello and C. De Concini. On a set of equations characterizing Riemann matrices. *Ann. of Math. (2)*, 120(1):119–140, 1984.

[6] H. F. Baker. *Abelian Functions.* Cambridge University Press, Cambridge, 1897. Reissued 1995.

[7] H. F. Baker. *An Introduction to the Theory of Multiply-Periodic Functions.* Cambridge University Press, Cambridge, 1907.

[8] F. J. Dyson. Fredholm determinants and inverse scattering. *Comm. Math. Phys.*, 47:171–183, 1976.

[9] N. Ercolani and H. P. McKean. Geometry of KdV (4). Abel sums, Jacobi variety, and theta function in the scattering case. *Invent. Math.*, 99(3):483–544, 1990.

[10] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Method for solving the Korteweg-de Vries equation. *Phys. Rev. Letters*, 19:1095–1097, 1967.

[11] M. Heins. Riemann surfaces of infinite genus. *Ann. of Math. (2)*, 55:296–317, 1952.

[12] A. R. Its and V. B. Matveev Hill operators with a finite number of lacunae. *Funkcional. Anal. i Priložen.*, 9(1):69–70, 1975.

[13] C. G. J. Jacobi. *Vorlesungen über analytische Mechanik.* Deutsche Mathematiker Vereinigung, Freiburg; Fried. Vieweg & Sohn, Braunschweig, 1848.

[14] G. Kempf. On the geometry of a theorem of Riemann. *Ann. of Math. (2)*, 98:178–185, 1973.

[15] S. Kowalevski. Sur le problème de la rotation d'un corps solide autour d'un point fixe. *Acta Math.*, 12(1):177–232, 1889.

[16] P. Lax. Integrals of non-linear equations and solitary waves. *Comm. Pure Appl. Math.*, 21:467–490, 1968.

[17] P. Lax. Periodic solutions of the Korteweg-de Vries equation. *Comm. Pure Appl. Math.*, 28:141–188, 1975.

[18] H. P. McKean and P. van Moerbeke. The spectrum of Hill's equation. *Inv. Math.*, 30:217–274, 1975.

[19] H. P. McKean and E. Trubowitz. Hill's operator and hyperelliptic function theory in the presence of infinitely many branch points. *Comm. Pure Appl. Math.*, 29:143–226, 1976.

[20] H. P. McKean and E. Trubowitz. Hill surfaces and their theta functions. *Bull. Amer. Math. Soc.*, 84:1042–1085, 1978.

[21] F. Merkl. A Riemann-Roch theorem for infinite genus Riemann surfaces. *Invent. Math.*, 139(2):391–437, 2000.

[22] P. J. Myrberg. Über transzendente hyperelliptische Integrale erster Gattung. *Ann. Acad. Sci. Fenn. Math.*, 14:1–32, 1943.

[23] P. J. Myrberg. Über analytische Funktionen auf transzendenten zweiblättrigen Riemannschen Flächen mit reellen Verzweigungspunkten. *Acta Math.*, 76:185–224, 1945.

[24] P. J. Myrberg. Über analytische Funktionen auf transzendenten Riemannschen Flächen. In *C. R. Dixième Congrès Math. Scandinaves 1946*, pages 77–96. Jul. Gjellerups Forlag, Copenhagen, 1947.

[25] C. Neumann. De problemate quodam mechanico, quod ad primam integralium ultraellipticorum classem revocatur. *J. Reine Angew. Math.*, 56:46–63, 1859.

[26] R. Nevanlinna. Quadratisch integrierbare Differentiale auf einer Riemannschen Mannigfaltigkeit. *Ann. Acad. Sci. Fenn. Math.*, 1941(1):1–34, 1941.

[27] S. P. Novikov. The periodic problem for the Korteweg-de Vries equation. *Funct. Anal. Appl.*, 8:236–246, 1974.

[28] M. U. Schmidt. Integrable systems and Riemann surfaces of infinite genus. *Mem. Amer. Math. Soc.*, 122(581):viii+111, 1996.

[29] T. Shiota. Characterization of Jacobian varieties in terms of soliton equations. *Inv. Math.*, 83:333–382, 1986.

[30] S. Venakides. The infinite period limit of the inverse formalism for periodic potentials. *Comm. Pure Appl. Math.*, 41:3–17, 1988.

[31] V. E. Zakharov and L. D. Faddeev. The Korteweg-de Vries equation is a fully integrable Hamiltonian system. *Funkcional Anal. i Priložen*, 5:18–27, 1974.

[9] Fredholm Determinants and the Camassa-Holm Hierarchy

[9] Fredholm Determinants and the Camassa-Holm Hierarchy. *Comm. Pure Appl. Math.* **56** (2003), 638–680.

9 Fredholm Determinants and the Camassa-Holm Hierarchy

Henry P. McKean Jr.[1]

9.1. Introduction

The equation of Camassa and Holm [2][2] is an approximate description of long waves in shallow water. It reads

$$\mathrm{CH}: \frac{\partial m}{\partial t} = -(mD + Dm)v$$

in which $D = \partial/\partial x$ and $m = v - v''$: in extenso,

$$\mathrm{CH}: \frac{\partial v}{\partial t} - \frac{\partial^3 v}{\partial t \partial x^2} + 3v\frac{\partial v}{\partial x} - 2\frac{\partial v}{\partial x}\frac{\partial^2 v}{\partial x^2} - v\frac{\partial^3 v}{\partial x^3} = 0.$$

Its Eulerian form is more attractive: In terms of Green's function $\mathrm{G} = (1-D^2)^{-1} = \frac{1}{2}e^{-|x-y|}$, it reads

$$\mathrm{CH}': \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \frac{\partial p}{\partial x} = 0 \quad \text{with the "pressure" } p = G\left[v^2 + \frac{1}{2}(v')^2\right].$$

It is also important to emphasize the Lagrangian standpoint, tracking the moving "fluid" in the natural scale $\overline{x} = \overline{x}(t,x)$ determined by $\partial\,\overline{x}/\partial t = v(t,\overline{x})$ with $\overline{x}(0,x) \equiv x$: In this language, the equation reads

$$\mathrm{CH}'': \frac{d}{dt}v(t,\overline{x}) + p'(\overline{x}) = 0.$$

The flow can break down: For example, if v is odd to start with, it stays odd and it is easy to see that $v'(t,0) \equiv s(t)$ satisfies $s^\bullet \le -\frac{1}{2}s^2$, which drives it down to $-\infty$ at some time $T < \infty$ if $s(0) < 0$. It is the special purpose of this paper to show that this type of unpleasantness is only apparent. More precisely. I will prove that if m is summable at the start together with m', then the Lagrangian form of the flow is perfectly fine for all time $0 \le t < \infty$. The flow is integrable[3]: In fact, I will integrate it explicitly in terms of certain theta-like Fredholm determinants, providing expressions of $v(t,\overline{x})$ and the scale $\overline{x}(t,x)$ that are always sensible for any $(t,x) \in [0,\infty) \times \mathbb{R}$; it is only $v'(t,\overline{x})$ that misbehaves, and this does not spoil the Lagrangian version CH″. Consequently, the Eulerian version CH′ is also fine: $v'(t,\mathrm{x})$ can be

[1]Courant Institute, 251 Mercer Street, New York, NY 10012-1185, USA, mckean@cims.nyu.edu.

[2]See also Camassa, Holm, and Hyman [3].

[3]I recall H. Flaschka's definition of integrability, the only honest one around: "You didn't think I could integrate that, but I can!"

infinite now and then, but this is not very serious; it is just that $\partial v/\partial t + v\partial v/\partial x + \partial p/\partial x = 0$ must be treated with the obvious precautions.

$H_2 = \frac{1}{2}\int mGm = \frac{1}{2}\int[(v')^2+v^2]$ serves as the Hamiltonian for the (back-wards) flow, and there is a whole series of further constants of motion obtained from the "spectral problem" $(\frac{1}{4} - D^2)f = \lambda mf$. Introduce the operators $K = (\frac{1}{2} - D)^{-1} = e^{x/2}\int_x^\infty e^{-y/2}$ and $K^\dagger = (\frac{1}{2} + D)^{-1} = e^{-x/2}\int_{-\infty}^x e^{y/2}$ mapping H^0 onto H^1. Then the spectral problem for $f \in H^1$ is equivalent to[4] $K^\dagger mKF = F/\lambda$ for $F = (\frac{1}{2} - D)f \in H^0$, and it is not hard to see that you have a simple spectrum[5] $1/\lambda_n : n \in \mathbb{Z} - 0$ serving as a complete list of (commuting) constants of motion, with finite (absolute) trace $\Sigma 1/|\lambda_n| \le \int |m|$. Section 9.2 explains all this together with the appropriate Hamiltonian formalism.

Section 9.3 investigates the "individual" flows based upon the Hamiltonians $H_n = 1/\lambda_n$ for $n \in \mathbb{Z} - 0$. The corresponding vector fields are $\mathbb{X}_n : m \to (mD + Dm)f_n^2$, in which f_n is the associated eigenfunction with $\|f_n\|^2 \equiv \int[(f_n')^2 + \frac{1}{4}f_n^2] = 1$, and it is the content of Sect. 9.3 that such a flow can be integrated in elementary terms with the help of the Lagrangian-like scale $\overline{x} = \overline{x}(t,x)$, determined by $\partial\,\overline{x}/\partial t = -f_n^2, (t, \overline{x})$ and $\overline{x}(0,x) \equiv x$, which has an elementary expression, too.[6] Section 9.4 carries over this piece of luck to the "composite" flow[7] $e^{t\mathbb{X}} \equiv \prod_{\mathbb{Z}-0} e^{t_n\mathbb{X}_n}$ with $t = [t_n : n \in \mathbb{Z} - 0]$ tame so as to avoid any technicalities. Everything is expressed in terms of three "theta functions" ϑ_-, ϑ_+, and ϑ, now to be described.

Let m have its initial value and let $f_n : n \in \mathbb{Z} - 0$ be the associated eigenfunctions. Then[8]

$$\left.\begin{matrix}\vartheta_- \\ \vartheta \\ \vartheta_+\end{matrix}\right\} = \det\left[1 + \left(e^{t_i} - 1\right)\int_{-\infty}^x \left(f_i'f_j' + \frac{1}{4}f_if_j\right)\left\{\begin{matrix} -\dfrac{1}{2}f_if_j(x) \\ -f_i'f_j(x) \\ +\dfrac{1}{2}f_if_j(x)\end{matrix}\right.\right]$$

in terms of which the updated $e^{t\mathbb{X}}m$ and $e^{t\mathbb{X}}f_n$ may be expressed with the help of the Lagrangian scale $\overline{x}(t,x)$, determined by $\partial\,\overline{x}/\partial t_n = -(e^{t\mathbb{X}}f_n^2)(\overline{x})$ and $\overline{x}(0,x) \equiv x$, as follows:

$$e^{\overline{x}} = \frac{e^x\vartheta_-}{\vartheta_+} \quad \text{or, what is more or less the same, } \overline{x}' = \frac{\vartheta^2}{\vartheta_-\vartheta_+}, \tag{9.1.1}$$

$$\left(e^{t\mathbb{X}}m\right)(\overline{x}) = \frac{m(x)}{(\overline{x}')^2}, \tag{9.1.2}$$

and

$$\left(e^{t\mathbb{X}}f\right)(\overline{x}) = \frac{e^{t/2}\vartheta}{\sqrt{\vartheta_-\vartheta_+}}\left[1 + \int_{-\infty}^x mf\otimes f\left(e^t - 1\right)\lambda\right]^{-1} f(x) \tag{9.1.3}$$

in which t is the vector $[t_n : n \in \mathbb{Z} - 0]$, etc., $f \otimes f$ is the matrix $[f_if_j : i,j \in \mathbb{Z} - 0]$, and m and f have their initial values as in the three theta functions. The formula for $\overline{x}'$ in (9.1.1) comes from the pretty identity $\vartheta^2 = \vartheta_-\vartheta_+ + \vartheta_-'\vartheta_+ - \vartheta_-\vartheta_+'$. Here, ϑ_- and ϑ_+ do not vanish, so $\overline{x}$ always makes sense. ϑ *can* vanish. Then (9.1.2) goes bad, but (9.1.3) does not: $\vartheta = \det[1 + (e^t - 1)\lambda\int_{-\infty}^x mf\otimes f]$ so $\vartheta[1 + \int_{-\infty}^x mf\otimes f(e^t-1)\lambda]^{-1}$ is a Fredholm *cofactor* and always makes sense.

[4] 0 is not in the spectrum.

[5] The indexing will be explained later.

[6] The computation can be found in McKean [12] I repeat it here for the reader's convenience.

[7] The individual flows commute so the order of the product is immaterial.

[8] McKean and Trubowitz [13] encountered similar theta-like determinants in connection with the isospectral class of the quantum-mechanical oscillator $-D^2 + x^2 - 1$.

Section 9.5 deals with technicalities when $t = [t_n : n \in \mathbb{Z} - 0]$ is not tame: as needed for application to CH: If both m and m' are summable and if $t_n = c/\lambda_n + 0(1/\lambda_n^2)$ for large n with a fixed constant c, then all three determinants are fine, representing functions that are real analytic with respect to t and of class C^2 or better with respect to x.

Section 9.6 applies the formulae of Sect. 9.4 to CH. The (forward) Hamiltonian is

$$H = -\frac{1}{2}\int \left[(v')^2 + v^2\right] = -\frac{1}{4}\,\mathrm{sp}\left(K^\dagger m K\right)^2 = -\frac{1}{4}\sum_{\mathbb{Z}-0}\lambda_n^{-2},$$

so the corresponding vector field $\mathbb{X}$ is nothing but $-\sum(2\lambda_n)^{-1}\mathbb{X}_n$, and if now $t \geq 0$ is the running time of the flow, then the parameters $t_n = -t/2\lambda_n : n \in \mathbb{Z} - 0$ obey the technical condition of Sect. 9.5. The three theta functions are

$$\left.\begin{matrix}\vartheta_-\\ \vartheta\\ \vartheta_+\end{matrix}\right\} = \det\left[1 + (e^{-t/2\lambda} - 1)\int_{-\infty}^{x}\left(f'\otimes f' + \frac{1}{4}f\otimes f\right)\left\{\begin{matrix}-\frac{1}{2}f\otimes f(x)\\ -f'\otimes f(x)\\ +\frac{1}{2}f\otimes f(x)\end{matrix}\right.\right],$$

so CH is integrated in the Lagrangian scale determined by $\partial\,\overline{x}/\partial t = (e^{t\mathbb{X}}v)(\overline{x})$ as follows:

$$e^{\overline{x}} = \frac{e^x\vartheta_-}{\vartheta_+} \quad \text{or, what is more or less the same,} \quad \overline{x}' = \frac{\vartheta^2}{\vartheta_-\vartheta_+}, \tag{9.1.4}$$

$$\left(e^{t\mathbb{X}}m\right)(\overline{x}) = \frac{m(x)}{(\overline{x}')^2}, \tag{9.1.5}$$

and

$$\left(e^{t\mathbb{X}}v\right)(\overline{x}) = \frac{\partial\,\overline{x}}{\partial t} = \left[\ln\frac{\vartheta_-}{\vartheta_+}\right]^\bullet = \frac{\vartheta_-^\bullet}{\vartheta_-} - \frac{\vartheta_+^\bullet}{\vartheta_+}. \tag{9.1.6}$$

(9.1.6) is the most gratifying aspect of the recipe: ϑ can vanish so (9.1.5) can go bad, but ϑ_- and ϑ_+ do not, so both $\overline{x}$ as in (9.1.4) and $e^{t\mathbb{X}}v$ as in (9.1.6) are perfectly fine for $t \geq 0$ and, indeed, for $t \leq 0$, too. It is only when you differentiate $e^{t\mathbb{X}}v$ that trouble starts:

$$\left(e^{t\mathbb{X}}v\right)'(\overline{x}) = \frac{\overline{x}^{\bullet\prime}}{x}(\ln\overline{x}')^\bullet = \left(\ln\frac{\vartheta^2}{\vartheta_-\vartheta_+}\right)^\bullet = \frac{2\vartheta^\bullet}{\vartheta} + \text{nice stuff}$$

which can (and will) be bad if ϑ vanishes. Actually, nothing bad can happen if m_- lies wholly to the left of m_+,[9] a fact more or less due to Constantin and Escher [4]. This has a very simple proof in the present theta format. McKean [11] proved the converse: If any part of m_+ is located to the left of some part of m_-, then the flow must break down, i.e., $(e^{t\mathbb{X}}v)' = -\infty$ someplace. I had hoped to find an equally simple theta-format proof of this fact but did not succeed. Be that as it may, it is a source of satisfaction that the Lagrangian version $\mathrm{CH}'' : (d/dt)(e^{t\mathbb{X}}v)(\overline{x}) + (\partial/\partial\,\overline{x})p(t,\overline{x}) = 0$ does not break down at all!

Xin and Zhang [15] have proved that a nice (weak) solution of CH may be obtained by introducing a dispersive term $(0+)\ \times\partial^3 v/\partial^3 x$. I presume their recipe has the same outcome as mine but did not prove it. I must also say that Beals, Sattinger, and Szmigielski [1] have adapted beautiful old formulae of Stieltjes [14] to solve CH in the "many soliton" case when $m\,dx$ reduces to a (finite) collection of signed point masses. The formulae in question permit

[9] m_+/m_- is the positive/negative part of m.

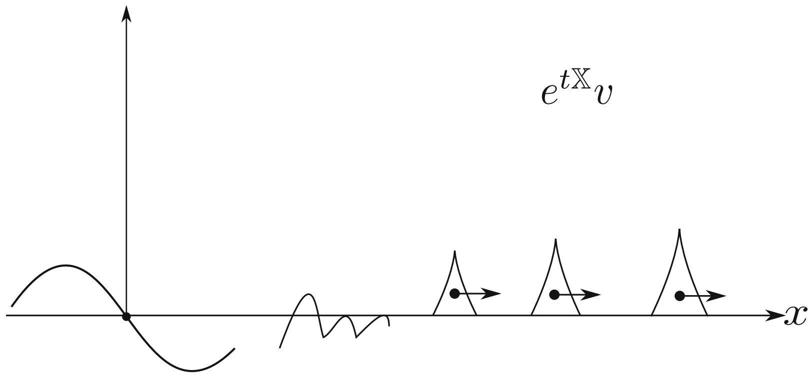

FIGURE 9.1.

the recovery, by means of certain determinants; of the numbers a and b appearing in the continued fraction

$$\frac{1 \rfloor}{\lceil \lambda a_1} + \frac{1 \rfloor}{\lceil b_1} + \frac{1 \rfloor}{\lceil \lambda a_2} + \frac{1 \rfloor}{\lceil b_2} + \cdots + \frac{1 \rfloor}{\lceil \lambda a_n} + \frac{1 \rfloor}{\lceil b_n} = \sum \frac{r_n}{\lambda - c_n}$$

from its poles c and residues r. Here, effectively, the a & b determine masses & locations while the c & r determine spectrum & norming constants.[10] This recipe is an extreme case of mine but it is prettier as it stands: Under the CH flow, the spectrum is fixed while the norming constants move simply as per $r_n(t) = \overline{e}^{t/2\lambda_n} r_n(0)$, and because the correspondence of a & b to c & r is purely algebraic, the recipe holds for all time regardless of the signatures of the masses; in particular, any breakdown of v' is evanescent. This is a nice example of the ubiquitous and still mysterious fact that, in every known example, integrability is accompanied by complex projective structure: It may lie deep and hidden but it is always there.

Section 9.7 presents a partial description of what happens when m is odd with $m(x) > 0$ for $x < 0$, $m(x) < 0$ for $x > 0$, and $m'(0) < 0$. Breakdown (of v') is sure to happen in these circumstances. ϑ $(x = 0)$ has an infinite number of (double) roots at times $0 < T_1 < T_2 <$ etc. $\uparrow \infty$. These roots are double in respect to x and immediately split into symmetrically disposed pairs of simple roots $\pm r_1(t)$, $\pm r_2(t)$, etc., of ϑ as a function of x, moving steadily to the right/left. At such a root, $(e^{t\mathbb{X}}v)(\overline{x})$ has a cusp of the form const $\times$ $|\overline{x}(t,x) - \overline{x}(t,r)|^{2/3} +$ something nice, as in the Fig. 9.1, and I believe but cannot yet prove that for large time these little cusps shape themselves into well-separated "solitons" of the form $(p/2)e^{-|x-q-(p/2)t|}$ moving at speeds $p/2 = 1/2\lambda_n : n \in \mathbb{Z} - 0$, with "phase shifts" q related to the norming constants cited before, meaning that

$$\big(e^{t\mathbb{X}}v\big)(x) \simeq \sum_{\mathbb{Z}-0} \frac{1}{2\lambda_n} e^{-|x-q_n-t/2\lambda_n|} \quad \text{as } t \uparrow \infty.$$

This would be in agreement with the exact result of Beals, Sattinger, and Szmigielski [1] for many solitons and with the numerical observations of Camassa [private communication] in more general circumstances. The discussion of the roots of ϑ relies upon the curious identity

$$\vartheta''_{-}\vartheta + \vartheta_{-}\vartheta'' + \vartheta'_{-}\vartheta - \vartheta_{-}\vartheta' - \frac{m'}{m}\vartheta'\vartheta_{-} - 2\vartheta'\vartheta'_{-} = 0.$$

Ercolani and McKean [7] encountered a similar identity for the determinant of Dyson [6] appearing in the solution of KdV. That is how I guessed the present one.

[10] $f_n(x) \simeq$ a "norming constant" $\times e^{-x/2}$ at $+\infty$.

Section 9.8 is in the nature of an appendix. The classical theta function of Riemann and its translates obey an astonishing number of identities and it is the same here: The identity $\vartheta^2 = \vartheta_-\vartheta_+ + \vartheta'_-\vartheta_+ - \vartheta_-\vartheta'_+$ cited in connection with (9.1.1) is an example; the last display is another—only now you must ask yourself what translation is and whether ϑ_- and ϑ_+ are related to ϑ in this way.

The answer is couched in terms of the "addition," so-called. introduced by McKean [10] in the context of KdV and revamped in McKean [12] for CH. It is properly defined only if m is of one sign, e.g. $m(x) > 0$. Fix $\Lambda < 0$. Then $(\frac{1}{4} - D^2)e = \Lambda me$ has two positive solutions: $e \simeq -e^{x/2}$ at $-\infty$ and $e_+- \simeq e^{-x/2}$ at $+\infty$. I write $e_\pm = e(\bullet, p)$ in which p records the eigenvalue Λ and a signature to indicate which function is meant. The corresponding "addition" $A^{\mathfrak{p}}$ is defined by its action $(A^{\mathfrak{p}}m)(\overline{x}) = m(x)/(\overline{x}')^2$ in the scale determined by

$$e^{\overline{x}} = e^x \frac{e' - \frac{1}{2}e}{e' + \frac{1}{2}e}$$

or, what is more or less the same,

$$\overline{x}' = \frac{-\Lambda me^2}{e'^2 - \frac{1}{4}e^2}.$$

The several additions are the proper counterpart of translation. They are invertible,[11] they commute, and what is most important, they are part of (commute with) the whole CH hierarchy; in fact, successive additions are more or less coextensive with the hierarchy in that any CH flow can be well approximated by them. The terminology and the identification of "addition" as translation stem from the fact that, for periodic m, $A^{\mathfrak{p}}$ corresponds to addition of $-\mathfrak{p} + \infty$ in the Jacobi variety of a (singular, transcendental) projective curve. I realize that this is all a little cryptic. I do not explain more here, but see McKean [10, 12].

Now addition also makes sense at $\Lambda = 0$ and, with $\mathfrak{o} = (0, \pm)$ in place of $\mathfrak{p} = (\Lambda, \pm)$, you have $(\Lambda^{\mathfrak{o}}\vartheta)\ (\overline{x}) = \vartheta_\pm(x)$ in the associated scale, which I take to mean that there is really only one theta function here, the others being produced from it by "translation," just as Jacobi's four theta functions are produced from anyone of them, by addition of half periods. The action on ϑ of the general addition $A^{\mathfrak{p}}$ is also amusing:

$$\begin{aligned} &e^{t\mathbb{X}}e(\bullet, p) \text{ taken in the scale for the composite flow } e^{t\mathbb{X}} \\ &= \frac{e(x, \mathfrak{p})}{\sqrt{\vartheta_-\vartheta_+(x)}} \times A^{\mathfrak{p}}\vartheta \text{ taken in the scale for the addition } A^{\mathfrak{p}}. \end{aligned}$$

This rule is a (relative) version of the formula of Its and Matveev [9] for the Baker-Akhiezer function associated to KdV; compare also Ercolani and McKean [7]. Section 9.8 closes with a "composition rule." Write ϑ with both $t = [t_n : n \in \mathbb{Z} - 0]$ and x displayed, as in $\vartheta(t, x)$. Then $\vartheta(t+s, x) = \vartheta(t, x)[e^{t\mathbb{X}}\vartheta(s, \bullet)](\overline{x})$ in the scale for the composite flow. In this respect and others, too, the present ϑ is reminiscent of the theta function of McKean and Trubowitz [13] associated with the isospectral class of the quantum-mechanical oscillator $-D^2 + x^2 - 1$.

Section 9.9 describes some open questions.

9.2. Preparations

I collect here the necessary information about the CH hierarchy in its particular Hamiltonian format; most of this can be found in the literature already; see, for instance, Camassa and Holm [2] and Camassa, Holm, and Hyman [3].

---

[11] $A^{\mathfrak{p}}$ inverse is $A^{-\mathfrak{p}}$, in which $-\mathfrak{p}$ is $\mathfrak{p}$ with signature reversed.

9.2.1. Brackets. Let A, B, C, etc., be nice functions of m. The Poisson bracket of A and B is declared to be $[A, B] = \int \nabla A J \nabla B$ with the gradient $\nabla = \partial/\partial m$ and the skew operator $J = mD + Dm$. The vector field $\mathbb{X}$ associated with the Hamiltonian H is $\mathbb{X} : C \to [C, H]$; in particular, $\mathbb{X}m = J\partial H/\partial m$. For example, if $H = \frac{1}{2}\int mGm = \frac{1}{2}\int[(v')^2 + v^2]$,[12] then $\partial H/\partial m = Gm = v$, and you will recognize the flow $\partial m/\partial t = \mathbb{X}m = (mD + Dm)v$ as CH run backwards.

9.2.2. Spectral Problem and Trace. The space H^1 is equipped with the nonstandard norm $\|f\|_1^2 = \int[(f')^2 + \frac{1}{4}f^2]$. It is mapped $1:1$ onto H^0 by $(\frac{1}{2} - D) : f \to F$, $\|f\|_1^2 = \|F\|_0^2 \equiv \int F^2$, and you have the simple inverse map $K : F \to f = e^{x/2}\int_x^\infty e^{-y/2}F(y)dy$; similarly, $(\frac{1}{2} + D) : H^1 \to H^0$ is inverted by the transposed $K^\dagger = e^{-x/2}\int_{-\infty}^x e^{y/2}$, and $K^\dagger K = KK^\dagger = (\frac{1}{4} - D^2)^{-1} = \exp(-\frac{1}{2}|x - y|)$. Now the spectral problem for CH is $(\frac{1}{4} - D^2)f = \lambda m f$ for $f \in H^1$. Write $f = KF$ with $F \in H^0$. Then $K^\dagger m K F = K^\dagger(\frac{1}{4} - D^2)KF/\lambda = F/\lambda$. This makes sense because $\lambda = 0$ does not belong to the spectrum. It is to be proved that $K^\dagger m K : H^0 \to H^0$ is compact with simple spectrum $1/\lambda_n : n \in \mathbb{Z} - 0$ and finite absolute trace $\sum 1/|\lambda_n| \leq \int |m|$. The spectrum is indexed so that $\lambda_n < 0$ for $n < 0$ and $\lambda_n > 0$ for $n > 0$, with the understanding that λ_{-1}, λ_{-2}, etc. $= -\infty$ if no negative spectrum is present, as for $m > 0$, and so on.

Step 1. $|ff'| \leq f'^2 + \frac{1}{4}f^2$, so $\|f\|_\infty \leq \sqrt{2}\|f\|_1$ and

$$\int F_1 K^\dagger m K F_2 = \int f_1 m f_2 \leq 2\|f_1\|_1\|f_2\|_1 \int |m|,$$

whence $\|K^\dagger m K\|_0 \leq 2\int |m|$.

Step 2. $(K^\dagger m K)^2$ is symmetric and nonnegative (-definite), by inspection. It is also of trace class: in fact, for nice m,

$$\begin{aligned} \operatorname{sp}\left(K^\dagger m K\right)^2 &= \operatorname{sp}\left(KK^\dagger m KK^\dagger m\right) \\ &= \iint e^{-\frac{1}{2}|x-y|}m(y)e^{-\frac{1}{2}|x-y|}m(x)dx\,dy \leq \left(\int |m|\right)^2. \end{aligned}$$

This carries over to the general (summable) m, so $K^\dagger m K$ is compact; in particular, it has pure point spectrum $1/\lambda_n : n \in \mathbb{Z} - 0$, subject to $\sum \lambda_n^{-2} < \infty$, and this must be simple since $(\frac{1}{4} - D^2)f = \lambda m f$ cannot have two independent solutions of class H^1: Their Wronskian $[f_1, f_2] = f_1' f_2 - f_1 f_2'$ would be constant and must vanish at $\pm\infty$.

Step 3. This step is preparatory: The (normalized) eigenfunctions $F_n : n \in \mathbb{Z} - 0$ span the closure L of $K^\dagger m K H^0$. This is the whole of H^0 if m does not vanish on any set of positive measure (and only then) since $K^\dagger m K$ has no null space in this circumstance. Be that as it may, for the (likewise) normalized eigenfunctions $f = KF$, you have

$$\sum_{\mathbb{Z}-0} f_n \otimes f_n = K \sum_{\mathbb{Z}-0} F_n \otimes F_n K^\dagger = KPK^\dagger,$$

P being the projection on L, so that $\sum f_n^2 \leq KK^\dagger$ on diagonal $\equiv 1$. This fact will be helpful in the next step. Note in passing that $\sum f_n \otimes f_n = KK^\dagger = e^{-\frac{1}{2}|x-y|}$ and $\sum f_n^2 \equiv 1$ if $P = I$ $(L = H^0)$.

[12] $G = (1 - D^2)^{-1} = \frac{1}{2}\exp(-|x - y|)$ as before.

STEP 4. This step elicits the trace. Fix $n > 0$. Then

$$\frac{1}{\lambda_n} = \int F_n K^\dagger m K F_n = \int m\big(KF_n\big)^2 = \int m f_n^2 \leq \int m_+ f_n^2,$$

whence

$$\sum_{n>0} \frac{1}{\lambda_n} \leq \int m_+ \sum_{n>0} f_n^2 \leq \int m_+ \text{ and similarly } \sum_{n<0} \frac{1}{\lambda_n} \geq \int m_-.$$

The proof is finished, but note still that the signature of the spectrum is predictable: If $m \geq 0$, it is all positive; if $m \leq 0$, it is all negative; and if m is capable of both signatures, then there is an infinite number of eigenvalues of both signs, both $K^\dagger m_+ K$ and $K^\dagger m_- K$ being of infinite rank.

9.2.3. Constants of Motion. The spectrum represents a "complete" list of commuting constants of motion of the CH hierarchy. I do not enter into the completeness here since it is irrelevant to the sequel, but commutativity is important in view of the form of the (backwards) CH Hamiltonian $H = \frac{1}{2}\int mGm = \frac{1}{4}\,\mathrm{sp}(K^\dagger m K)^2 = \frac{1}{4}\sum \lambda_n^{-2}$ and the inherent possibility, pointed out in Sect. 9.1, of building up CH out of the individual flows associated with the Hamiltonians $H_n = 1/\lambda_n : n \in \mathbb{Z} - 0$ as they compose in a simpler manner if they commute. The proof is easy[13]: $\partial(1/\lambda_n)/\partial m = f_n^2$ and $\lambda_n J f_n^2 = \frac{1}{2}D(1-D^2)f_n^2$ by routine computation,[14] so

$$\begin{aligned}\lambda_j\left[\frac{1}{\lambda_i}, \frac{1}{\lambda_j}\right] &= \int f_i^2 \lambda_j J f_j^2 = \int f_i^2 \frac{1}{2} D(1-D^2) f_j^2 \\ &= -\int f_j^2 \frac{1}{2} D(1-D^2) f_i^2 = -\lambda_i \int f_j^2 J f_i^2 \\ &= \lambda_i \int f_i^2 J f_j^2 = \lambda_i \left[\frac{1}{\lambda_i}, \frac{1}{\lambda_j}\right]\end{aligned}$$

must vanish if $i \neq j$. The fact that the numbers $\lambda_n : n \in \mathbb{Z} - 0$ really are constants of motion for CH is now obvious from the form of $H = \frac{1}{4}\sum \lambda_n^{-2}$.

9.2.4. More Constants of Motion. Let H be any nice function of m and define (1) a flow $e^{t\mathbb{X}}$ by $\partial m/\partial t = \mathbb{X}m = J\partial H/\partial m$ in the standard scale x and (2) a new (Lagrangian) scale $\overline{x}$ by $\partial\,\overline{x}/\partial t = -(\partial H/\partial m)(\overline{x})$ with $\overline{x}(0,x) \equiv x$. Then $(e^{t\mathbb{X}}m)(\overline{x})\overline{x}'^2$ is a constant of motion $\equiv m(x)$, as you will readily check. This observation plays an essential role below. But what do these new constants of motion represent? They are "foreign" to the hierarchy since they do not commute with the spectrum, nor do they commute among themselves. In the special case of CH flow, they may be likened to the vorticity for two-dimensional, incompressible Eulerian flow because they have similar brackets. but this is both far fetched and unhelpful, and because none of it matters for the sequel, let's forget about it and press on.

9.3. Individual Flows

Let $H = 1/\lambda_0$ be the reciprocal of any eigenvalue of the spectral problem and let f_0 be the associated eigenfunction with $\|f_0\|^2 = \lambda_0 \int m f_0^2 = 1$. $\partial H/\partial m = f_0^2$ as noted in Sect. 9.2, so the flow is regulated by $\partial m/\partial t = \mathbb{X}m = (mD + Dm)f_0^2$, while the associated scale obeys $\partial\,\overline{x}/\partial t = -f_0^2(\overline{x})$ with $\overline{x}(0,x) \equiv x$.

[13] Here and below f_n is always normalized as in $\|f_n\|^2 = \int[(f_n')^2 + \frac{1}{4}f_n^2] = \lambda_n \int m f_n^2 = 1$.
[14] $J = mD + Dm$ as before.

9.3.1. Computing the Flow. It is desired to integrate this flow explicitly in terms of the initial values m^0 and f_0^0 of $m \equiv e^{t\mathbb{X}}m^0$ and $f \equiv e^{t\mathbb{X}}f_0$. The discussion is broken up into eight little steps, with a summary at the end.

STEP 1. $\mathbb{X}f_0 = -\lambda_0 f_0 \int_{-\infty}^x mf_0^2 + f_0(x)/2$. Let $\bullet$ denote the action of the vector field $\mathbb{X}$. A routine computation shows that the stated function $f_0^\bullet \equiv \mathbb{X}f_0$ solves $(\frac{1}{4} - D^2)f_0^\bullet = (\lambda_0^\bullet = 0) \times mf_0 + \lambda_0 m^\bullet f_0 + \lambda_0 m f_0^\bullet$, as it should. The added $f_0/2$ is killed. Its role is to keep $\|f_0\|^2 = \lambda_0 \int mf_0^2 \equiv 1$; Indeed, with $N = \lambda_0 \int mf_0^2$,

$$N^\bullet = \lambda_0 \int f_0^2(mD + Dm)f_0^2 + \lambda_0 \int m2f_0\Big[-\lambda_0 f_0 \int_{-\infty}^x mf_0^2 + \frac{f_0(x)}{2}\Big]$$

$$^{15} = 0 - N^2 + N = N(1-N),$$

of which the only solution with $N(0) = 1$ is $N \equiv 1$.

STEP 2. Now consider the scaled integral $I(\overline{x}) = \lambda_0 \int_{-\infty}^{\overline{x}} mf_0^2$ and look for a piece of luck!

$$\frac{d}{dt}I(\overline{x}) = \lambda_0 mf_0^2(\overline{x})\overline{x}^\bullet + \lambda_0 \int_{-\infty}^{\overline{x}} f_0^2(mD + Dm)f_0^2 - I^2 + I,$$

in which $-I^2 + I$ comes out just like $-N^2 + N$ in Step 1. But also $\overline{x}^\bullet = -f_0^2(\overline{x})$ and

$$\int_{-\infty}^{\overline{x}} f_0^2(mD + Dm)f_0^2 = \int_{-\infty}^{\overline{x}} \big(m'f_0^4 + 4mf_0^3 f_0'\big) = \int_{-\infty}^{\overline{x}} \big(mf_0^4\big)' = mf_0^4(\overline{x}),$$

so $I^\bullet = 1(1 - I)$, plain, and you can integrate back to obtain

$$I(t) = \frac{e^t I(0)}{1 + (e^t - 1)I(0)} \equiv \frac{\vartheta^\bullet}{\vartheta}$$

$$\text{with } \vartheta = 1 + \big(e^t - 1\big)I(0) = 1 + (e - 1)\lambda_0 \int_{-\infty}^x m^0 f_0^{02}.$$

This happy outcome should convince you of the utility of the scale $\overline{x}$.

STEP 3. This step is to differentiate the formula for I with respect to the original scale x, keeping in mind that $m(\overline{x})\overline{x}'^2 \equiv m^0(x)$, as per Sect. 9.2.4.[16] The routine computation produces

$$f_0(\overline{x}) = \frac{e^{t/2}\sqrt{\overline{x}'}\, f_0^0(x)}{\vartheta}$$

with the same function as in Step 2. Naturally, this may not make sense for all $t \geq 0$: ϑ and/or $\overline{x}'$ might vanish, but not for small times, so let's continue and see what happens.

STEP 4. Here we combine $\partial\,\overline{x}/\partial t = -f_0^2(\overline{x})$ with Step 3 to obtain

$$\frac{\partial \overline{x}}{\partial t} + \left[e^t \frac{f_0^{02}(x)}{\vartheta^2}\right] \frac{\partial\,\overline{x}}{\partial x} = 0.$$

In principle, this serves to determine $\overline{x}$ fully, $\overline{x}(0, x) \equiv x$ being known, but is not really helpful. A better way is to think of $e^{\overline{x}/2}$ as the update of the (improper) eigenfunction $e^{x/2}$ associated with eigenvalue $\lambda = 0$ and to imitate Steps 2 and 3 for the new integral

$$J(x) = \lambda_0 \int_{-\infty}^x mf_0 e^{y/2} = \int_{-\infty}^x e^{y/2}\left(\frac{1}{4} - D^2\right) f_0 = e^{x/2}\left(\frac{1}{2}f_0 - f_0'\right)(x).$$

[15] $mD + Dm$ is skew.

[16] The computation is the same.

STEP 5. This step imitates Step 2: In the scale $\overline{x}$,

$$\frac{d}{dt}J(\overline{x}) = -\lambda_0 m f_0^3(\overline{x}) e^{\overline{x}/2} + \lambda_0 \int_{-\infty}^{\overline{x}} m^\bullet f_0 e^{y/2} + \lambda_0 \int_{-\infty}^{\overline{x}} m f_0^\bullet e^{y/2}$$

$$= \lambda_0 \int_{-\infty}^{\overline{x}} \Big[-m' f_0^3 - 3m f_0^2 f_0' - \frac{1}{2} m f_0^3 + m' f_0^3 + 4m f_0^2 f_0'$$

$$- \lambda_0 m f_0 \int_{-\infty}^{y} m f_0^2 + \frac{m f_0}{2} \Big] e^{y/2}$$

$$\overset{17}{=} -\int_{-\infty}^{\overline{x}} \left[\left(\lambda_0 \int_{-\infty}^{y} m f_0^2 \right)' J + \lambda_0 \int_{-\infty}^{y} m f_0^2 \times J' \right] + \frac{1}{2} J(\overline{x})$$

$$= \left(\frac{1}{2} - \frac{\vartheta^\bullet}{\vartheta} \right) J(\overline{x}),$$

by Step 2, or what is the same, $[\ln J(\overline{x})]^\bullet = \frac{1}{2} - \vartheta^\bullet/\vartheta$ and so also

$$\int_{-\infty}^{\overline{x}} m f_0 e^{y/2} = \frac{J(\overline{x})}{\lambda_0} = \frac{e^{t/2} J^0(x)}{\lambda_0 \vartheta} = \frac{e^{t/2} \int_{-\infty}^{x} m^0 f_0^0 e^{y/2}}{1 + (e^t - 1)\lambda_0 \int_{-\infty}^{x} m^0 f_0^{02}}.$$

STEP 6. This step imitates Step 3 in differentiating the present display with respect to x: Assuming m^0, f_0^0, and $\overline{x}'$ do not vanish, it develops after some cancellation that

$$e^{\overline{x}/2} = -\frac{\sqrt{\overline{x}'}}{\vartheta} \times e^{x/2} \Big[1 + (e^t - 1)\lambda_0 \int_{-\infty}^{x} m^0 f_0^{02} - (e^t - 1)\lambda_0 f_0^0 \int_{-\infty}^{x} m^0 f_0^0 e^{y/2} \Big]$$

$$\overset{18}{=} \frac{\sqrt{\overline{x}'}}{\vartheta} \times e^{x/2} \times 1 + (e^t - 1) \int_{-\infty}^{x} \left[\left(\frac{1}{2} - D \right) f_0^0 \right]^2 \equiv \frac{\sqrt{\overline{x}'}}{\vartheta} \times e^{x/2} \times \vartheta_-.$$

STEP 7. Here we finish the determination of $\overline{x}$ by integrating

$$e^{\overline{x}/2} = \frac{\sqrt{\overline{x}'} e^{x/2} \vartheta_-}{\vartheta}$$

with the aid of a companion to the function ϑ_- introduced two lines above, to wit, $\vartheta_+ = 1 + (e^t - 1) \int_{-\infty}^{x} [(\frac{1}{2} + D) f_0^0]^2$. The identity $\vartheta^2 = \vartheta_- \vartheta_+ + \vartheta_-' \vartheta_+ - \vartheta_- \vartheta_+'$ is readily checked. Then

$$e^{-\overline{x}} \overline{x}' = e^{-x} \frac{\vartheta^2}{\vartheta_-^2} = e^{-x} \left[\frac{\vartheta_+}{\vartheta_-} - \left(\frac{\vartheta_+}{\vartheta_-} \right)' \right] = -\left(e^{-x} \frac{\vartheta_+}{\vartheta_-} \right)',$$

whence $e^{-\overline{x}} = e^{-x} \vartheta_+ / \vartheta_-$ plus an additive constant depending upon time alone. This must vanish if $\overline{x}(\infty) = +\infty$ since both ϑ_- and ϑ_+ reduce to e^t out there, so I propose $e^{\overline{x}} = e^x \vartheta_- / \vartheta_+$.

STEP 8. Now, if you like, this is only an educated guess, but if you will observe that $\vartheta_\pm$ cannot vanish and put the proposed $\overline{x} = x + \ln(\vartheta_- / \vartheta_+)$ back into the partial differential equation of Step 4, you will have complete success, $\overline{x}(0, x) \equiv x$ included. This ends the computation of the individual flow.

[17] The preliminary evaluation of J from the previous display is used.
[18] $\lambda_0 m_0 f_0 = (\frac{1}{4} - D^2) f_0$ is used to reduce the integrals to their final form.

Summary

I replace the initial values m^0 and f_0^0 by m and f_0, plain, and rewrite the recipe for the individual flow in terms of the three "theta functions":

$$\begin{aligned}\vartheta_- &= 1 + (e^t - 1)\int_{-\infty}^x \left[\left(\frac{1}{2} - D\right) f_0\right]^2 \\ &= 1 + (e^t - 1)\left[\int_{-\infty}^x \left[(f_0')^2 + \frac{1}{4} f_0^2\right] - \frac{1}{2} f_0^2(x)\right] \\ \vartheta &= 1 + (e^t - 1)\lambda_0 \int_{-\infty}^x m f_0^2 \\ &= 1 + (e^t - 1)\left[\int_{-\infty}^x \left[(f_0')^2 + \frac{1}{4} f_0^2\right] - f_0 f_0'(x)\right] \\ \vartheta_+ &= 1 + (e^t - 1)\int_{-\infty}^x \left[\left(\frac{1}{2} + D\right) f_0\right]^2 \\ &= 1 + (e^t - 1)\left[\int_{-\infty}^x \left[(f_0')^2 + \frac{1}{4} f_0^2\right] + \frac{1}{2} f_0^2(x)\right].\end{aligned}$$

The rules are

(1) $e^{\overline{x}} = \dfrac{e^x \vartheta_-}{\vartheta_+}$ or, what is more or less the same, $\overline{x}' = \dfrac{\vartheta^2}{\vartheta_- \vartheta_+}$,

(2) $(e^{t\mathbb{X}} m)(\overline{x}) = \dfrac{m(x)}{(\overline{x}')^2}$, and

(3) $(e^{t\mathbb{X}} f_0)(\overline{x}) = e^{t/2} f_0(x) \times \dfrac{\sqrt{\overline{x}'}}{\vartheta} = \dfrac{e^{t/2} f_0(x)}{\sqrt{\vartheta_- \vartheta_+(x)}}$.

It is a tiresome but comforting exercise to check directly that (3) solves the updated spectral problem. I do not repeat it here.

9.3.2. When ϑ Vanishes. This is all very fine if ϑ (and with it $\overline{x}'$) does not vanish: (1) and (3) still make sense, but (2) can go bad, though even that does not spoil the application to CH, as will be seen in Sect. 9.6. The fact that ϑ *can* vanish is simply illustrated by the extreme soliton/antisoliton case when $m\,dx$ reduces to point masses $\pm\, 1$ at $x = \mp 1$; see Sect. 9.7.2 for details. Then there are just two eigenvalues $\pm\Lambda$, with eigenfunctions $f_\pm$, so with $\lambda_0 = \lambda_{-1} = -\Lambda$, the function $\vartheta = 1 - (e^t - 1)\Lambda \int_{-\infty}^x m f_-^2$ reduces

to 1	for $x < -1$,
to $1 - \big(e^t - 1\big)\Lambda f_-^2(-1)$	for $-1 < x < +1$,[19]
to e^t	for $x > +1$.

Here, $f_-(-1)$ cannot vanish,[20] so ϑ vanishes for $|x| < 1$ at time $T = \ln[1 + 1/\Lambda f_-^2(-1)]$ and only then; at this moment, $\overline{x}' \equiv 0$ in the whole interval $|x| < 1$; i.e., $\overline{x}$ collapses the interval to a single point, ceasing to be a diffeomorphism. Then it recovers and there is no further trouble.

[19] $-\Lambda \int_{-\infty}^{+\infty} m f_-^2 = 1$.

[20] $f_-(-1) = 0$ implies $f_- \equiv 0$. which is not so.

9.3.3. Other Eigenfunctions. These evolve a little differently from f_0: In the notation of Steps 1 through 8 above, $\mathbb{X} f_n = -\lambda_n f_0 \int_{-\infty}^{x} m f_0 f_n$, without a corrective term, and you may imitate Sect. 9.3.1 to obtain

$$(e^{t\mathbb{X}} f_n)(\overline{x}) = \frac{\vartheta f_n(x) + (e^t - 1)(1 - \lambda_n/\lambda_0)^{-1}[f_0, f_n] f_0(x)}{\sqrt{\vartheta_- \vartheta_+(x)}},$$

normalization and all, where now the right-hand eigenfunctions have their initial values and $[f_0, f_n]$ is the Wronskian $f_0' f_n - f_0 f_n'$. It is a nice exercise to check this directly: $\|e^{t\mathbb{X}} f_n\|^2 = \lambda_n \int (e^{t\mathbb{X}} m f_n^2)(\overline{x}) d\overline{x}$ is maintained $(\equiv 1)$ and everything works out, as you may confirm with the help of the identity $[f_0, f_n] = (\lambda_n - \lambda_0) \int_{-\infty}^{x} m f_n f_0$.

9.4. Composite Flows

The several individual flows $e^{t\mathbb{X}_n}$ based upon the Hamiltonians $H_n = 1/\lambda_n$ commute since $[H_i, H_j] = 0$ $(i \neq j)$, and it is not hard to integrate the "composite" flow $e^{t\mathbb{X}} \equiv \prod e^{t_n \mathbb{X}_n}$ with many parameters $t = [t_n : n \in \mathbb{Z} - 0]$. In principle, this could be done already, by iteration of the rules of Sect. 9.3, but to express the result in an agreeable way it is simpler to start over, following the pattern of Sect. 9.3. Here, it is technically necessary that the individual parameters t_n vanish fast enough at $n = \pm\infty$. What that should mean will be clarified in Sect. 9.5. For the present, it is better to take t tame and not to worry. Temporarily. it is convenient to denote the action of $t\mathbb{X} \equiv \Sigma t_n \mathbb{X}_n$ by a black spot $(\bullet)$ and to write $m = e^{t\mathbb{X}} m^0$, as in the bulk of Sect. 9.3. A more explicit notation for subsequent use is adopted in the final summary.

STEP 1. $f_i^\bullet = \sum t_j \mathbb{X}_j f_i = \sum t_j [-\lambda_i f_j \int_{-\infty}^{x} m f_i f_j + f_i(x)/2 \text{ if } i = j]$ is immediate from Step 1 of Sect. 9.3.1 and from Sect. 9.3.3 above.

STEP 2. This step is to compute $I(\overline{x}) = \int_{-\infty}^{\overline{x}} m f \otimes f$ as in Step 2 of Sect. 9.3.1[21]:

$$\begin{aligned}
\big[I_{ij}(\overline{x})\big]^\bullet &= m f_i f_j(\overline{x}) \times \Big[\overline{x}^\bullet = -\sum t_k f_k^2(\overline{x})\Big] \\
&\quad + \int_{-\infty}^{\overline{x}} \Big[(m' + 2mD) \sum t_k f_k^2\Big] f_i f_j \\
&\quad + \int_{-\infty}^{\overline{x}} m f_j \sum t_k \Big[-\lambda_i f_k \int_{-\infty}^{y} m f_i f_k + \frac{f_i(y)}{2} \text{ if } i = k\Big] \\
&\quad + \int_{-\infty}^{\overline{x}} m f_i \sum t_k \Big[-\lambda_j f_k \int_{-\infty}^{y} m f_j f_k + \frac{f_j(y)}{2} \text{ if } j = k\Big] \\
&= (1) + (2) + (3) + (4) \\
&= \int_{-\infty}^{\overline{x}} \Big[m f_i f_j \Big(\sum t_k f_k^2\Big)' - m \sum t_k f_k^2 (f_i f_j)'\Big] \\
&= (1) + (2) - \int_{-\infty}^{\overline{x}} m \sum \lambda_i t_k f_j f_k - \int_{-\infty}^{y} m f_i f_k + \frac{1}{2} t_i \int_{-\infty}^{\overline{x}} m f_i f_j
\end{aligned}$$

[21] $f \otimes f$ is the matrix $[f_i f_j : i, j \in \mathbb{Z} - 0]$.

$$= (3) - \int_{-\infty}^{\overline{x}} m \sum \lambda_j t_k f_i f_k \int_{-\infty}^{y} m f_j f_k + \frac{1}{2} t_j \int_{-\infty}^{\overline{x}} m f_i f_j = (4)$$

$$= \int_{-\infty}^{\overline{x}} m \sum t_k f_k (f_k' f_i - f_i' f_k) f_j + \int_{-\infty}^{\overline{x}} m \sum t_k f_k (f_k' f_j - f_j' f_k) f_i$$
$$+ (3) + (4)$$

$$= \int_{-\infty}^{\overline{x}} m \sum t_k f_k f_j (\lambda_i - \lambda_k) \int_{-\infty}^{y} m f_i f_k$$
$$+ \int_{-\infty}^{\overline{x}} m \sum t_k f_k f_i (\lambda_j - \lambda_k) \int_{-\infty}^{y} m f_j f_k + (3) + (4).$$

Here $(3)+(4)$ cancel out the terms in λ_i, and λ_j, leaving only

$$- \sum t_k \lambda_k \int_{-\infty}^{\overline{x}} m \left[f_j f_k \int_{-\infty}^{y} m f_i f_{k+} f_i f_k \int_{-\infty}^{y} m f_j f_k \right]$$
$$+ \frac{1}{2} t_i \int_{-\infty}^{\overline{x}} m f_i f_j + \frac{1}{2} t_j \int_{-\infty}^{\overline{x}} m f_i f_j$$
$$= - \sum t_k \lambda_k I_{ik} I_{jk} + \frac{1}{2} (t_i I_{ij} + I_{ij} t_j);$$

in short, $I^{\bullet} = -It\lambda I + \frac{1}{2}(tI + It)$ in a self-explanatory notation.[22] This can be integrated back to obtain $I = \int_{-\infty}^{\overline{x}} mf \otimes f$ in terms of its initial value $I(0) = \int_{-\infty}^{x} m^0 f^0 \otimes f^0 \equiv M$, to wit, $I = e^{t/2} M [1 + (e' - 1)\lambda M]^{-1} e^{t/2}$. The verification is routine, assuming $\det[1 + (e^t - 1)\lambda M] \equiv \vartheta$ does not vanish; but watch out: The notation is a little tricky in that $(e^t)^{\bullet} = te^t$, you will also need to keep in mind that $A(I + BA)^{-1} = (I + AB)^{-1} A$ for any matrices A and B.

STEP 3. This step is to differentiate $I(\overline{x})$ by the original scale x to obtain the updated eigenfunctions $f = [f_n : n \in \mathbb{Z} - 0]$: With the abbreviation $(e^t - 1)\lambda \equiv C$ and the help of $m(\overline{x})(\overline{x}')^2 \equiv m^0(x)$, you find

$$\begin{aligned}
\frac{m^0(x)}{\overline{x}'} \times f \otimes f(\overline{x}) &= (mf \otimes f)(\overline{x})\overline{x}' = I' \\
&= e^{t/2} m^0 f^0 \otimes f^0(x)(1 + CM)^{-1} e^{t/2} \\
&\quad - e^{t/2} M (1 + CM)^{-1} C m^0 f^0 \otimes f^0 (1 + CM)^{-1} e^{t/2} \\
&\overset{23}{=} m^0(x) e^{t/2} (1 + MC)^{-1} f^0 \otimes f^0 (1 + CM)^{-1} e^{t/2},
\end{aligned}$$

which is to say that if $m^0(x)$, $\overline{x}'$, and $\vartheta = \det(1 + CM) = \det(1 + MC)$ do not vanish, then

$$f(\overline{x}) = \sqrt{\overline{x}'} e^{t/2} (1 + MC)^{-1} f^0(x).$$

This will be expressed in a more robust form presently.

STEP 4. This step would be to imitate the partial differential equation for the scale $\overline{x}$ from Step 4 of Sect. 9.3.1, but this is not really needed.

[22] t is the diagonal matrix $t_n : n \in \mathbb{Z} - 0$, and similarly for λ.

[23] $1 - M(1 + CM)^{-1} C = (1 + MC)^{-1}$, $(1 + CM)^{-1} C$ being symmetric.

STEP 5. The serious computation of the scale begins by evaluating $J(\overline{x}) \equiv \int_{-\infty}^{\overline{x}} mfe^{y/2}$. $[J(\overline{x})]^\bullet = \frac{1}{2}tJ - It\lambda J$, much as in Step 2, and now the known value $I = e^{t/2}M(1+CM)^{-1}e^{t/2}$ leads quickly to $J(\overline{x}) = e^{t/2}(1+MC)^{-1}J^0(x)$ with $J^0(x) \equiv \int_{-\infty}^{x} m^0 f^0 e^{y/2}$.

STEP 6. This step is to differentiate $J(\overline{x})$ with respect to x to obtain

$$\begin{aligned}\big[J(\overline{x})\big]' &= mf(\overline{x})e^{\overline{x}/2}\overline{x}' \\ &= -e^{t/2}\big[(1+MC)^{-1}m^0 f^0 \otimes f^0 C(1+MC)^{-1}\big]J^0(x) \\ &\quad + e^{t/2}(1+MC)^{-1}m^0 f^0 e^{x/2}.\end{aligned}$$

Now use $m(\overline{x})(\overline{x}')^2 \equiv m^0(x)$, Step 3, and a bit of cancellation to produce

$$e^{\overline{x}/2} = \sqrt{\overline{x}'}\big[e^{x/2} - Cf^0(x) \bullet (1+MC)^{-1}J^0\big].$$

This may be put into a better form with the help of the three theta functions:

$$\left.\begin{matrix}\vartheta_- \\ \vartheta \\ \vartheta_+\end{matrix}\right\} = \det\left[1 + \big(e^t - 1\big)\int_{-\infty}^{x}\left(f^{0\prime} \otimes f^{0\prime} + \frac{1}{4}f^0 \otimes f^0\right)\begin{cases} -\dfrac{1}{2}f^0 \otimes f^0(x) \\ -f^{0\prime} \otimes f^0(x) \\ +\dfrac{1}{2}f^0 \otimes f^0(x)\end{cases}\right],$$

the middle ϑ being nothing but $\det(1+CM) = \det(1+MC)$. Here ϑ_- and ϑ_+ cannot vanish, as will be seen in Sect. 9.5, and with the notation $F = \frac{1}{2}f - f'$ and the proviso that ϑ does not vanish either, you find

$$\begin{aligned}\frac{\vartheta_-}{\vartheta} &= \frac{\det\big[1 + CM - \big(e^t - 1\big)F^0 \otimes f^0\big]}{\det(1+CM)} \\ &= \det\big[1 - f^0 \otimes (1+CM)^{-1}C\lambda^{-1}F^0\big] \\ &= 1 - f^0 \bullet C(1+MC)^{-1}e^{-x/2}\int_{-\infty}^{x} m^0 f^0 e^{y/2} \\ &= e^{-x/2}\big[e^{x/2} - f^0 \bullet C(1+MC)^{-1}J^0\big] \\ &= \frac{e^{\overline{x}/2}e^{-x/2}}{\sqrt{\overline{x}'}}.\end{aligned}$$

This identity was the aim of Step 6.

STEP 7. This step prepares the identity $\vartheta^2 = \vartheta_-\vartheta_+ + \vartheta_-'\vartheta_+ - \vartheta_-\vartheta_\perp'$ for the final determination of $\overline{x}$. This is a purely algebraic fact, so it is harmless to assume the invertibility of

$$Q \equiv 1 + \big(e^t - 1\big)\int_{-\infty}^{x}\left[f^{0\prime} \otimes f^{0\prime} + \frac{1}{4}f^0 \otimes f^0\right].$$

Then

$$\left.\begin{matrix}\vartheta_- \\ \vartheta \\ \vartheta_+\end{matrix}\right\} = (\det Q) \times \begin{cases} 1 - \dfrac{1}{2}\xi \bullet Q^{-1}\eta \\ 1 - \xi \bullet Q^{-1}\eta' \\ 1 + \dfrac{1}{2}\xi \bullet Q^{-1}\eta\end{cases}$$

with $\xi = f^0$ and $\eta = (e^t - 1)f^0$. In this language, $\vartheta_-\vartheta_+ + \vartheta_-'\vartheta_+ - \vartheta_-\vartheta_+'$ is the product of $(\det Q)^2$ and

$$1 - \frac{1}{4}\big(\xi \bullet Q^{-1}\eta\big)^2 - \frac{1}{2}\big(\xi \bullet Q^{-1}\eta\big)'\left[1 + \frac{1}{2}\xi \bullet Q^{-1}\eta + 1 - \frac{1}{2}\xi \bullet Q^{-1}\eta\right]$$

$$= 1 - \frac{1}{4}\big(\xi \bullet Q^{-1}\eta\big)^2 - \xi' \bullet Q^{-1}\eta - \xi \bullet Q^{-1}\eta'$$

$$+ \xi \bullet Q^{-1}\left(\eta' \otimes \xi' + \frac{1}{4}\eta \otimes \xi\right)Q^{-1}\eta$$

$$= \big(1 - \xi' \bullet Q^{-1}\eta\big)\big(1 - \xi \bullet Q^{-1}\eta'\big) =^{24} \big(1 - \xi' \bullet Q^{-1}\eta\big)^2 = \vartheta^2.$$

Step 8. This step puts it all together: By Steps 6 and 7,

$$e^{-\overline{x}}\overline{x}' = e^{-x}\frac{\vartheta^2}{\vartheta_-^2} = e^{-x}\left[\frac{\vartheta_+}{\vartheta_-} - \left(\frac{\vartheta_+}{\vartheta_-}\right)'\right] = -\left(e^{-x}\frac{\vartheta_+}{\vartheta_-}\right)',$$

so $e^{-\overline{x}} = e^{-x}\vartheta_+/\vartheta_-$, up to an additive constant depending possibly upon time. This constant vanishes. One proof employs $\partial\overline{x}/\partial t_n = -f_n^2(\overline{x})$ and $\overline{x}(0,x) \equiv x$. A better way is to note that $\overline{x}(\infty) = \infty$ is inherited by the composite flow from the individual flows of Sect. 9.3. This kills the constant since, for tame t, you have

$$\vartheta_\pm(x = \infty) = \det\left[1 + \left(e^t - 1\right)\int_{-\infty}^{\infty}\left(\frac{1}{2}f^0 \pm f^{0\prime}\right) \otimes \left(\frac{1}{2}f^0 \pm f^{0\prime}\right)\right]$$

$$= \prod e^{t_n} > 0.$$

Summary

I change notation as in Sect. 9.3, replacing the initial values m^0 and f^0 by m and f, plain, and writing

$$\left.\begin{matrix}\vartheta_- \\ \vartheta \\ \vartheta_+\end{matrix}\right\} = \det\left[1 + \left(e^t - 1\right)\int_{-\infty}^{x}\left(f' \otimes f' + \frac{1}{4}f \otimes f\right)\left\{\begin{matrix} -\frac{1}{2}f \otimes f(x) \\ -f' \otimes f(x) \\ +\frac{1}{2}f \otimes f(x)\end{matrix}\right.\right].$$

In this language, the rules for composite flow are

(1) $e^{\overline{x}} = \dfrac{e^x\vartheta_-}{\vartheta_+}$ or, what is more or less the same, $\overline{x}' = \dfrac{\vartheta^2}{\vartheta_-\vartheta_+}$,

(2) $(e^{t\mathbb{X}}m)(\overline{x}) = \dfrac{m(x)}{(\overline{x}')^2} = \dfrac{m(x)\vartheta_-^2\vartheta_+^2}{\vartheta^4}$,

(3) $(e^{t\mathbb{X}}f)(\overline{x}) = \dfrac{e^{t/2}}{\sqrt{\vartheta_-\vartheta_+(x)}}\vartheta[1 + M(e^t - 1)\lambda]^{-1}f(x)$.

Note that $\overline{x} = x + \ln\vartheta_-/\vartheta_+$ always makes sense since $\vartheta_\pm$ do not vanish[25]; another nice feature is that $\overline{x}(-\infty) = -\infty$ and $\overline{x}(+\infty) = +\infty$ in view of $\vartheta_\pm(-\infty) = 1$ and $\vartheta_\pm(+\infty) = \prod e^{t_n}$. ϑ can and often will vanish as pointed out in Sect. 9.3 already, but (3) always makes sense,

[24] $Q^{-1}(e^t - 1)$ is symmetric and $\eta = (e^t - 1)\xi$.

[25] See Sect. 9.5.2.

$\vartheta[1+M(e^t-1)\lambda]^{-1}$ being a Fredholm co-factor; it is only the skimpy notation that is at fault. It is a nice exercise to recover the individual recipe of Sect. 9.3 from the present composite one.

9.5. Technicalities About Theta

The three theta functions ϑ_-, ϑ_+, and ϑ are written, here and below, as in the summary to Sect. 9.4. I collect some technical facts about them, anticipating the discussion of CH in Sects. 9.6 and 9.7: There, $t_n = -t/2\lambda_n$, $t \geq 0$ being the "running time" of the flow, so I do not apologize for taking $t_n = c/\lambda_n + O(1/\lambda_n^2)$ here. It is convenient to assume, in addition to the summability of m, that it be smooth and that m' be summable, too.

9.5.1. Convergence of $\vartheta_\pm$. Let $F = \frac{1}{2}f - f'$. Then $\vartheta_- = \det[1 + (e^t-1)P]$ with $P = \int_{-\infty}^{x} F \otimes F$, and from $P^2 \leq P \leq 1$, you conclude that the spectrum of $(e^t-1)P[(e^t-1)P]^\dagger \leq (e^t-1) \otimes (e^t-1)$ is majorized by $(e^{t_n}-1)^2 : n \in \mathbb{Z} - 0$ and that $(e^t-1)P$ has absolute trace $\mathrm{sp}\,|(e^t-1)P| \leq \sum |e^{t_n}-1| < \infty$, $e^t - 1$ being comparable to $1/\lambda$, which is summable as per Sect. 9.2. The convergence of the determinant ϑ_+ is similar.

9.5.2. Positivity of $\vartheta\pm$. Write $(e^t-1) = \sigma C^2$ will $\sigma_n = +1$ or -1 according as $t_n \geq 0$ or $t_n < 0$, and introduce the projections $\xi \to \xi^\pm = \frac{1}{2}(1 \pm \sigma)\xi$. $\vartheta_\pm$ is the determinant of $I + \sigma CPC$, and since CPC has finite absolute trace, it suffices to confirm the positivity of the associated quadratic form:

$$\begin{aligned}
\xi(1+\sigma CPC)\xi &= \xi^2 + \xi^+ CPC\xi - \xi^- CPC\xi \\
&= \xi^2 + \xi^+ CPC\xi^+ - \xi^- CPC\xi^- \\
&\geq \xi^2 - \xi^- C^2 \xi^- \\
&= \sum_{\sigma=+1} \xi_n^2 + \sum_{\sigma=-1} (I - C_n^2)\xi_n^2 \\
&= \sum_{\sigma=+1} \xi_n^2 + \sum_{\sigma=-1} e^{t_n}\xi_n^2.
\end{aligned}$$

The finer result $\vartheta_\pm \geq \prod e^{\min(t_n,0)}$ is pretty and not hard to prove. It may be reduced to a simple statement in the $(d < \infty)$-dimensional space:

$$\det\left[1 + \left(e^t - 1\right)P\right] \geq \prod e^{\min(t_n,0)}$$

for any $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$ and $P : \mathbb{R}^d \to \mathbb{R}^d$ subject to $0 < P = P^\dagger \leq 1$.

PROOF. Let $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, etc., be the standard basis of $\mathbb{R}^d$ and take $t_1 \geq 0$. Then $Q \equiv 1 + (e^{t_1}-1)e_1 \otimes Pe_1$ has determinant $1 + (e^{t_1}-1)e_1 \bullet Pe_1 \geq 1$ and inverse $Q^{-1} = 1 - (\det Q)^{-1}(e^{t_1}-1)e_1 \otimes Pe_1$. Now

$$\begin{aligned}
1 + \left(e^t - 1\right)P &= Q + \left(e^{t_2}-1\right)e_2 \otimes Pe_2 + \text{etc.} \\
&= Q \times \left[1 + Q^{-1}\left(e^{t_2}-1\right)e_2 \otimes Pe_2 + \text{etc.}\right] \\
&= Q \times P^{-1}\left[1 + \left(e^{t_2}-1\right)e_2 \otimes PQ^{-1}e_2 + \text{etc.}\right]^\dagger P,
\end{aligned}$$

so

$$\det\left[1 + \left(e^t-1\right)P\right] \geq \det\left[1 + \left(e^{t_2}-1\right)e_2 \otimes P'e_2 + \text{etc.}\right]$$

in which $P' \equiv PQ^{-1}$ is subject to the same conditions as for P: In fact,

$$\det Q \times \xi P'\eta = \xi \bullet P\big[1 + \big(e^t - 1\big)e_1 P e_1 - \big(e^{t_1} - 1\big)e_1 \otimes P e_1\big]\eta$$
$$= \xi \bullet P\eta + \big(e^{t_1} - 1\big)\big[(e_1 P e_1)(\xi P\eta) - \big(\xi P e_1\big)\big(\eta P e_1\big)\big],$$

from which the symmetry of P' is plain, while the quadratic form $\det Q \times \xi P'\xi$ is minorized by $\xi P\xi > 0$ and majorized by

$$\xi P\xi\big[1 + \big(e^{t_1} - 1\big)e_1 \bullet P e_1\big] = \xi P\xi \times \det Q \le \xi^2 \det Q;$$

in short, $t_1 \ge 0$ permits the reduction of $\det[1 + (e^t - 1)P]$ to the product of $\det Q \ge 1$ times a determinant $\det[1 + (e^t - 1)P']$ of the same type with $t_1 = 0$. This means that it is harmless to take $t_1, \dots, t_d < 0$ from the start. But then

$$\det\big[1 + \big(e^t - 1\big)P\big] = \det[1 - CPC] \quad \text{with } C^2 = 1 - e^t,$$

and the estimate

$$\xi(1 - CPC)\xi = \xi^2 - \xi CPC\xi \ge \xi^2 - (C\xi)^2 = \sum \big(1 - C_n^2\big)\xi_n^2 = \sum e^{t_n}\xi_n^2,$$

shows, by min/max, that the (positive) spectrum of $1 - CPC$ is minorized by $e^{t_n} : 1 \le n \le d$, whence

$$\det\big[1 + \big(e^t - 1\big)P\big] \ge \prod e^{\min(t_n, 0)},$$

as required. □

9.5.3. Convergence of ϑ. Let $F = \frac{1}{2}f - f'$ and note that

$$\vartheta = \det\left[1 + \big(e^t - 1\big)\left[\int_{-\infty}^{x}\left(f' \otimes f' + \frac{1}{4}f \otimes f\right) - f' \otimes f(x)\right]\right]$$
$$= \det\left[1 + \big(e^t - 1\big)\left[\int_{-\infty}^{x} F \otimes F + F \otimes f(x)\right]\right]$$
$$= \det\big[1 + \big(e^t - 1\big)P + \big(e^t - 1\big)F \otimes f\big]$$
$$= \det\big[Q + \big(e^t - 1\big)F \otimes f\big]$$

with P as before and $Q = 1 + (e^t - 1)P$. Here $\det Q = \vartheta_-$ does not vanish, so Q is invertible and ϑ is or ought to be

$$\vartheta_- \times \det\big[1 + Q^{-1}\big(e^t - t\big)F \otimes f\big] = \vartheta_- \times \big[1 + Q^{-1}\big(e^t - 1\big)F \bullet f\big].$$

That does the trick.

9.5.4. Differentiability of $\vartheta\pm$. Write $Q = 1 + (e^t - 1)P$ with $P = \int_{-\infty}^{x} F \otimes F$ and $F = \frac{1}{2}f \pm f'$ as in Sect. 9.1. Formally,

$$\frac{\vartheta'_\pm}{\vartheta_\pm} = \operatorname{sp}\big(Q^{-1}Q'\big) = \operatorname{sp} Q^{-1}\big(e^t - 1\big)F \otimes F$$
$$\overset{26}{=} \operatorname{sp}\big(e^t - 1\big)F \otimes F - \operatorname{sp}\big(e^t - 1\big)PQ^{-1}\big(e^t - 1\big)F \otimes F.$$

The second piece is not troublesome: It is $PQ^{-1}(e^t - 1)F \bullet (e' - 1)F$ and is controlled by[27] $|(e^t - 1)F|^2$; compare Sect. 9.3.

[26] $Q^{-1} = 1 - (Q - 1)Q^{-1}$.
[27] PQ^{-1} is bounded.

It remains to deal with the first piece: $\mathrm{sp}(e^t-1)F\otimes F = F\bullet(e^t-1)F$. This is where the summability of m' comes in: Indeed,

$$\frac{F_n^2}{\lambda_n} = -mf_n^2 + e^{-x}\int_{-\infty}^{x}(m'+2m)f_n^2e^y,$$

as you may check by application of $1+D$ to both sides, so

$$\sum\frac{F_n^2}{|\lambda_n|} \le |m| + \int_{-\infty}^{x}\big[|m'|+2|m|\big] < \infty$$

in view of $f^2\le 1$. I will need the actual value: $\sum F_n^2/\lambda_n = e^{-x}\int_{-\infty}^{x} me^y$. This is easily obtained from the display for F^2/λ if $m\ne 0$ on any interval: Then $f^2=1$ and

$$\sum\frac{F_n^2}{\lambda_n} = -m + e^{-x}\int_{-\infty}^{x}(m'+2m)e^y = e^{-x}\int_{-\infty}^{x}me^y.$$

Obviously, the proviso on m has nothing to do with the result.

9.5.5. Higher Derivatives of $\vartheta\pm$. I deal with ϑ_-. The discussion of ϑ_+ is similar. Now $F=\frac{1}{2}f-f'$ and formal differentiation of $\vartheta_-'/\vartheta_- = F\bullet Q^{-1}(e^t-1)F$ produces

$$\begin{aligned}\frac{\vartheta_-''}{\vartheta_-} &= \left(\frac{\vartheta_-'}{\vartheta_-}\right)^2 + F'\bullet Q^{-1}(e^t-1)F \\ &\quad - F\bullet\big[Q^{-1}(e^t-1)F\otimes FQ^{-1}(e^t-1)F\big] + F\bullet Q^{-1}(e^t-1)F' \\ &\equiv (1)+(2)-(3)+(4),\end{aligned}$$

in which you recognize (3) as $[(1)]^2$ and (2) as the same as (4), $Q^{-1}(e^t-1)$ being symmetric. But

$$\begin{aligned}(2)=(4) &= F'\bullet Q^{-1}(e^t-1)F \\ &= \left(-\frac{1}{2}F+\lambda mf\right)\bullet Q^{-1}(e^t-1)F \\ &= -\frac{1}{2}\left(\frac{\vartheta_-'}{\vartheta_-}\right) + mF\bullet Q^{-1}(e^t-1)\lambda f \\ \text{[28]}\quad &= -\frac{1}{2}\left(\frac{\vartheta_-'}{\vartheta_-}\right) + mF\bullet(e^t-1)\lambda f - mF\bullet(e^t-1)PQ^{-1}(e^t-1)\lambda f \\ &\equiv (5)+(6)-(7),\end{aligned}$$

in which (5) is fine, as is (7) since $|f|\le 1$, $|(e^t-1)F|<\infty$, and PQ^{-1} is bounded, leaving only $(6)=mF\bullet(e^t-1)\lambda f$ for serious consideration. Here, $(e^t-1)\lambda f = cf + 0(1/\lambda)$, permitting the reduction of (6) to $c\times F\bullet f$, plus a piece under the safe control of $|f|$ and $|F/\lambda|$. But what, in fact, is $F\bullet f$? The answer is contained in the identity

$$\lambda_n\int_a^b mF_nf_n = \int_a^b F_n\left(\frac{1}{2}+D\right)F_n = \frac{1}{2}\int_a^b F_n^2 + \frac{1}{2}F_n^2\Big|_a^b$$

[28] $Q^{-1} = 1-(Q-1)Q^{-1}$.

whence

$$\begin{aligned}\int_a^b mF\bullet f &= \Sigma\int_a^b mF_nf_n \\ &= \frac{1}{2}\int_a^b m\Sigma F_n^2/\lambda_n + \frac{1}{2}\Sigma F_n^2/\lambda_n\bigg|_a^b \\ &= \frac{1}{2}\int_a^b e^{-x}\int_\infty^x me^y + \frac{1}{2}e^{-x}\int_{-\infty}^X me^y\bigg|_a^b,\end{aligned}$$

with the interpretation

$$mF\bullet f = \frac{1}{2}e^{-x}\int_{-\infty}^x me^y + \frac{1}{2}\left(e^{-x}\int_{-\infty}^x me^y\right)' = \frac{1}{2}m.$$

Indeed, inspection of the proof shows that this rule produces a perfectly correct expression for ϑ_-'', and more: that ϑ_-'' is a continuous function of $x\in\mathbb{R}$. Quicker but more formal is the derivation $F\bullet f = (f/2-f')\bullet f = (1-D)f^2/2 = 1/2$ in case m does not vanish on any interval ($f^2\equiv 1$).

9.5.6. Derivatives of ϑ. Write $\vartheta = \vartheta_-[1+f\bullet Q^{-1}(e^t-1)F]$, as in Sect. 9.3, with $Q=1+(e^t-1)P$, $P=\int_{-\infty}^x F\otimes F$, and $F=\frac{1}{2}f-f'$, reducing the differentiability of ϑ to that of $(0)\equiv f\bullet Q^{-1}(e^t-1)F$. Formally,

$$\begin{aligned}\big(f\bullet Q^{-1}(e^t-1)F\big)' &= \left[f'=\frac{1}{2}f-F\right]\bullet Q^{-1}(e^t-1)F \\ &\quad - f\bullet\big[Q^{-1}(e^t-1)F\otimes FQ^{-1}(e^t-1)F\big] \\ &\quad + f\bullet Q^{-1}(e^t-1)\left[F'=-\frac{1}{2}F+\lambda mf\right] \\ &= -F\bullet Q^{-1}(e^t-1)F \\ &\quad - \big[f\bullet Q^{-1}(e^t-1)F\big]\times\big[F\bullet Q^{-1}(e^t-1)F\big] \\ &\quad + mf\bullet Q^{-1}(e^t-1)\lambda f \\ &\equiv -(1)-(0)\times(1)+m\times(2).\end{aligned}$$

But (2) takes care of itself and $(1)=\vartheta_-'/\vartheta_-$ as per Sect. 9.4—in short, ϑ' is fine.

ϑ'' requires only a little more talk: $(0)'$ was just dealt with, Sect. 9.5 takes care of the present $(1)'$ and $m'\times(2)$ is fine, leaving for further consideration only the product of m and

$$\begin{aligned}(2)' &= \big[f\bullet Q^{-1}(e^t-1)\lambda f\big]' \\ &= \left(\frac{1}{2}f-F\right)\bullet Q^{-1}(e^t-1)\lambda f \\ &\quad - f\bullet\big[Q^{-1}(e^t-1)F\otimes FQ^{-1}(e^t-1)\lambda f\big] \\ &\quad + f\bullet Q^{-1}(e^t-1)\lambda\left(\frac{1}{2}f-F\right)\end{aligned}$$

$$= \frac{1}{2} \times (2) - F \bullet Q^{-1}\big(e^t - 1\big)\lambda f - \big[f \bullet Q^{-1}\big(e^t - 1\big)F\big] \times$$
$$\times \big[F \bullet Q^{-1}\big(e^t - 1\big)\lambda f\big]$$
$$+ \frac{1}{2} \times (2) - f \bullet Q^{-1}\big(e^t - 1\big)\lambda F$$
$$\equiv (2) - (3) - (4) \times (3) - (5),$$

in which (2) needs no more discussion, (3) is related to the (2)=(4)=(5)−(6)−(7) of Sect. 9.5, and (4) takes care of itself, so only $m \times (5)$ is left. But, up to well-controlled connections, $m \times (5) = mf \bullet Q^{-1}(e^t - 1)\lambda F = c \times mf \bullet Q^{-1}F$, and $mf \bullet Q^{-1}F$ succumbs to the trick applied to $mF \bullet f = m/2$ in Sect. 9.5. Indeed, you have the formal identity

$$f \bullet Q^{-1}F = f \bullet F - f \bullet \big(e^t - 1\big)PQ^{-1}F,$$

in which the third piece is perfectly fine—in short, all is well with ϑ'', too, first on a formal level and then in fact. It is even a continuous function of $x \in \mathbb{R}$, just like ϑ''_- and ϑ''_+, as inspection of the proof will confirm.

Summary

If m is smooth, if both m and m' are summable, and if $t = c/\lambda + O(1/\lambda^2)$, then ϑ_-, ϑ_+, and ϑ are all of class $C^2(\mathbb{R})$. Obviously, these conditions could be much relaxed, but that is not my purpose here.

9.6. Application to Camassa-Holm

The Hamiltonian for Camassa-Holm run forward is[29]

$$H = -\frac{1}{2}\int mGm = -\frac{1}{4}\,\mathrm{sp}\,\big(K^\dagger mK\big)^2 = -\frac{1}{4}\Sigma\lambda_n^{-2},$$

so the corresponding vector field $\mathbb{X} : m \to -(mD + Dm)v$ is a combination of the individual fields $\mathbb{X}_n : m \to (mD + Dm)f_n^2$ of Sect. 9.3:

$$\mathbb{X} = \sum_{\mathbb{Z}-0} \frac{-1}{2\lambda_n}\mathbb{X}_n,$$

the flow being $e^{t\mathbb{X}} = \prod e^{-(t/2\lambda_n)\mathbb{X}_n}$ with "running time" $t \geq 0$. The parameters $t_n = -t/2\lambda_n$ obey the technical condition of Sect. 9.5, so if the initial value of $m = v - v''$ is smooth and if both m and m' are summable, then the three theta functions

$$\vartheta_\pm = \det\left[1 + \big(e^{-t/2\lambda} - 1\big)\int_{-\infty}^{x}\left(\frac{1}{2}f \pm f'\right) \otimes \left(\frac{1}{2}f \pm f'\right)\right]$$

and

$$\vartheta = \det\left[1 + \big(e^{-t/2\lambda} - 1\big)\lambda\int_{-\infty}^{x} mf \otimes f\right]$$

are twice differentiable in respect to x; they are also real analytic functions of the running time $t \geq 0$, as you will readily check. Here, the Lagrangian scale obeys $\partial\,\overline{x}/\partial t = (e^{t\mathbb{X}}v)(\overline{x})$, so from the general rules for composite flows in Sect. 9.4, you find $\overline{x} = x + \ln\vartheta_-/\vartheta_+$ and

$$\big(e^{t\mathbb{X}}v\big)(\overline{x}) = \left[\ln\frac{\vartheta_-}{\vartheta_+}\right]^\bullet = \frac{\vartheta_-^\bullet}{\vartheta_-} - \frac{\vartheta_+^\bullet}{\vartheta_+}.$$

[29] $G = (1 - D^2)^{-1} = \frac{1}{2}\exp(-|x - y|)$ as before.

This is of immediate interest: It makes sense for any $t \geq 0$ since $\vartheta_\pm$ does not vanish. What *does* occasionally break down is $(e^{t\mathbb{X}}v)'$: In fact, $\overline{x}' = \vartheta^2/(\vartheta_-\vartheta_+)$, as usual, and

$$\big(e^{t\mathbb{X}}v\big)'(\overline{x}) = \frac{\overline{x}'^\bullet}{\overline{x}'} = (\ln \overline{x}')^\bullet = \frac{2\vartheta^\bullet}{\vartheta} - (\ln \vartheta_-\vartheta_+)^\bullet,$$

in which the second piece is always nice, but $\vartheta^\bullet/\vartheta$ can and will be bad if ϑ vanishes. Be that as it may, the formalism of Sect. 9.4 provides an effective continuation of CH past any conventional "breakdown." Most of Sect. 9.6 is devoted to a more detailed description of what is really happening when ϑ vanishes.

9.6.1. CH Is Solved. The fact that $e^{t\mathbb{X}}v$ solves CH in its original form $[\partial v/\partial t - \partial v''/\partial t + 3vv' - 2v'v'' - vv''' = 0]$ is a tautology if $\overline{x}' \neq 0$: In the scale $\overline{x}$, $e^{t\mathbb{X}}m$ is nothing but $m(x)/(\overline{x}')^2$ and $\overline{x}^\bullet = (e^{t\mathbb{X}}v)(\overline{x})$, so

$$\begin{aligned}\frac{d}{dt}\big[\big(e^{t\mathbb{X}}m\big)(\overline{x})\big] + 2\big(e^{t\mathbb{X}}mv'\big)(\overline{x}) &= m(x)\big[\big(\overline{x}'\big)^{-2}\big]^\bullet + 2m(x)(\overline{x}')^{-2}\overline{x}^{\bullet\prime}(\overline{x}')^{-1}\\ &= m(x)(\overline{x}')^{-3}\big[-2\overline{x}'^\bullet + 2(\overline{x}-1)^\bullet\big]\\ &= 0,\end{aligned}$$

which is the same thing. This can be improved so that any vanishing of ϑ causes no difficulty. To illustrate the point in its simplest aspect, let's look at the (backward) Hamiltonian $H = \frac{1}{2}\int[(v')^2 + v^2]$: With $e^{t\mathbb{X}}v$ in place of v,

$$\begin{aligned}H &= \frac{1}{2}\int \big[\big(e^{t\mathbb{X}}v\big)'^2(\overline{x}) + \big(e^{t\mathbb{X}}v\big)^2(\overline{x})\big]d\overline{x}\\ &= \frac{1}{2}\int\left[\left(\frac{\overline{x}'^\bullet}{\overline{x}'}\right)^2 + (\overline{x}^\bullet)^2\right]\overline{x}'dx\\ &= \int\left[2\Big(\sqrt{\overline{x}'}\Big)^{\bullet 2} + \frac{1}{2}\big(\overline{x}^\bullet\big)^2\overline{x}'\right]dx\\ &= \int\left[2\left(\frac{\vartheta}{\sqrt{\vartheta_-\vartheta_+}}\right)^{\bullet 2} + \frac{1}{2}\left(\ln\frac{\vartheta_-}{\vartheta_+}\right)^{\bullet 2}\frac{\vartheta^2}{\vartheta_-\vartheta_+}\right]dx\end{aligned}$$

is a perfectly sensible integral ($\leq \infty$) for all $t \geq 0$.[30] This is constant in time, regardless of any vanishing of ϑ, being (real) analytic for $t \geq 0$ and constant when t is small ($\vartheta \neq 0$). Now back to CH, but write it in its Lagrangian form relative to the scale $\overline{x}$:

$$\frac{d}{dt}\big(e^{t\mathbb{X}}v\big)(\overline{x}) + \big(e^{t\mathbb{X}}p\big)'(\overline{x}) = 0,$$

i.e.,

$$\begin{aligned}\frac{d}{dt}\big[\big(e^{t\mathbb{X}}v\big)(\overline{x})\big] &= \frac{1}{2}e^{-\overline{x}}\int_{-\infty}^{\overline{x}} e^{\overline{y}}e^{t\mathbb{X}}\left[v^2 + \frac{1}{2}(v')^2\right](\overline{y})d\overline{y}\\ &\quad - \frac{1}{2}e^{\overline{x}}\int_{\overline{x}}^{\infty} e^{-\overline{y}}e^{t\mathbb{X}}\left[v^2 + \frac{1}{2}(v')^2\right](\overline{y})d\overline{y},\end{aligned}$$

[30] You may think line 3 is a little shaky but you can check line 4 directly from line 2.

which is to say

$$\left(\ln\frac{\vartheta_-}{\vartheta_+}\right)^{\bullet\bullet} = \frac{1}{2}e^{-x}\frac{\vartheta_+}{\vartheta_-}\int_{-\infty}^{x} e^{y}\frac{\vartheta_-}{\vartheta_+}\left[\left(\ln\frac{\vartheta_-}{\vartheta_+}\right)^{\bullet 2} + \frac{1}{2}\left(\ln\frac{\vartheta^2}{\vartheta_-\vartheta_+}\right)^{\bullet 2}\right]\frac{\vartheta^2}{\vartheta_-\vartheta_+}dy$$
$$-\frac{1}{2}e^{x}\frac{\vartheta_-}{\vartheta_+}\int_{x}^{\infty} e^{-y}\frac{\vartheta_+}{\vartheta_-}\left[\left(\ln\frac{\vartheta_-}{\vartheta_+}\right)^{\bullet 2} + \frac{1}{2}\left(\ln\frac{\vartheta^2}{\vartheta_-\vartheta_+}\right)^{\bullet 2}\right]\frac{\vartheta^2}{\vartheta_-\vartheta_+}dy.$$

It is needless to spell it out further: The integrals to the right are perfectly sensible after some cancellation of ϑ's above and below, and the identity is valid for all $t \geq 0$ by the same reasoning employed in respect to H: In short, $e^{t\mathbb{X}}v$ solves the Lagrangian form of CH that does not break down at all, regardless of any vanishing of ϑ!

9.6.2. No Breakdown. Constantin and Escher [4] proved that forward CH does not break down if m is positive (or negative), and their method extends to the case when m has positive part m_+ situated wholly to the right of its negative part m_- [private communication]. This fact is readily verified in the present format. The question is: Does ϑ vanish or not? Let m be ≤ 0 for $x \leq 0$ and ≥ 0 for $x \geq 0$ and observe that $-C_n^2 = (e^{-t/2\lambda_n} - 1)\lambda_n$ is negative for every $n \in \mathbb{Z} - 0$. $\vartheta = \det[1 - C\int_{-\infty}^{x} mf \otimes fC]$ is now seen to be: (a) equal to 1 at $-\infty$, (b) increasing for $x \leq 0$, and (c) decreasing for $x \geq 0$ until it vanishes (if it ever does). But, in fact, it cannot vanish: As $x \uparrow \infty$, $1 - C\int_{-\infty}^{x} mf \otimes fC$ decreases to $1 - C^2/\lambda = e^{-t/2\lambda} > 0$,[31] so ϑ stays positive:

$$\vartheta \geq \exp\left[-\left(\frac{t}{2}\right)\sum\frac{1}{\lambda_n}\right] = \exp\left[-\left(\frac{t}{2}\right)\int m\right].$$

Thus, the good behavior of $e^{t\mathbb{X}}v$ is assured, slope and all; indeed, it is even of class $C^3(\mathbb{R})$ under the present assumptions.

9.6.3. Breakdown. I had hoped to find an equally simple proof of the result of McKean [11] that breakdown must occur in the opposite case, i.e., if any part of m_+ lies to the left of some part of m_-. I did not succeed but record here a few simple indications. The condition $\vartheta \neq 0$ for the absence of breakdown means that the spectrum of $C\int_{-\infty}^{x} mf \otimes fC$ lies properly below 1 for every $t \geq 0$ and $x \in \mathbb{R}$.[32] This requires that the associated quadratic form be positive: $\xi^2 - \int_{-\infty}^{x} m|\sum \xi_n C_n f_n|^2 > 0$ unless $\xi = 0$. Now $C_n^2 \simeq t/2$ as n tends to $\pm\infty$, so the condition for no breakdown can be restated as

$$\int_{-\infty}^{x} m\left|\sum \xi_n f_n\right|^2 < \sum\frac{\xi_n^2}{C_n^2} \quad \text{for every } t \geq 0 \text{ and } x \in \mathbb{R},$$

and since C_n^2 increases as $t \uparrow \infty$, to λ_n or to $+\infty$ according as $n > 0$ or $n < 0$, you have the equivalent condition[33]

$$\int_{-\infty}^{x} mf^2 \leq \sum_{n>0}\frac{1}{\lambda_n}\langle f, f_n\rangle^2 \quad \text{for any } x \in \mathbb{R} \text{ and } f \in H^1.$$

This must be the same as to say that m_- lies wholly to the left of m_+. I could not see it in general, but here are two simple examples where I can.

[31] $\lambda\int_{-\infty}^{\infty} mf \otimes f =$ the identity.

[32] $C\int_{-\infty}^{x} mf \otimes fC$ is compact with finite absolute trace; compare Sect. 9.5.3.

[33] $\langle f_1, f_2\rangle$ is the inner product $\int_{-\infty}^{\infty}(f_1' f_2' + \frac{1}{4}f_1 f_2)$.

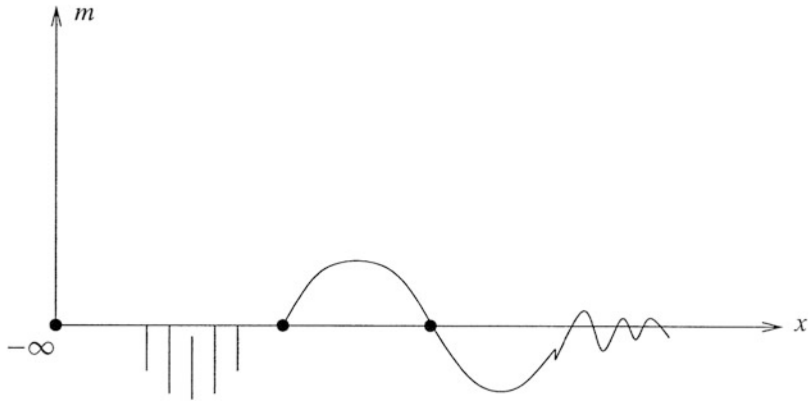

FIGURE 9.2.

EXAMPLE 9.6.1. Let m be positive for $x < 0$ and let it be negative someplace to the right. Then there are plenty of negative eigenvalues and the condition for no breakdown requires $\int_{-\infty}^{0} mf_{-1}^2 = 0$, and so also $f_{-1} \equiv 0$ unless $m \equiv 0$ for $x \leq 0$; i.e., if that is not the case, then breakdown must occur.

EXAMPLE 9.6.2. Now let m be as in Fig. 9.2 and take the artificial case indicated there when $mdx : x < 0$ reduces to a finite number of negative point masses $p_1, \ldots, p_5$, say, placed at $q_1 < q_2 < \cdots < q_5 < 0$. The negative part of m situated to the right of $x = 1$ ensures the existence of ≥ 6 (indeed, of infinitely many) negative eigenvalues, and you may choose numbers $c_1, \ldots, c_6$, not all zero, so that $f = c_1 f_{-1} + \cdots + c_6 f_{-6}$ vanishes at $x = q_1, \ldots, q_5$. Then the condition for no breakdown requires $\int_0^1 mf^2 = 0$, whence $f \equiv 0$ for $0 < x < I$ if $m > 0$ there. But this is not possible in the face of

$$\left[\frac{1}{m}\left(\frac{1}{4} - D^2\right)\right]^{j-1} f = \sum_{i=1}^{6} \lambda_{-i}^{j-1} c_i f_{-i} = 0 \quad \text{for } j = 1, \ldots, 6$$

and the nonvanishing of the Vandermonde determinant

$$\det\left[t_{-i}^{j-1} : 1 \leq i, j \leq 6\right] = \prod_{i>j} \left(\lambda_{-i} - \lambda_{-j}\right).$$

No individual eigenfunction can vanish on any interval, i.e., breakdown occurs.

9.6.4. A Differential Equation for ϑ. To investigate $(e^{t\mathbb{X}}v)(\overline{x})$ more fully, I need the curious fact that

$$\vartheta''\vartheta_- + \vartheta\vartheta''_- + \vartheta'_-\vartheta - \vartheta_-\vartheta' - \frac{m'}{m}\vartheta'\vartheta_- - 2\vartheta'\vartheta'_- = 0 \quad \text{when } m \neq 0.$$

This is valid for general parameters $t = [t_n : n \in \mathbb{Z} - 0]$ subject to the technical condition of Sect. 9.5 that $t = c/\lambda + 0(1/\lambda^2)$. Be warned that, for the present section only, t *is not the "running time" of Sects. 9.1, 9.2, and 9.3 above*. Ercolani and McKean [7, p. 536] suggested the identity. The proof can be made in a number of ways; indeed, it is all implicit in the computations of Sect. 9.5, but it would be tiresome to unscramble all that. Better to start over and, recognizing that the question is a purely algebraic one, to take t tame so as to finesse all technical considerations.

PROOF. Write $\vartheta = \det[Q + (e^t - 1)F \otimes f]$ with $Q = 1 + (e^t - 1)P$, $P = \int_{-\infty}^x F \otimes F$, and $F = \frac{1}{2}f - f'$ as in Sect. 9.5. Then

$$\vartheta = \vartheta_- \times \big[1 + f \bullet Q^{-1}(e^t - 1)F\big] \tag{0}$$

as before, and

$$\frac{\vartheta_-'}{\vartheta_-} = \operatorname{sp} Q^{-1}(e^t - 1)F \otimes F = F \bullet Q^{-1}(e^t - 1)F, \tag{1}$$

$$\begin{aligned} \frac{\vartheta_-''}{\vartheta_-} &= \left(\frac{\vartheta_-'}{\vartheta_-}\right)^2 + 2F \bullet Q^{-1}(e^t - 1)F' \\ &\quad - F \bullet \big[Q^{-1}(e^t - 1)F \otimes FQ^{-1}(e^t - 1)F\big] \\ &\overset{34}{=} 2F \bullet Q^{-1}(e^t - 1)\Big(-\frac{1}{2}F + \lambda m f\Big) \\ &= -\frac{\vartheta_-'}{\vartheta_-} + 2F \bullet Q^{-1}(e^t - 1)\lambda m f, \end{aligned} \tag{2}$$

$$\begin{aligned} \frac{\vartheta'}{\vartheta_-} &= \frac{\vartheta_-'}{\vartheta_-}\big[1 + f \bullet Q^{-1}(e^t - 1)F\big] + \big[f \bullet Q^{-1}(e^t - 1)F\big]' \\ &= F \bullet Q^{-1}(e^t - 1)F \times \big[1 + f \bullet Q^{-1}(e^t - 1)F\big] \\ &\quad + \Big(\frac{1}{2}f - F\Big) \bullet Q^{-1}(e^t - 1)F \\ &\quad - f \bullet \big[Q^{-1}(e^t - 1)F \otimes FQ^{-1}(e^t - 1)F\big] \\ &\quad + f \bullet Q^{-1}(e^t - 1)\Big(-\frac{1}{2}F + \lambda m f\Big) \\ &\overset{35}{=} f \bullet Q^{-1}(e^t - 1)\lambda m f, \end{aligned} \tag{3}$$

and

$$\begin{aligned} \frac{\vartheta''}{\vartheta_-} &= \frac{\vartheta'\vartheta_-'}{\vartheta_-^2} + \big[f \bullet Q^{-1}(e^t - 1)\lambda m f\big]' \\ &= \big[f \bullet Q^{-1}(e^t - 1)\lambda m f\big] \times \big[F \bullet Q^{-1}(e^t - 1)F\big] \\ &\quad + \Big[\frac{1}{2}f - F\Big] \bullet Q^{-1}(e^t - 1)\lambda m f \\ &\quad - f \bullet \big[Q^{-1}(e^t - 1)F \otimes FQ^{-1}(e^t - 1)\lambda m f\big] \\ &\quad + f \bullet Q^{-1}(e^t - 1)\lambda m' f \end{aligned} \tag{4}$$

[34] The rest cancels.
[35] The rest cancels.

$$+ f \bullet Q^{-1}\big(e^t - 1\big)\lambda m\left(\frac{1}{2}f - F\right)$$

$$= (5) + (6) + (7) + (8) + (9),$$

in which

$$(5) = \frac{\vartheta'\vartheta'_-}{\vartheta_-^2},$$

$$(6) = \frac{1}{2}\frac{\vartheta'}{\vartheta_-} - \frac{1}{2}\left(\frac{\vartheta''_-}{\vartheta_-} + \frac{\vartheta'_-}{\vartheta_-}\right) \quad \text{by (2) and (3)},$$

$$(7) = \left(1 - \frac{\vartheta}{\vartheta_-}\right) \times \frac{1}{2}\left(\frac{\vartheta''_-}{\vartheta_-} + \frac{\vartheta'}{\vartheta_-}\right) \quad \text{by (0) and (2)},$$

$$(8) = \frac{m'}{m}\frac{\vartheta'}{\vartheta_-} \quad \text{by (3)},$$

and

$$(9) = \frac{1}{2}\left(\frac{\vartheta'}{\vartheta_-}\right) - \frac{1}{2}\left(\frac{\vartheta''_-}{\vartheta_-} + \frac{\vartheta'_-}{\vartheta_-}\right)\frac{\vartheta}{\vartheta_-} + \frac{\vartheta'\vartheta'_-}{\vartheta_-^2} \quad \text{by (0), (1), (2), and (3).}$$

This is all obvious except for the last two pieces of the formula for (9), representing the product of m and $-f \bullet Q^{-1}(e^t - 1)\lambda F$, which requires a little trick. Let

$$R = 1 + \big(e^t - 1\big)\lambda \int_{-\infty}^{x} mf \otimes f = Q + \big(e^t - 1\big)F \otimes f(x) = Q(1 + \xi \otimes \eta)$$

with $\xi = Q^{-1}(e^t - 1)F$ and $\eta = f$, and suppose, what is harmless, that $\vartheta \neq 0$ so that R is invertible. Then $Q^{-1} = (1 + \xi \otimes \eta)R^{-1}$ and $R^{-1}(e^t - 1)\lambda$ is symmetric (which is the point of this maneuver), permitting you to compute as follows:

$$f \bullet Q^{-1}\big(e^t - 1\big)\lambda F = f \bullet (1 + \xi \otimes \eta)R^{-1}\big(e^t - 1\big)\lambda F$$

$$= F \bullet R^{-1}\big(e^t - 1\big)\lambda(1 + \eta \otimes \xi)f$$

$$\overset{36}{=} F \bullet \left[1 - \frac{\xi \otimes \eta}{1 + \xi \bullet \eta}\right]Q^{-1}\big(e^t - 1\big)\lambda(1 + \eta \otimes \xi)f.$$

This is reduced, after a little manipulation, to

$$F \bullet Q^{-1}\big(e^t - 1\big)\lambda f \times \big[1 + f \bullet Q^{-1}\big(e^t - 1\big)F\big]$$

$$- F \bullet Q^{-1}\big(e^t - 1\big)F \times f \bullet Q^{-1}\big(e^t - 1\big)\lambda f.$$

The evaluation of (9) now follows from (1), (2), and (3). It remains only to put (4)= (5) + $\cdots$ + (9) together to obtain the differential equation for ϑ stated at the start. ☐

9.6.5. Application to the Roots of $\vartheta = 0$. To make life easy, let $m(x) = 0$ have only a finite number of simple roots. It is to be proved that the roots of $\vartheta(x) = 0$ are likewise isolated and either double or simple according as $m(x) = 0$ there or not.

Proof. Let $\vartheta = 0$ have a double root at a point where $m \neq 0$. Then, by the differential equation of Sect. 9.6.4, $\vartheta \equiv 0$ out to the next root of $m = 0$ to the left or right. This cannot happen if the way is clear ($m \neq 0$) out to $-\infty$ or to $+\infty$ since $\vartheta(-\infty) = 1$ and $\vartheta(+\infty) = \det[1 + (e^t - 1)] = \prod e^{t_n} > 0$. Thus, $m = 0$ must have a root at $x = 0$, say, where ϑ,

[36] $(1 + \xi \otimes \eta)^{-1} = 1 - (1 + \xi \bullet \eta)^{-1}\xi \otimes \eta$.

ϑ', and ϑ'' all vanish. This cannot happen either: The root of m is simple, so that $m'/m - 1/x$ is smooth nearby; then the differential equation of Sect. 9.6.4 above takes the form

$$\left(\frac{\vartheta'}{x}\right)' = \frac{\vartheta''}{x} - \frac{\vartheta'}{x^2} = A\frac{\vartheta'}{x} + B\frac{\vartheta}{x}$$

with continuous

$$A = \frac{m'}{m} - \frac{1}{x} + 1 + \frac{2\vartheta'_-}{\vartheta_-} \quad \textit{and} \quad -B = \frac{\vartheta''_-}{\vartheta_-} + \frac{\vartheta'_-}{\vartheta_-},$$

and $\vartheta'/x = o(1)$ implies

$$\left|\frac{\vartheta'}{x}\right| \leq \int_0^x \left|A\frac{\vartheta'}{y} + \frac{B}{y}\int_0^y \vartheta'\right| dy \leq \text{constant} \times \int_0^x \left|\frac{\vartheta'}{y}\right| dy.$$

This forces ϑ'/x and, with it, ϑ itself to vanish in a whole neighborhood of $x = 0$; in short, the vanishing of ϑ propagates across any intervening roots of $m = 0$ all the way out to $\pm\infty$, contradicting $\vartheta(\pm\infty) > 0$; i.e., the original root of $\vartheta = 0$ must have been simple. Now let $\vartheta = 0$ at a point where $m = 0$ (simply). Then $m'\vartheta'\vartheta_- = 0$ by the differential equation multiplied by m, so $\vartheta' = 0$, too, and the vanishing of ϑ'' is ruled out by the same reasoning as before; i.e., the root is double. This is all I need for the sequel. □

9.6.6. How $(e^{t\mathbb{X}}v)'$ Breaks Down. This takes place when $\vartheta = 0$ in accordance with

$$\left(e^{t\mathbb{X}}v\right)'(\overline{x}) = \left(\ln\frac{\vartheta^2}{\vartheta_-\vartheta_+}\right)^\bullet = \frac{2\vartheta^\bullet}{\vartheta} + \text{nice stuff.}$$

I keep the assumption of Sect. 9.6.5 that $m(x) = 0$ has only simple roots. If $\vartheta = 0$ and $m \neq 0$, then $\vartheta' \neq 0$—i.e., the root is simple—but what about $\vartheta^\bullet$? The fact that $(e^{t\mathbb{X}}m)(\overline{x})(\overline{x}')^2 \equiv m(x)$ is a constant of motion provides the answer:

$$\left(e^{t\mathbb{X}}v\right)''(\overline{x}) = \left(\ln\frac{\vartheta^2}{\vartheta_-\vartheta_+}\right)^{\bullet\prime} \times \frac{1}{\overline{x}'} = \left[\frac{2\vartheta^\bullet}{\vartheta} - \left(\ln\vartheta_-\vartheta_+\right)^\bullet\right]'\frac{\vartheta_-\vartheta_+}{\vartheta^2}$$

so, putting the root at $x = 0$, say, you have

$$\begin{aligned} m(x) &= \left[e^{t\mathbb{X}}v - \left(e^{t\mathbb{X}}v\right)''\right](\overline{x}) \times (\overline{x}')^2 \\ &= \left[\left(\ln\frac{\vartheta_-}{\vartheta_+}\right)^\bullet - \left[\frac{2\vartheta^\bullet}{\vartheta} - \left(\ln\vartheta_-\vartheta_+\right)^\bullet\right]'\frac{\vartheta_-\vartheta_+}{\vartheta^2}\right] \times \frac{\vartheta^4}{\vartheta_-^2\vartheta_+^2} \\ &\simeq \frac{-2}{\vartheta_-\vartheta_+}\left[\vartheta^{\bullet\prime}\vartheta - \vartheta^\bullet\vartheta'\right] \quad \text{near the root,} \\ &= \frac{2\vartheta^\bullet\vartheta'}{\vartheta_-\vartheta_+} \quad \text{at the root;} \end{aligned}$$

i.e., $\vartheta^\bullet\vartheta'(0) = \frac{1}{2}m(0)\vartheta_-\vartheta_+(0) \neq 0$ has the same sign as $m(0)$. The behavior of $e^{t\mathbb{X}}v$ is clarified thereby: Taking $\overline{x}(0) = 0$ for ease of writing, you find

$$\overline{x}' = \frac{\vartheta^2}{\vartheta_-\vartheta_+} \simeq \frac{x^2\left[\vartheta'(0)\right]^2}{\vartheta_-\vartheta_+(0)} \quad \text{so that} \quad \overline{x} - \simeq \frac{x^3}{3}\frac{\left[\vartheta'(0)\right]^2}{\vartheta_-\vartheta_+(0)}$$

and

$$\big(e^{t\mathbb{X}}v\big)'(\overline{x}) \simeq \frac{2\vartheta^\bullet}{\vartheta} \simeq \frac{m(0)\vartheta_-\vartheta_+}{\big[\vartheta'(0)\big]^2 x} \simeq \text{a positive multiple of } m(0)(\overline{x})^{-1/3}$$

whence $(e^{t\mathbb{X}}v)(\overline{x})$ is of the approximate form $a + bm(0)(\overline{x})^{2/3}$ with $b > 0$, i.e., a cusp is seen rising/falling to the left/right of $x = 0$ in case $m(0) < 0$; in particular, up to uninteresting positive multipliers,

$$\big[(e^{t\mathbb{X}}v)'(\overline{x})\big]^2\overline{x}' \simeq (\overline{x})^{-2/3}\vartheta^2 \simeq x^{-2} \times x^2 = 1$$

is summable across the root, in agreement with the fact that $H = \frac{1}{2}\int[(v')^2v^2]$ is a constant of motion.

Now let $\vartheta = 0$ have a double root at $x = 0$. Then $m(0) = 0$, $\vartheta'(0) = 0$, $\vartheta''(0) \neq 0$, and

$$\begin{aligned} m'(x) &= \frac{2}{\vartheta_-\vartheta_+}\big[\vartheta''\vartheta^\bullet - \vartheta^{\bullet\prime}\vartheta'\big] + o(1) \quad \text{near the root} \\ &= \frac{2\vartheta^\bullet\vartheta''}{\vartheta_-\vartheta_+} \quad \text{at the root} \end{aligned}$$

by the preceding display for m, whence $\vartheta^\bullet\vartheta''(0) = \frac{1}{2}m'(0)\vartheta_-\vartheta_+(0) \neq 0$ has the same sign as $m'(0)$. Besides, up to harmless positive multipliers, $\overline{x}' \simeq x^4$, so that $\overline{x} \simeq x^5$, $(e^{t\mathbb{X}}v)(\overline{x}) \simeq 2\vartheta^\bullet/\vartheta \simeq m'(0)x^{-2} \simeq m'(0)(\overline{x})^{-2/5}$, and $e^{t\mathbb{X}}v(\overline{x}) \simeq a + bm'(0)(\overline{x})^{3/5}$ exhibits a slightly different type of cusp with summable

$$\big[\big(e^{t\mathbb{X}}v\big)'(\overline{x})\big]^2\overline{x}' \simeq (\overline{x})^{-4/5}x^4 \simeq x^{-4} \times x^4 = 1,$$

as before.

9.7. Breakdown and Soliton Train

I want to follow the breakdown of CH in as much detail as I currently know how in the simple case when m is odd and positive/negative to the left/right of the origin with $m'(0) < 0$.

9.7.1. Solitons Reviewed. CH has "solitons" of the form

$$(e^{t\mathbb{X}}v(x) = (p/2)e^{-|x-q-(p/2)t|},$$

moving to the right or left, according as $p > 0$ or $p < 0$, and it is reasonable to expect that, as $t \uparrow \infty$, quite a general solution will separate out into a "train" of such solutions, moving at different speeds and so widely separated, with a little trailing "radiation" behind, i.e., to a first approximation.

$$e^{t\mathbb{X}}v\big)(x) \simeq \sum \frac{p_n}{2}e^{-|x-q_n-(p_n/2)t|}.$$

In the "pure soliton" case, the initial m is a sum of (signed) point masses p located at points q, and the initial v is an exact sum $\sum(p_n/2)e^{-|x-q_n|}$ of this type. Remarkably enough, $e^{t\mathbb{X}}v$ retains this shape under the flow, only now the p's and q's move in accordance with the classical Hamiltonian flow $q^\bullet = \partial H/\partial p$, $p^\bullet = -\partial H/\partial q$ with $H = -\frac{1}{4}\sum p_ip_je^{-|q_i-q_j|}$[2, 3]. I omit the verification in favor of a formal discussion of what happens as $t \uparrow \infty$. The conventional wisdom already quoted has it that solitons ought to "disperse," i.e., to move at different speeds, so that $|q_i - q_j|$ should be large if $i \neq j$, with the following effect on the constants of motion[37]:

$$\sum \frac{1}{\lambda_n} = \operatorname{sp} K^\dagger m K = \operatorname{sp} K K^\dagger m = \int m = \sum p_n,$$

$$\sum \frac{1}{\lambda_n^2} = \operatorname{sp} \left(K^\dagger m K\right)^2 = \operatorname{sp} \left(K K^\dagger m K K^\dagger m\right) = \sum p_i p_j e^{-|q_i - q_j|} \simeq \sum p_n^2,$$

$$\sum \frac{1}{\lambda_n^3} = \operatorname{sp} \left(K^\dagger m K\right)^3 = \sum p_i p_j p_k e^{-\frac{1}{2}|q_i - q_j|} e^{-\frac{1}{2}|q_j - q_k|} e^{-\frac{1}{2}|q_k - q_i|} \simeq \sum p_n^3,$$

and so on. Thus, the p's must tend to the $1/\lambda$'s (in some order), $H \simeq -\frac{1}{4}\sum p_n^2$, and $q^\bullet \simeq -\frac{1}{2}p \simeq -1/2\lambda$, whence

$$\left(e^{t\mathbb{X}}v\right)(x) \simeq \sum \frac{1}{2\lambda_n} e^{-|x - q_n - t/2\lambda_n|}$$

with some new constants q, these being in the nature of "phase shifts." Beals, Sattinger, and Szmigielski [1] present the full computation in their format; compare Sect. 9.7.6 below.

9.7.2. Two Solitons. It is instructive to see exactly what happens in the "soliton/antisoliton" collision when $m\,dx$ is a positive mass $p(0)$ placed at $-q(0) < 0$ and a symmetrically disposed negative mass $-p(0)$ placed at $+q(0)$. The flow of n solitons $\sum(p_i/2)\overline{e}|x - q_i|$ is regulated by $q^\bullet = -\partial H/\partial p$ and $p^\bullet = +\partial H/\partial q$ with Hamiltonian $-H = -\frac{1}{4}\sum p_i p_j \exp(-|q_i - q_j|)$,[38] so for $n = 2$, $q_1 = -q$, $p_1 = +p$, $q_2 = +q$, and $p_2 = -p$ you find (1) $q^\bullet = -\frac{p}{2}(1 - e^{-2q}) = -H/p$ and (2) $p^\bullet = \frac{1}{2}p^2 e^{-2q} = \frac{1}{2}p^2 - H$ with $H = \frac{1}{2}p^2(1 - e^{-2q})$, the symmetries $q_2 = -q_1$ and $p_2 = -p_1$ being maintained. (2) is integrated to obtain

$$p(t) = \sqrt{2H}\frac{1 + e^{(t-T)\sqrt{2H}}}{1 - e^{(t-T)\sqrt{2H}}},$$

T being the "breakdown time." At this moment, $q = 0$ and $p = \infty$, so (2) implies

$$T = \int_{p(0)}^{\infty} \frac{dp}{\frac{1}{2}p^2 - H} = \frac{1}{\sqrt{2H}} \ln \frac{p(0) + \sqrt{2H}}{p(0) - \sqrt{2H}} = \frac{1}{\sqrt{2H}} \ln \frac{1 + \sqrt{1 - e^{2q(0)}}}{1 - \sqrt{1 - e^{2q(0)}}}.$$

(1) can now be integrated to obtain[39]

$$q(t) = \ln \operatorname{ch} \left[(t - T)\sqrt{\frac{H}{2}}\right].$$

Taking the formulae literally, you see that q decreases to 0 at time T and then increases, while p increases to $+\infty$ and comes back with the opposite sign—in short, the two solitons pass through each other.[40]

The function $(e^{t\mathbb{X}}v)(x) = \frac{p}{2}e^{-|x+q|} - \frac{p}{2}e^{-|x-q|}$ is, of course, odd: For $x > q$ it is $-pe^{-x}\operatorname{sh} q$, while for $0 < x < q$ it is $-pe^{-q}\operatorname{sh} x$, so that, as $t \uparrow T$, $|e^{t\mathbb{X}}v|$ is comparable to $pq \simeq H/p \simeq H/p \downarrow 0$; likewise, for $x > q$, $(e^{t\mathbb{X}}v)' = -e^{t\mathbb{X}}v$ vanishes as $t \uparrow T$, but $(e^{t\mathbb{X}}v)'(0) = -pe^{-q} \downarrow -\infty$, as it should. Next comes the scale $\overline{x}$: If $\overline{x} = q$, then $\overline{x}^\bullet = (e^{t\mathbb{X}}v)(\overline{x}) = -pe^{-q}\operatorname{sh} q = -\frac{p}{2}(1 -$

---

[37] $K K^\dagger = e^{-\frac{1}{2}|x-y|}$.

[38] See [2] and/or [3].

[39] $q(0) = \ln \operatorname{ch}(T\sqrt{H/2})$ is used; it comes from the second version of T.

[40] If you think of the two solitons as particles, you might prefer to say they "bounce off each other," but all such descriptions are just a manner of speaking, not the thing itself.

$e^{-2q}) = q^\bullet$, so $\overline{x}[t, q(0)] = q(t)$; in particular, $\overline{x}(t,x) > q(t)$ if $x > q(0)$ and only then, whence

$$\overline{x}^\bullet = -p\,\mathrm{sh}\,q e^{-\overline{x}} = \sqrt{\frac{H}{2}}\,\mathrm{sh}\left[(t-T)\sqrt{\frac{H}{2}}\right]e^{-\overline{x}} \quad \text{if } \infty > x > q(0)$$

$$= -pe^{-q}\,\mathrm{sh}\,\overline{x} = 2\sqrt{\frac{H}{2}}\frac{1}{\mathrm{sh}\left[(t-T)\sqrt{\frac{H}{2}}\right]}\mathrm{sh}\,\overline{x} \quad \text{if } 0 < x < q(0),$$

which can be integrated, too:

$$e^{\overline{x}} = e^x + \mathrm{ch}\left[(t-T)\sqrt{\frac{H}{2}}\right] - \mathrm{ch}\left[T\sqrt{\frac{H}{2}}\right] \quad \text{if } \infty > x > q(0)$$

$$\frac{e^{\overline{x}}-1}{e^{\overline{x}}+1} = \frac{e^x-1}{e^x+1}\left[\frac{1-e^{(t-T)\sqrt{H/2}}}{1+e^{(t-T)\sqrt{H/2}}}\,\frac{1+e^{T\sqrt{H/2}}}{1-e^{T\sqrt{H/2}}}\right]^2 \quad \text{if } 0 < x < q(0).$$

$\overline{x}$ ceases to be a proper scale at breakdown: At this moment (and only then) it collapses the whole interval $|x| \le q(0)$ to the single point $\overline{x} = 0$. Otherwise, it is perfectly fine: $\overline{x}'(t,x) = 0$ only if $t = T$ and $|x| < q(0)$, and likewise for ϑ, it is needless to know the actual values of $\vartheta_\pm$ and ϑ, but observe that if $\pm\Lambda$ are the two eigenvalues of the spectral problem and if $f_\pm$ are the values of the associated eigenfunctions at $x = -q(0)$, then, with $p(0) = p$ for short,

$$\vartheta(x=0) = \det\begin{bmatrix} 1 + \left(e^{-t/2\Lambda}-1\right)\Lambda p f_+^2 & \left(e^{-t/2\Lambda}-1\right)\Lambda p f_- f_+ \\ -\left(e^{t/2\Lambda}-1\right)\Lambda p f_- f_+ & 1 - \left(e^{-t/2\Lambda}-1\right)\Lambda p f_-^2 \end{bmatrix}$$

$$= 1 + \left(e^{-t/2\Lambda}-1\right)\Lambda p f_+^2 - \left(e^{t/2\Lambda}-1\right)\Lambda p f_-^2$$

decreases steadily to $-\infty$; compare Sect. 9.7.5 below. Now back to the main story with (smooth) odd m, positive/negative to the left/right of $x = 0$, with $m'(0) < 0$.

9.7.3. Parity of ϑ. The parity of m implies that the spectrum $\lambda_n : n \in \mathbb{Z} - 0$ is also odd in respect to n and that $f_n(-x) = f_{-n}(x)$. It follows that ϑ is an even function of x: Indeed.

$$\vartheta(-x) = \det\left[1 + \left(e^{-t/2\lambda_i}-1\right)\lambda_i \int_{-\infty}^{-x} m f_i f_j\right]$$

$$= \det\left[1 - \left(e^{-t/2\lambda_i}-1\right)\lambda_i \int_{x}^{\infty} m f_{-i} f_{-j}\right]$$

$$\overset{41}{=} \det\left[1 + \left(e^{t/2\lambda_i}-1\right)\lambda_i\left(1 - \lambda_i \int_{-\infty}^{x} m f_i f_j\right)\right]$$

$$= \det\left[e^{t/2\lambda_i} + \left(1 - e^{t/2\lambda_i}\right)\int_{-\infty}^{x} m f_i f_j\right]$$

$$= \prod_{\mathbb{Z}-0} e^{t/2\lambda_n} \det\left[1 + \left(e^{-t/2\lambda_i}-1\right)\int_{-\infty}^{x} m f_i f_j\right]$$

$$= \vartheta(x).$$

The same type of manipulation shows that $\vartheta_+(-x) = \vartheta_-(x)$.

[41] Change n to $-n$ and use $\lambda \int_{-\infty}^{\infty} m f \otimes f = 1$.

9.7.4. Roots of ϑ $(x = 0)$. Write $\vartheta(0) = \det(1 - CMC)$ with $M = \int_{-\infty}^{0} mf \otimes f$ and $C^2 = (1 - e^{-t/2\lambda})\lambda > 0$. Here CMC is positive since $m(x) > 0$ to the left of 0; it is also of finite trace

$$\int_{-\infty}^{0} m \sum C_n^2 f_n^2 \le \max_{\mathbb{Z}-0} C_n^2 \times \int_{-\infty}^{0} m,$$

and so you may write $\vartheta(0) = \prod_{n=1}^{\infty}(1 - \mu_n)$, in which $\mu_1 \ge \mu_2 \ge$ etc. are the (strictly) positive eigenvalues of CMC. $\vartheta(0)$ vanishes when one of these numbers hits the level $+1$, and these events are isolated in time since $\vartheta(0)$ is an analytic function of $t \ge 0$, with initial value $+1$. Now it is known from Sect. 9.6.6 that $\vartheta^\bullet \vartheta''(0) = \frac{1}{2}m'(0)\vartheta_-\vartheta_+(0) < 0$ when $\vartheta(0)$ vanishes, so the roots are double with respect to x and simple with respect to t. I aim to prove that there is an infinite number of such roots: More precisely, that CMC has an ∞-dimensional null space and also an infinite number of nontrivial eigenvalues $\mu_1 \ge \mu_2 >$ etc. > 0, crossing the level $+1$ one at a time, μ_1 at time $T_1 > 0$, μ_2 at time $T_2 > T_1$, etc., passing on to $+\infty$ as $t \uparrow \infty$; actually, $T_n \uparrow +\infty$. I believe that the individual eigenvalues are always simple and increase with time but could not prove it. The proof is broken up into four little steps.

Step 1. Fix $t > 0$. The numbers $C_n^2 = (1 - e^{-t/2\lambda_n})\lambda_n$ tend to $t/2$ as n tends to $\pm\infty$, so[42] any function $h \in H^1$ vanishing for $x \le 0$ (of which there are many) may be written $h = \sum \xi_n C_n f_n$ with $\xi^2 < \infty$, and you have $\xi \bullet CMC\xi = \int_{-\infty}^{0} mh^2 = 0$; i.e., CMC has an ∞-dimensional null space. But it is also obvious that it is of infinite rank and, as such, must have an infinite number of positive eigenvalues $\mu_1 \ge \mu_2 \ge$ etc. $\downarrow 0$.

Step 2. This step shows that these positive eigenvalues cross level $+1$ one at a time (if at all). In fact, if, let us say, three of them cross at the same instant $t_0 > 0$, then one of three things must happen: Either (1) they are all (local) analytic functions of $T = t - t_0$, or (2) one of them is such while the other two branch as in

$$\left.\begin{matrix}\mu_2\\ \mu_3\end{matrix}\right\} = 1 + \sum_{n=1}^{\infty} R_n T^{n/2} \times \left\{\begin{matrix}1\\ (-1)^n,\end{matrix}\right.$$

or else (3) you see a triple branching as in

$$\left.\begin{matrix}\mu_1\\ \mu_2\\ \mu_3\end{matrix}\right\} = 1 + \sum_{n=1}^{\infty} R_n T^{n/3} \times \left\{\begin{matrix}1\\ \omega^n\\ \omega^{2n}\end{matrix}\right. \quad \text{with } \omega = \text{the cube root of unity } e^{2\pi\sqrt{-1}/3}.$$

Both (1) and (2) violate $\vartheta^\bullet \ne 0$ at the root, and (3) is not possible either since you cannot keep all three roots real without having the leading term of the sum be $T^{n/3}$ with $n \equiv 0(3)$, in which case ϑ vanishes triply at $T = 0$. This is contradictory.

Step 3. This step reveals that once it has crossed the level $+1$, no eigenvalue returns: Near the crossing time, the eigenvalue μ is simple, and both it and its associated normalized eigenvector ξ are smooth. Then a routine calculation produces

$$\mu^\bullet = 2\mu\xi C^{-1}C^\bullet\xi = \frac{1}{2}\sum e^{-t/2\lambda_n}\mu_n C_n^{-2}\xi_n^2 > 0.$$

That does the trick.

Step 4. This step shows that each eigenvalue really does cross the level $+1$, passing on to $+\infty$ with increasing time. This follows from min/max. Let $\eta_1, \dots, \eta_{n-1}$ be unit vectors in general position, and let the unit vector ξ be perpendicular to them. It is the content of min/max that $\mu_n = \min_\eta \max_\xi \xi \bullet CMC\xi$. Now the matrix $[\int_{-\infty}^{0} mf_i f_j : -1 \ge i, j \ge -n]$

[42] $m \ne 0$ on any interval, so the functions $f_n : n \in \mathbb{Z} - 0$ span H^1; see Sect. 9.2.2.

is positive definite with spectrum $\geq \Lambda > 0$. Pick ξ with vanishing components except for $\xi_{-1}, \dots \xi_{-n}$ and make it perpendicular to the η's, as you may, having one extra degree of freedom. Then

$$\mu_n \geq \Lambda \sum_1^n \xi_{-i}^2 C_{-i}^2 \geq \Lambda \min_{1 \leq i \leq n} C_{-i}^2 \geq \Lambda \lambda_{-1} \big(e^{-t/2\lambda_n} - 1 \big) \uparrow \infty$$

with t, as desired.

9.7.5. Shape of $\vartheta(x)$. The information obtained to date already tells a lot. Let $0 < T_1 < T_2 < \text{etc.} \uparrow \infty$ be the roots of $\vartheta(0) = 0$, alias the crossing times. $\vartheta(x) = \det[1 - C \int_{-\infty}^{x} mf \otimes fC]$ is seen at various stages of its development in Fig. 9.3. Figure 9.3a shows it before time T_1: ϑ decreases to its minimum at $x = 0$, rising symmetrically to $\vartheta(\infty) = 1$. Figure 9.3b shows it at time T_1: At $x = 0$, $\vartheta^\bullet \vartheta'' < 0$, so $\vartheta^\bullet < 0$ and $\vartheta'' > 0$. Thus, $\vartheta(0)$ continues to decrease as in Fig. 9.3c, and the old root at $x = 0$ disappears, or rather it splits into two symmetrically disposed simple roots $\pm r_1(t)$, one to the right and one to the left. The appearance of any other roots is ruled out for a while: If there were another (simple) root r_2 to the right of r_1, there would have to be yet a third root $r_3 > r_2$ since $\vartheta(\infty) = 1$, and, running every thing backwards, the simplicity of these roots would force ϑ to vanish sixfold at $x = 0$ at time T_1!

Thus: Fig. 9.3c is realistic; moreover, at $x = r_1 > 0$, $\vartheta^\bullet \vartheta'$ has the same sign as $m(r_1) < 0$, so from $0 = [\vartheta(r_1)]^\bullet = \vartheta^\bullet + \vartheta' r_1^\bullet$, you find $r_1^\bullet = -\vartheta^\bullet / \vartheta' > 0$, i.e., r_1 moves steadily to the right, as in the picture. Once μ_1 crosses the level $+1$, it never comes back, passing on to $+\infty$, but pretty soon μ_2 crosses, at time $T_2 > T_1$. This moment is the next root of $\vartheta(0) = 0$ at which ϑ has the shape seen in Fig. 9.3d: At $x = 0$, ϑ vanishes and $\vartheta^\bullet \vartheta''$ is negative in accord with $m'(0) < 0$, so $\vartheta''(0) < 0$, $\vartheta^\bullet(0) > 0$, and $\vartheta(0)$ is rising. This rise produces two new simple roots $\pm r_2$ moving symmetrically to right and left as in Fig. 9.3e, and so on; i.e., there is a continual production of double roots at $x = 0$ that split into pairs of simple roots, moving steadily right and left. Section 9.6 explained what that means: Near each such simple root $r > 0$, $(e^{t\mathbb{X}} v)[\overline{x}(t, x)]$ exhibits a little cusp of the approximate form $a - bm(r)|\overline{x}(t, x) - \overline{x}(t, r)|^{2/3}$ with $b > 0$, so there is produced a whole train of cusps as in Fig. 9.4.

9.7.6. The Soliton Train. Now it will be clear what must be happening. Though this part is mere conjecture, it is hard to disbelieve. The roots $r_1 > r_2 > r_3 >$ etc. should tend to $+\infty$, $\overline{x}(r_1) > \overline{x}(r_2) > \overline{x}(r_3) >$ etc. should follow suit, and if, as I believe, $e^{t\mathbb{X}} v$ is a train of cusps as seen in Fig. 9.4, plus a little trailing radiation, and if these move at diminishing speeds, the leader being fastest, then the individual cusps should shape themselves into solitary traveling waves and what should these be but solitons? And, on the evidence of Sect. 9.1, what should their speeds be but $1/(2\lambda_1)$, $1/(2\lambda_2)$, etc.; in short, for $t \uparrow \infty$, large positive x, and suitable "phase shifts" $q_n : n \geq 1$, you should see a "soliton train"

$$\big(e^{t\mathbb{X}} v\big)(x) \simeq \sum_{n=1}^{\infty} \frac{1}{2\lambda_n} e^{-|x - q_n - t/2\lambda_n|}.$$

Here, m was odd and positive/negative to the left/right with $m'(0) < 0$, but I believe that ultimately a similar picture will be seen for general m. I do not know how to prove *any* of this and can merely record that it is in agreement with numerical calculations of R. Camassa [private communication]. Note. however, that the conjecture is exact for any m vanishing near

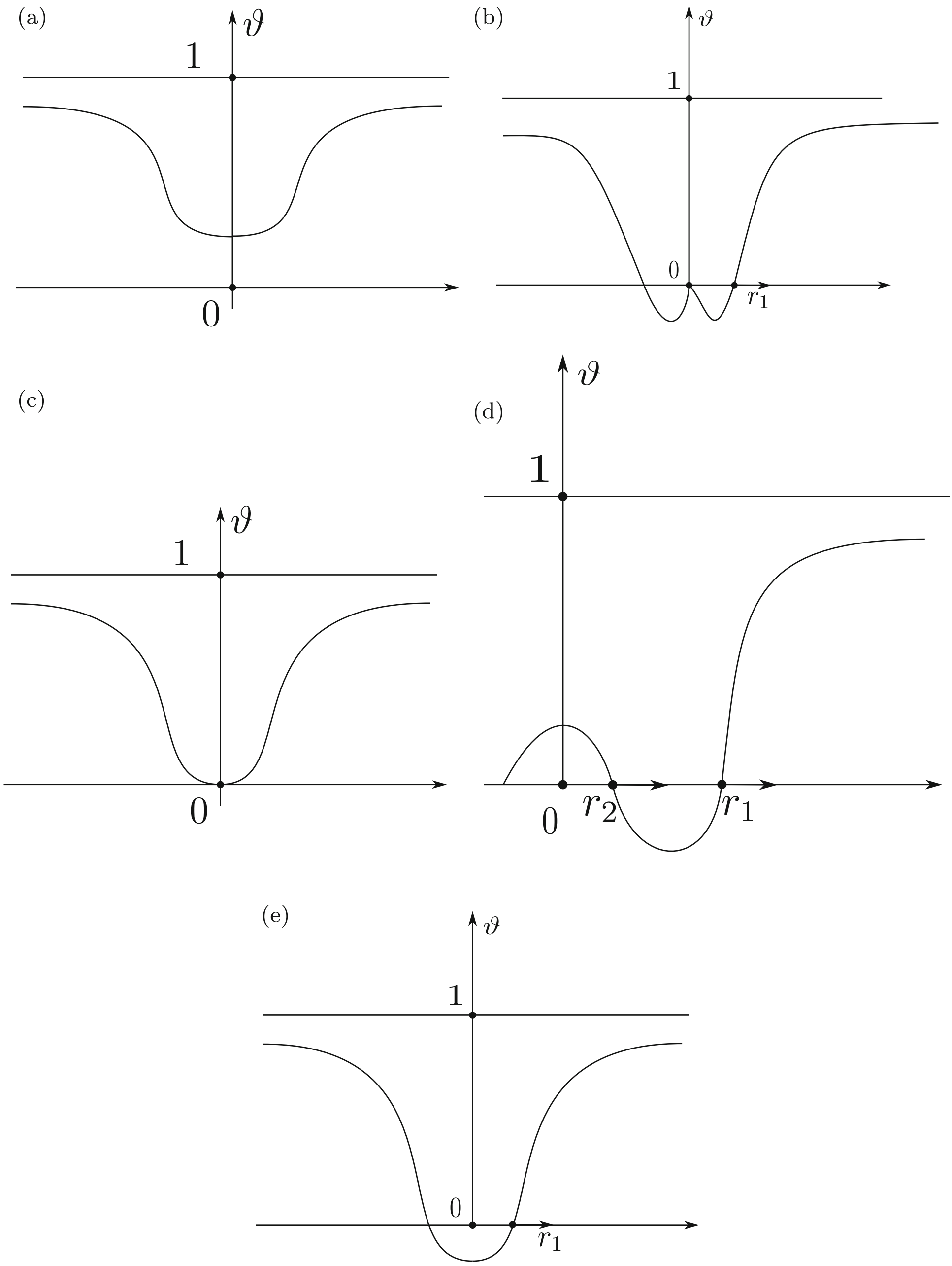

FIGURE 9.3.

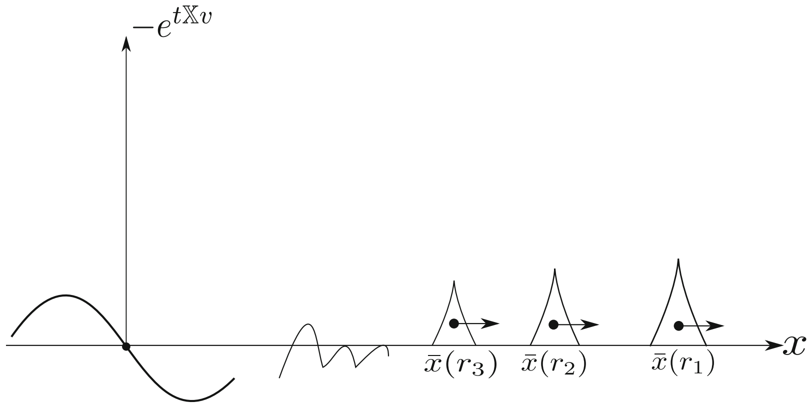

FIGURE 9.4.

$+\infty$ and x way out there: In fact, near $x = \infty$, $f_n,(x) =$ some norming constant $c_n \times e^{-x/2}$,

$$\vartheta_+(x) = \det\left[1 + \left(e^{-t/2\lambda} - 1\right)\int_{-\infty}^{x}\left(\frac{1}{2}f + f'\right)\otimes\left(\frac{1}{2}f + f'\right)\right]$$

$$= \vartheta_+(\infty) = \prod_{\mathbb{Z}-0} e^{-t/2\lambda_n} = 1^{43}$$

and likewise

$$\vartheta(x) = \det\left[1 + \left(e^{-t/2\lambda} - 1\right)\lambda\int_{-\infty}^{x} mf\otimes f\right] = \vartheta(\infty) = 1,$$

while

$$\vartheta_-(x) = \det\left[1 + \left(e^{-t/2\lambda} - 1\right)\left[1 - \int_{x}^{\infty}\left(\frac{1}{2}f - f'\right)\otimes\left(\frac{1}{2}f - f'\right)\right]\right]$$

$$= \det\left[e^{-t/2\lambda} + \left(1 - e^{-t/2\lambda}\right)\int_{x}^{\infty} c\otimes ce^{-y}\right]$$

$$= \prod_{\mathbb{Z}-0} e^{-t/2\lambda_n}\det\left[1 + \left(e^{t/2\lambda} - 1\right)c\otimes ce^{-x}\right]$$

$$= 1 + e^{-x}\sum c_n^2\left(e^{t/2\lambda_n} - 1\right),$$

so that

$$e^{\overline{x}} = e^{x}\times\left[1 + e^{-x}\sum_{\mathbb{Z}-0} c_n^2\left(e^{t/2\lambda_n} - 1\right)\right]$$

[43]The spectrum is odd.

and

$$\left(e^{t\mathbb{X}}v\right)(\overline{x}) = \frac{\vartheta_-^\bullet}{\vartheta_-} = \frac{e^{-x}\sum_{\mathbb{Z}-0} c_n^2/2\lambda_n e^{t/2\lambda_n}}{1+e^{-x}\sum_{\mathbb{Z}-0} c_n^2\left(e^{t/2\lambda_n}-1\right)}$$

$$\overset{44}{=} \sum_{\mathbb{Z}-0} \frac{1}{2\lambda_n} e^{-\overline{x}+q_n+t/2\lambda_n},$$

as advertised.

9.8. Appendix on Theta

I revert to general parameters $t = [t_n : n \in \mathbb{Z} - 0]$ as in Sects. 9.4 and 9.5 to describe some amusing features of ϑ. Jacobi's theta function and its three companions, produced by addition of half periods, obey an astonishing number of identities, and this is only a hint of what happens for Riemann's function and its translates. A similar thing happens here, only now it is not at first apparent what translation means. This is explained in the next section.

9.8.1. Addition. I take $m > 0$, the "addition" to be described below being ill-defined otherwise, and I fix $\Lambda < 0$ below the spectrum. Then $(\frac{1}{4} - D^2)e = \Lambda m e$ has two positive, increasing/decreasing solutions e_-/e_+ imitating $e^{x/2}/e^{-x/2}$ at $-\infty/+\infty$. I write $e_\pm = e(\bullet, \mathfrak{p})$ in which $\mathfrak{p}$ records the eigenvalue Λ together with a signature to indicate which function is to be employed. The associated "addition" is the map $A^\mathfrak{p} : m(x) \to A^\mathfrak{p} m(\overline{x}) \equiv m(x)/(\overline{x}')^2$ with scale determined by

$$e^{\overline{x}} = e^x \frac{e' - \frac{1}{2}e}{e' + \frac{1}{2}e}$$

or, what is more or less the same,

$$\overline{x}' = \frac{-\Lambda m e^2}{e'^2 - \frac{1}{4}e^2}.$$

The analogy with the formulae of Sect. 9.4 will be plain if you write, just for the moment, $\vartheta_\pm = e' \pm \frac{1}{2}e$ and $\vartheta^2 = -\Lambda m e^2$; for instance, $\vartheta^2 = \vartheta_-\vartheta_+ + \vartheta_-'\vartheta_+ - \vartheta_-\vartheta_+'$.

The several additions commute, and $A^\mathfrak{p} A^{-\mathfrak{p}} = 1$ in which $-\mathfrak{p}$ is $\mathfrak{p}$ with its signature reversed. Addition is part of the CH hierarchy in that the two commute; in fact, addition is more or less coextensive with the hierarchy in that any CH flow can be approximated by successive applications of $A^\mathfrak{p} : \Lambda < 0$. The name stems from the fact that, in the case of periodic m, addition is reflected in the Jacobi variety of the associated "multiplier curve" where the analogue of $\mathfrak{p}$ now lives, by subtraction of $\mathfrak{p}$ and addition of ∞—in short, "addition" is some kind of translation. McKean [10] explains this type of thing in the context of KdV. The present addition is explained independently in McKean [12]. For the moment, I record only the familiar-looking $A^\mathfrak{p} m(\overline{x}) = m(x)/(\overline{x}')^2$ and the updated eigenfunctions:

$$\left(A^\mathfrak{p} f_n\right)(\overline{x}) = \left(1 - \frac{\lambda_n}{\Lambda}\right)^{-1/2} \frac{\left[e(x,\mathfrak{p}), f_n(x)\right]^{45}}{\sqrt{\left[(e')^2 - \frac{1}{4}e^2\right](x,\mathfrak{p})}}$$

9.8.2. How ϑ_- and ϑ_+ Are Reduced to ϑ. I want to convince you that there is really only one theta function here, namely ϑ itself. The other two, ϑ_- and ϑ_+, may be produced from it by "addition" with $\Lambda = 0$; i.e., they are "translates" of ϑ. Indeed, for $\Lambda = 0$, the

[44] $q = \ln c^2$.

[45] $[a, b]$ is the Wronskian $a'b - ab'$.

action of addition is expressed as[46]

$$A^{\mathfrak{o}}m(\overline{x}) = \frac{m(x)}{(\overline{x}')^2} \quad \text{and} \quad A^{\mathfrak{o}}f_n(\overline{x}) = \frac{\left[e^{\mp x/2}, f_n(x)\right]}{\sqrt{\lambda_n}} e^{\pm\overline{x}/2},$$

in which $\mathfrak{o} = (0, \pm 1)$, the scale $\overline{x}$ is $\ln \int_{-\infty}^{x} me^{y}$ or $-\ln \int_{x}^{\infty} me^{-y}$ according as the signature is -1 or $+1$, and you have the pretty rule $A^{\mathfrak{o}}\vartheta(\overline{x}) = \vartheta_{\pm}(x)$. The proof is easy: For example, with signature -1,

$$\begin{aligned}
A^{\mathfrak{o}}\vartheta(\overline{x}) &= A^{\mathfrak{o}} \det\left[1 + \left(e^t - 1\right)\lambda \int_{-\infty}^{x} mf \otimes f\, dy\right] \quad \text{taken at } \overline{x} \\
&= \det\left[1 + \left(e^t - 1\right)\lambda \int_{-\infty}^{\overline{x}} A^{\mathfrak{o}}(mf \otimes f)(y) dy\right] \\
&= \det\left[1 + \left(e^t - 1\right)\lambda \int_{-\infty}^{x} A^{\mathfrak{o}}(mf \otimes f)(\overline{y}) d\overline{y}\right] \\
&= \det\left[1 + \left(e^t - 1\right)\lambda \int_{-\infty}^{x} \frac{m(y)}{\overline{y}'^2} \frac{\left(\frac{1}{2}f - f'\right)e^{y/2}}{\sqrt{\lambda}e^{\overline{y}/2}} \otimes \frac{\left(\frac{1}{2}f - f'\right)e^{y/2}}{\sqrt{\lambda}e^{\overline{y}/2}} \overline{y}' dy\right] \\
{}^{47} &= \det\left[1 + \left(e^t - 1\right) \int_{-\infty}^{x} \left(\frac{1}{2}f - f'\right) \otimes \left(\frac{1}{2}f - f'\right) dy\right] \\
&= \vartheta_{-}(x).
\end{aligned}$$

The computation with signature $+1$ is similar.

9.8.3. The General Addition Applied to ϑ. Fix p with $\Lambda < 0$ and let $e(\bullet, \mathfrak{p})$ be the associated eigenfunction. Then, with $\overline{x}$ = the scale for the addition $A^{\mathfrak{p}}$ and $\overline{\overline{x}}$ = the scale for the composite flow $e^{t\mathbb{X}}$,

$$e^{t\mathbb{X}}e(\bullet, \mathfrak{p})(\overline{\overline{x}}) = \frac{e(x, \mathfrak{p})A^{\mathfrak{p}}\vartheta(\overline{x})}{\sqrt{\vartheta_{-}\vartheta_{+}(x)}}.$$

The proof employs the method of Sect. 9.4 to obtain

$$e^{t\mathbb{X}}e(\bullet, \mathfrak{p})(\overline{\overline{x}}) = \frac{\vartheta(x)}{\sqrt{\vartheta_{-}\vartheta_{+}(x)}}\left[e(x, \mathfrak{p}) - (1 + MC)^{-1}f(x) \bullet C \int_{-\infty}^{x} mef\right]$$

with the usual $f = [f_n : n \in \mathbb{Z} - 0]$, $C = (e^t - 1)\lambda$, and $M = \int_{-\infty}^{x} mf \otimes f$. Now apply $A^{\mathfrak{p}}$ to ϑ: In the associated scale $\overline{x}$, you find, much as in Sect. 9.2 above,

$$\begin{aligned}
A^{\mathfrak{p}}\vartheta(\overline{x}) &= \det\left[1 + C \int_{-\infty}^{x} A^{\mathfrak{p}}\left(mf \otimes f\right)(\overline{y}) d\overline{y}\right] \\
&= \det\left[1 + C \int_{-\infty}^{x} \frac{m}{\overline{y}'^2} \frac{[e, f]}{\sqrt{1 - \frac{\lambda}{\Lambda}}} \otimes \frac{[e, f]}{\sqrt{1 - \frac{\lambda}{\Lambda}}} \frac{\overline{y}' dy}{e'^2 - \frac{1}{4}e^2}\right] \\
{}^{48} &= \det\left[1 + \frac{C}{\lambda - \Lambda} \int_{-\infty}^{x} \frac{\left(e'f - ef'\right) \otimes \left(e'f - ef'\right)}{e^2} dy\right].
\end{aligned}$$

[46] The scale $\overline{x}$ must be adjusted by an additive $-\ln(-\Lambda)$ before making $\Lambda \uparrow 0$.

[47] $\overline{x}' = me^{x}/e^{\overline{x}}$.

[48] $m/\overline{y}' = -(e'^2 - \frac{1}{4}e^2)/\Lambda e^2$.

Now the inner integral is

$$\int_{-\infty}^{x}\left[\left(\frac{e'}{e}\right)^2 f\otimes f-\frac{e'}{e}(f\otimes f)'+f'\otimes f'\right]$$

$$=\int_{-\infty}^{x}\left[f'\otimes f'+\frac{e''}{e}f\otimes f\right]-\frac{e'}{e}f\otimes f(x)$$

$$=(\lambda-\Lambda)\left[\int_{-\infty}^{x} mf\otimes f\,dy-\frac{1}{e(x)}\int_{-\infty}^{x} mef\,dy\otimes f(x)\right],$$

as you will readily check, so the whole reduces to

$$\det\left[1+C\int_{-\infty}^{x} mf\otimes f-C\frac{1}{e}\int_{-\infty}^{x} mef\otimes f(x)\right]$$

$$=\frac{\vartheta}{e}\left[e-f\bullet(1+CM)^{-1}C\int_{-\infty}^{x} mef\right].$$

It remains only to compare the start and the finish.

9.8.4. A Composition Rule.[49]

ϑ is now written $\vartheta(t,x)$ with the parameters $t=[t_n : A\in\mathbb{Z}-0]$ displayed. The rule is

$$\vartheta(t+s,x)=\vartheta(t,x)\times\left[e^{t\mathbb{X}}\vartheta(s,\bullet)\right](\overline{x})$$

with the scale $\overline{x}$ for the composite flow $e^{t\mathbb{X}}$. The proof is simple enough: With the usual $C=(e^t-1)\lambda$,

$$e^{t\mathbb{X}}f_n(\overline{x})=e^{t/2}\vartheta(1+MC)^{-1}f_n/\sqrt{\vartheta_-\vartheta_+}\quad\text{in the original scale } x,$$

so

$$\left[e^{t\mathbb{X}}\vartheta(s,\bullet)\right](\overline{x})$$

$$=\det\left[1+\left(e^s-1\right)\lambda\int_{-\infty}^{x}\frac{m}{\overline{y}'^2}\frac{e^{t/2}\vartheta(1+MC)^{-1}f}{\sqrt{\vartheta_-\vartheta_+}}\otimes\frac{e^{t/2}\vartheta(1+MC)^{-1}f}{\sqrt{\vartheta_-\vartheta_+}}\overline{y}'\,dy\right]$$

[50] $$=\det\left[1+\left(e^s-1\right)\lambda e^t\int_{-\infty}^{x} m(1+MC)^{-1}f\otimes(1+MC)^{-1}f\right].$$

But[51]

$$\left[\left(1+MC\right)^{-1}\right]'=-m\left(1+MC\right)^{-1}f\otimes(1+MC)^{-1}fC,$$

[49] McKean and Trubowitz [13] served as a model.

[50] $\overline{x}'=\vartheta^2/\vartheta_-\vartheta_+$.

[51] $(1+cM)^{-1}C$ is symmetric.

whence

$$\begin{aligned}\left[e^{t\mathbb{X}}\vartheta(s,\bullet)\right](\overline{x}) &= \det\left[1-(e^s-1)e^t\lambda\int_{-\infty}^{x}\left[(1+MC)^{-1}\right]'C^{-1}\right]\\ &= \det\left[1-\left(e^s-1\right)e^t\lambda\left[(1+MC)^{-1}-1\right]C^{-1}\right]\\ &= \det\left[1+\left(e^s-1\right)e^t\lambda(1+MC)^{-1}M\right].\end{aligned}$$

Now multiply from the right by $\vartheta = \det(1+CM)$ to obtain

$$\begin{aligned}&\det\left[1+CM+\left(e^s-1\right)e^t\lambda(1+MC)^{-1}M(1+CM)\right]\\ &\quad= \det\left[1+\left(e^t-1\right)\lambda M+\left(e^s-1\right)e^t\lambda M\right]\\ &\quad= \det\left[1+\left(e^{t+s}-1\right)\lambda M\right]\\ &\quad= \vartheta(t+s,x),\end{aligned}$$

as advertised.

9.9. Some Open Questions

A task for the future is to verify the formation of the soliton train described in Sect. 9.7. KdV also produces soliton trains, but with this difference: Except for reflectionless v, the KdV soliton train accounts only for the part of the total energy ascribed to the bound states of the associated spectral problem, the deficiency being carried by the evanescent radiation ascribed to the continuous spectrum. Now for summable m, CH has *only* bound states, the individual soliton $v_n \equiv (1/(2\lambda_n))\exp(-|x-t/(2\lambda_n)|)$ carries energy $\frac{1}{2}\int[(v_n')^2+v_n^2] = 1/(4\lambda_n^2)$, and these individual energies add up to the whole: $H_2 = \sum 1/(4\lambda_n^2)$, so here nothing is lost. A special, more primitive version of soliton formation is seen when v is run by one of the individual flows of Sect. 9.3 with Hamiltonian $H = 1/\lambda_n$. Now a single soliton v_n is kicked out and runs away to $\pm\infty$, leaving behind a "residual" v_∞ with the same spectrum as for v but with λ_n removed. Doubtless the map $v \to v_\infty$ is the counterpart of the standard Darboux transformation for removing one bound state for KdV and can be expressed much like the "addition" cited in Sect. 9.8. These questions are under current investigation by Sr. E. Loubet.

CH is the special case $c = 2$ of $m^\bullet + vm' + cv'm = 0$, which is integrable for $c = 3$ as well but not otherwise; moreover, it exhibits solitons for any $c > 1$.[52] What happens then? Does my method apply at all? I do not know.

The introduction of the term $2c\partial v/\partial x$ into standard CH raises further questions. This is equivalent to taking a solution w of standard CH that vanishes at $\pm\infty$ and putting $v(t,x) = \mathrm{w}(t,x-ct)+c$ so that $v(\pm\infty) = c \neq 0$. It introduces continuous spectrum $[1/(4c),+\infty)$ or $(-\infty,1/(4c)]$ according as c is positive or negative, producing radiation over and above the solitons associated with any bound states, and it may be that some version of the present theta formalism applies. I do not know about that either.

Acknowledgment

It is a pleasure to thank Sr. E. Loubet for correcting, innumerable mistakes and awkward bits.

[52]See Holm and Staley [8] and Degasperis, Holm, and Hone [5].

Bibliography

[1] R. Beals, D. H. Sattinger, and J. Szmigielski. Multipeakons and the classical moment problem. *Adv. Math.*, 154:229–257, 2000.

[2] R. Camassa and D. Holm. An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett.*, 71(11):1661–1664, 1993.

[3] R. Camassa, D. D. Holm, and M. Hyman. A new integrable shallow water equation. *Adv. Appl. Math.*, 31:1–33, 1993.

[4] A. Constantin and J. Escher. Well-posedness, global existence, and blow up phenomena for a periodic quasi-linear hyperbolic equation. *Comm. Pure Appl. Math.*, 51:475–504, 1998.

[5] A. Degasperis, D. D. Holm, and A. N. N. Hone. A new integrable equation with peakon solutions. *Theoretical and Mathematical Physics*, 133:1463–1474, 2002.

[6] F. J. Dyson. Fredholm determinants and inverse scattering. *Comm. Math. Phys.*, 47:171–183, 1976.

[7] N. Ercolani and H. P. McKean. Geometry of KdV (4). Abel sums, Jacobi variety, and theta function in the scattering case. *Invent. Math.*, 99(3):483–544, 1990.

[8] D. Holm and M. F. Staley. Wave structure and nonlinear balances in a family of evolutionary PDEs. *SIAM J. Appl. Dyn. Syst.*, 2:323–380, 2003.

[9] A. R. Its and V. B. Matveev. Hill operators with a finite number of lacunae. *Funkcional. Anal. i Priložen.*, 9(1):69–70, 1975.

[10] H. P. McKean. Geometry of KdV (1): addition and the unimodal spectral class. *Rev. Math. Iberoamer.*, 2:235–261, 1986.

[11] H. P. McKean. Breakdown of a shallow water equation. *Asian J. Math.*, 2:867–874, 1998.

[12] H. P. McKean. Addition for the acoustic equation. *Comm. Pure Appl. Math.*, 54:1271–1288, 2001.

[13] H. P. McKean and E. Trubowitz. The spectral class of the quantum-mechanical harmonic oscillator. *Comm. Math. Phys.*, 82(4):471–495, 1981/82.

[14] T. J. Stieltjes. *Recherches sur les fractions continues, Oeuvres complétes 2.* Nordhoff, Gröningen, 1918.

[15] Z. Xin and P. Zhang. On the weak solutions to a shallow water equation. *Comm. Pure Appl. Math.*, 53(11):1411–1433, 2000.

[10] Breakdown of the Camassa-Holm Equation

[10] Breakdown of the Camassa-Holm Equation. *Comm. Pure Appl. Math.* **57** (2004), 416–418.

10 Breakdown of the Camassa-Holm Equation

Henry P. McKean Jr.[1]

10.1. Introduction

The Camassa-Holm equation[2] is written in its Eulerian form as

$$\text{CH:} \quad \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \frac{\partial p}{\partial x} = 0,$$

with "pressure" $p = G[v^2 + \frac{1}{2}(v')^2]$. Here $G = (1 - D^2)^{-1} = \frac{1}{2}e^{-|x-y|}$. It is easy to see that if, at time $t = 0$, v is odd with $v(x) > 0$ for $x < 0$ and $v'(0) < 0$, then the slope $s(t) = v'(t, 0)$ satisfies $s^\bullet < -\frac{1}{2}s^2$ and so is driven down to $-\infty$ at some time $T \leq -2/s(0)$: in short, the flow may "break down." This is the "steepening lemma" of Camassa and Holm [2]. Nothing worse happens: $v(t, x)$ itself cannot jump like the usual kind of shock. Actually, it is not the obvious v but rather the auxiliary function $m = v - v''$ that controls this; m is a caricature of the vorticity that controls the behavior of honest Euler in dimensions 2 and 3; see Bertozzi and Majda [1] for a full account.

McKean [6] proved that CH breaks down if and only if some portion of the positive part of m lies to the left of some portion of its negative part. The proof was made under the condition that m be smooth and summable, with a finite number of simple roots. It was a little complicated and even wrong at one point, as Xin Zhouping kindly pointed out to me; see McKean [5] for the correction. Here, I want to present a very simple proof using the machinery of McKean [4], summarized in Sect. 10.2. The prior condition on the roots of m is dropped.

10.2. Solving CH by Fredholm Determinants

CH is integrated by means of the eigenfunctions $f_n : n \in \mathbb{Z} - 0$ of the associated spectral problem $-f'' + \frac{1}{4}f = \lambda m f$ considered at time $t = 0$. These are labeled so that the corresponding eigenvalues $\lambda_n : n \in \mathbb{Z} - 0$ are positive (negative) for $n > 0$ ($n < 0$). They form a unit perpendicular basis of H^1 with nonstandard norm $|f| = (\int[(f')^2 + \frac{1}{4}f^2])^{1/2}$. The finiteness

[1] Courant Institute, 251 Mercer Street, New York, NY 10012-1185, USA, mckean@cims.nyu.edu.
[2] [2]; see also [3].

of the absolute trace $\sum_{\mathbb{Z}-0} |\lambda_n|^{-1} \le \int |m|$ permits formation of the 3 "theta functions":

$$\left.\begin{array}{l}\vartheta_-\\ \vartheta\\ \vartheta_+\end{array}\right\} = \det \left[I + \left(e^{-t/2\lambda_i} - 1\right) \times \int_{-\infty}^{x} \left(f_i' f_j' + \frac{1}{4} f_i f_j \right) dy \left\{ \begin{array}{l} -\,\frac{1}{2} f_i f_j(x) \\ -\, f_i' f_j'(x) \\ +\,\frac{1}{2} f_i f_j(x) \end{array} \right. \right].$$

ϑ_- and ϑ_+, cannot vanish, but ϑ may and that causes breakdown. The solution of CH with initial data $v = Gm$ is expressed in the associated Lagrangian scale $\bar{x}(t,x)$, determined by $\partial\bar{x}/\partial t = v(t,\bar{x})$ with $\bar{x}(0+,x) \equiv x$, thus:

$$\bar{x} = x + \ln\left(\frac{\vartheta_-}{\vartheta_+}\right) \quad \text{and} \quad v(t,\bar{x}) = \bar{x}^\bullet = \left[\ln\left(\frac{\vartheta_-}{\vartheta_+}\right)\right]^\bullet.$$

This makes perfect sense for all times $0 \le t < \infty$ due to the nonvanishing of $\vartheta_\pm$, and CH is solved thereby in its Lagrangian form $(d/dt)v(t,\bar{x}) + p'(\bar{x}) = 0$. It is only v' that can go bad, and this at isolated instants only. The identity $\vartheta^2 = \vartheta_-\vartheta_+ \vartheta_-' \vartheta_+ - \vartheta_-\vartheta_+'$ leads to $\bar{x}' = \vartheta^2/\vartheta_-\vartheta_+$, whence

$$v'(t,\bar{x}) = \frac{1}{\bar{x}'}\frac{d}{dx}v(t,\bar{x}) = \frac{1}{\bar{x}'}\frac{d}{dx}\bar{x}^\bullet = (\ln \bar{x}')^\bullet = \frac{2\vartheta^\bullet}{\vartheta} + \text{nice stuff}.$$

Here, $\vartheta^\bullet$ cannot vanish when ϑ does, so v' misbehaves precisely when $\vartheta = 0$. Now ϑ can be re-expressed as

$$\vartheta = \det\left[I - C_i \int_{-\infty}^{x} m f_i f_j C_j\right] \quad \text{with } C_n = \sqrt{\lambda_n\left(1 - e^{-t/2\lambda_n}\right)} > 0,$$

from which it is easy to see that ϑ does not vanish (for any t or x) if and only if you always have[3] $\int_{-\infty}^{x} mh^2 < \sum_{\mathbb{Z}-0} C_n^{-2}\langle h, f_n\rangle^2$ for every h of class H^1; moreover, C_n^2 increases with $t \uparrow \infty$ to $+\infty$ if $n < 0$ and to λ_n if $n > 0$, leading to the final condition for absence of breakdown:

$$\max_{x\in\mathbb{R}} \int_{-\infty}^{x} mh^2 \le \sum_{n>0} \lambda_n^{-1}\langle h, f_n\rangle^2 \quad \text{for every } h \in H^1.$$

10.3. Discussion of Breakdown

The condition is met if the negative part of m lies wholly to the left of its positive part (Fig. 10.1). Then $\int_{-\infty}^{x} mh^2$ is at first ≤ 0 and then increases to its maximum at $x = \infty$. But $\lambda_i \int_{-\infty} m f_i f_j = \langle f_i, f_j\rangle = 1$ or 0 according as $i = j$ or not, so for the general function h,

$$\int_{-\infty}^{\infty} mh^2 = \sum_{\mathbb{Z}-0} \lambda_n^{-1}\langle h, f_n\rangle^2 \le \sum_{n>0} \lambda_n^{-1}\langle h, f_n\rangle^2.$$

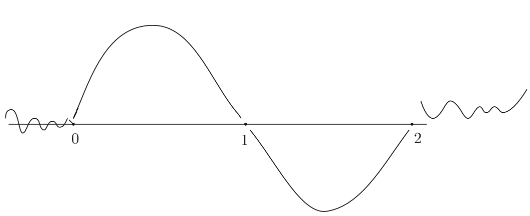

FIGURE 10.1.

[3] $\langle h, f\rangle$ is the inner product $\int (h'f' + \frac{1}{4}hf)$.

Now let m be as in the figure (which is merely illustrative) and look at the reduced spectral problem $-h'' + \frac{1}{4}h = \mu mh$ for h of class H^1 vanishing to the left of $x = 0$. The presence of negative m between $x = 1$ and $x = 2$ ensures the existence of infinity of many negative eigenvalues $\mu \downarrow -\infty$. The associated eigenfunctions h will violate the condition of Sect. 10.2. Indeed, with $h(0) = 0$ and $h'(0) = 1$, say, you have

$$\langle h, f_n\rangle = hf_n'\big|_0^\infty + \int_0^\infty h\Big(-f_n'' + \frac{1}{4}f_n\Big) = \lambda_n \int_0^\infty mhf_n$$

and also

$$\langle h, f_n\rangle = h'f_n\big|_0^\infty + \int_0^\infty \Big(-h'' + \frac{1}{4}h\Big)f_n = -f_n(0) + \mu\int_0^\infty mhf_n,$$

i.e.,

$$\langle h, f_n\rangle = -f_n(0)\Big(1 - \frac{\mu}{\lambda_n}\Big)^{-1},$$

providing the bound

$$\sum_{n>0} \lambda_n^{-1}\langle h, f_n\rangle^2 \le \sum_{n>0}\lambda_n^{-1}f_n^2(0) \le \lambda_1^{-1}\sum_{n>0}\langle e, f_n\rangle^2 \le \lambda_1^{-1}|e|^2$$

with $e = \exp(-\frac{1}{2}|x|)$, independently of $\mu < 0$. But also m is positive between $x = 0$ and $x = 1$ so that $h'' = (\frac{1}{4} - \mu m)h > 0$ there, and if, for example, $m(x) \ge 1$ between $x = \frac{1}{3}$ and $x = \frac{2}{3}$, then $h(x)$ will be $\ge \frac{1}{\sqrt{-\mu}}8h\sqrt{-\mu}(x - \frac{1}{3})$ in *that* interval, causing $\int_{-\infty}^1 mh^2$ to be large for $\mu \downarrow -\infty$. That's all there is to it.

Bibliography

[1] Bertozzi, A; Majda, A. Vorticity and incompressible flow. Cambridge University Press, Cambridge, 2002.

[2] Camassa, R.; Holm, D. An integrable shallow water equation with peaked solitons. Phys. Rev. Lett. 71 (1993), 1661–1164.

[3] Camassa, R.; Holm, D; Hyman, M. A new integrable shallow water equation. Adv. Appl. Math. 31 (1994), 1–33.

[4] McKean, H. P. Breakdown of a shallow water equation. Asian J. Math. 2 (1998), 767–774.

[5] McKean, H. P. Correction to 'Breakdown of a shallow water equation'. Asian J. Math. 3 (1999).

[6] McKean, H. P. Fredholm determinants and the Camassa-Holm hierarchy. Comm. Pure Appl. Math. 61 (2003), 638–680.

[11] Rational Theory of Warrant Pricing

[11] Rational Theory of Warrant Pricing. *Ind. Management Rev.* **6** (1965), 13–32.

Rational Theory of Warrant Pricing

11

Paul A. Samuelson[1]

11.1. Introduction

This is a compact report on desultory researches stretching over more than a decade.

In connection with stock market fluctuations, L. Bachelier [3], a French mathematician, discovered the mathematical theory of Brownian motion five years before Einstein's classic 1905 paper, Bachelier gave the same formula for the value of a warrant (or "call" or put) based upon this "absolute" or "arithmetic" process that Dr. R. Kruizenga [20, 21] developed years later in a thesis under my direction. Under this formula, the value of a warrant grows proportionally with the square-root of the time to go before elapsing; this is a good approximation to actual pricing of short-lived warrants, but it leads to the anomalous result that a long-lived warrant will increase in price indefinitely, coming even to exceed the price of the common stock itself—even though ownership of the stock is equivalent to a perpetual warrant exercisable at zero price!

The anomaly apparently came because Bachelier had forgotten that stocks possess limited liability and thus cannot become negative, as is implied by the arithmetic Brownian process, To correct this, I introduced the "geometric" or "economic Brownian motion," with the property that every dollar of market value is subject to the same multiplicational or percentage fluctuations per unit time regardless of the absolute price of the stock. This led to the log-normal process for which the value of a call or warrant has these two desired properties: for short times, the $\sqrt{t}$ law holds with good approximation; and for $t \to \infty$, the value of the call approaches the value of the common stock. (All the above assumes that stock-price changes represent a "fair-game" or martingale—or certain trivial generalizations thereof to allow for a fair return. In an unpublished paper and lecture, I made explicit the derivation of this property from the consideration that, if everyone could "know" that a stock would rise in price, it would *already* be bid up in price to make that impossible. See my companion paper appearing in this same issue, entitled "Proof That Properly Anticipated Prices Fluctuate Randomly.")

[1] Massachusetts Institute of Technology, Cambridge, MA, USA.
Acknowledgment is made to the Carnegie Corporation for research aid, but sole responsibility for the results is mine.

The above results, which have been presented in lectures since 1953 at M.I.T., Yale, Carnegie, the American Philosophical Society, and elsewhere have also been presented by such writers as Osborne [25, 26], Sprenkle [27], Boness [4], Alexander [1, 2, 16, 17] and no doubt others.

However, the theory is incomplete and unsatisfactory in the following respects:

1. It assumes, explicitly or implicitly, that the mean rate of return on the warrant is no more than on the common stock itself, despite the fact that the common stock may be paying a dividend and that the warrant may have a different riskiness from the common stock.
2. In consequence of the above, the theory implies that warrants (or calls) will never be converted prior to their elapsing date. Necessarily, therefore, no proper theory is provided for the conditions under which warrants will cease to be outstanding.
3. The existing theory, in effect, assumes that the privilege of converting the warrant at any time in the interval (rather than at the end of the period) is worth literally nothing at all.
4. Finally, the theory leads to the mentioned result, that the price of a perpetual warrant should be literally equal to the stock itself—a paradoxical result, and one that does not agree with the observed facts of life (for example, the fact that perpetual Tri-Continental Warrants sell for less than their equivalent amount of common stock, and are in fact being continuously converted into stock in some positive volume).

The present paper publishes, I believe for the first time, the more difficult theory of rationally evaluating a warrant, taking account of the extra worth of the right to convert at *any time in the interval* and deducing the value of the common stock above which it will pay to exercise the warrant. I am glad to acknowledge the valuable contribution of Professor Henry P. McKean, Jr. of the M.I.T. Department of Mathematics, in effecting certain exact solutions and in proving the properties of the general solutions. His analysis appears as a self-contained mathematical appendix. It will be clear that there still remain many unsolved problems. (For example, *exact* explicit solutions are now known in the case of perpetual warrants only for three cases: the log-normal, the log-Poisson, and the case where the only two possibilities are those of instantaneous complete loss or of a gain growing exponentially in time. Only for this last case is an exact explicit solution known for the finite-time warrant. These exact solutions, which are all due to McKean, correspond to various intuitive conjectures and empirical patterns and can be approximated by the solutions to the simpler problem of discrete, albeit small, time periods.)

11.2. The Postulated Model

Let the price of a particular common stock be defined for all time and be denoted by X_t. If we stand at the time t, we know with certainty X_t (and all of its past values X_{t-T}). Its future price X_{t+T} is knowable only in some probability sense, its probability distribution being in the most general case a function of the whole past profile of X_{t-T}. A special simplification involves postulating a Markov property to the process, so that future X_{t+T} has a distribution depending only on present X_t—namely

$$\text{Prob}\left\{X_{t+T} \le X \mid X_t = x\right\} = P(X, x; T). \tag{1}$$

Obviously, (1) involves the critical assumption of a "stationary time series."

I further posit that each dollar of present value must be expected to have some mean gain per unit time, α, where α may perhaps be zero or more likely will be a positive quantity whose magnitude depends on the dispersion riskiness of X, and the typical investor's utility aversion to risk. (A deeper theory would posit concave utility and deduce the value of α for each category of stocks.) This expected-returns axiom says

$$\text{E}\left[X_{t+T} \mid X_t\right] = \int_0^\infty X\, dP\big(X, X_t; T\big) = X_t e^{\alpha T}, \quad \alpha \geq 0. \tag{2}$$

(Since money bears the safe return of zero, α cannot be less than zero for risk averters; indeed, it cannot be less than the safe return or pure interest on funds, if such exists. If utility were convex rather than concave, people might be willing to pay for riskiness, and α might be permitted to be negative—but not here.)

The integral in (2) is the usual Stieltjes integral: if the probability distribution $P(X, X_t; T)$ has a regular probability density $\partial P(X, X_t; T)/\partial X = p(X, X_t; T)$, we have the usual Riemann integral $\int_0^\infty X p(X, X_t; T) dX$; if only discrete probabilities are involved, at $X = X_1$ with probabilities $P_1(X_t; T)$, the integral of (2) becomes the sum $\Sigma X_1 P_1(X_t; T)$, which may involve a finite or countably-infinite number of terms. The reader can use the modern notation $\int_0^\infty XP(dX, X_t; T)$ rather than that of (2) if he prefers.

In (2) the limit of integration is given as 0 rather than $-\infty$, because of the important phenomenon of limited liability. A man cannot lose more than his original investment: General Motors stock can drop to zero, but not below.

If the probability of a future price X_{t+T} depends solely on knowledge of X_t alone, having the Markov property of being independent of further knowledge of past prices such as X_{t-s}, then

$$P\big(X_{t+T} \mid X_t, X_{t-s}\big) \equiv P\big(X_{t+T} \mid X_t\big) \tag{3}$$

and (1) will satisfy the so-called Chapman-Kolmogorov equation

$$P\big(X_{t+T}, X_t; T\big) \equiv \int_0^\infty P\big(X_{t+T}, x; T - S\big) dP\big(x, X_t; S\big), \quad 0 \leq S \leq T. \tag{4}$$

11.3. Remarks About Alternative Axioms

To see the meaning of this, suppose t takes on only discrete integral values. Then, without the Markov property (3), (1) would have the general form

$$\text{Prob}\left\{X_{t+k} \leq X \mid X_t, X_{t-1}, \dots\right\} = P\big(X, X_t, X_{t-1}, \dots; k\big) \tag{5}$$

with

$$\text{E}\left[X_{t+1} \mid X_t, X_{t-1}, \dots\right] = \int_0^\infty X\, dP\big(X, X_t, X_{t-1}, \dots; 1\big) = e^{\alpha} X_t. \tag{2'}$$

Instead of (4), we would have

$$\begin{aligned} &P\big(X_{t+2}, X_t, X_{t-1}, \dots; 2\big) \\ &= \int_0^\infty P\big(X_{t+2}, X_{t+1}, X_t, \dots; 1\big) dP\big(X_{t+1}, X_t, X_{t-1}, \dots; 1\big) \end{aligned} \tag{4'}$$

where the integration is over X_{t+1} and where X_t is seen to enter in the first factor of the integrand. Even without the Markov axiom of (3), from (2′) applied to the next period's gains, we could deduce the truth of (2) for two periods' gains as well and, by induction, for all-periods' gains—namely

$$\begin{aligned}
&\mathrm{E}\left[X_{t+2} \mid X_t, X_{t-1}, \dots\right] \\
&= \int_0^\infty X\,dP\big(X, X_t, X_{t-1}, \dots; 2\big) \\
&= \int_0^\infty X\,d\int_0^2 P\big(X, x, X_t, X_{t-1}, \dots; 1\big)dP\big(x, X_t, X_{t-1}, \dots; 1\big) \\
&= \int_0^\infty e^{\alpha} x\,dP\big(x, X_t, X_{t-1}, \dots; 1\big) = e^{2\alpha} X_t.
\end{aligned} \tag{6}$$

Then, by induction, (2) or

$$\mathrm{E}\left[X_{t+k} \mid X_t, X_{t-1}, \dots\right] = e^{k\alpha} X_t$$

follows from the weak assumption of (5) and (2′) alone even when the Markov property (3) and Chapman-Kolmogorov property (4) do not necessarily hold.

However, I shall assume (3), and a fortiori (4), in order that the rational price of a warrant be a function of current common stock price X_t alone and not be (at this level of approximation) a functional of all past values X_{t-T}. A more elaborate theory would introduce such past values, if only to take account of the fact that the numerical value of α will presumably depend upon the estimate from past data that risk averters make of the riskiness they are getting into when holding the stock.

I might finally note that Bachelier assumed implicitly or explicitly

$$P(X, x; T) \equiv P(X - x; T), \quad \alpha = 0 \tag{7}$$

so that an absolute Brownian motion or random walk was involved. He thought that he could deduce from these assumptions alone the familiar Gaussian distribution—or, as we would say since 1923, a Wiener process—but his lack of rigor prevented him from seeing that his form of (4):

$$P(X - x; T) \equiv \int_{-\infty}^{\infty} P(X - x - y; T - S)dP(y; S), \quad 0 \le S \le T \tag{8}$$

does have for solutions, along with the Gaussian distribution, all the other members of the Lévy-Khintchin family of infinitely-divisible distributions [22, 12]. including the stable distribution of Lévy-Pareto, the Poisson distribution, and various combinations of Poisson distributions.

11.4. The "Geometric or Relative Economic Brownian Motion"

As mentioned, Bachelier's absolute Brownian motion of (7) leads to negative values for X_{t+T} with strong probabilities. Hence, a better hypothesis for an economic model than $P(X, x; T) = P(X - x; T)$ is the following

$$\begin{aligned}
P(X, x; T) &\equiv P\left(\frac{X}{x}; T\right), \quad x > 0 \\
P(X, 0; T) &\equiv 1 \quad \text{for all } X > 0.
\end{aligned} \tag{9}$$

By working with ratios instead of algebraic differences, we consider logarithmic or percentage changes to be subject to uniform probabilities. This means that the first differences of the logarithms of prices are distributed in the usual absolute Brownian way. Since the arithmetic mean of logs in the geometric mean of actual prices, this modified random walk can be called the geometric Brownian motion in contrast to the absolute or arithmetic Brownian motion.

The log-normal distribution bears to the geometric Brownian motion the same relation that the normal distribution does to the ordinary Brownian motion. As the writings of Mandelbrot [23, 24] and Fama [9, 10] remind us, there are non-log-normal stable Pareto-Lévy distributions (of logs) satisfying the following form of (4):

$$P\left(\frac{X}{x};T\right) \equiv \int_0^\infty P\left(\frac{X}{y};T-S\right) dP\left(\frac{y}{x},S\right). \tag{10}$$

Some of our general results require only that (1), (2), and (4) hold. But most of our explicit solutions are for multiplicative processes, in which (9), (10) and the following hold:

$$\mathrm{E}\left[X_{t+T} \mid X_t\right] = \int_0^\infty X\, dP\left(\frac{X}{X_t};T\right) = X_t e^{\alpha T}, \quad \alpha \geq 0. \tag{11}$$

Actually, (9) and (10) alone require that the family $P(X;T)$ is determined once a single admissible function $P(X;T_1) = P(X)$ is given, as for $T_1 = 1$. Then if α is defined by

$$e^{\alpha T_1} = \mathrm{E}\left[X_{t+T_1}/X_t\right] = \int_0^\infty X\, dP(X), \tag{12}$$

(11) is provable as a theorem and need not be posited as an axiom. McKean's appendix assumes the truth of (9) and (10) throughout. It is known from the theory of infinitely-divisible processes that $P(X)$ above cannot be an arbitrary distribution but must have the characteristic function for its log, $Y = \log X$, of the Lévy-Khintchin form:

$$\begin{aligned} \mathrm{E}\left[e^{i\lambda Y}\right] &= \int_{-\infty}^\infty e^{t\lambda Y} dP(e^Y) = e^{g(\lambda)} \\ g(\lambda) &= \mu i\lambda + \int \left(e^{t\lambda z} - 1 - \frac{i\lambda z}{1+z^2}\right)\frac{1+z^2}{z^2}\, d\psi(z), \end{aligned} \tag{13}$$

where $\psi(z)$ is itself a distribution function. In the special cases of the log-normal distribution, the log-Poisson distribution, and the log-Lévy distribution, we have respectively

$$\begin{aligned} g(\lambda) &= \mu i\lambda - \frac{\sigma^2}{2}\lambda^2 \\ g(\lambda) &= e^{t\lambda\mu} - i \\ g(\lambda) &= \mu i\lambda - \gamma|\lambda|^{\alpha^0}\left[1 + i\beta\big(\lambda/|\lambda|\big)\tan\big(\alpha^0\pi/2\big)\right], \quad 0 \leq \alpha^0 \leq 2. \end{aligned} \tag{14}$$

All of (14) is on the assumption that

$$\lim_{X\to 0} P(X) = P(0) = 0.$$

If $P(0) > 0$, there is a finite probability of complete ruin in any time interval, and as the interval approaches infinity that probability approaches 1. An example (the only one for which *exact* formulas for rational warrant pricing of all durations are known) is given by

$$\begin{aligned} \mathrm{Prob}\left\{X_{t+T} = X_t e^{aT}\right\} &= e^{-bT} \quad a, b > 0 \\ \mathrm{Prob}\left\{X_{t+T} = 0\right\} &= 1 - e^{-bT} \end{aligned} \tag{15}$$

where $\alpha = a - b \geq 0$.

Letting $w(X;T)$ be an infinitely-divisible (multiplicative) function satisfying (13), the most general pattern would be one where

$$(16)\qquad \begin{aligned} P(0,T) &= 1 - e^{-bT} \quad b > 0 \\ P(X,T) &= e^{-bT} w(X,t) + P(0,T) \end{aligned}$$

with $P(\infty, T) = e^{-bT}1 + 1 - e^{-bT} = 1$.

One final remark. Osborne, by an obscure argument that appeals to Weber-Fechner and to clearing-of-free-markets reasoning, purports to deduce, or make plausible, the axiom that the geometric mean of the distribution $P(X/x;T)$ is to be unity or that the expected value of the logarithmic difference is to be a random walk without mean bias or drift. Actually, if $\alpha = 0$ in (2), so that absolute price is an unbiased martingale, the logarithmic difference must have a negative drift. For α sufficiently positive, and depending on the dispersion of the log-normal process, the logarithmic difference can have any algebraic sign for its mean bias. Only if one could be sure that $P(X/X;T) = P(1;T) \equiv \frac{1}{2}$, so that the chance of a rise in price could be known to be always the same as the chance of a fall in price, would the gratuitous Osborne condition turn out to be true.

If $P(X,x;T)$ corresponds to a martingale or "fair game," with $\alpha = 0$ as in the Bachelier case, the arithmetic mean of the ratio X/x in always exactly 1 and the geometric mean, being less than the arithmetic mean if P has any dispersion at all, is less than 1. Its logarithm, the mean or expected value of log X_{t+T}/X_t is then negative, and the whole drift of probability for $P(X,x;T)$ shifts leftward or downward through time. In long enough time, the probability approaches certainty that the investor will be left with less than 1 cent of net worth—i.e., $P(0+,x;\infty) = 1$. This virtual certainty of almost-complete ruin bothers many writers. They forget, or are not consoled by, the fact that the gains of those (increasingly few) people who are not ruined grow prodigiously large—in order to balance the complete ruin of the many losers. Therefore, many writers are tempted by Osborne's condition, which makes the expected median of price X_{t+T} neither grow above nor decline below X_t.

However, in terms of present discounted value of future price, $X_{t+T}e^{-\alpha T}$, where the mean yield α is used as the discount factor, most people's net worth does go to zero, and this occurs *in every case of* $\alpha \geq 0$. Relative to the expected growth of X_{t+T}—i.e., relative to $X_t e^{\alpha T}$, X_{t+T} does become negligible with great probability. I call this condition "relative ruin," with the warning that a man may be comfortably off and still be ruined in this sense. And I now state the following general theorem:

THEOREM 11.4.1. *Let $P(X,x;T)$ have non-zero dispersion, satisfying*

$$\int_0^\infty X\, dP(X,x;T) \equiv e^{\alpha T},$$

$$P(X,x;T) = \int_0^\infty P(X,y;T-S)dP(y,x;S), \quad \alpha \geq 0$$

as in (2) and (4). Then

$$\lim_{T\to\infty} P(Xe^{-\alpha T},x;T) = 1 \quad \textit{for all } (X,x) > 0.$$

In the multiplicative-process case, $P(X,x;T) = P(X/x;T)$ and the theorem follows almost directly from the fact that the geometric mean is less than the arithmetic mean.

In words, the theorem says that, with the passage of ever longer time, it becomes more and more certain that the stock will be at a level whose present discounted value (discounted at the expected yield α of the stock) will be less than 1 cent, or one-trillionth of a cent.

As is discussed on page 220, we can replace relative ruin by absolute ruin whenever the dispersion of the log-normal process becomes sufficiently large. Thus, even if $\alpha > 0$ in accordance with positive expected yield, whenever the parameter of dispersion $\sigma^2 > 2\alpha$, there is virtual certainty of absolute ruin. Indeed, for the log-normal case we can sharpen the theorem to read

$$\lim_{T\to\infty} P(0+, x; T) = 1, \quad \sigma^2 > 2\alpha.$$

11.5. Summary of Probability Model

The X_{t+T} price of the common stock is assumed to follow a probability distribution dependent in Markov fashion on its X_t price alone and on the elapsed time:

$$\text{(1)} \qquad \text{Prob}\{X_{t+T} \leq X \mid X_t\} = P(X, X_t; T)$$

$$\text{(4)} \qquad P(X_{t+T}, X_t; T) \equiv \int_0^\infty P(X_{t+T}, x; T-S)\,dP(x, X_t; S), \quad 0 \leq S \leq T$$

with the expected value of price assumed to have a constant mean percentage growth per unit time of α, or

$$\text{(2)} \qquad \mathrm{E}\left[X_{t+T} \mid X_t\right] = X_t e^{\alpha T} = \int_0^\infty X\,dP(X, X_t; T), \quad \alpha \geq 0.$$

In many cases $P(X, x; T)$ will be assumed to be a multiplicative process, with the ratio X_{t+T}/X_t independent of all X_{t-w}. Then we can write

$$P(X, x; T) \equiv P\left(\frac{X}{x}; T\right),$$

where $P(u; T)$ belongs to the special family of infinitely-divisible (multiplicative) distributions of which the log-normal, log-Poisson, and log-Lévy functions are special cases. (If the Lévy coefficient α^0 in (14), which must not be confused with α of (2), were not 2 as in the log-normal case, we can show that α in (2) will be infinite. Ruling out that case will rule out the Lévy-Pareto-Mandelbrot distributions.)

11.6. Arbitrage Conditions on Warrant Prices

A warrant is a contract that permits one to buy one share of a given common stock at some stipulated exercise price X^0 (here assumed to be unchangeable through time, unlike certain real-life changing-terms contracts) at any time during the warrant's remaining length of life of T time periods. Thus, a warrant to buy Kelly, Douglas stock at \$4.75 per share until November, 1965, has $X^0 = \$4.75$ and (in March, 1965) has $T = 7/12$ years. A perpetual warrant to buy Allegheny Corporation at \$3.75 per share has $X^0 = \$3.75$ and $T = \infty$.

When a warrant is about to expire and its $T = 0$, its value is only its actual conversion value. If the stock now has $X_t = X^0$, with the common selling at the exercise price to anyone whether or not he has a warrant, the warrant is of no value. If $X_t < X^0$, a fortiori it is worth nothing to have the privilege of buying the stock at *more* than current market price, and the warrant is again worthless. Only if $X_t > X^0$ is the expiring warrant of any value, and—brokerage charges being always ignored—it is then worth the positive difference $X_t - X^0$.

In short, arbitrage alone gives the rational price of an expiring warrant with $T = 0$, as the following function of the common price known to be $X_t = X$, $F(X, T) = F(X, 0)$, where $F(X, 0) = \text{Max}[0, X - X^0]$.

A warrant good for $T_1 > 0$ periods is worth at least as much as one good only for $T_2 < T_1$, periods and generally is worth more. Hence, arbitrage will ensure that the rational price for a warrant with T_1 time to go, denoted by $F(X, T_1)$, will satisfy

$$F\big(X, T_1\big) \geq F\big(X, T_2\big) \quad \text{if } T_1 \geq T_2.$$

A perpetual warrant is one for which $T = \infty$. But recall that outright ownership of the common stock, aside from giving the owner any dividends the stock declares, is equivalent to having a perpetual warrant to buy the stock (from himself!) at a zero exercise price. Hence, a perpetual warrant cannot now sell for more than the current price of the common stock. Or, in general, arbitrage requires that

$$X \geq F(X, \infty) \geq F\big(X, T_1\big) \geq F\big(X, T_2\big) \geq F(X, 0) = \text{Max}\left[0, X - X^0\right] \tag{17}$$

where

$$\infty \geq T_1 \geq T_2 \geq 0.$$

In all that follows we shall, by an admissible choice of conventional units, be able to assume that the exercise price is $X^0 = 1$. Thus, instead of working with the price of one actual Kelly, Douglas warrant, which gives the right to buy one share of Kelly, Douglas common stock at \$4.75, we work with the standardized variable $X/X^0 = X/4.75$—the number of shares purchasable at \$1, which is of course 1/4.75 actual shares; correspondingly, the warrant price Y, we work with is not the actual Y_t but the standardized variable Y_t/X^0, which represents the price of a warrant that enables the holder to buy 1/4.75 actual shares at the exercise price of \$1. We are able to do this by the following homogeneity property of competitive arbitrage:

$$\frac{F(X, X^0; T)}{X^0} \equiv F\left(\frac{X}{X^0}, 1; T\right), \tag{18}$$

a property that says no more than that two shares always cost just twice one share. Wherever we write $F(X; T)$, we shall really be meaning (18). (Note that Tri-Continental perpetual warrants involve the right to buy 1.27 shares at \$17.76 per share. In calculating $X/X^0 = X/17.76$, we use for X the price of 1.27 shares, not of one share.)

Our conventions with respect to units ought to be adopted by advisory services dealing with warrants, to spare the reader the need to calculate X/X^0 and Y/Y^0. All this being understood, we can rewrite the fundamental inequalities of arbitrage shown in (17) as follows:

$$\begin{aligned} X \geq F(X, \infty) \geq F\big(X, T_1\big) \geq F\big(X, T_2\big) \geq F(X, 0) \\ = \text{Max}\big(0, X - 1\big), \quad \infty \geq T_1 \geq T_2 \geq 0. \end{aligned} \tag{19}$$

In Fig. 11.1a, b, the outer limits are shown in heavy black: OAB is the familiar function $\text{Max}(0, X - 1)$. (In McKean's appendix, this is written in the notation $(X - 1)^+$.) The 45° line OZ represents the locus whose warrant price equals X, the price of the common stock itself.

11.7. Axiom of Expected Warrant Gain

Mere arbitrage can take us no further than (19). The rest must be experience—the recorded facts of life. Figure 11.1a shows one possible pattern of warrant pricing. The expiring warrant, with $T = 0$, must be on the locus given by OAB. If positive length of life remains, $T > 0$, Fig. 11.1a shows the warrant always to be worth more than its exercise price: thus, OCD lies above OAB for all positive X; because OEF has four times the length of life of OCD, its value at $X = 1$ is about twice as great—in accordance with the rule-of-thumb $\sqrt{T}$ approximation; because T is assumed small, and $P(X/x; T)$ approximately symmetrical around $X/x = 1$, the slope at C is about 1/2—in accordance with the rule-of-thumb approximation that if two warrants differ only in their exercise price X^0, the owner should pay \$1/2 for each \$1 reduction in X^0, this being justifiable by the reasoning that there in only a half-chance that he will end up exercising at all and benefitting by the X^0 reduction. Note that all the curves in the figures are convex (from below) and all but the OAB and OZ limits are drawn to be strictly convex (as would be the case if $P(u;1)$ were log-normal or a distribution with continuous probability density). Our task is to demonstrate rigorously that the functions shown in the figures are indeed the only possible rational pricing patterns.

The pricing of a warrant becomes definite once we know the probability distribution of its common stock $P(X, x; T)$ if we pin down buyers' reactions to the implied probability distribution for the warrant's price $Y_t(T_t)$, in the form of the following axiom:

Axiom of mean expectation. Whereas the common stock is priced so that its mean expected percentage growth rate per unit time is a non-negative constant α, the warrant is priced so that it, too, will have a constant mean expected percentage growth rate per unit time for as long as it pays to hold it, the value of the constant being at least as great as that for the stock—or $\beta \geq \alpha$. Mathematically

$$\mathrm{E}\left[Y_{t+T}(T_t - T) \mid Y_t(T_t)\right] = e^{\beta t} Y(T_t) \tag{20}$$

for all times T it pays to hold the warrant, where

$$\beta \geq \alpha = \log_e \int_0^\infty X\, dP(X, x; 1) \geq 0. \tag{21}$$

The reader should be warned that the expected value for the warrant in (20) is more complicated than the expected value of the stock in (2). The latter holds for any prescribed time period; but in (20), the time period T must be one in which it pays to have the warrant held rather than converted. (In the appendix, McKean's corresponding expectation is given in 2.8 and in 4.8.) It is precisely when the warrant has risen so high in price (above C_T in Fig. 11.1b) that it can no longer earn a stipulated positive excess $\beta - \alpha$ over the stock that it *has* to be converted. Actually if β is stipulated to equal α, we are in Fig. 11.1a rather than Fig. 11.1b: there is never a need to convert before the end of life, and hence all points like $C_1, \ldots, C_T$ are at infinity; as we shall see, the conventional linear integral equations enable us easily to compute the resulting functions in Fig. 11.1a.

Warrants, unlike calls, are not protected against the payment of dividends by the common stock. Hence, for any stock that pays a positive dividend, say at the instantaneous rate of δ times its market value, the warrant will have to have a $\beta > \alpha$ if it is to represent as good a buy as the stock itself. Taxes and peculiar subjective reactions to the riskiness patterns of the two securities aside, at the least $\beta = \alpha + \delta > \alpha$. However, even if $\delta = 0$ and there is no dividend, buyers may feel that the volatility pattern of warrants is such that owners must be paid a greater mean return to hold warrants than to hold stocks. I do not pretend to give a theory from which one can deduce the relative values of β and α. Here, I merely postulate that they are constants (independent, incidentally, of T, the life span of the warrant).

My whole theory rests on the axiomatic hypotheses:

1. The stock price is a definite probability distribution, $P(X, x; T)$, with constant mean expected growth per unit time $\alpha \geq 0$.
2. The warrant's price, derivable from the stock price, must earn a constant mean expected growth per unit time, $\beta \geq \alpha \geq 0$.

Once these axioms, the numbers α, β, and the form of $P(X, x; T)$ are given, it becomes a determinate mathematical problem to work out the rational warrant price functions $Y_t(T_t) = F(X_t, T_t)$ for all non-negative T_t, including the perpetual warrant $F(X_t, \infty)$.

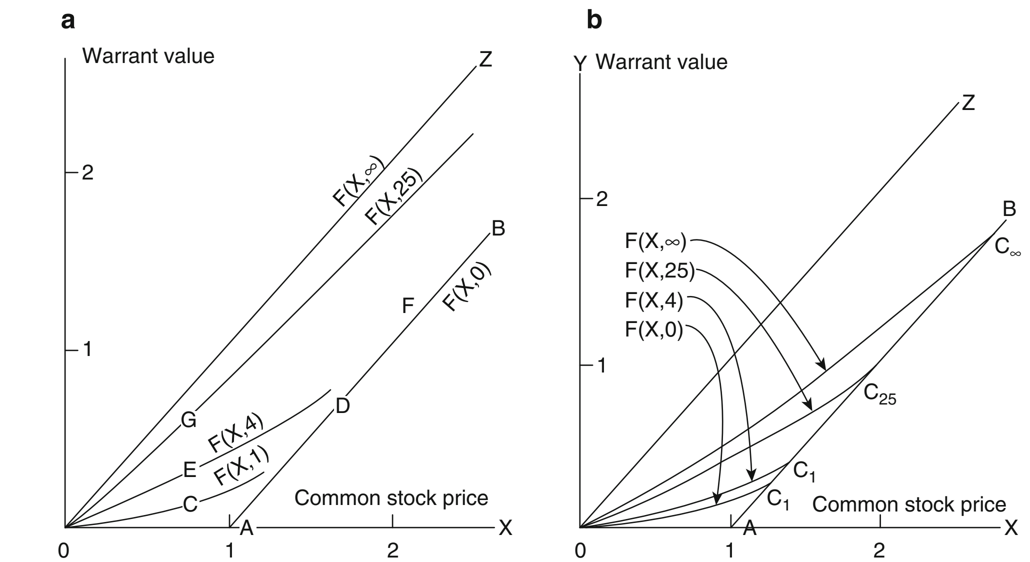

FIGURE 11.1. Rational warrant pricing[2]

11.8. Some Intuitive Demonstrations

Before giving the mathematical solutions, I shall indicate how one can deduce the paradoxical result that a perpetual warrant must have the same price as the common stock if they both have to earn the same mean yield. The reader may want to think of the fair-game case where $\beta = \alpha = 0$, a case which has a disproportionate fascination for economists because they wrongly think that if prices were known to be biased toward rising in the future, that fact would already be "discounted" and the price would already have risen to the point where α can be expected to be zero. (What is forgotten here by Bachelier and others—but not by Keynes, Houthakker, Cootner [18, 13, 14, 6, 5]. See my cited companion paper in this issue and other exponents of "normal backwardation"—is that time may involve money, opportunity cost, and risk aversion.)

A warrant is said to involve "leverage" in comparison with the common stock, and in the real world where brokerage charges and imperfect capital rationing are involved, leverage can make a difference. The exact meaning of leverage is not always clear, and writers use the term in two distinct senses. The usual sense is merely one of *percentage* volatility. Suppose a stock is equally likely to go from \$10 to \$11 or to \$9. Suppose its warrant is equally likely to go from \$5 to \$6 or to \$4. Both are subject to a \$1 swing in either direction; but \$1 on \$5 is twice the percentage swing of \$1 on \$10, as will be seen if equal dollars were invested in each security. In this sense, the warrant would be said to have twice the leverage of the stock. Leverage in the sense of mere enhanced percentage variability is a two-edged sword: as much as it works for you on the upside, it works against you on the downside. It is perfectly compatible with

[2]These graphs show the general pattern of warrant pricing as a function of the common stock price (where units have been standardized to make the exercise price unity). The longer the warrant's life T, the higher is $F(X,T)$. For fixed T, $T(X,T)$ is a convex function of X. In Fig. 11.1a, the perpetual warrant's price is equal to that of the stock, with $F(X,\infty)$ falling on OZ; it never pays to exercise such a warrant. In Fig. 11.1b, the points C_1, C_4, C_{25}, and C_∞ on AB are the points at which it pays to convert a warrant with $T = 1, 4, 25$ and ∞ years to run. Note that $F(X,\infty)$ is much less than X in this case. The pattern of Fig. 11.1b will later be shown to result from the hypothesis that a warrant must have a mean yield β greater than the stock's mean yield α.

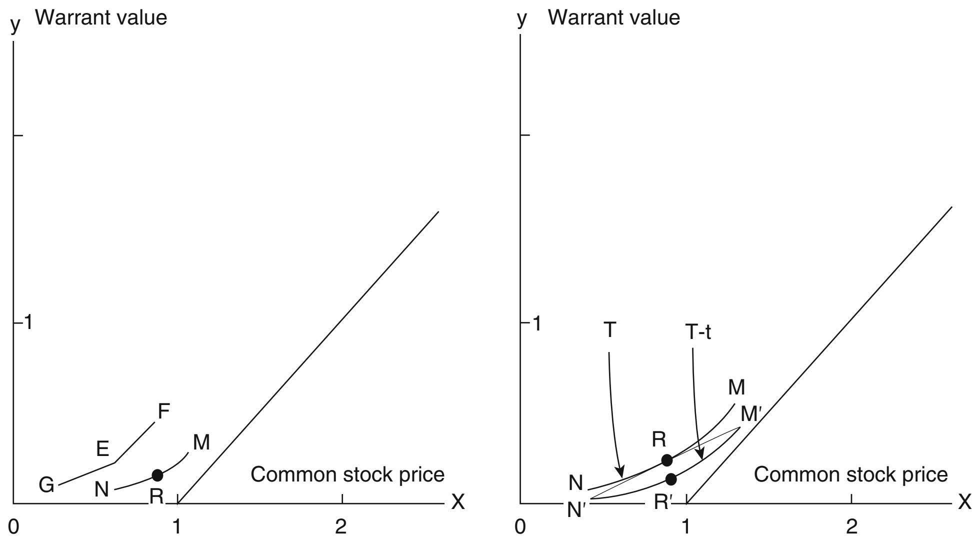

FIGURE 11.2. Warrant pricing—the perpetual case

$\alpha = \beta = 0$. (However, if there were a two-thirds chance of each security's going up \$1 and a one-third chance of its going down \$1, the warrant's β would be definitely greater than stock's α, since a mean expected return of $+33\,1/3$ ¢ on \$5 is twice that of $+33\,1/3$ ¢ on \$10; and this impinges on the second sense of leverage.)

The second sense of the term leverage is merely enhanced expected yield from the warrant in comparison with the common stock. Here is an example from the R.H.M. Warrant and Stock Survey of February 25, 1965: Newconex Holdings Warrant, Toronto Exchange, "would rise about 2.25 times as fast as the common stock on the upside and decline no faster than the common on the downside." To one who believes this, the warrant offers very good value or "leverage" in this second sense of the term. Indeed, by selling one common short and buying one warrant, one could presumably break even if the stock went down in price and make money if the stock rises—a sure-thing hedge that cannot lose if one believes the stated probability judgment.

Figure 11.2 shows for a hypothetical perpetual warrant a convex corner at the existing price E, with EF steeper for a rise than EG for a fall. Obviously, GEF could not persist if the warrant's β gain were to be no bigger than the stock's α gain. Similarly, the strongly convex NRM could not persist with $\beta = \alpha$. What pattern for a perpetual warrant could persist? Only a straight-line pattern, since for any convexity at all the mean of points along a curve must lie above the curve itself.

What straight line can be fitted in between OZ and OAB of Fig. 11.1? Obviously, only the line OZ itself—proving that the only rational price for a perpetual warrant must be that of the common stock itself when $\alpha = \beta$. (Any straight line not parallel to OZ and AB will intersect one or both of them; any intermediate line parallel to OZ and AB will hit the zero axis at positive X and then develop a corner there. So OZ alone remains as the formula for $F(X, \infty) \equiv X$.)

The curve of $F(X, T)$ for finite T can and will be convex. But as time passes, one does not move up and down the curve itself—say from R to M if X rises or from R to N if X falls. Instead, as time passes T diminishes, and one moves from R to a point below M or N on the new $F(X, T - t)$ curve; and if the two convex curves have been sketched correctly and placed in the proper shift relationship to each other, it will be found that the mean expectation of gain from the warrant is precisely that from the stock.

The moral of this is not that surveys are wrong when they recommend a bargain. It is rather that one recognizes correct or rational pricing and the absence of bargains when the warrants are priced in a certain way relative to the common stock. It is only as people act to take advantage of transient bargain opportunities that the bargains disappear. When I speak of rational or correct pricing, I imply no normative approval of any particular pattern but merely describe that pattern which (if it were to come into existence and were known to prevail) would continue to reproduce itself while fulfilling the postulated mean expectations in the form of α in (2) and β in (20). It would be a valuable empirical exercise to measure the α for different stocks at different times and deduce the value of β that the warrants earn *ex* post and that can rationalize the observed scatter of warrant and stock prices.

Intuition can carry us a bit further and throw light on the case where $\beta > \alpha$. With the warrant having to produce a better gain than the common, the curve for a perpetual warrant becomes strictly convex—as in Fig. 11.1b and in contrast to Fig. 11.1a. Furthermore, when the common price becomes very high compared to the exercise price—i.e., when $X/1$ is very large—the conversion value of the warrant becomes negligibly less than the common—i.e., $(X-1)/X = 1$. If in the period ahead the warrant can rise at most \$1 more in price than the common rises, the warrant's gain will approach indefinitely close to the common's α. But that contradicts the assumption that $\beta > \alpha$. So for X high enough, $X > C_\infty < \infty$, it will never pay to hold the warrant in the expectation of getting $\beta > \alpha$; above this C_∞, cut-off point, the warrant must be converted. What has been demonstrated here for perpetual warrants holds *a fortiori* for finite warrants with finite T. Even sooner, at $C_T < C_\infty$, it will pay to convert since with the clock running on and running out, there will be even less advantage in holding the warrant for an additional period when the stock and it have become very large.

11.9. Linear Analysis Where $\beta = \alpha \geq 0$

If the expected yields of common and warrant are to be the same in (2) and (20), there is never any advantage in converting the warrant before the end of its life. That is

$$F(X,T) > F(X,0) = \text{Max}(0, X-1), \quad T > 0;\ \beta = \alpha \geq 0. \tag{22}$$

Equation (20), postulating that the warrant have an expected gain per unit time of β, can therefore be written, for all times S,

$$\begin{aligned} &\mathrm{E}\left[Y_{t+S}(T-S) = F\big(X_{t+S}, T-S\big) \mid Y_t(T) = F\big(X_t, T\big)\right] \\ &\qquad = e^{\beta S} F\big(X_t, T\big) = \int_0^\infty F(X, T-S)\, dP\big(X, X_t; S\big) \end{aligned} \tag{23}$$

or

$$\begin{aligned} F(x,T) &\equiv e^{-\beta S} \int_0^\infty F(X, T-S) dP(X, x; S) \\ &= e^{-\beta T} \int_0^\infty F(X, 0) dP(X, x; T) \\ &= e^{-\beta T} \int_1^\infty (X-1) dP(X, x; T). \end{aligned} \tag{24}$$

This last integral equation provides, by a quadrature, the solution of our problem. From the fact that $P(X, x; 0) = 1$, $X > x$ and $= 0$, $X < x$, it is evident that

$$\lim_{T\to 0} F(x,T) = F(x,0) = \text{Max}(0, x-1).$$

We can now prove that

$$\lim_{T\to\infty} F(x,T) = x = F(x,\infty), \quad \beta = \alpha \geq 0. \tag{25}$$

Substitute $F(x,\infty) = F(x)$ into both sides of (24) to get a self-determining integral equation for $F(x)$,

$$F(x) = e^{-\beta S} \int_0^\infty F(X) dP(X, x; S). \tag{26}$$

The substitution $F(x) = x$ does satisfy (26), since by (2)

$$\begin{aligned} x &= e^{-\beta S} \int_0^\infty X\, dP(X, x; S) \\ &= e^{-\beta S} e^{\alpha S} x = e^{(\alpha-\beta)S} x = x, \quad \beta = \alpha. \end{aligned} \tag{27}$$

Any kx would also satisfy (26), but only for $k = 1$ do we satisfy

$$x \geq F(x) = kx \geq \mathrm{Max}(0, x-1).$$

To prove that the stationary solution of (26) does in fact fulfill the limit of (25), rewrite (24)

$$\begin{aligned} F(x,T) &= e^{-\beta T} \int_1^\infty (X-1) dP(X, x; T) \\ &= e^{-\beta T} \int_0^\infty (X-1) dP(X, x; T) \\ &\quad + e^{-\beta T} \frac{(1-X) dP(X,x;T)}{\int_0^1 dP(X,x;T)} \int_0^1 dP(X,x;T) \\ &= e^{-\beta T} e^{\alpha T} x - e^{-\beta T} + e^{-\beta T} \theta_1(x,T)\theta_2(x,T), \quad \text{where } |\theta_1| \leq 1. \end{aligned} \tag{28}$$

Obviously, if $\beta = \alpha > 0$, $F(x,\infty) = x + 0$, as was to be proved. For $\alpha = 0$

$$\begin{aligned} \lim_{t\to\infty} \theta_2(x,T) &= \int_0^1 dP(X,x;\infty) = 1, \quad \text{since } P(0+, x; \infty) \equiv 0 \\ \lim_{t\to\infty} \theta_1(x,T) &= \frac{\int_0^1 (1-x) dP(X,x;\infty)}{\int_0^1 dP(X,x;\infty)} = 1, \quad \text{since } P(0+, x; \infty) \equiv 0. \end{aligned} \tag{29}$$

Hence, for $\alpha = 0 = \beta$, $F(x,\infty) = x - 1 + 1 = x$, as required.

Now that (24) gives the explicit solution in the case $\alpha = \beta$, we can put in for $P(X, x; T)$ any specialization, such as

$$\begin{aligned} &P(X, x; T) = P(X/x; T) \text{ log-normal with } P(x;T) = N\big(\log x; \mu t, \sigma\sqrt{t}\big) \\ &\text{where } N(y; 0, 1) = N(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-u^2/2} du; \end{aligned} \tag{30}$$

or

$$\begin{aligned} &\mathrm{Prob}\left\{\frac{X}{x} = e^{at}\right\} = e^{-bt} \qquad \mathrm{Prob}\left\{\frac{X}{x} = 0\right\} = 1 - e^{-bt} \\ &a - b = \alpha = \beta; \quad a, b > 0. \end{aligned} \tag{31}$$

For this last case, (24) calculates out to

$$F(x,T) = \text{Max}\left(0, x - e^{-aT}\right). \tag{32}$$

Note that the $\sqrt{T}$ law does not hold true here for small T, but rather, at $x = 1$

$$F(x,T) = F(1,T) = 1 - e^{-at} = 1 - 1 + aT + \text{remainder}\left(T^2\right) = aT. \tag{33}$$

Hence, a warrant for twice the duration of a short-lived warrant should be worth about twice as much when (31) holds—even though the ratios X_{t+T}/X_t are strictly independent.

11.10. Valuation of End-of-Period Warrants

The exact solution of (24) holds only for the case $\beta = \alpha \geq 0$. It will be shown that new formulas must handle the case of $\beta > \alpha$. However, the simple integral (24) does give a solution under all cases to the simpler case of a warrant that can be *exercised* only at the end of the period T. We might call this a "European warrant" by analogy with the "European call," which, unlike the American call that is exercisable at any time from now to T, is exercisable only at a specified terminal date.

Obviously, the additional American option of early conversion can do the owner no harm, and it may help him. Denote the rational price of a European warrant by $f(x,T)$, in contrast to $F(x,T)$ of the American type warrant. Then

$$f(x,T) \leqq F(x,T), \quad 0 \leq T \tag{34}$$

and our axiom of expected gain (20) in now applicable in the form that gives the last version of (24), namely

$$\begin{aligned} f(x,T) &= e^{-\beta T}\int_0^\infty \text{Max}(0, X-1)dP(X,x;T) \\ &= e^{-\beta T}\int_1^\infty (X-1)dP(X,x;T), \quad \beta \geq \alpha \geq 0. \end{aligned} \tag{35}$$

Since this is the same formula as held in (24) for $F(x,T)$ when $\beta = \alpha$, we note that in such a case the American warrant's early conversion options are actually of no market value; or

$$f(x,T) \equiv F(x,T) \quad \text{if } \beta = \alpha. \tag{36}$$

When $\beta > \alpha$, (35) still holds. But now

$$f(x,T) < F(x,T) \tag{37}$$

for all or some positive (x,T). In the log-normal case, the strong inequality must always hold.

There seems to be a misapprehension concerning this inequality. Thus some people argue that the owner of a European call or warrant can in effect exercise it early by selling the stock short, thereby putting himself in the position of the owner of an American warrant. If this view were valid there would be no penalty to be subtracted from $F(x,T)$ to get true $f(x,T)$. Such a view is simply wrong—as wrong as the naive view that giving your broker a stop-loss order gives you the same protection as buying a put. (The fallacy here has naught to do with the realistic fact that in a bad market break your broker will not be able to execute your stop-loss order at the stipulated price; waive that point. Suppose I buy a stock at \$100 and protect it by buying (say for \$10) a six-month put on it at exercise price of \$100. You buy the stock and merely give your broker a stop-loss sell order at just below \$100. If the stock drops below \$100 at some intermediate time during the next six months, you are sold out without loss; but you do as well as I do only if the stock never subsequently rises to above \$100; and the \$10 cost of the put is precisely the market value of my opportunity to make a differential profit over you in case the stock does end up at more than \$100, after at least once dipping below \$100.) By the vector calculus that Kruizenga and I worked out for various options, after one sells a stock short and still holds a European call or warrant on it, he is not for the remainder of the time T in the position of a man who has sold out his American warrant; instead he is in the net position of holding a put on the stock. (If $(1,0)$ and $(0,1)$ represent holding a call and put respectively, the owner of an American warrant goes through the cycle $(+1,0)$ and—in midstream—$(-1,0)$, ending up with $(0,0)$. The holder of the European warrant goes through the cycle $(+1,0)$ and—in midstream—$(-1,+1)$, leaving him for the remainder of the period with $(0,+1)$.)

To see that (37) does hold when $\beta > \alpha$, recall that $F(x,T)$ cannot decrease with T. But applying to (35) the version of (24) given in (28), we can see that a long-lived European warrant does ultimately approach zero in value as $T \to \infty$. Thus, by (28) applied to $f(x,T)$,

$$\begin{aligned} f(x,T) &= e^{-\beta T}\int_1^\infty (X-1)dP(X,x;T) \\ &= e^{-\beta T}e^{\alpha T}x - e^{-\beta T} + e^{-\beta T}\theta_1\theta_2, \quad |\theta_1| < 1 \end{aligned} \tag{38}$$

$$\lim_{T\to\infty} f(x,T) = f(x,\infty) = f(x) = e^{-(\beta-\alpha)\infty}x = 0, \quad \beta > \alpha \geq 0.$$

11.11. General Formula for $\beta > \alpha \geq 0$

The last section's demonstration that $f(x,T) < F(x,T)$ when $\beta > \alpha$ provides a rigorous proof that the linear integral equation of (24) cannot apply to the proper $F(x,T)$ for this case. Hence $\beta > \alpha$ does imply that a warrant cannot possibly be worth holding at very high prices. I.e., the inequality

$$F(x,T) \geq x-1, \quad x \geq 1 \tag{39}$$

must for sufficiently high x become the equality

$$F(x,T) = x-1, \quad x > C_\infty(T;\beta,\alpha) < \infty,\ \beta > \alpha \tag{40}$$

where $\partial C_\infty/\partial T \geq 0$, $\partial C_\infty/\partial\alpha \geq 0$, $\partial C_\infty/\partial\beta \leq 0$ (McKean's appendix also proves this fact, in 2.8 and 4.7.)

In place of the integral equation (24), we have the following basic inequality to define $F(x,T)$ where $\beta > \alpha$:

$$x \geq F(x,T) \geq \text{Max}\left[0, x-1, e^{-\beta S}\int_0^\infty F(X,T-S)dP(X,x;S)\right]. \tag{41}$$

McKean's appendix terms any solution of this relation an "excessive function," and he seeks as the solution to the problem the *minimum* function that belongs to this class. Rather than arbitrarily postulate that it is the minimum function which constitutes the desired solution, I deduce from my axiom of expected gain (20) the only solution which satisfies it and which satisfies the basic inequality. It follows as a provable theorem that this does indeed give the minimum of the excessive functions. That is, any excessive function which is not the minimum will fail to earn β per unit time whenever it is being held.

How shall we find the simultaneous solution to (20) and (41)? I begin from the intuitive consideration that splitting up continuous time into small enough finite intervals will approach (from below) the correct solution for the continuous case. If a warrant can be converted only every hour, its value will be a bit less than one that can be converted at any time—less because an extra privilege is presumably worth something, only a little less because not much of a price change is to be expected in a time period so short as an hour. The approximation will be even better if we split time up into discrete minutes and still better if we use seconds. In the limit, we get the exact solution.

Let $\Delta T = h$ and define recursively in (41) for fixed h and integral n

$$\begin{aligned} F_{n+1}(x;h) &= \text{Max}\left[\left(0, x-1, e^{\beta h}\int_0^\infty F_n(X;h)dP(X,x;h)\right)\right] \\ F_0(x;h) &= \text{Max}(0, x-1). \end{aligned} \tag{42}$$

Then

$$\lim_{h=\frac{T}{n}\to 0} F_n(x;h) = F(x,T), \tag{43}$$

the desired exact solution to our problem as formulated by (20) and (41). In principle, by enough integrations, the degree of approximation can be made as close as we like.

The general properties of the solution can also be established by this procedure. Thus, if $P(X,x;1) = P(X/x;1)$ is a multiplicative process—or even if some weaker conditions are put on the way that P shrinks with an increase in x—we begin with a convex function $F_0(x;h)$ and end at each stage with a convex expectation function. Hence, by induction $F(x,T)$ and $F(x) = F(x,\infty)$ must be convex. ($F(x,T)$ will be strictly convex if $P(X/x;1)$ is log-normal or similarly smooth.) Where the slope $\partial F_n(x;h)/\partial x$ exists it can be shown inductively that its value must lie in the closed interval $[0,1]$, a property which must hold for $F(x,T)$. At the critical conversion point C_T, where $C_T - 1 = F(C_T,T)$, one expects the slopes of the two equal branches to be equal.

It will be instructive to work through an example in which time itself is divided into small, discrete intervals $t = 0,1,2,\ldots$, etc. And suppose that $P(X,x;1)$ corresponds to a simple, multiplicative random walk of martingale type, where

$$\begin{aligned} \text{Prob}\left\{\frac{X_{t+1}}{X_t} = \lambda > 1\right\} &= p > 0, \\ \text{Prob}\left\{\frac{X_{t+1}}{X_t} = \lambda^{-1}\right\} &= 1 - p = q > 0. \end{aligned}$$

The gain per unit time is now given by

$$e^{\alpha} = p\lambda + q\lambda^{-1} = 1, \quad \text{where } \lambda = \frac{1-p}{p}.$$

It will help to keep some simple numbers in mind: e.g. $p = 1/3$, $q = 2/3$, $\lambda = 2$, making $\alpha = 0$ and the $[X_t]$ sequence a "fair game" or martingale, with zero net expected yield.

If β is also set equal to zero, so that it never pays to exercise the warrant, (41) reduces to the simple form (35), and we are left with the familiar partial-difference equation of the classical random walk (but in terms of $\log X_t$, not X_t itself). Specifically, $\log X_t/X_0$ will take on only integral values for $t > 0$; if later we make λ nearer and nearer to 1, the fineness of the grid of integral values will increase; and it will cause little loss of generality to suppose that initially $X_0 = \lambda^k$, where k is a positive or negative integer. This being assured, a two-way $F(X,m) = F(\lambda^n, m)$ can always be written as a two-way sequence F_{nm}, where m denotes non-negative integers and n integers that can be positive, negative, or zero. Corresponding to (35), we now have:

$$F_{n0} = \text{Max}(0, \lambda^n - 1)$$
$$F_{n1} = pF_{n+1,0} + qF_{n-1,0}, \quad p + q = 1$$
$$\dots\dots\dots\dots\dots\dots\dots$$
$$F_{n,m+1} = pF_{n+1,m} + qF_{n-1,m}$$
$$\dots\dots\dots\dots\dots\dots\dots$$
$$F_{n\infty} = F_n = pF_{n+1} + qF_{n-1}.$$

The last of these is an ordinary second-order difference equation with constant coefficients, whose characteristic polynomial is seen to be

$$p\sigma^2 - \sigma + (1-p) = p(\sigma - 1)(\sigma - \sigma_2), \quad \text{where } \sigma_2 = \frac{1-p}{p} = \lambda > 1.$$

Write the general solution for F_n as

$$F_n = e_1(1)^n + e_2\sigma_2^n.$$

Since $F_n \to 0$ as $n \to -\infty$, we must have $e_1 = 0$. Since

$$\text{Max}(0, X-1) = \text{Max}\big(0, \lambda^n - 1\big) \leq F(X) = e_2\lambda^n \leq \lambda^n = X,$$

we must have

$$e_2 = 1, \qquad F_n = \lambda^n, \qquad F(X) = X$$

verifying the general derivation of (25).

Now drop the assumption that $\alpha = 0$, but still keep $\beta = \alpha$. The above partial-difference equations are unchanged except that now (p, q) are replaced by (Bp, Bq) where

$$B^{-1} = e^{\alpha} = p\lambda + q\lambda^{-1} > 1.$$

Again it can be shown that $\sigma_2 = \lambda$ is a root of the characteristic polynomial, and that only if $(e_1, e_2) = (0, 1)$ can the boundary condition be satisfied. Again we confirm (35)'s $F(X) = X$ solution for $\alpha = \beta$.

Now let $B^{-1} = e^{\beta} > e^{\alpha} = p\lambda + q\lambda^{-1} = \Phi(\lambda) \geq 1$.

For $m \leq \infty$, there will exist critical integral constants n_m, equal (except for the coarseness of the integral grid) to $\log C_m$, above which warrant conversion is mandatory. The partial-difference equations derivable from (41) now become

$$\begin{aligned}
F_{n,m} &= \lambda^n - 1, \quad n \geq n_m > 0 \\
F_{n,m} &= BpF_{n-1,m-1} + BqF_{n-1,m-1} \leq \lambda^n - 1, \quad n < n_m, \ (m = 1, 2, \dots) \\
F_{n,\infty} &= F_n = BpF_{n+1} + BqF_{n-1}, \quad n < n_1 \\
&= \lambda^n - 1, \quad n \geq n_1.
\end{aligned}$$

These relations define the sequence (n_m) recursively—e.g., n_1 is the lowest integer for which

$$\lambda^{n_1} - 1 \geq Bp\big(\lambda^{n_1}\lambda - 1\big) + Bq\big(\lambda^{n_1}\lambda^{-1} - 1\big).$$

With n_1 known, we have initial conditions to the right to determine F_{n+1} for $n \leq n_1$. The difference equation for $F_{n,1}$, which can be written symbolically in terms of the operators E

and E^{-1} defined by $EF_{n,m} = F_{n+1,m}$, $E^{-1}F_{nm} = F_{n-1,m}$ as $\Phi(E)F_{n,1} = 0$ then determines F_{n1}. With this known we determine n_2 as the smallest integer for which

$$\lambda^{n_2} - 1 \geq B\Phi(E)F_{n,1};$$

then determine $F_{n,2}$ by $\Phi(E)F_{n,2} = 0$, etc.

The constant n, can be determined along with $F_{n,\infty}$ by the following relations

$$\Phi(E)F_n = 0; \qquad F_n = e_1\sigma_1^n + e_2\sigma_2^n,$$

where the characteristic polynomial can be shown to be

$$\sigma\Phi(\sigma) - \sigma = Bp\big(\sigma - \sigma_1\big)\big(\sigma - \sigma_2\big),$$

where

$$0 < \sigma_1 < 1 < \lambda < \sigma_2 = \lambda^\gamma, \quad \gamma > 1.$$

If $F_n \to 0$ as $n \to -\infty$, $e_1 = 0$; to determine e_2, and $n_2 = a$ for short, we set

$$e_2\lambda^{\gamma a}\lambda^\gamma = \lambda^a\lambda - 1, \qquad e_2\lambda^{\gamma a} = \lambda^a - 1,$$

or

$$\big(\lambda^a - 1\big)\lambda^\gamma = \lambda^a\lambda - 1, \qquad \big(\lambda^\gamma - \lambda\big)\lambda^a = \gamma - 1$$

$$a = \log(\gamma - 1) - \log\left(\lambda^\gamma - \lambda\right), \qquad e_2 = (\lambda^a - 1)\lambda^{-\gamma a},$$

where of course γ is a function of α and β through its dependence on the coefficients Bp and λ. $F_n = \lambda^{\gamma n}$ means in terms of X, the antilog of n, that $F(X) = eX^\gamma$, $\gamma \geq 1$ as our first general answer.

We can always convert one-period partial difference equations into N-period equations. When we do this (p, q) are replaced by $(p^3, 2pq, q^2), \ldots$ and by $(p^x, {}_xC_1p^{x-1}q, \ldots, q^x)$ where ${}_xC_1$ are the familiar binomial coefficients. By the usual central limit theorem, these approach the normal distribution. But since these coefficients apply to the $F_{n,m}$, which refer to the logarithms of X, we arrive at the log-normal distribution. Hence, if we can prove that the partial-difference equation, not merely for $\Phi(E)F_n = (BpE + BqE^{-1})F_n$ but for any general set of probabilities

$$\Phi(E)F_n = B\sum_{-k}^{\infty} p_jE^jF_n = F_n, \qquad \sum_{-k}^{\infty} p_j = 1,$$

satisfies the $F(X) = cX^\gamma$ power law, we have strong heuristic evidence that this will he the exact case for the log-normal case—as McKean has rigorously proved in the Appendix. The characteristic polynomial of this last becomes

$$-1 + \sigma^k\Phi(\sigma) = \big(\sigma - \sigma_1\big)\big(\sigma - \lambda^\gamma\big)\Phi_2(\sigma),$$

where, as before

$$0 < \sigma_1 < \lambda \leq \lambda^\gamma, \quad \gamma \geq 1\Phi_2(\sigma),$$

and $\Phi_2(\sigma)$ is a polynomial with no roots greater than 1 in absolute value. Hence, in the general solution

$$F_n = \Sigma c_1\sigma_1^n = c_2\lambda^\gamma + \text{Remainder},$$

all the c's except c_2 must vanish if $F_n \to 0$. The value of c_2 and the critical conversion point n_2, is determined just as in the simple (p, q) case. If the grid in very fine because $\lambda \to 1$, $\lambda^{n_2} = c_2 = \gamma/(\gamma - 1)$ to an increasingly good approximation.

As a preview to McKean's exact result for the continuous-time case, I shall sketch the usual Bachelier-Einstein derivation of the partial differential equations of probability diffusion—of so-called Fokker-Planck type—by applying a limit process to the discrete partial difference equations. From now on consider $n = \log x$ as if it were a continuous rather than integral variable. Bachelier wrote in 1900

$$p_{n,t} = \frac{1}{2}p_{n+t,t-1} + \frac{1}{2}p_{n-t,t-1},$$

or

$$p_{n,t+\Delta t} = \frac{1}{2}p_{n+\Delta n,t} + \frac{1}{2}p_{n+\Delta n,t}$$

$$\frac{\Delta t}{(\Delta n)^2}\frac{p_{n,t+\Delta t} - p_{n,t}}{\Delta t} = \frac{1}{2}\frac{\left(p_{n+\Delta n,t} - p_{n,t}\right)}{(\Delta n)^2} + \frac{1}{2}\frac{\left(p_{n-\Delta n,t} - p_{n,t}\right)}{(\Delta n)^2}.$$

Now if $\Delta t \to 0$, with $\Delta t/(\Delta n)^2 \to 2c^2$, we get the Fourier parabolic equation

$$c^2\frac{\partial p(n,t)}{\partial t} = \frac{\partial^2 p(n,t)}{\partial n^2}.$$

Bachelier assumed a fair game with probabilities of unit steps in either direction equal to 1/2. If we replace $(1/2, 1/2)$ by (p, q) so that the random walk has a biassed drift of α as its expected instantaneous rate of growth, we find $p(n - \alpha t, t)$ satisfying the above equation and hence the requisite distribution $r(n,t) \equiv p(n - \alpha t, t)$ satisfies

$$\frac{\partial^2 r(n,t)}{\partial n^2} = c^2\frac{\partial r(n,t)}{\partial t} + c^2\alpha\frac{\partial r(n,t)}{\partial n}.$$

Bachelier and Einstein were talking about the diffusion of probabilities. But we have seen that the warrant prices $F_{n,t}$, now written as $F(e^n, t) = \psi(n,t)$, satisfy similar partial-difference equations, the only difference being (i) that the coefficients add up to less than 1 when $\beta > \alpha$; and (ii) the boundary conditions for c_1 become rather complicated. Just as we had a simple second-order (partial) difference equation $E\Phi(E)F_{n,t} = EF_{n,t}$, we derive in the limit—as McKean shows in 3, and 5, drawing on the work of E. B. Dynkin—a simple (partial) second-order differential equation for $\psi(n,t)$, which in terms of $\log x = n$ becomes,

$$\frac{\sigma^2}{2}\frac{\partial^2\Psi(n,t)}{\partial n^2} + \delta\frac{\partial\psi(n,t)}{\partial n} - \frac{\partial\Psi(n,t)}{\partial t} - \beta\Psi(n,t) \equiv 0, \quad \delta = \alpha - \frac{\sigma^2}{2}$$

$$\Psi(n,0) = \text{Max}\left(0, e^n - 1\right)$$

$$\Psi(e^{n_t}, t) = e^{n_t} - 1.$$

It is understood that the equation holds for (n,t) to the left of $n = e^{n_t}$ and that $\psi(-\infty, t) \equiv 0$. However, it is a difficult task to compute the e_t, function, even using the high contact property $\partial F(c_t, t)/\partial x = 1$.

The perpetual warrant is much simpler, since then $\psi(n,\infty) = \psi(n)$ with $\partial\psi(n,\infty)/\partial t = 0$, giving the *ordinary* differential equation

$$\frac{\sigma^2}{2}\psi''(n) + \delta\psi'(n) - \beta\psi(n) = 0, \quad n < c,$$

$$\psi(-\infty) = 0, \qquad \psi(c_2) = c_2 - 1, \quad \psi'(c_2) = e^{n_2}.$$

The general solution can be written as a sum of two exponentials, in terms of the roots of the characteristic polynomial

$$\frac{\sigma^2}{2}\rho^2 + \delta\rho - \beta = \frac{\sigma^2}{2}\left(\rho - \rho_1\right)\left(\rho - \rho_2\right), \quad \rho_1 = \gamma > 1 > \rho_2.$$

If the boundary conditions are to be realized, the ρ_2 root must be suppressed and we are left with

$$\psi(n) = \big(c_2 - 1\big)\frac{e^{n\gamma}}{c_2}, \quad \text{or}$$
$$F(x) = (c_2 - 1)\left(\frac{x}{c_2}\right)^{\gamma}$$
$$\gamma = \frac{c_\infty}{c_2 - 1}.$$

11.12. Intuitive Proofs from Arbitrage

Equation (18), which related the rational price of a warrant with any exercise price X° to the formula for a warrant with $X^\circ \gtrless 1$, can be used directly to deduce restrictions on the way $F(x,T;X^\circ)$ varies with X°. Because $F(x,T)$ has been shown to be convex with numerical slope on the closed interval $[0,1]$, (18) can deduce that the numerical slope of $F(x,T;X^\circ)$ with respect to X° must be on the closed interval $[-1,0]$—i.e.,

$$-1 \le \frac{F\big(x,T;X^\circ + \Delta X^\circ\big) - F\big(x,T;X^\circ\big)}{\Delta X^\circ} \le 0,$$

or

$$-1 \le \frac{\partial F\big(x,T;X^\circ\big)}{\partial X^\circ} \le 0, \tag{44}$$

where the last partial derivatives, if they do not exist at certain corners, can be interpreted as either left-hand or right-hand derivatives.

One proves (44) directly by differentiating (18) with respect to X°, to get

$$\begin{aligned}\frac{\partial F\big(x,T;X^\circ\big)}{\partial X^\circ} &= \frac{\partial}{\partial X^\circ}\left\{X^\circ F\left(\frac{x}{X^\circ},T\right)\right\}\\ &= F\left(\frac{x}{X^\circ},T\right) - \frac{x}{X^\circ}\frac{\partial F\big(\frac{x}{X^\circ},T\big)}{\partial\big(x/X^\circ\big)}.\end{aligned} \tag{45}$$

That the right-hand expression in (45) is non-positive follows directly from the definition of convexity of $F(x,T)$ when $F(0,T) \equiv 0$. That it is not algebraically less than -1 follows from the fact that $F(x,T) \ge \text{Max}(0, x-1)$.

Intuitive economic arguments provide an alternative demonstration that

$$-1 \le \frac{\partial F\big(x,T;X^\circ\big)}{\partial X^\circ} \le 0. \tag{46}$$

An increase in the exercise price X° must, if anything, lower the value of the warrant since it then entails a higher future payment. But a fall of \$1 in X° can never be worth more than \$1, since stapling a \$1 bill to a warrant with X° exercise price is a possible way of making it the full equivalent of a warrant exercisable at $X^\circ - \$1$. Hence, we have established (46).

The condition for high contact at a conversion point C_T, namely $\partial F(x,T)/\partial x \to 1$ as $x \to C_T$, seems intuitively related to realization of left-hand equality in (46) as $x \to C_T/X^\circ$, which in turn seems intuitively related to the probability that, when x is already near C_T, x will be reaching C_t in a sufficiently short future time. For the log-normal Brownian motion of (30) and the special case of (31), these conditions for high contact will be realized. But for any solution of the Chapman-Kolmogorov equation (4) of log-Poisson type, like that discussed by McKean and involving jumps, high contact will definitely fail. If we rule out combinations of Poisson jumps, only (30) and (31) and combinations of them like that shown in (16) would seem to be relevant. For them high contact is indeed ensured. And for both of these types an exact power-of-x solution for the perpetual warrant has been shown by McKean to hold.

11.13. Final Exact Formula for Perpetual Warrant in Log-Normal Case

McKean has proved in (30) the following exact smooth formula for $F(x, \infty) = F(x)$, for the log-normal case

$$\begin{aligned} F(x) &= \frac{(\gamma-1)^{\gamma-1}}{\gamma^\gamma} x^\gamma = (c-1)\left(\frac{x}{c}\right)^\gamma, \\ x \le c &= \frac{\gamma}{\gamma-1} > 1 = x - 1, \quad x \ge c, \quad \gamma = \frac{c}{c-1} > 1. \end{aligned} \tag{47}$$

This has the nice property of high contact, with $F'(c) = 1$ from either direction. Examples of (47) for different values of γ would be

$$F(x) = 3\left(\frac{x}{4}\right)^{4/3}, \qquad \gamma = 4/3, c = 4$$

$$F(x) = 2\left(\frac{x}{3}\right)^{3/2}, \qquad \gamma = 3/2, c = 3$$

$$F(x) = \frac{1}{4}x^2, \qquad \gamma = 2, c = 2.$$

The last of these formulas has been proposed, in different notation, on a purely ad hoc empirical basis by Guigère [11]. The notation there, of course, needs to be related to my notation involving $\frac{X}{X^\circ}$ and $\frac{Y}{Y^\circ}$, as in Fig. 11.3.

I append a brief *table* of *values* (Fig. 11.3) of $F(x)$ for what would seem to be empirically relevant values of γ.[3] Figure 11.4 plots as straight lines on double-log paper $F(x)$ for various values of γ.

x	0	.25	.50	.75	1.0	1.25	1.50	1.75	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$\frac{1}{4}x^2$	0	.016	.062	.141	.250	.391	.562	.766	1.0°	1.5°	2.0°	2.5°	3.0°	3.5°	4.0°
$2\left(\frac{x}{3}\right)^{3/2}$	0	.048	.136	.250	.385	.538	.707	.891	1.088	1.521	2.0°	2.5°	3.0°	3.5°	4.0°
$3\left(\frac{x}{4}\right)^{4/3}$	0	.074	.188	.322	.472	.636	.811	.996	1.190	1.603	2.044	2.511	3.0°	3.5°	4.0°
$4\left(\frac{x}{5}\right)^{5/4}$	0	.094	.225	.373	.535	.707	.888	1.078	1.572	1.682	2.112	2.561	3.026	3.506	4.0°

Explanation: $x = X/X^\circ$, the common stock price ÷ exercise price, $y = Y/Y^\circ$, warrant price ÷ exercise price is given by $y = (c-1)(x/c)^\gamma$ where $\gamma = c/(c-1)$; value of γ depends on α/σ^2 and β/σ^2 as given in Equation (48).

° Warrant at conversion value, $x - 1$.

FIGURE 11.3. Rational price for perpetual warrant in log-normal model

[3] Acknowledgment is made to F. Skilmore for these computations.

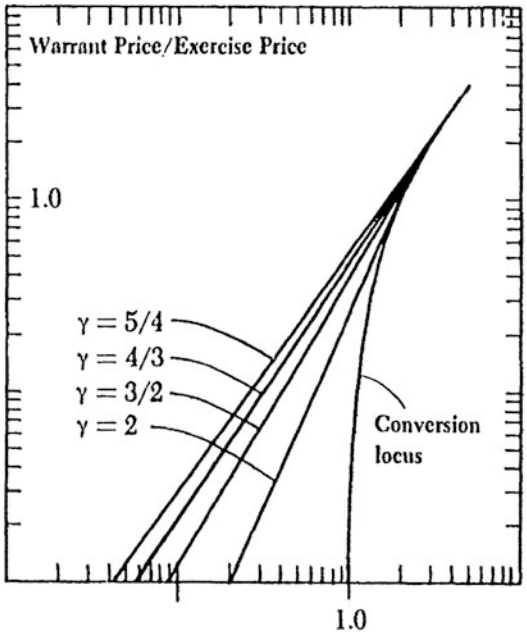

FIGURE 11.4. Rational price for perpetual warrant in log-normal model

To relate γ to α, β, and the dispersion parameter σ^2 in the log-normal distribution, I rewrite McKean's formula for γ (in the Appendix) as

$$\gamma = \left(\frac{1}{2} - \frac{\alpha}{\sigma^2}\right) + \sqrt{\left[\frac{1}{2} + \frac{\alpha}{\sigma^2}\right]^2 + 2\left[\frac{\beta}{\sigma^2} - \frac{\alpha}{\sigma^2}\right]}. \tag{48}$$

That γ is a function of $(\alpha/\sigma^2, \beta/\sigma^2)$ follows from the invariance of the problem under transformations of the unit used to measure time. Similar ratios of parameters occur in the log-Poisson process and the multiplicative-translation-with-absorbtion process of (15).

It is instructive to hold (α, β) fixed in (48), and examine how γ varies with the dispersion parameter σ^2 of the log-normal process for the stock. When $\sigma^2 \to \infty$, the difference $(\beta - \alpha)/\sigma^2 \to 0$ and $\gamma \to 1$, the case where the warrant never gets prematurely converted. Such a large value for the dispersion parameter σ^2 would create a very large α if the drift of $\log X_t$ were not strongly negative. Any such negative drift implies that it is "almost certain" that the holder of the stock will be "eventually" ("almost completely") ruined—even though the stock does have a positive mean capital gain. Note the tricky statement involving a triple limit, as in the earlier theorem on (virtually) certain (relative) ruin.

We will see in (50) that $\gamma = \sqrt{\beta/\alpha}$ when $\sigma^2 = 2\alpha$ and there is no drift at all to $\log X_t$ and hence to X_t. In this knife-edge case of Osborne, where the geometric mean of future X_{t+T} just equals X_t, the probability of a future capital loss (or gain) is exactly one half. At the other limit, where the dispersion $\sigma^2 \to 0$, we put $(\alpha/\sigma^2, \beta/\sigma^2) = (\infty, \infty)$ in (48) and find $\gamma \to \beta/\alpha$. This can be verified by substituting into $y = (c-1)(X/c)^\gamma$ the now-*certain* path $X(t) = X_0 e^{\alpha T}$ and deducing $Y(t) = Y_0 e^{\beta t} = Y_0 e^{\gamma \alpha T}$, with $\gamma = \beta/\alpha$.

To estimate γ empirically, one might regress log warrant price against log common price, γ being the regression coefficient. Then α might be estimated statistically by calculating the mean percentage gain per unit time of the common, or by computing $E[X_{t+1}]/X_t = e^\alpha$. Then β will be determined by the formula (48) for γ once one has an estimate of σ^2. Since σ is the standard deviation of $\log(X_{t+1}/X_t)$, it can be estimated from the sample variance of this last variate. The consistency of the model with the facts could then be checked by calculating β separately as the mean value of the warrant's gain, or by

$$E\big[Y_{t+T}\big]/Y_T = e^{\beta T}$$

where T is always less than the time after t when it pays to convert the warrant. A further check on the log-normality model comes from the fact that, when the "instantaneous variance per unit of time" of X_t is σ^2, the instantaneous variance for unit time of Y_t should work out to be $\gamma^2\sigma^2$, greater than σ^2 by the factor $\gamma^2 > 1$.

I am not presenting any empirical results here. But I shall draw upon some findings of others by way of illustrating the theory. (Incidentally, they suggest remarkably high β/α, giving the warrants a suspiciously favorable return.)

Osborne [25, p. 108] finds some empirical warrant for his theoretically dubious axiom that $\log X_t$ takes an unbiased random walk, with neither upward nor downward drift. If μt represents the net drift of $\log X_t$, we have

$$\mu = \alpha - \frac{\sigma^2}{2} = 0, \qquad \frac{\alpha}{\sigma^2} = \frac{1}{2}. \tag{49}$$

Substituting these values into (48) gives

$$\gamma = \sqrt{\frac{\beta}{\alpha}}, \quad \text{when } \frac{\alpha}{\sigma^2} = \frac{1}{2}. \tag{50}$$

Osborne and many investigators report average capital gains on a stock of three to five per cent per year. So set $\alpha = .04$. Finally Giguère in the cited paper [11] infers $\gamma = 2$ from empirical scatters of perpetual warrant prices against their common stock prices. (My casual econometric measurements suggest $\gamma = 2$ is much too high: these days one can rarely buy a long-lived warrant for only one-fourth of the common when the common is selling near its exercise price. But accept $\gamma = 4$ for the sake of the demonstration.) Combining $\mu = 0$, $\alpha = .04$, $\gamma = 2$, we get for the mean return per year for holding the warrant no less than 16 per cent!—i.e., $\beta = \gamma^2\alpha = 4(.04) = .16$.

This does seem to be a handsome return, and one would expect it to be whittled away over time—unless people are exceptionally averse to extra risk. The high β return would be whittled away as people bid up the prices of perpetual warrants until they approached the value of the common stock itself—at which point $\beta,\ \alpha,\ \gamma = 1$, and $e = \infty$. There is no other way. Yet this does not seem to happen. Why not? One obvious explanation is that whenever a stock pays a regular dividend of δ per period, β will, taxes aside, naturally come to exceed α by at least that much. But there are stocks that pay no dividend which still sell much above their perpetual warrants. Perhaps a departure from our assumption of a stationary time series, in the form of a supposition that there will later be a regular dividend, can help explain away the paradox. Coming events do cast their shadow before them.

I should like now to sketch a theory to explain why $\beta - \alpha$ cannot become too large. If $\beta > \alpha$ so that $\gamma > 1$, hedging will stand to yield a sure-thing positive net capital gain (commission and interest charges on capital aside!). This follows from the concept of leverage as curvature in Fig. 11.2. Let the stock be initially at X_0 with the warrant at $F(X_0) = Y_0$. Then buying \$1 long of the warrant and selling \$1 short of the common gives the new hedged variate $Z = Y/Y_n - X/X_n$. Whether X goes up or down, Z is sure to end up greater than 1, with a positive gain. Indeed, its expected gain per unit time is $\beta - \alpha$. But there will be a variance per unit time around this mean value that works out to $(\gamma - 1)^2\sigma^2$. This variance will be quite small when γ is near to 1, but with $\gamma = 1$ it is likely that the difference $\beta - \alpha$ will also be small.

In the example worked out earlier from the data of Osborne and Giguère, a hedger would have the same variance as would a buyer of the common stock; but instead of earning 4 per cent a year, he would earn 12 per cent a year. And, commissions aside, he would have no risk of a positive loss. This would seem like almost too much of a good thing. Under the stock exchange rules, I believe he would have to put up about the same amount of money as margin to engage in the hedged transaction as to buy a dollar's worth of the warrant or stock outright; he would not need margin money for each side of the hedged transaction. So he would have to reckon in the opportunity cost of the safe interest rate per unit time of money itself, ρ. Presumably though, the buyer of the common stock has already felt that its $\alpha = .04$ return was adjusted to compensate for that ρ. (If the stock pays a percentage dividend, δ, the excess $\beta - \alpha$ includes compensation for δ, ρ and for extra riskiness. Actually, if the excess of β over α comes only from the fact of the dividend δ, there is no advantage to be gained from the hedge; this is because the man who sells the common short must make good the dividend, and that will reduce the apparent profit of the hedge to zero. Hence in what follows, I deal only with the excess of β over α that in unrelated to dividends, and I ignore all dividends.)

If hedging arbitrage *alone* is counted on to keep $\beta - \alpha$ small, under present margin requirements we should expect $\beta - \alpha = \rho$ if riskiness were not a consideration. Since there is some aversion to dispersion around the mean gain from the hedge, we should not expect from hedging arbitrage *alone* that $\beta - \alpha < \rho$. On the other hand, if people are risk averters and $\gamma < 2$, as seems realistic, it is hard to see how one could get $\beta - \alpha > \alpha$, since people would shift from holding X outright to holding a hedged position Z if the latter had the greater return, less variance, and no chance of loss. One could, in principle, learn from stock exchange records how much hedging is in fact being done, since a rational hedger will minimize margins by dealing with one firm on both sides of his hedge. It is my impression that not much warrant hedging is in fact done, although in convertible bonds there does seem to be a greater volume of hedging. Still if γ and $\beta - \alpha$ threatened to become too large, potential hedgers would become actual hedgers. Hence, the limits derived above do have some relevance, particularly

$$\beta - \alpha < \alpha. \tag{51}$$

11.14. Conclusion

The methods outlined here can be extended by the reader to cases of calls and puts, where the dividend receives special treatment different from the case of warrants, and to the case of convertible bonds.

Paul A. Samuelson, Ph.D., Professor of *Economics, Massachusetts Institute of Technology,* Formerly consultant to the Natural Resources Planning Board (1945) and the U.S. Treasury (1945–52). Guggenheim Fellow, 1948–49, Ford Faculty Research Fellow, 1958–59. Social Science Research Council Predoctoral Fellow, 1935–37. Society of Fellows, Harvard University, 1937–40. Member of the American Academy of Arts and Sciences, the American Philosophical Society, and National Commission on Money and Credit Advisory Board (1958–60); President's National Goals Commission Adviser (1960). Honorary Degrees from the University of Chicago, Oberlin College, Ripon College, and Boston College. President of the American Economic Association (1961) and of the Econometric Society (1951). Author of *Foundations of Economic Analysis*, 1947; *Economics, An Introductory Analysis*, 1948; *Readings in Economics*, 1955; *Linear Programming and Economic Analysis*, 1958.

Appendix. *A free boundary problem for the heat equation arising from a Problem of Mathematical Economics.*

Henry P. McKean, Jr., M.I.T. [4]

11.15. Introduction

Paul Samuelson has developed a model of warrant pricing from the economic standpoint; the purpose of the present article is to add some mathematical complements.

Samuelson supposes that the motion of the price $x(t) \geq 0$ of the common stock is a (multiplicative) differential process; this means that for each $s \geq 0$, the (scaled) future motion $x(t+s)/x(s) : t \geq 0$ is independent of the past $x(t) : t \leq s$ and has the same statistics as $x(t)/x(0) : t \geq 0$. Define $P_1(B)[E_1(f)]$ to be the chance of the event B [expectation of the function f] for prices starting at $x(0) = 1$ and impose the condition $E_1(x) < \infty$. $E_1(x) = e^{\alpha t}$ follows; it is assumed that $\alpha \geq 0$.

Define $h = h(t, \xi)$[5] to be the "correct" price of a warrant to purchase the common stock at unit price, as a function of the time of purchase $t \geq 0$ and of the current price $\xi \geq 0$, subject to the additional condition that the warrant appreciate at the rate $\beta \geq \alpha$ up to such time as it becomes unprofitable still to hold it. The problem of computing h has the following mathematical expression: *find the smallest solution $f = h$ of*

$$f(t, \xi) \geq e^{-\beta s} E_1\big[f\big(t-s, \xi x(s)\big)\big] \quad (s \leq t,\ \xi \geq 0)$$

that lies above $(\xi - 1)^+ \equiv$ *the greater of* $\xi - 1$ *and* 0; the simpler problem of finding the "correct" price $h(\infty, \cdot)$ of the perpetual warrant can be expressed in the same language as follows: *find the smallest solution $f = h$ of*

$$f(\xi) \geq e^{-\beta t} E_1\big[f\big(\xi x(t)\big)\big] \quad (t \geq 0,\ \xi \geq 0)$$

that lies above $(\xi - 1)^+$.

The existence of h is proved and its simplest properties discussed in Sects. 11.16 and 11.18 below: if $\beta > \alpha$, h turns out to be an increasing convex function of ξ up to a point $\xi = c(t) > 1$ [the corner], to the right of which it coincides with $\xi - 1$; c and h increase with time to $c(\infty) < \infty$ and $h(\infty, \xi) < \xi$. The latter is computed in Sect. 11.17 for a (multiplicative) Brownian motion of prices $[h = (c-1)(\xi/c)^\gamma,\ c = \gamma/(\gamma - 1)]$ and also for a (multiplicative) Poisson process of prices [$h =$ a broken line], and, in Sect. 11.19, h is computed for $t \leq \infty$ and a (multiplicative) translation of prices with possible absorbtion at 0. A partial solution of the problem for $t < \infty$ and a (multiplicative) Brownian motion of prices is described in Sect. 11.20: it leads to a free boundary problem for the heat equation, the free boundary being a solution of an unfortunately intractable integral equation due to I. I. Kolodner [19].

An unsolved problem is to find a nice condition on the prices that will make $h^-(c) =$ the left slope at the corner be 1, as in the Brownian case of Sect. 11.20. $h^-(c) \leq 1$ is automatic. Samuelson has conjectured that this will be the case if $Q = P_1[x(t) \leq 1,\ t \downarrow 0] = 0$ [the alternative is $Q = 1$], but I could not prove it in general. Another inviting unsolved problem is presented by the integral equation for the free boundary of Sect. 11.20.

I must not end without thanking Professor Samuelson for posing me this problem and for several helpful conversations about it.

[4]The partial support of the Office of Naval Research and of the National Science Foundation, NSF G-19684, is gratefully acknowledged.

[5]Samuelson's notation for this is $F(X, T)$.

11.16. Perpetual Warrants

Consider a (multiplicative) differential process with sample paths $t \to x(t) = x(t+) \geq 0$, probabilities $P_1(B)$, and expectations $E_1(f)$ for paths starting at $x(0) = 1$, i.e., let $P_1[x(0) = 1] = 1$ and, conditional on $x(t_1) > 0$, let $x(t_2)/x(t_1)$ be independent of $x(s) : s \leq t_1$ and identical in law to $x(t_2 - t_1)$ for each choice of $t_2 \geq t_1 \geq 0$. $P_a(B)$ and $E_a(f)$ denote probabilities and expectations for the motion starting at $x(0) = a \geq 0$; this motion is identical in law to $[ax, P_1]$; *esp.*, $P_0[x(t) = 0,\ t \geq 0] = 1$ and $P_a[x(t) < b] = P_1[x(t) < b/a]$ for $b, a > 0$. $[x, P]$ *begins afresh at stopping times.* A *stopping time* is a non-negative function $T \leq \infty$ of the sample path, such as $T = 1$ or the exit time $T = \inf(t : x(t) > 1)$, such that for each $t \geq 0$ the event $(T < t)$ depends upon $x(s) : s \leq t$ alone. *Beginning afresh* means that if B_T is the field of events B such that $B \cap (T < t)$ is measurable over $x(s) : s \leq t$, then *conditional on the present* $a = x(T)$ *and on the event* $T < \infty$, *the future* $x(t + T) : t \geq 0$ *is independent of the past* $B_T =$ *the field of* $x(t) : t \leq T$, *and identical in law to* $[x, P_a]$:

$$P_\bullet\big[x(t+T)\epsilon\, db \mid B_T\big] = P_a\big[x(t)\epsilon\, db\big] \quad \text{if } T < \infty;$$

see G. Hunt [15] for a complete explanation of stopping times for (additive) differential processes.

$E_1[x(t)] = f(t)$ is a solution of $f(t-s)f(s) = f(t)$ $(s \leq t)$ and $0 < f \leq \infty$, so $E_1(x) = e^{\alpha t}$ for some $-\infty < \alpha \leq \infty$. $P_1[x = 1,\ t \geq 0] = 0$ or 1 because $P_1[x(s) = 1,\ s \leq t] = f(t)$ is a solution of $f(t-s)f(s) = f(t)$ $(s \leq t)$ and $0 \leq f \leq 1$, so that $f = e^{-\gamma t}$ for some $0 \leq \gamma \leq \infty$ and $f(\infty) = 0$ or 1 according as $\gamma > 0$ or not. $P_1[x = 1,\ t \geq 0] = 0$ *is assumed below.*

A non-negative function $f(x) \not\equiv \infty$ defined on $[0, \infty)$ is (β) excessive if $e^{-\beta t}E_\bullet[f(x)] \uparrow f$ as $t \downarrow 0$; in this language *the problem of the perpetual warrant is to find the smallest excessive function* $h \geq (\xi - 1)^+$ *in case* $\infty \geq \beta \geq \alpha \geq 0$. h is constructed and its simplest properties derived in a series of brief articles.

1. Define $h^0 = (\xi - 1)^+$ and $h^n = \sup_{t \geq 0} e^{-\beta t}\ E_\xi[h^{n-1}(x)]$ *for* $n \geq 1$; *then* $(\xi - 1)^+ \leq h^n \uparrow h \leq \xi$ *as* $n \uparrow \infty$.

PROOF. $h^{n-1}(\xi) \leq h^n(\xi) = \sup_{t\geq 0} e^{-\beta t}E_\xi[h^{n-1}(x)] \leq \sup_{t \geq 0} e^{-\beta t}E_\xi(x) = \sup_{t\geq 0} e^{-\beta t}e^{\alpha t}\xi = \xi$ if $h^{n-1} \leq \xi$, and the obvious induction completes the proof. □

2. h is *increasing, convex (and so continuous), and its slope is* ≤ 1.

PROOF. $h^n(\xi) = \sup_{t \geq 0} e^{-\beta t}E_t[h^{n-1}(\xi x)]$ inherits all the desired properties from h^{n-1}; now use induction and let $n \uparrow \infty$. □

3. *h is the smallest excessive function* $\geq (\xi - 1)^+$.

PROOF. $e^{-\beta t}E_\bullet[h(x)] \leq h$ is obvious from 1. Then the differential character of $x(t)$ shows that the left side decreases as t increases, and since $h \in C[0, \infty)$ (2), an application of Fatou's lemma implies $\lim_{t \downarrow 0} e^{-\beta t}E_\bullet[h(x)] \geq E_\bullet[\liminf h(x)] = h$, completing the proof that h is excessive. Also, $h \geq (\xi - 1)^+$, and if j is another such excessive function, then the obvious induction supplies us with the underestimate $j \geq h^n \uparrow h(n \uparrow \infty)$. □

4. $h = (\xi - 1)^+$ *to the right of some point* $1 < c \leq \infty$. $h > (\xi - 1)^+$ *to the left.*

PROOF. Given $s \leq t$ and $a, b > 0$,

$$P_1\big[x(t) \geq ab\big] \geq P_1\big[x(s) \geq a,\ x(t)/x(s) \geq b\big] = P_1\big[x(s) \geq a\big]P_1\big[x(t-s) \geq b\big]$$

so that $P_1[x(nt) \geq d^n] \geq P_1[x(t) \geq d]^n$, and *either* $P_1[x \leq 1] \equiv 1$ $(t \geq 0)$ violating $P_1[x = 1,\ t \geq 0] = 0$ (use $E_1(x) \geq 1$) or $P_1[x(nt) \geq d^n] > 0$ for some $t > 0$, $d > 1$, and each $n \geq 1$. But in the second case, $h(\xi) \geq e^{-\beta nt} \cdot E_\xi[(x(nt) - 1)^+] \geqq e^{-\beta nt}(\xi d^n - 1)^+ P_1[x(nt) > d^n]$ is positive for large n and either $h(\xi) > (\xi - 1)^t$ always $c = d$ or else $h(\xi) > (\xi - 1)^+$ for $\xi \leqq 1$, $h(\xi) > \xi - 1$ has a first root $1 < c < \infty$, and agrees with $\xi - 1$ to the right, by 2. □

5. $h \equiv \xi$ *if* $\beta = \alpha \geq 0$.

PROOF. $\xi \geq h \geq e^{-\beta t} E_\xi[x-1] = \xi(1-e^{-\beta t}) \uparrow \xi$ as $t \uparrow \infty$ if $\beta = \alpha > 0$, while if $\beta = \alpha = 0$, then $E_\xi[(x-1)^+] \geq (\xi - 1)^+$ from which it is easy to see that $h = \lim_{t\uparrow\infty} E_\bullet[(x-1)^+] = E_\bullet[h(x)]$. Because h is convex (2), its 1-sided slope h^+ is an increasing function,

$$h(\xi) = h(1) + (\xi - 1)h^+(1) + \int_1^\xi \big[h^+(\eta) - h^*(1)\big]d\eta,$$

and putting $\xi = x$ and taking expectations (E_1) on both sides, it follows that $h^+(\xi) = h^+(1)$ between $0 < a < 1$ and $b > 1$ if $0 < P_1[x \leq a]P_1[x \geq b]$ for some $t > 0$. But $P_1[x(nt) \geq d^n] > 0$ for some $t > 0$, $d > 1$, and each $n \geq 1$ as in 4, and using the same method, it is also possible to make $P_1[x(nt) \leq d^{-n}] > 0$ for the same $t > 0$, some (perhaps smaller) $d > 1$, and each $n \geq 1$ (use $E_1(x) = 1$). $h^+ \equiv h^+(1)$ is immediate and $h \equiv \xi$ follows from the bounds $\xi - 1 \leq h \leq \xi$ and the fact that $h(0+) = 0$. □

Warning: $\beta > \alpha \geq 0$ *until the end of the next section.*

6. *Given a closed interval* $0 < a \leq \xi \leq b < \infty$ *with exit time* $T = T_{ab} \equiv \inf(t : x < a \text{ or } x > b)$ *and exit place* $X = x(T)$, $P_\bullet[T < \infty] \equiv 1$ *and* $j = j_{ab} \equiv E_\bullet[e^{-\beta T}h(X)]$ *lies under* h.

PROOF. Adapted from E. B. Dynkin [8]; $P_\bullet[T < \infty] \equiv 1$ since in the opposite case, $0 < p(\xi) = P_\xi[a \leq x \leq b,\ t \geq 0]$ for some $a \leq \xi \leq b$, and putting $p_{ab} = \sup_{ab} p(\xi)$, the bound $p(\xi) \leq p_{ab}P_\xi[a \leq x \leq b,\ t \leq n]$ decreases to $p_{ab}p(\xi)$ as $n \uparrow \infty$, proving $p_{ab} = 1$. But $p_{ab} = \sup_{ab} P_1[a/\xi \leq x \leq b/\xi,\ t \geq 0] \leq P_1[a/b \leq x \leq b/a,\ t \geq 0]$, and this cannot be 1 without violating the estimate $P_1[x(nt) \geq d^n] > 0$ of 4. Define $G_\gamma f = E_\bullet[\int_0^\infty e^{-\gamma t} f(x)dt]$ for non-negative f and $\gamma \geq 0$. $G_\gamma h \equiv u < \infty$ if $\gamma \geq \beta$ and $G_\gamma = G_\beta[1 + (\beta - \gamma)G_\gamma]$, so that if $v = h + (\beta - \gamma)u$, then $u = G_\beta v = E_\bullet[\int_0^\infty e^{-\beta t}v(x)dt]$.

Because h is excessive and $\beta - \gamma \leq 0$, $v \geq h + (\beta - \gamma)\int_0^\infty e^{-\gamma t}dt e^{\beta t}h = 0$; it follows that

$$u \geq E_\bullet\left[\int_T^\infty e^{-\beta t}v(x)dt\right] = E_\bullet\left[e^{-\beta T}\int_0^\infty e^{-\beta t}v\big[x(t+T)\big]dt\right],$$

and since x begins afresh at the stopping time T while T itself is measurable over B_T,

$$u \geq E_\bullet\big[e^{-\beta T}G_\beta v(X)\big] = E_\bullet\big[e^{-\beta T}u(X)\big]$$

with $X = x(T)$. Now use the fact that $(\gamma - \beta)u \uparrow h$ as $\gamma \uparrow \infty$. □

7. j *is excessive.*

PROOF. Because x begins afresh at time $t \geq 0$,

$$e^{-\beta t}E_\bullet\big[j(x)\big] = E_\bullet\big[e^{-\beta T^\circ}h\big(X^\circ\big)\big] \equiv j^\circ$$

with T° defined as the next exit time from $a \leq \xi \leq b$ after time t and $X^\circ = x(T^\circ)$. Using the notation and method of proof of 6,

$$E_\bullet\big[e^{-\beta T^\circ}u\big(X^\circ\big)\big] = E_\bullet\left[\int_{T^\circ}^\infty e^{-\beta t}v(x)dt\right]$$

$$\leq E_\bullet\left[\int_T^\infty e^{-\beta t}v(x)dt\right] = E_\bullet\big[e^{-\beta T}u(X)\big],$$

and since $(\gamma-\beta)u \uparrow h$ as $\gamma \uparrow \infty$, it follows that $j^\circ \le j$. But also $T^\circ \downarrow T$ as $t \downarrow 0$ and $x(t+) = x(t)$, so Fatou's lemma implies

$$\lim_{t\downarrow 0} j^\circ \ge E_\bullet\big[\liminf e^{-\beta T^\circ} h\big(X^\circ\big)\big] = j,$$

completing the proof. □

8. *$c < \infty$ and $E_\bullet[e^{-\beta T}h(X)] \equiv h$ with $T = \min(t : x \ge c)$, $X = x(T)$, and $e^{-\beta T}h(X) \equiv 0$ if $T = \infty$, in case $P_1[t_{ab} \downarrow 0$ as $a \uparrow 1$ and $b \downarrow 1] = 1$.*

Proof. Define for the moment $T = T_{ab}$ and $X = x(T_{ab})$. Because x is differential and $h > (\xi-1)^+$ near $\xi = 1$, it is possible to choose $a < 1 < b$ so as to make $j_{ab} \ge (\xi-1)^+$. But $j_{ab} \le h$ is excessive while h is the smallest excessive function $\ge (\xi-1)^+$, so $j_{ab} \equiv h$ for this choice of $a < 1 < b$. Given 2 overlapping closed intervals $a_1 \le \xi \le b_1$ and $a_2 \le \xi \le b_2$ with $0 < a = a_1 < a_2 < b_1 < b_2 \equiv b < \infty$ and corresponding functions $j \equiv h$, it is to be proved that $j_{ab} \equiv h$ also. Consider for the proof paths starting at $a_1 \le x(0) = \xi \le b_1$ and define stopping times

$$\begin{gathered} T_1 = \textit{the exit time from } a_1b_1, \\ T_2 = T_1 \textit{ or the next exit time from } a_2b_2 \textit{ according as } T_1 = T_{ab} \textit{ or not}, \\ T_3 = T_2 \textit{ or the next exit time from } a_1b_1 \textit{ according as } T_2 = T_{ab} \textit{ or not}, \\ \textit{etc.} \end{gathered}$$

$T_1 \le T_2 \le$ etc. $\le T_n$ is constant $(= T = T_{ab})$ from some smallest $n = m$ on, and putting $X_n = x(T_n)$ for $n \le m$ and $X = x(T)$, a simple induction justifies

$$h(\xi) = E_\xi\big[e^{-\beta Tn}h\big(X_n\big)\big] = E_\xi\big[e^{-\beta T_n}h(X_n),\ n \ge m\big] + E_\xi\big[e^{-\beta Tn}h\big(X_n\big),\ n < m\big].$$

As $n \uparrow \infty$, this tends to j_{ab} since $P_\xi[m < \infty] = 1$ while $h(X_n) \le b < \infty$ on $(n < m)$. $j_{ab} \equiv h$ follows at once. Now choose $0 < a < 1 < b < c$ so that $j_{ab} = h$. Repeating the first part of the proof, it is clear that the function j associated with a small neighborhood of b is identical to h, and using the second part, it follows that $j_{ab} \equiv h$ for a little bigger b. Because a can be diminished for the same reason, it is clear that if $0 < b < c$ (or if $b = c$ in ease $c < \infty$), then it is possible to find closed intervals $0 < a_n \le \xi \le b_n \le b$ increasing to $0 < \xi < b$ with $j_n \equiv h$. But for paths starting at $0 \le x(0) = \xi \le b$ and $n \uparrow \infty$, the exit times T_n from $a_n \le \xi \le b_n$ increase to the exit time $T = \min(t : x = 0$ *or* $x \ge b)$ while $X_n = x(T_n)$ tends to $X = x(T)$ (see G. Hunt [15]), so

$$h(\xi) = \lim_{n\uparrow\infty} j_n(\xi) = \lim_{n\uparrow\infty} E_\xi\big[e^{-\beta T_n}h\big(X_n\big)\big] = E_\xi\big[e^{-\beta T}h(X)\big]$$

because of the bound

$$\begin{aligned} e^{-\beta T_n}h\big(X_n\big) &= e^{-\beta T}h(X) && \textit{if } X_n \ge b \\ &< b && \textit{if } X_n < b, \end{aligned}$$

and to complete the proof, it suffices to replace T by $T = \min(t : x \ge b)$ and to prove $c < \infty$. As to T, the replacement is obvious since $h(0) = 0$. As to the proof that $c < \infty$, if $c = \infty$, then $h = E_\bullet[e^{-\beta T_n}h(X_n)]$ with $T_n = \inf(t : x > n)$ and $X_n = x(T_n)$. Because

$$\xi - 1 \le h(\xi) \le E_\xi\big[e^{\beta T_n}x\big(T_n\big)\big] = E_1\big[e^{-\beta T_{n/\xi}}\xi x\big(T_{n/\xi}\big)\big]$$

for $n > \xi$, $1 \le E_1[e^{-\beta T_2}x(T_2)]$ as follows on putting $n/\xi = 2$ and letting $n \uparrow \infty$. Because $\beta > \alpha$, $E_1[e^{-\beta T_2}x(T_2)] < E_1[e^{-\alpha T_2}x(T_2)]$, and adapting the proof of 6 to the (α) excessive function $f \equiv \xi$, one finds $E_\xi[e^{-\alpha T_2}x(T_2)] \le \xi$. But this leads to a contradiction: $1 < E_1[e^{-\alpha T_2}x(T_2)] \le 1$. □

9. $c < \infty$ *and* $E_\bullet[e^{-\beta T}h(X)] \equiv h$ *with* $T = \min(t : x \geq c)$ *and* $X = x(T)$ *in general.*

PROOF. 8 covers the case $P_1[t_{ab} \downarrow 0$ *as* $a \uparrow 1$ *and* $b \downarrow 1] = 1$; otherwise,

$$P_1\left[\lim_{\substack{a\uparrow 1\\ b\downarrow 1}} t_{ab} > 0\right] = 1$$

according to Kolmogorov's 01 law, so the particle moves by *jumps* with exponential holding times between. Consider the modified motion $x^\circ = e^{\epsilon t}x$ with so small a positive ϵ that $\beta > \alpha^\circ \equiv \alpha + \epsilon$ and let h° and c° be the analogues of h and c. Because $e^{-\beta t}E_\bullet[h^\circ(x)] \leq e^{-\beta t}E_\bullet[h^\circ(x^\circ)] \leq h^\circ$, it is clear that $h^\circ \geq h$ and $c^\circ \geq c$. As $\epsilon \downarrow 0$, $h^\circ \downarrow j \geq h$ and c° decreases to some number $b \geq c$. Because x° satisfies the conditions of 8, $h^\circ = E_\bullet[e^{-\beta T^\circ}h^\circ(X^\circ)]$ with $T^\circ = \min(t : x^\circ \geq b)$ and $X^\circ = x^\circ(T^\circ)$. Now an unmodified path starting at $x(0) = \xi < b$ *jumps* out of $[0, b)$ landing at $X \geq b$; this means that $T^\circ = T$ and $X^\circ = e^{\epsilon T}X$ for some $\epsilon < \beta$, so $e^{-\beta T^\circ}h^\circ(X^\circ) \downarrow e^{-\beta T}j(X)$ as $\epsilon \downarrow 0$ for a class of paths with as large a probability as desired, while on the complement of this class $(T^\circ < T)$, $e^{-\beta T^\circ}h^\circ(X^\circ) \leq e^{-\beta T^\circ}e^{\epsilon T^\circ}b \leq b$. Because $h = j = \xi - 1$ $(\xi \geq b)$ and $x \geq b$, it follows that

$$j(\xi) = \lim_{\epsilon\downarrow 0} E_\xi\left[e^{-\beta T^\circ}h^\circ\big(X^\circ\big)\right] = E_\xi\left[e^{-\beta T}j(x)\right] = E_\xi\left[e^{-\beta T}h(X)\right] \leq h$$

for $\xi < b$, i.e., $j \equiv h$, and since $b \geq c$, the result follows after a moment's reflection. □

Summing up: if $\beta = \alpha \geq 0$, then $h \equiv \xi$, while if $\beta > \alpha \geq 0$, then h is convex with slope $0 \leq h^+ \leq 1$, $h > (\xi - 1)^+$ to the left of some point $1 < c < \infty$, $h \equiv \xi - 1$ to the right of c, and $h = E_\bullet[e^{-\beta T}h(X)]$ with $T = \min(t : x \geq c)$, $X = x(T)$, and the usual $e^{-\beta T}h(X) \equiv 0$ if $T = \infty$.

11.17. Two Examples

Consider the multiplicative Brownian motion with drift $x(t) = \exp[\sigma b + \delta t]$ with $\sigma > 0$, $b = b(t)$ a standard (additive) Brownian motion, and $-\infty < \delta < \infty \cdot E_1(x) = \exp[\sigma^2/2 + \delta t]$ so $\alpha = \sigma^2/2 + \delta$. Because $h = E_\bullet[e^{-\beta T}h(X)]$ with $T = \min(t : x = c)$, it follows from a formula of E. B. Dynkin [7] that if G is the generator of $[x, P_.]$:

$$Gf(\xi) = \big(\sigma^2/2\big)\xi^2 f''(\xi) + \big(\sigma^2/2 + \delta\big)\xi f'(\xi),$$

then $Gh = \beta h$ to the left of c. Now solve for $h = (c-1)(\xi/c)^\gamma$ with an adjustable γ and find $(\sigma^2/2)\gamma^2 + \delta\gamma - \beta = 0$, or, what is the same,

$$\gamma = -\delta/\sigma^2 + \sqrt{2\beta/\sigma^2 + \delta^2/\sigma^t} > 1$$

(the negative radical is excluded). Besides the above formula for h, the solution requires us to locate the corner c. Consider for this purpose G expressed in terms of the new scale $ds = \xi^{-1-2\delta/\sigma^2}\,d\xi$ and the so-called speed measure $m(d\xi) = 2\sigma^{-2}\xi^{-1+2\delta/\sigma^2}\,d\xi$: $Gf = df^+/dm$ with f^+ computed relative to the new scale.

In this language, the fact that h is excessive is expressed by $dh^+ - \beta h\,dm \leq 0$ and computing the mass that this distribution attributes to the corner c, you find the old left slope $h^-(c) = (c-1)\gamma/c$ matches the old right slope $h^+(c) = 1$, which is to say $c = \gamma/(\gamma-1)$. The reader can easily compute all desired probabilities for this Brownian model with the help of the formulas:

$$P_\xi\left[x(t) \in d\eta\right] = \big(2\pi\sigma^2 t\big)^{-1/2} e^{-(\lg\eta/\xi - \delta t)^2/2\sigma^2 t}\,d\eta/\eta,$$

$$P_\xi[T \in dt] = \big(2\pi\sigma^6 t^3\big)^{-1/2}(\xi/c)^{-\delta/\sigma^2}e^{-\delta^2 t/2\sigma^2}\lg(\xi/c)e^{-(\lg\xi/c)^2/2\sigma^2 t}\sigma^2 dt,$$

and

$$P_\xi\big[x(t)\varepsilon\, d\eta,\ t<T\big] = P_\xi\big[x(t)\varepsilon\, d\eta\big]$$
$$- (\eta/\xi)^{\delta/\sigma^2} e^{-\delta^2 t/2\sigma^2}\big(2\pi\sigma^2 t\big)^{-1/2} e^{-(\lg \xi\eta/c^2)^2/2\sigma^2 t}\, d\eta/\eta;$$

in the first formula t, ξ, $\eta > 0$, while in the second and third, $t > 0$, $0 < \xi$, $\eta < c$.

Consider as a second example, the (multiplicative) Poisson process $x(t) = \exp[p(\sigma t)]$ in which p is a standard (additive) Poisson process with jump size 1 and unit rate, i.e., $P[p(t) - p(0) = n] = t^n e^{-t}/n!$. $E_1(x) = \exp[\sigma(e-1)t]$ so $\alpha = \sigma(e-1)$. Given $c/e \le \xi < c$ with exit time $T = \inf(t : x \ne \xi)$ and exit place $X = x(T) = e\xi \ge c$,

$$h(\xi) = E_\xi\big[e^{-\beta T}h(X)\big] = \int_0^\infty \sigma e^{-\sigma t} dt e^{-\beta t} h(e\xi) = \frac{h(e\xi)}{1+\beta/\sigma},$$

esp., $h(\xi) = (e\xi - 1)(1+\beta/\sigma)^{-1}$ for $c/e \le \xi < c$, and letting $\xi \uparrow c$ and solving for c, one finds $c = (1-\alpha/\beta)^{-1}$. h itself is a broken line with corners at $e^{-n}c$ $(n \ge 0)$, *esp.* $h^+(c) = 1 > h^-(c) = (\sigma+\alpha)(\sigma+\beta)^{-1}$.

11.18. General Warrants

Now the problem is to find *the smallest excessive function* $h \ge (\xi - 1)^+$ *for the stopped space-time motion*

$$z(s) = \big[t-s,\ x(s)\big] \qquad (s \le t)$$
$$= \big[0,\ x(t)\big] \qquad (s > t),$$

i.e., the smallest function $h(t,\xi) \ge (\xi-1)^+$ such that $e^{-\beta t}E_\xi[h(t-s,\ x(s))] \uparrow h(t,\xi)$ as $s \downarrow 0$ for each $(t,\xi)\epsilon[0,\infty)\times[0,\infty)$.

1. *Define* $h^0 = (\xi-1)^+$ *and* $h^n = \sup_{s\le t} e^{-\beta s}E_\bullet[h^{n-1}(t-s,\ x(s))]$ *for* $n \ge 1$; *then* $(\xi-1)^+ \le h^n \uparrow h \le \xi$ *as* $n \uparrow \infty$.

PROOF. As before. □

2. *h is a convex function of* $\xi \ge 0$ *with slope* $0 \le h^+ \le 1$.

PROOF. As before. □

3. *h is an increasing function of* $t \ge 0$,

PROOF. h^0 is independent of $t \ge 0$, and

$$h^n\big(t_2,\xi\big) = \sup_{s\le t_2} e^{-\beta s}E_\xi\big[h^{n-1}\big(t_2 - s,\ x(s)\big)\big]$$
$$\ge \sup_{s\le t_1} e^{-\beta s}E_\xi\big[h^{n-1}\big(t_1 - s,\ x(s)\big)\big] = h^n\big(t_1,\xi\big)$$

if h^{n-1} is an increasing function of $t \ge 0$; now use induction and let $n \uparrow \infty$. □

4. *h is the smallest (space-time) excessive function* $\ge (\xi-1)^+$; *it is continuous from below as function of* $t > 0$.

PROOF. $h \ge e^{-\beta s}E_\bullet[h(t-s,\ x(s))]$ $(s \le t)$ is obvious. Now

$$\lim_{s\downarrow 0} e^{-\beta s}E_\xi\big[h\big(t-s,\ x(s)\big)\big] \ge h(t-,\xi) \ge (\xi-1)^+ \quad \text{for } t > 0,$$

and since

$$j = (\xi - 1)^+ \qquad (t = 0)$$
$$= h(t-, \xi) \qquad (t > 0)$$

is a (space-time) excessive function $\geq (\xi - 1)^+$, it is enough to prove that h is the smallest solution $\geq (\xi - 1)^+$ of $j \geq e^{-\beta s} E_\bullet[j(t - s, x(s))]$. But this is obvious. □

5.

$$h(0+, \xi) = \lim_{t \downarrow 0} h(t, \xi) = (\xi - 1)^+.$$

PROOF. $k(t, \xi) = E_\xi[(x(t) - 1)^+] \geq (\xi - 1)^+$, and since $k = E_\bullet[k(t-s,\ x(s))]$, $e^{-\beta s} E_\bullet[k(t-s, x(s))]$ increases to k as $s \downarrow 0$, proving $k \geq h$. Now as $t \downarrow 0$,

$$k(t, \xi) = e^{\alpha t}\xi - 1 + E_\xi\big[1 - x(t), x(t) < 1\big]$$

tends to $\xi - 1$ if $\xi > 1$. But $0 \leq k(0+, \xi) = \lim_{t \downarrow 0} E_1[(\xi x - 1)^+]$ is increasing, so the proof is complete. □

6. $h(\infty, \xi) = \lim_{t \uparrow \infty} h(t, \xi)$ *coincides with the perpetual warrant.*

PROOF. $h(\infty, \xi)$ is continuous (its slope falls between 0 and 1), so $e^{-\beta s} E_\xi[h(\infty, x(s))] \leq h(\infty, \xi)$ increases to $h(\infty, \xi)$ as $s \downarrow 0$ i.e., $h(\infty, \xi)$ is excessive; that it is the smallest excessive function $\geq (\xi - 1)^+$ is obvious. □

7. $h \equiv \xi - 1$ to *the right of some point* $1 < c = c(t) < \infty$ *for* $0 \leq t \leq \infty$. c *is increasing,* $c(t-) = c(t)$, *and* $c(\infty) < \infty$. $h > (\xi - 1)^+$ *between* c *and* $d = d(t) < 1$. d *is decreasing,* $d(t-) = d(t)$, *and* $d(\infty) = 0$. $h \equiv 0$ *to the left of* d. $d = e^{-t} > 0$ if $x(t) = e^t$. $d \equiv 0$ *if* $x(t)$ *is a multiplicative Brownian motion.*

PROOF. Use the information above and $c(\infty) < \infty$ (2.9). □

8. $h(t, \xi) = E_\xi[e^{-\beta T} h(t - T, X)]$ *if* T is *the (space-time) exit time from the region*

$$R : 0 < s \leq t, \qquad 0 < \xi < c(s)$$

and $X = x(T)$ *is the exit place; see Fig. 11.5 for* R *and* t.

PROOF. As before with some (mild) technical complications. □

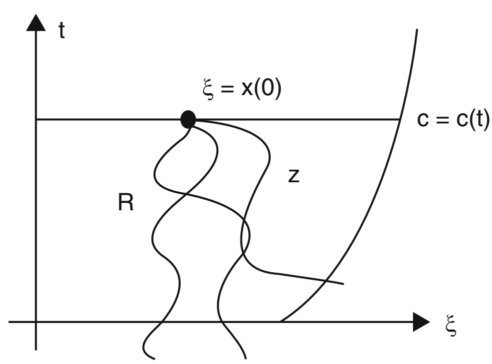

Figure 11.5

11.19. General Warrant for a Multiplicative Translation with Absorbtion

Consider the motion of translation $x(t) = \xi \exp[(\alpha+\delta)t]$ with absorbtion at a rate $\delta \ge 0$, i.e., let

$$P_1\big[x = \xi e^{(\alpha+\delta)t}\big] = e^{-\delta t} = 1 - P_1[x=0],$$

and let us prove that

$$\begin{aligned} h(t,\xi) &= e^{-(\beta+\delta)t}\big[\xi e^{(\alpha+\delta)t} - 1\big]^+ \quad && \text{when } \xi \le ce^{-(\alpha+\delta)t} \\ &= (\xi/c)^\gamma (c-1) && \text{when } ce^{-(\alpha+\delta)t} \le \xi \le c \end{aligned}$$

with $\gamma = (\beta+\delta)(\alpha+\delta)^{-1}$ and $c \equiv c(\infty) = \gamma(\gamma-1)^{-1}$.

Point 9 in Sect. 11.16 implies that the perpetual warrant is a solution of

$$e^{-(\beta+\delta)t} h\big[\xi e^{(\alpha+\delta)t}\big] = e^{-\beta t} E_\xi\big[h(x)\big] = h(\xi)$$

for $t \ge 0$ and $\xi \exp[(\alpha+\delta)t] \le c$, or, and this is the same, a solution of

$$Gh(\xi) = \xi(\alpha+\delta)h'(\xi) - \delta h(\xi) = \beta h(\xi) \quad (\xi < c).$$

Now solve and find $h(\xi) = (\xi/c)^\gamma(c-1)$ with $\gamma = (\beta+\delta)(\alpha+\delta)^{-1}$ and an unknown corner $c \ge 1$. Because $h^-(c) \le 1$, $(\gamma/c)(c-1) \le 1$, while from the fact that h is excessive, it follows that

$$e^{-(\beta+\delta)t}\big[\xi e^{(\alpha+\delta)t} - 1\big] = e^{-\beta t} E_\xi\big[h(x)\big] \le h(\xi) = \xi - 1 \quad (\xi \ge c),$$

and this cannot hold for $\xi = c$ and $t \downarrow 0$ unless $(\gamma/c)(c-1) \ge 1$, i.e., unless $c = \gamma(\gamma-1)^{-1} = (\beta+\delta)(\beta-\alpha)^{-1}$.

As to the general warrant, if $\xi \ge c(t)$, then

$$e^{-(\beta+\delta)s}\big[\xi e^{(\alpha+\delta)s} - 1\big] = e^{-\beta s} E_\xi\big[h\big(t-s, x(s)\big)\big] \le h(t,\xi) = \xi - 1,$$

and solving for $\xi = c(t)$, one finds $c(t) \ge \gamma(\gamma-1)^{-1} = c = c(\infty)$, i.e., $c(t) \equiv c(\infty)$. Now if $c\exp[-(\alpha+\delta)t] \le \xi \le c$ and if $s \le t$ is chosen so that $\xi \exp[(\alpha+\delta)s] = c$, then

$$\begin{aligned} h(\infty,\xi) \ge h(t,\xi) &\ge e^{-\beta s} E_\xi\big[h\big(t-s, x(s)\big)\big] \\ &= e^{-(\beta+\delta)s} h\big[t-s, e^{(\alpha+\delta)s}\big] = e^{-(\beta+\delta)s}(c-1) = (\xi/c)^\gamma(c-1), \end{aligned}$$

so that $h(t,\xi) = (\xi/c)^\gamma(c-1)$, while if $\xi \le c\exp[-(\alpha+\delta)t]$, then in view of 4.8,

$$\begin{aligned} h(t,\xi) &= e^{-\beta s} E_\xi\big[h\big(t-s, x(s)\big)\big] = e^{-(\beta+\delta)s} h\big(t-s, \xi e^{(\alpha+\delta)s}\big) \\ &= e^{-(\beta+\delta)t} h\big(0+, \xi e^{(\alpha+\delta)t}\big) \quad (s = t-) \\ &= e^{-(\beta+\delta)t}\big[\xi e^{(\alpha+\delta)t} - 1\big]^+, \end{aligned}$$

as stated. Note that $h^+(t,\xi)$ jumps at $\xi = \exp[-(\alpha+\delta)t]$ but not at $\xi = c$.

11.20. General Warrant for a Multiplicative Brownian Motion with Drift

Now let us compute as far as possible the general warrant for the multiplicative Brownian motion $x(t) = \exp[\sigma b + \delta t]$ of Sect. 11.17, granting that c and the left slope $h^-(t,c)$ are continuous, that $c(0+) = 1$, and that c has a continuous slope $c^\bullet$ for $t > 0$, consider

$$Gf(\xi) = \big(\sigma^2/2\big)\xi^2 f''(\xi) + \big(\sigma^2/2 + \delta\big)\xi f'(\xi) = f\phi(d\xi)/c(d\xi)$$

as in 3 and let us prove that h is a solution of the *free boundary problem:*

$$
\begin{array}{ll}
(G - \partial/\partial t)h = \beta h & \text{on the region } R: t > 0,\ 0 < \xi < c(t) \\
h(t, 0+) = 0 & (t > 0) \\
h(0+, \xi) = 0 & (\xi \leq 1) \\
h(t, c-) = c - 1 & (t > 0) \\
h^-(t, c) = 1 & (t > 0).
\end{array}
$$

R (or, what is the same, the free boundary c) is unknown, and it is the extra (flux) condition $h^- = 1$ that makes it possible to solve for *both* R and h. Point 8 in Sect. 11.18 implies the partial differential equation, and the evaluations of h on the three sides of ∂R follow from points 1, 5, and 7 in Sect. 11.17. As to the flux condition $h^-(t, c) = 1$, recall how G was expressed in Sect. 11.17: $Gf = df^+/dm$ with slope f^+ taken relative to the new scale $ds = \xi^{-\gamma}\, d\xi$ ($\gamma = 1 + 2\delta/\sigma^2$) and $dm = \xi^{\gamma-2} d\xi$. In this language $dh^t/dm - \beta h \leq \partial h/\partial t$ is the (formal) expression of the fact that h is (space-time) excessive. Note that h is still increasing and even convex in the new scale since h^+ (new) $= \xi^\gamma h^+$ (old) and $\gamma > 1$. Now integrate as follows:

$$
\begin{aligned}
&\int_{\frac{k-1}{n}}^{\frac{k}{n}} \left[h^+\big(t, c(t)\big) - h^-\big(t, c(t)\big)\right] dt \\
&\quad \leq \int_{\frac{k-1}{n}}^{\frac{k}{n}} \left[h^+\left(t, c\left(\frac{k}{n}\right)\right) - h^-\left(t, c\left(\frac{k-1}{n}\right)\right)\right] dt \\
&\quad = \int_{\frac{k-1}{n}}^{\frac{k}{n}} dt \int_{c(\frac{k-1}{n})-}^{c(\frac{k}{n})+} dh^+ \\
&\quad \leq \int_{\frac{k-1}{n}}^{\frac{k}{n}} dt \int_{c(\frac{k-1}{n})-}^{c(\frac{k}{n})+} \left[\frac{\partial h}{\partial t} + \beta h\right] dm \\
&\quad = \int_{c(\frac{k-1}{n})}^{c(\frac{k}{n})} dm \left[h\left(\frac{k}{n}, \xi\right) - h\left(\frac{k-1}{n}, \xi\right) + \beta \int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t, \xi)\, dt\right] \\
&\quad \leq \int_{c(\frac{k-1}{n})}^{c(\frac{k}{n})} dm \left[h\left(\frac{k}{n}, c\left(\frac{k}{n}\right)\right) - h\left(\frac{k-1}{n}, c\left(\frac{k}{n}\right)\right) + \beta \int_{\frac{k-1}{n}}^{\frac{k}{n}} h(t, \xi)\, dt\right] \\
&\quad = \int_{c(\frac{k-1}{n})}^{c(\frac{k}{n})} \left[c\left(\frac{k}{n}\right) - c\left(\frac{k-1}{n}\right) + O\left(\frac{1}{n}\right)\right] dm \\
&\quad = O\left(\frac{1}{n}\right)
\end{aligned}
$$

under the assumptions on $c(t)$. But in the old scale (ξ), the first integral is just

$$
\int_{\frac{k-1}{n}}^{\frac{k}{n}} c^\gamma \left[1 - h^-(t, c)\right]
$$

and the flux condition $h^-(t, c) = 1$ follows.

Now transform the free boundary problem by the substitution $h = e^{-\beta t} w(t, \sigma^{-1}[\lg \xi + \delta t])$:

$$\begin{aligned}
&\frac{\partial w}{\partial t} = \frac{1}{2}\frac{\partial^2 w}{\partial \xi^2} \text{ on the region:} && t > 0,\ -\infty < \xi < b(t) \equiv \sigma^{-1}[\lg c + \delta t] \\
&w(t, -\infty) = 0 && (t > 0) \\
&w(0+, \xi) = 0 && (\xi \le 0) \\
&w(t, b) = e^{\beta t}(c - 1) && (t > 0) \\
&w^{-}(t, b) = e^{\beta t}\delta c && (t > 0).
\end{aligned}$$

Because

$$w(t, \xi) \le e^{\beta t} h\big(\infty, c^{\sigma\xi - \delta t}\big) = e^{\beta t}\big[c(\infty) - 1\big] c(\infty)^{-\gamma} e^{\gamma[\sigma\xi - \delta t]}$$

to the left of $\xi = b$, it is legitimate to take a Fourier transform $\widehat{w}(t, \eta) = \int_{-\infty}^{b} e^{i\xi\eta} w(t, \xi) d\xi$. $c(0+) = 1$ implies $\widehat{w}(0+, \cdot) \equiv 0$; this leads at once to

$$\widehat{w}(t, \eta) = \int_0^t e^{-\eta(t-s)/2} e^{\beta s} e^{i\eta b(s)} \left[\frac{c(s)}{2} + \left(\overset{\bullet}{b}(s) - \frac{i\eta}{2}\right)\big(c(s) - 1\big)\right] ds,$$

or, what is the same,

$$\begin{aligned}
&\int_0^t \frac{e^{-[\xi + b(t) - b(s)]^2/2(t-s)}}{\sqrt{2\pi(t-s)}} e^{\beta s} \left[\frac{c}{2} + \left[\overset{\bullet}{b}(s) - \frac{\xi + b(t) - b(s)}{2(t-s)}\right](c(s) - 1)\right] ds \\
&\quad = w\big(t, \xi + b(t)\big) \quad (\xi < 0) \\
&\quad = 0 \quad (\xi > 0),
\end{aligned}$$

and from this it is possible to deduce an infinite series of integral equations for the free boundary c by (a) evaluation at $\xi = 0+$, (b) evaluation of the slope at $\xi = 0+$, etc.:

$$\text{(a)} \quad \frac{c - 1}{2} = \int_0^t \frac{e^{-[b(t) - b(s)]^2/2(t-s)}}{\sqrt{2\pi(t-s)}} e^{-\beta(t-s)} \left[\frac{c}{2} + \left[\overset{\bullet}{b}(s) - \frac{b(t) - b(s)}{2(t-s)}\right](c - 1)\right] ds,$$

$$\text{(b)} \quad \frac{c}{2} = \int_0^t \frac{e^{-[b(t) - b(s)]^2/2(t-s)}}{\sqrt{2\pi(t-s)}} e^{-\beta(t-s)} \left[\overset{\bullet}{b}(s) + \beta(c(s) - 1) - \frac{b(t) - b(s)}{2(t-s)} c(s)\right] ds,$$

etc.

I. I. Kolodner [19] treated such free boundary problems and derived (a) and (b) by a more complicated method. Unfortunately, it is not possible to obtain explicit solutions, though machine computation should be feasible; as a matter of fact, even the existence and uniqueness of solutions is still unproved.

Henry P. McKean, Jr., Ph.D., *Professor of Mathematics, Massachusetts Institute of Technology.* Author of "Diffusion Processes and Their Sample Paths" (with K. Itô), *Grundlehren der Math. Wiss.* 125, Berlin, 1965; and other papers.

Bibliography

[1] S. S. Alexander. Rice movements in speculative markets: Trends or random walks. *Industrial Management Review*, 2:7–26, 1961.

[2] S. S. Alexander. Some movements in speculative markets: trends or random walks. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 338–375. MIT Press, Cambridge, 1964.

[3] L. Bachelier. Theory of speculation. Translation of the 1900 french edition. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 17–78. MIT Press, Cambridge, MA, 1964.

[4] A. J. Boness. Some evidence on the profitability of trading in put and call options. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 475–496. MIT Press, Cambridge, 1964.

[5] P. H. Cootner. Rejoinder. *J. Political Econ.*, 68:415–418, 1960.

[6] P. H. Cootner. Returns to speculators: Telser versus Keynes. *J. Political Econ.*, 68:396–404, 1960.

[7] E. B. Dynkin. Infinitesimal operators of Markov processes (Russian). *Teor. Veroyatnost. i Primenen*, 1:38–60, 1956.

[8] E. B. Dynkin. The natural topology and excessive functions connected with a Markov process. *Dokl. Akad. Nauk SSSR*, 127:17–19, 1959.

[9] E. F. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36:420–429, 1963.

[10] E. F. Fama. Mandelbrot and the stable Paretian hypothesis. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 297–306. MIT Press, Cambridge, 1964.

[11] G. Giguère. Warrants: a mathematical method of evaluation. *Analysts Journal*, 14:17–25, 1958.

[12] B. V. Gnedenko and A. N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley Publishing Company, Inc., Cambridge, 1954.

[13] H. Houthakker. The scope and limits of future trading. In M. Abramowitz, editor, *Allocation of Economic Resources*, pages 134–159, 1959.

[14] H. Houthakker. Systematic and random elements in short-term price movements. *American Economic Review*, 51:164–172, 1961.

[15] G. Hunt. Some theorems concerning Brownian motion. *Trans. AMS*, 81:294–319, 1956.

[16] M. G. Kendall. The analysis of economic time series—Part I: prices. *Journal of the Royal Statistical Society*, 96:11–25, 1963.

[17] M. G. Kendall. The analysis of economic time series—Part I: prices. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 85–99. MIT Press, Cambridge, 1964.

[18] J. M. Keynes. *A treatise on money–II: The Applied Theory of Money.* MacMillan and Company, London, 1930.

[19] I. I. Kolodner. Free boundary problem for the heat equation with applications to problems of change of phase. I. General method of solution. *Comm. Pure Appl. Math.*, 9:1–31, 1956.

[20] R. J. Kruizenga. Introduction to the option contract. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 277–391. MIT Press, Cambridge, 1964.

[21] R. J. Kruizenga. Profit returns from purchasing puts and calls. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 392–411. MIT Press, Cambridge, 1964.

[22] P. Lévy. *Calcul des probabilitès.* Librarie du Bureau des Longitudes de l'Ecole Polytechnique, Paris, 1925.

[23] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

[24] B. Mandelbrot. The variation of certain speculative prices. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 307–332. MIT Press, Cambridge, 1964.

[25] M. F. M. Osborne. Brownian motion in the stock market. *Operations Res.*, 7:145–173, 1959.

[26] M. F. M. Osborne. Brownian motion in the stock market. In P. Cootner, editor, *The Random Character of Stock Market Prices*, pages 100–128. MIT Press, Cambridge, 1964.

[27] C. M. Sprenkle. Warrant prices as indicators of expectations and preferences. *Yale Economic Essays*, 1:178–231, 1961.

[12] Geometry of KdV (1): Addition and the Unimodular Spectral Classes

[12] Geometry of KdV (1): Addition and the Unimodular Spectral Classes. *Rev. Math. Iberoamer* **2** (1986), 235–261.

Geometry of KdV (1): Addition and the Unimodular Spectral Classes

12

Henry P. McKean Jr.[1,2]

To Alberto Calderón for his 65th birthday

12.1. Introduction

This is the first of three papers on the geometry of KDV. It presents what purports to be a foliation of an extensive function space into which all known invariant manifolds of KDV fit naturally as special leaves. The two main themes are *addition* (each leaf has its private one) and *unimodular spectral classes* (each leaf has a spectral interpretation), but first a bit of background.

Darboux's Transformation: Scattering Case. Let[3] $Q = -D^2 + q(x)$ be a Schrödinger operator on $\mathbb{R}$ with potential of scattering class[4] $C^\infty_\downarrow$. The spectrum of Q acting in $L^2(\mathbb{R})$ comprises $0 \le g < \infty$ simple *bound states* $-k_1^2 < \cdots < -k_g^2 < 0$, plus the *continuum* $[0, \infty)$ of multiplicity 2. The ground state $-k_1^2$ (if present) has an eigenfunction e_1 of one sign; it is removed by the transformation $Q \to Q - 2D^2 \lg e_1 = Q^-$, the bound states of the latter being the same as for Q with $-k_1^2$ left out. Bound states can also be added: if $-k_0^2$ is any number below $\operatorname{spec} Q$, then $Qh = -k_0^2 h$ has positive solutions $h_- \in L^2(-\infty, 0]$ with $\int_0^\infty h_-^2 = \infty$ and $h_+ \in L^2[0, \infty)$ with $\int_{-\infty}^0 h_+^2 = \infty$, and if $e_0 = (1-c)h_- + ch_+$ with $0 < c < 1$, then $Q^+ = Q - 2D^2 \lg e_0$ has bound states $-k_0^2 < -k_1^2 < \cdots < -k_g^2$. This type of transformation stems from [4]; see also [1, 3, 7]. It can be expressed in other ways: for example, if e is any positive solution of $Qe = -k^2 e$, if $AQ = Q - 2D^2 \lg e$, and if $p = e'/e$, then $Q = -D^2 + p' + p^2 - k^2$ while $AQ = -D^2 - p' + p^2 - k^2$, so that the Darboux transformation A is identified with the Bäcklund transformation B of KDV in the form discovered by Miura [20], *to wit*, $B : p' + p^2 - k^2 \to -p' + p^2 - k^2$. A variant is to express Q as[5] $(eDe^{-1})^\dagger (eDe^{-1}) - k^2$ and to exchange the factors: $(eDe^{-1})(eDe^{-1})^\dagger - k^2 = AQ$; see [5].

Addition Defined. I make two small but important changes in the Darboux transformation: *I insist that* $-k^2$ *be to the left of* $\operatorname{spec} Q$ *and I take for* e *always* h_- *or* h_+, which was

[1]Courant Institute, New York University, New York, NY, USA, mckean@cims.nyu.edu.

[2]The work presented in this paper was performed at the Courant Institute of Mathematical Sciences with the support of the National Science Foundation under Grant NSF-MCS-76-07039.

[3]D signifies differentiation with respect to $x \in \mathbb{R}$.

[4]$C^\infty_\downarrow$ is the class of infinitely differentiable real-valued functions vanishing rapidly at $\pm\infty$.

[5]The *dagger* signifies the *transposed*.

not done before. Let $\mathfrak{p} = (\lambda, -)$ or $(\lambda, +)$ with $\lambda = -k^2$ and take $e(x, \mathfrak{p}) = h_-(x)$ or $h_+(x)$ in accordance with this choice. The map

$$A^{\mathfrak{p}} : Q \longrightarrow Q - 2D^2 \lg e(x, \mathfrak{p})$$

is called *addition* of $\mathfrak{p}$ for reasons to be explained presently; unlike the previous maps of its kind, it is always *isospectral*, and a great deal more besides, as you will see. The composition rule[6]

$$A^{\mathfrak{p}_1} \cdots A^{\mathfrak{p}_n} = Q - 2D^2 \lg \big[e\big(x, \mathfrak{p}_1\big), \ldots, e\big(x, \mathfrak{p}_n\big) \big]$$

shows that addition is *commutative*, and its specialization to $n = 2$ with $\mathfrak{p}_1 = \mathfrak{p} = (\lambda, \pm)$ and $\mathfrak{p}_2 = -\mathfrak{p} = (\lambda, \mp)$ shows that $A^{-\mathfrak{p}}$ is inverse to $A^{\mathfrak{p}}$, so that repeated additions produce a commutative group of transformations. The point now to be stressed is that addition makes sense in a very wide class of operators: the only thing you really need is that $\operatorname{spec} Q$ *is bounded away* from $-\infty$! The corresponding class of (smooth) potentials is the *extensive function space* alluded to before.

Addition Explained: Hill's Case. The name *addition*[7] has a better justification than its mere commutativity. Let Q be a Hill's operator with potential of class[8] C_1^∞. Then the spectrum of Q acting in $L^2(\mathbb{R})$ consists of bands

$$\big[\lambda_0^+, \lambda_1^-\big] \cup \big[\lambda_1^+, \lambda_2^-\big] \cup \big[\lambda_2^+, \lambda_3^-\big] \cup \cdots$$

marked off by the periodic/anti-periodic spectrum of Q:

$$-\infty < \lambda_0^+ < \lambda_1^- \le \lambda_1^+ < \lambda_2^- \le \lambda_2^+ < \lambda_3^- \le \cdots \uparrow \infty.$$

Now the class of Hill's operators Q having one and the same spectrum as some fixed specimen is faithfully represented by divisors in real position on the (nonclassical) multiplier curve M determined by the irrationality

$$\sqrt{\Delta^2 - 1} = \sqrt{\lambda_0^+ - \lambda} \prod_{n=1}^{\infty} n^{-2}\pi^{-2} \sqrt{\big(\lambda_n^+ - \lambda\big)\big(\lambda_n^- - \lambda\big)}.$$

I pause to explain what all that means. The points of the curve are pairs $\mathfrak{p} = (\lambda, \pm)$ comprising a *projection* $\lambda(\mathfrak{p}) \in \mathbb{C}$ and a *signature* (of the irrationality). The function $e(x, \mathfrak{p})$ introduced before agrees, to the left of $\operatorname{spec} Q$, with the *Baker-Akhiezer function* of M specified by

(a) $Qe = \lambda(\mathfrak{p})e$;
(b) $e(x + 1) = m(\mathfrak{p})e(x)$, $m(\mathfrak{p})$ being the *multiplier* $\Delta - \sqrt{\Delta^2 - 1}$; and
(c) $e(0) = 1$.

The pole divisor $\mathfrak{p}_1 + \mathfrak{p}_2 + \cdots$ of $e(x, \mathfrak{p})$ is independent of $0 < x < 1$. It is the *divisor of* Q and is in *real position* in the sense that its projections fall one into each of the spectral lacunae: $\lambda_n^- \le \lambda(\mathfrak{p}_n) \le \lambda_n^+$ $(n \ge 1)$. The association of Q to its divisor is 1:1. The latter form an ∞-dimensional torus J which is (the real part of) the *Jacobi variety* of M. In this language, *the addition*[9] *of* $\mathfrak{p}_0$ *with* $\lambda(\mathfrak{p}_0) < \lambda_0^+$ *to the left of* $\operatorname{spec} Q$ *is effected by the following recipe*

$$\begin{aligned} Q &\longrightarrow \mathfrak{p}_1 + \mathfrak{p}_2 + \cdots \\ &\longrightarrow -\mathfrak{p}_0 + \mathfrak{p}_1 + \mathfrak{p}_2 + \cdots \\ &\longrightarrow \infty + \mathfrak{p}_1' + \mathfrak{p}_2' + \cdots \\ &\longrightarrow Q' \end{aligned}$$

[6]Crum [3]. $[e_1, \ldots, e_n]$ is Wronski's determinant.
[7]McKean and Trubowitz [15] is cited for background.
[8]C_1^∞ is the class of infinitely differentiable real-valued functions of period 1.
[9]It would be more accurate to speak of *subtraction* but never mind.

in which $-\mathfrak{p}_0$ is $\mathfrak{p}_0$ reckoned with the opposite signature; $\mathfrak{p}_1' + \mathfrak{p}_2' + \cdots$ is a new divisor in real position; the divisors of lines 2 and 3 are equal in J, meaning that the one comprises the roots and the other the poles of a function of rational character on M; and $Q' = A^{\mathfrak{p}_0}Q$ is the new Hill's operator with divisor $\mathfrak{p}_1' + \mathfrak{p}_2' + \cdots$. The proof is presented below; compare [14].

NOTE. The special addition (not permitted by the present recipe) of the point of ramification $\mathfrak{p}_0 = \lambda_0^+$ is an involution[10] of the Jacobi variety corresponding to the addition (in J) of the sum of all real half-periods; see [19].

KDV-Type Fields. The next item is the connection between addition and KDV. Take $\mathfrak{p} = (\lambda, +)$ and $\mathfrak{p}' = (\lambda + \Delta\lambda, +)$ for λ below $\operatorname{spec} Q$. Then

$$A^{\mathfrak{p}'} A^{-\mathfrak{p}} Q = Q - XQ\Delta\lambda + \textit{etc.}$$

with[11]

$$XQ = 2G'_{xx}(\lambda),$$

$G_{xy}(\lambda)$ being the Green's function $(Q-\lambda)^{-1}_{xy}$. The vector field X, familiar from Hill's equation,[12] appears in this way as an *infinitesimal addition.* Now XQ may be expanded as $\lambda \downarrow -\infty$ in diminishing half-integral powers of $-\lambda$:

$$XQ = \sum_{n=1}^{\infty} (-\lambda)^{-(1/2+n)} X_n Q,$$

in which $X_1Q = q'$, $X_2Q = 3qq' - (1/2)q'''$, *etc.* are the conventional KDV fields up to unimportant constants, the point being that, *in the full generality envisaged here, addition provides a substitute for the flows of the KDV hierarchy even when the latter have no existence.*[13] This realization prompts the formation of the *additive class* produced by closing up the operators Q obtained from a fixed specimen by repeated additions of points $\mathfrak{p}$ to the left of $\operatorname{spec} Q$. *In this way, the space of operators with spectrum not extending to $-\infty$ is foliated*[14] *by classes which fill the office of the invariant manifolds of KDV, each class having its private addition.*

NOTE. The vector fields $XQ = 2G'_{xx}(\lambda)$, taken for λ to the left of $\operatorname{spec} Q$, may be integrated without obstruction to produce commuting, class-preserving flows. This can be done by explicit formulae involving Fredholm determinants, much as in [17]; see the third paper in this series. Mumford [23, 24] has studied these flows in a special case.

Unimodular Spectral Classes. It is known that the invariant manifolds of KDV have a spectral basis: for example, in the scattering case, the transmission coefficient is *the* invariant specifying the manifold, while in the Hill's case, it is the periodic/anti-periodic spectrum, alias the discriminant, that is fixed. You will ask: *what is the corresponding invariant for the general additive class*? I present a conjecture verified in three examples cited below. Let $dF(\lambda)$ be the (2×2) *spectral weight*[15] of Q. Then Q is isospectral (= unitarily equivalent) to a second such operator Q' if and only if the spectral weight of the latter is related to the former weight as[16] $G\,dF\,G^\dagger$ with a factor $G = G(\lambda)$ taking values in $\mathrm{GL}(2, \mathbb{R})$. Now it is easy

[10] $\mathfrak{p}_0 = -\mathfrak{p}_0$.
[11] The *prime* signifies differentiation on diagonal.
[12] McKean-Trubowitz [15, 16].
[13] KDV cannot be balanced with initial data x^2, for example.
[14] The usage is informal as the dimensionality of the leaf varies from 0 to ∞!
[15] Kodaira [12] is cited for background; compare art. 2 below.
[16] The *dagger* signifies the *transposed*, as before.

to compute the spectral weight $dF^{\mathfrak{p}} = G\,dF\,G^{\dagger}$ of $A^{\mathfrak{p}}Q$: isospectrality prevails, the factor being

$$G(\lambda) = \frac{1}{\sqrt{\lambda - \lambda(\mathfrak{p})}} \begin{bmatrix} c & -1 \\ \lambda - \lambda(\mathfrak{p}) - c^2 & c \end{bmatrix} \quad \text{with } c = \frac{e'(0,\mathfrak{p})}{e(0,\mathfrak{p})},$$

and you notice that $\det G = 1$! This prompts a definition: *two operators Q belong to the same unimodular spectral class if they are unitarily equivalent and* if the factor G takes its *value not just in* $\mathrm{GL}(2,\mathbb{R})$ *but in* $\mathrm{SL}(2,\mathbb{R})$; evidently, the additive class of Q is part of its unimodular spectral class. I conjecture that, with the proper technical precautions, *these classes are always one and the same*. The second paper of this series verifies this conjecture in three examples:

(a) the scattering case $C^{\infty}_{\downarrow}$[17];
(b) the Hill's case C^{∞}_{1};
(c) if the additive class is of finite dimension.

Under (a), the class has fixed transmission coefficient, the phase of the reflection coefficient together with the logarithmic norming constants serving as additive coordinates. Under (b), the periodic/anti-periodic spectrum is fixed and the class is identified as the real part of the Jacobi variety, the addition of the latter falling in with the addition of the class. Under (c), the class is a leaf of the Neumann system and every leaf appears in this way.[18]

Second Measure Class. The two measure classes determining the conventional isospectral class of Q are typified by[19]

$$\mathrm{sp}\,dF = df_{11} + df_{22}$$

and[20]

$$\sqrt{\det dF} = \sqrt{f'_{11}f'_{22} - \left(f'_{12}\right)^2}\,\mathrm{sp}\,dF.$$

The latter is of special importance in its role of *unimodular spectral invariant*. It is always smaller than $d\lambda$ and the density $D = \sqrt{\det dF}/d\lambda$ can be interpreted as the modulus of a (*mean*) *transmission coefficient*; in fact, it is precisely $|s_{11}|$ in the scattering case. The evaluation of the gradient:

$$\frac{\partial \lg D}{\partial q(x)} = -\text{the real part of } G_{xx}(\lambda + \sqrt{-1}\,0+) \quad \text{if } D > 0$$

hints at an attractive connection between the unimodular invariant and the vector fields $XQ = 2G'_{xx}(\lambda)$, but this has not been fully understood.

Acknowledgement. I wish to thank L. Menezes: her comments led to several improvements of the original presentation.

12.2. Preliminary Spectral Theory

I collect for future use a number of standard facts about Q under the sole assumption that its spectrum is bounded from $-\infty$.

[17] $Q = -D^2$ is the simplest instance; it is settled in item 7, art. 5.
[18] McKean [13] and Moser [21] are cited for background.
[19] $dF = [df_{ij} : 1 \le i,j \le 2]$.
[20] $f'_{ij} = df_{ij}/\,\mathrm{sp}\,dF$ $(1 \le i,j \le 2)$.

Eigendifferential Expansions. [21]
Fix a complex number λ outside $\operatorname{spec} Q$ so that $Qh = \lambda h$ has independent solutions

$$h_- \in L^2(-\infty, 0] \quad \text{with} \int_0^\infty |h_-|^2 = \infty$$

and

$$h_+ \in L^2[0, \infty) \quad \text{with} \int_{-\infty}^0 |h_+|^2 = \infty.$$

The Wronskian $[h_-, h_+] = h_-' h_+ - h_- h_+'$ is always taken as 1. Then Green's function $(Q - \lambda)^{-1}_{xy} = G_{xy}(\lambda)$ is expressed as $h_-(x)h_+(y)$ if $x \le y$. Let the bottom of $\operatorname{spec} Q$ be placed at 0 for simplicity. The *fundamental matrix*

$$M(\lambda) = \left[m_{ij}(\lambda) : 1 \le i, j \le 2\right] = \begin{bmatrix} 2h_-h_+ & (h_-h_+)' \\ (h_-h_+)' & 2h_-'h_+' \end{bmatrix} \quad \text{evaluated at } x = 0$$

is analytic in the cut plane $\mathbb{C} - [0, \infty)$; it is also symmetric and of determinant -1. It is real for $\lambda < 0$; most important, its imaginary part is positive ($-$definite) in the open upper half-plane[22] and so has there the presentation

$$\operatorname{imag} M(\lambda) = \frac{b}{\pi} \int_0^\infty \left[(\lambda' - a)^2 + b^2\right]^{-1} dF(\lambda')$$

with $\lambda = a + \sqrt{-1}b$, $b > 0$, and a 2×2, real, symmetric, positive[23] *spectral weight*

$$dF(\lambda = a) = \lim_{b \downarrow 0} \operatorname{imag} M(a + \sqrt{-1}b)da = \left[df_{ij}(\lambda) : 1 \le i, j \le 2\right].$$

The spectral theorem

$$Q = \int_0^\infty \lambda \, d\mathfrak{P}(\lambda)$$

is implemented thereby: if $E(x, \lambda) = (e_1, e_2)$ with $Qe = \lambda e$, $e_1(0) = e_2'(0) = 1$, and $e_1'(0) = e_2(0) = 0$, then the kernel of the projection $d\mathfrak{P}(\lambda)$ is

$$d\mathfrak{P}_{xy}(\lambda) = (2\pi)^{-1} E(x, \lambda) dF(\lambda) E^\dagger(y, \lambda).$$

Measure Classes. The function $f_1(x)$ is taken as $h_-(x)h_+(0)$ if $x < 0$ and as $h_-(0)h_+(x)$ if $x \ge 0$ for fixed $\lambda = -1$, say, to the left of $\operatorname{spec} Q$; similarly, $f_2(x)$ is taken as $h_-(x)h_+'(0)$ if $x < 0$ and as $h_-'(0)h_+(x)$ if $x \ge 0$. Then[24]

$$\left[\left(f_i, d\mathfrak{P}(\lambda) f_j\right) : 1 \le i, j \le 2\right] = (2\pi)^{-1}(\lambda + 1)^{-2} dF(\lambda);$$

also, the families $d\mathfrak{P} f_1$ and $d\mathfrak{P} f_2$ together span the whole of $L^2(\mathbb{R})$. Now df_{12} is dominated by[25] df_{11} and/or df_{22}, so every spectral measure $(f, d\mathfrak{P} f)$ is dominated by

$$\begin{aligned} &\left(f_1, d\mathfrak{P} f_1\right) + \left(f_2, \text{coprojection of } d\mathfrak{P} f_2 \text{ upon } d\mathfrak{P} f_1\right) \\ &= \left(f_1, d\mathfrak{P} f_1\right) + \left(f_2, d\mathfrak{P} f_2\right) - r\left(f_2, d\mathfrak{P} f_1\right) \end{aligned}$$

---

[21] Weyl [27] but see [12] for the present more elegant version.
[22] This is easily seen from the identity

$$\operatorname{imag} G_{xy}(\lambda) = \operatorname{imag} \lambda \times \left[(Q - \lambda)^{-1}(Q - \lambda^*)^{-1}\right]_{xy} = \int G_{x\bullet}(\lambda) G^*_{y\bullet}(\lambda).$$

[23] The adjective means that $f_{11}(\Delta) \ge 0$, $f_{22}(\Delta) \ge 0$, and $f_{12}^2(\Delta) \le f_{11}(\Delta) f_{22}(\Delta)$ for arbitrary sets Δ.
[24] Kodaira [12]. The *parenthesis* is the inner product.
[25] $f_{11} f_{22} \ge f_{12}^2$.

with $r = df_{12}/df_{11}$, i.e., by $df_{11} + df_{22} - r\,df_{12}$. The corresponding *top measure class* is typified by the *trace* $\mathrm{sp}\,dF = df_{11} + df_{12}$; indeed, by the positivity of dF, $|df_{12}| \le \sqrt{df_{11}df_{22}}$,[26] so that $|r\,df_{12}| \le df_{22}$, and the vanishing of $f_{11}(\Delta) + f_{22}(\Delta) - \int_\Delta r\,df_{12}$ implies $f_{11}(\Delta) = 0$, $|f_{12}|(\Delta) = 0$, and $f_{22}(\Delta) = 0$, in that order. The top measure class is now seen to be based upon the family $(1-r)d\mathfrak{P}f_1 + d\mathfrak{P}f_2$: in fact, the associated spectral measure is just $df_{11} + df_{22} - r\,df_{12}$. The second class is based upon the perpendicular family

$$\frac{df_{22} + (1-r)df_{12}}{df_{11} + df_{22} - r\,df_{12}}d\mathfrak{P}f_1 - \frac{df_{11}}{df_{11} + df_{22} - r\,df_{12}}d\mathfrak{P}f_2,$$

the associated measure being

$$\frac{df_{11}df_{22} - \big(df_{12}\big)^2}{df_{11} + df_{22} - r\,df_{12}} = \frac{f'_{11}f'_{22} - \big(f'_{12}\big)^2}{f_{11} + f'_{22} - r\,df'_{12}} \times \mathrm{sp}\,dF;$$

in particular, *the second class is typified by the measure*[27]

$$\sqrt{\det dF} = \sqrt{df_{11}df_{22} - \big(df_{12}\big)^2} = \sqrt{f'_{11}f'_{22} - \big(f'_{12}\big)^2}\,\mathrm{sp}\,dF.$$

NOTE. Masani-Wiener [28, 29] introduced $\sqrt{\det dF}$ in a different context and in a different way: it is an elementary fact that the interval function $D(I) = \sqrt{\det F(I)}$ is superadditive: $D(A \cup B) \ge D(A) + D(B)$ if $A \cap B$ is void. This permits the alternative definition:

$$\int_\Delta \sqrt{\det dF} = \text{the infimum of the sum of } D(I),$$

the infimum being taken over countable covers of Δ by intervals I. This will be helpful later on.

Unitary Equivalence. The main fact about unitary equivalence can now be stated in a convenient form[28]: *two operators Q_1 and Q_2 are isospectral if and only if they determine the same top and the same bottom measure classes*, or, what is the same, *if and only if their spectral weights are related as $dF_2 = G\,dF_1\,G^\dagger$, the 2×2 factor $G = G(\lambda)$ taking its values in* $\mathrm{GL}(2,\mathbb{R})$. I could not find the second criterion stated in just this form, though it is an easy consequence of the first, which is standard.

PROOF. Let Q_1 and Q_2 be isospectral. The weight dF_1 can be expressed as $OD_1O^\dagger dm$ with $O \in \mathrm{SO}(2)$, 2×2 non-negative diagonal D_1, and a positive (numerical) measure dm. Now as dF_2 has a similar expression with the *same* dm, it suffices to produce $G = [g_{ij} : 1 \le i, j \le 2]$ taking its values in $\mathrm{GL}(2,\mathbb{R})$ so as to make $D_2 = GD_1G^\dagger$ under the condition that $\mathrm{sp}\,D_2$ vanishes simultaneously with $\mathrm{sp}\,D_1$, and $\det D_2$ with $\det D_1$. The rest will be plain. □

NOTE 1. dF is not intrinsic. The fact is, it depends upon the choice of origin $x = 0$, and if the latter is displaced to $x = c$, then the former weight dF is changed to $G\,dF\,G^\dagger$ with the factor

$$G = \begin{bmatrix} e_1(c) & e_2(c) \\ e'_1(c) & e'_2(c) \end{bmatrix}.$$

G is unimodular, so $\sqrt{\det dF}$ is not changed. This already commends it to special attention.

[26] This type of expression always signifies what is must: here, the radical is $\sqrt{f'_{11}f'_{22}} \times \mathrm{sp}\,dF$, the primed densities being taken relative to the trace.

[27] Compare the preceding footnote.

[28] Stone [26].

NOTE 2. The *unimodular spectral class* of Q_1 is the subclass of isospectral operators Q_2 having one and the same *invariant* $\sqrt{\det dF_2} = \sqrt{\det dF_1}$; equivalently, the factors G cited above take their values *not just in* $\mathrm{GL}(2,\mathbb{R})$ *but in* $\mathrm{SL}(2,\mathbb{R})$. The distinction is non-existent at bound states: if $dF_2 = G\, dF_1\, G^\dagger$ with $G \in \mathrm{GL}(2,\mathbb{R})$ at a jump of dF_1, then the simplicity of bound states of Q implies that dF_1 and dF_2 are of rank 1. Now G may be chosen from $\mathrm{SL}(2,\mathbb{R})$, as you will easily check.

Side Operators. These are employed infrequently and can be skipped for now, as can the rest of this article. The *side operator* Q_+^0 is the restriction of Q to functions on the half-line $x \geq 0$ vanishing at $x = 0$; its spectrum is confined to $[0,\infty)$, in agreement with the form of its Green's function

$$G^0_{xy}(\lambda) = e_2(x)h_+(y)/h_+(0) \quad (x \leq y)$$

and the fact that $h_+(0)$ *is root-free off the cut* $[0,\infty)$.[29] The imaginary part of $G^0_{xx}(\lambda)$ is positive in the upper half-plane, so

$$\lim_{x\downarrow 0} x^{-2}\,\mathrm{imag}\, G^0_{xx}(\lambda) = \mathrm{imag}\,\frac{h'_+(0)}{h_+(0)} = \frac{b}{\pi}\int_0^\infty \big[(\lambda'-a)^2+b^2\big]^{-1} df_+^0(\lambda'),$$

in which $\lambda = a + \sqrt{-1}b$, $b > 0$, and the spectral weight df_+^0 is non-negative. The other side operator of interest is the restriction Q_+^∞ of Q to functions on the half-line with vanishing slope at $x = 0$; its spectrum is likewise confined to $[0,\infty)$, with the possible exception of an isolated ground state to the left of 0, in agreement with the form of its Green's function

$$G^\infty_{xy}(\lambda) = -e_1(x)h_+(y)/h'_+(0) \quad (x \leq y)$$

and the fact that $h'_+(0) = 0$ *has at most one* (*real, simple, negative*) *root off the cut* $[0,\infty)$.[30] The imaginary part of $G^\infty_{00}(\lambda)$ is positive in the upper half-plane, so

$$\mathrm{imag}\, G^\infty_{00}(\lambda) = -\,\mathrm{imag}\,\frac{h_+(0)}{h'_+(0)} = \frac{b}{\pi}\int \big[(\lambda'-a)^2+b^2\big]^{-1} df_+^\infty(\lambda'),$$

in which $\lambda = a + \sqrt{-1}b$, $b > 0$, and the non-negative spectral weight df_+^∞ vanishes off $[0,\infty)$ + ground state, if any. The analogous side operators Q_-^0 and Q_-^∞ for the left half-line also play a small role below.

NOTE. The identity for $m_{22}^\bullet$ of the last footnote shows that if $h'_+(0) = 0$ has a root < 0, then $h'_-(0) = 0$ does not, and it is easy to deduce that if the ground state of Q_+^∞ is below 0, then the bottom of $\mathrm{spec}\, Q_-^\infty$ is *at* 0.

Inverse Spectral Problem. The association of Q to its 2×2 spectral weight dF is 1:1. This is made plausible by a count: 2 degrees of freedom in Q, one for $(-\infty, 0]$ and one for $[0,\infty)$, *versus* 3 degrees of freedom in the symmetric 2×2 weight, *less* 1 to account for $\det M = -1$.[31] The proof is easy, too: dF determines M and so also

$$-h'_-(0)/h_-(0) = \big(1/m_{11}\big)\big(-1-m_{12}\big), \quad h'_+(0)/h_+(0) = \big(1/m_{11}\big)\big(-1+m_{12}\big),$$

---

[29] $h_+(0) = 0$ off the cut means that h_+ is an eigenfunction of the side operator Q_+^0. Then λ must be real and negative, and $\int_0^\infty h_+ Q h_+ = \lambda \int_0^\infty h_+^2 < 0$ violates $\mathrm{spec}\, Q \subset [0,\infty)$.

[30] $h'_+(0) = 0$ off the cut means that h_+ is an eigenfunction of the side operator Q_+^∞. Then $\lambda < 0$ as before, and

$$2\big[h'_-(0)h'_+(0)\big]^\bullet = m_{22}^\bullet(\lambda) = \frac{1}{\pi}\int (\lambda'-\lambda)^{-2} df_{22}(\lambda')$$

does the rest, the *spot* signifying differentiation with regard to λ.

[31] $\det M = -1$ looks like the loss of 2 degrees of freedom but, as the real part is conjugate to the imaginary, the count is only 1.

which determine the side operators Q^0_- and Q^0_+ by the recipe of [9]. Newton [25] proves this directly in the 2×2 format, of which the only drawback is that one does not know what $\det M = -1$ signifies for dF, i.e., one cannot (to date) satisfactorily describe what are the 2×2 spectral weights.

12.3. Addition

The letter $\mathfrak{p}_0$ denotes a pair $(\lambda_0, \pm)$ comprising a *projection* λ_0 to the left of $\operatorname{spec} Q \subset [0,\infty)$ and a *signature* indicating which function $e(x,\mathfrak{p}_0) = h_-$ or h_+ is to be employed. The operation of *adding* $\mathfrak{p}_0$ *to* (*the divisor of*) Q is the map

$$A^{\mathfrak{p}_0} : Q \longrightarrow Q - 2D^2 \lg e\big(x,\mathfrak{p}_0\big) = Q^{\mathfrak{p}_0}.$$

Its simplest properties will be elicited below.

ITEM 1. $e_0(x) = e(x,\mathfrak{p}_0)$ cannot vanish so that *addition is well-defined.*

ITEM 2. *$Q = P^\dagger P + \lambda_0$ in which $P = e_0 D e_0^{-1}$ and $P^\dagger$ is its transpose $-e_0^{-1} D e_0$. $Q^{\mathfrak{p}_0}$ is produced by exchange of factors: $Q^{\mathfrak{p}_0} = PP^\dagger + \lambda_0$; in particular, $P : h \to e_0^{-1}[h, e_0]$ maps solutions of $Qh = \lambda h$ into* solutions of $Q^{\mathfrak{p}_0} h = \lambda h$. The proof is routine.

ITEM 3. *Ph_- belongs to $L^2(-\infty, 0]$ if λ is not on the cut $[0,\infty)$; similarly, Ph_+ belongs to $L^2[0,\infty)$.*

PROOF. h_+ is typical[32]:

$$\int_0^x \big|Ph_+\big|^2 = \int_0^x \big[h_+, e_0\big]^* D\big(h_+/e_0\big) = \frac{h_+}{e_0}\big[h_+, e_0\big]^*\Big|_0^x - \int_0^x \frac{h_+}{e_0} D\big[h_+, e_0\big]^*$$
$$= \frac{h_+}{e_0}\big[h_+, e_0\big]^*\Big|_0^x + \big(\lambda - \lambda_0\big)^* \int_0^x \big|h_+\big|^2 \equiv r e^{\sqrt{-1}\theta} + O(1)$$

so that *either $Ph_+ \in L^2[0,\infty)$ or else* $(h_+/e_0)[h_+, e_0]^* = re^{\sqrt{-1}\theta}$ tends to ∞ as $x \uparrow \infty$ in such a way that $r \to +\infty$ and $\theta \to 0$. In the second case,

$$\big|h_+\big|^{-2} = r^{-2}\big|Ph_+\big|^2 = r^{-2}\Big[Dre^{\sqrt{-1}\theta} + \big(\lambda - \lambda_0\big)^*\big|h_+\big|^2\Big]$$
$$= \left[\frac{r'}{r^2} + \frac{\sqrt{-1}\theta'}{r}\right] e^{\sqrt{-1}\theta} + \frac{\big(\lambda - \lambda_0\big)^*}{r^2}\big|h_+\big|^2$$
$$= \frac{r'}{r^2}\cos\theta - \frac{\theta'}{r}\sin\theta + \text{a summable function},$$

upon taking the real part. Now the imaginary part of the first formula reveals that $r\sin\theta$ is monotone and bounded so that $(r\sin\theta)'$ is summable. It follows that

$$\frac{\cos\theta}{\big|h_+\big|^2} = \frac{r'}{r^2} - \frac{\sin\theta}{r^2}(r\sin\theta)' + \text{a summable function}$$

is itself summable, so $\int_0^\infty |h_+|^{-2} < \infty$, contradicting $\int_0^\infty |h_+|^2 < \infty$. The proof is finished. □

ITEM 4. *The addition of $\mathfrak{p}_0$ is a unimodular isospectral transformation.*

[32] The *star* means complex conjugate. D stands for differentiation with regard to x.

PROOF. By Item 3, Ph_- and Ph_+ play the role of h_- and h_+ for $Q^{\mathfrak{p}_0}$, after accounting for the fact that $[Ph_-, Ph_+] = \lambda - \lambda_0$ is not unity; in particular, $\operatorname{spec} Q^{\mathfrak{p}_0}$ is confined to $[0, \infty)$, and

$$M^{\mathfrak{p}_0} = \frac{1}{\lambda - \lambda_0} \begin{bmatrix} 2Ph_-Ph_+ & \big(Ph_-Ph_+\big)' \\ \big(Ph_-Ph_+\big)' & 2\big(Ph_-\big)'\big(Ph_+\big)' \end{bmatrix} \quad \text{taken at } x = 0$$

is the fundamental matrix of $Q^{\mathfrak{p}_0}$, permitting the evaluation of its 2×2 spectral weight as

$$dF^{\mathfrak{p}_0}(\lambda = a) = \lim_{b \downarrow 0} \operatorname{imag} M^{\mathfrak{p}_0}(a + \sqrt{-1}b)da = G\, dF\, G^\dagger$$

with the unimodular factor

$$G = \frac{1}{\sqrt{\lambda - \lambda_0}} \begin{bmatrix} c & -1 \\ \lambda - \lambda_0 - c^2 & c \end{bmatrix}, \quad c = \frac{e_0'(0)}{e_0(0)}.$$

The computation is routine in view of $M^{\mathfrak{p}_0} = GMG^\dagger$. □

EXAMPLE 12.3.1. If $Q = -D^2$, then $dF(\lambda)$ is the diagonal matrix $[\lambda^{-1/2}, \lambda^{+1/2}]d\lambda$ and $G\, dF\, G^\dagger = dF$, in agreement with $D^2 \lg e_0 = 0$: in brief, *addition has no effect.*

EXAMPLE 12.3.2. The proviso that $\lambda(\mathfrak{p}_0)$ lies to the left of $\operatorname{spec} Q$ may be essential for isospectrality, as the example $Q = -D^2 + x^2 - 1$ shows. $Qe_0 = 0$ with $e_0(x) = \exp(-x^2/2) =$ the ground state, h_- and h_+ being coincident, and the addition of $\mathfrak{p}_0 = (0, \pm)$ shifts the whole spectrum 2 units: $Q \to Q - 2D^2 \lg e_0 = -D^2 + x^2 + 1$. Contrariwise, if $Q = -D^2 +$ a positive compact function, then $\operatorname{spec} Q = [0, \infty)$, $Qh = 0$ has 2 independent positive solutions $h_- = 1$ near $-\infty$ and $h_+ = 1$ near $+\infty$, and both additions $(0, \pm)$ are unimodular isospectral, as in Item 4.

ITEM 5. The effect of repeated additions is described next. Let $\mathfrak{p}_1$ and $\mathfrak{p}_2$ be distinct points. The function $e(x, \mathfrak{p}_2)$ associated to $Q^{\mathfrak{p}_1}$ is proportional to $e_1^{-1}[e_2, e_1]$ in which $e_1 = e(x, \mathfrak{p}_1)$ and $e_2 = e(x, \mathfrak{p}_2)$ are now formed for Q itself. It follows that

$$A^{\mathfrak{p}_2} A^{\mathfrak{p}_1} Q = Q - 2D^2 \lg e_1 - 2D^2 \lg e_1^{-1}\big[e_2, e_1\big] = Q - 2D^2 \lg \big[e_1, e_2\big].$$

The more general formula[33]

$$A^{\mathfrak{p}_1} \cdots A^{\mathfrak{p}_n} Q = Q - 2D^2 \lg \big[e\big(x, \mathfrak{p}_1\big), \ldots, e\big(x, \mathfrak{p}_n\big)\big]$$

is obtained by induction; in particular, addition is *commutative* and *invertible*, the inverse to $A^{\mathfrak{p}}$ being $A^{-\mathfrak{p}}$ formed with the point $-\mathfrak{p}$ having the opposite signature to $\mathfrak{p}$ but the same projection.

ITEM 6. $A^{\mathfrak{p}}$ *approximates the identity as the projection* $\lambda(\mathfrak{p})$ *tends to* $-\infty$.

PROOF. Let $Q = -D^2 + q(x)$ and take $\mathfrak{p} = (\lambda, +)$ for instance. Then $D^2 \lg e = q(x) - \lambda - (h_+'/h_+)^2$, so what is needed is the development

$$-\frac{h_+'(x)}{h_+(x)} = (-\lambda)^{1/2} + \frac{1}{2} q(x)(-\lambda)^{-1/2} + etc. \quad (\lambda \downarrow -\infty).$$

The idea is to write the Green's function G_{00}^∞ in the form of a Brownian integral:

$$G_{00}^\infty(\lambda) = -\frac{h_+(0)}{h_+'(0)} = \int_0^\infty e^{\lambda t}(\pi t)^{-1/2} E_{00}\big(e^{-\mathfrak{Q}}\big)dt,$$

in which E_{00} is the expectation for the reflecting Brownian motion $\mathfrak{x}(t) : t \geq 0$ with infinitesimal operator D^2, conditional on $\mathfrak{x}(0) = 0$ and $\mathfrak{x}(t) = 0$, and $\mathfrak{Q}$ is the integral $\int_0^t q[\mathfrak{x}(t')]dt'$. The

[33] $[e_1, \ldots, e_n]$ is Wronski's determinant.

computation is localized by checking that paths with $1 \le \max[\mathfrak{x}(t') : t' \le t]$ do not contribute to the development. I omit the routine details.[34] $\square$

ITEM 7. *Let* $\mathfrak{p}$ *be any point* $(\lambda, +)$ *to the left of* $\operatorname{spec} Q$ *and let* $\mathfrak{p}' = (\lambda + \Delta\lambda, +)$. *Then* $A^{\mathfrak{p}'} A^{-\mathfrak{p}} Q = Q - XQ\Delta\lambda + etc.$ *with* $XQ = 2G'_{xx}(\lambda)$.

PROOF. The composite addition produces $Q - 2H'$ with[35]

$$\begin{aligned} H &= D \lg \big[h_-(x,\lambda), h_+(x, \lambda + \Delta\lambda)\big] = D \lg \big(1 + \big[h_-, h_+\big]\Delta\lambda + etc.\big) \\ &= \big(h''_- h^\bullet_+ - h_- h^{\bullet\prime\prime}_+\big)\Delta\lambda = \big(h_- Q h^\bullet_+ - h^\bullet_+ Q h_-\big)\Delta\lambda = h_- h_+ \Delta\lambda, \end{aligned}$$

to leading order, by Item 5. $\square$

NOTE 1. The vector fields $X : Q \to 2G'_{xx}(\lambda)$ inherit the commutativity of addition.

NOTE 2. X appears in Item 7 as an *infinitesimal addition*. This is of particular interest in view of the fact that, as in the scattering case, $G_{xx}(\lambda)$ can be developed, as in Item 6, in diminishing half-integral powers of $\lambda \downarrow -\infty$, with the conventional KDV fields as coefficients. It follows that *addition commutes with, and shares the invariant manifolds of, the KDV flows when these have any existence; in particular, addition can be viewed as a substitute for the KDV hierarchy under the sole condition that* $\operatorname{spec} Q$ *is bounded from* $-\infty$.

NOTE 3. X can be expressed in commutator format as $XQ = [K, Q]$ in which K is the infinitesimal skew operator[36] $[(3/2)G'_{xx} - DG_{xx}]G$. The skewness of K is equivalent to the vanishing of

$$(Q - \lambda)\big(K + K^\dagger\big)(Q - \lambda) = qD + Dq - (1/2)D^3 - 2\lambda D \quad \text{acting on } G_{xx},$$

which is well known and easily checked. This facilitates the evaluation of the commutator:

$$[K, Q] = K(Q - \lambda) + (Q - \lambda)K^\dagger = \frac{3}{2}G'_{xx} - DG_{xx} + \frac{3}{2}G'_{xx} + G_{xx}D = 2G'_{xx}.$$

NOTE 4. The gradient $\partial G_{xx}(\lambda)/\partial q(y) = -G^2_{xy}(\lambda)$ is easily computed; it is symmetric in x and y so that *in the small,* $2G_{xx}(\lambda)$ *is itself a gradient* $\partial H/\partial q(x)$, by Poincaré's lemma, and the field X has the conventional KDV form $XQ = (\partial H/\partial q)'$. H is termed an *integral*. The discussion indicates but does not prove that the additive class of Q should be determined by fixing the values of these integrals for every λ to the left of $\operatorname{spec} Q$. Unfortunately, the application of Poincaré's lemma is only formal in the present very wide generality. The integrals are morally equivalent to the *additive invariant* $D = \sqrt{\det dF}/d\lambda$, or rather to its logarithm, as attested by Item 6, art. 4:

$$\partial D/\partial q(x) = \text{the real part of } G_{xx}(\bullet + \sqrt{-1}0+) \times D,$$

but this looks more satisfactory than it really is. For example, in Hill's case, $G_{xx}(\bullet + \sqrt{-1}0+)$ is imaginary on $\operatorname{spec} Q$ where D lives, and the formula is without content.

ITEM 8. Concerns an *additive duality*[37] which exchanges the spectral weights of the side operators Q^0_+ and Q^∞_+, and likewise the weights of the side operators Q^0_- and Q^∞_- for the left half-line. The weights are recalled from art. 2: df^0_+ represents $\operatorname{imag} h'_+(0)/h_+(0)$, df^∞_+ represents $-\operatorname{imag} h_+(0)/h'_+(0)$, and so on. I take[38] the ground state $\lambda_0 = \text{bottom spec}\, Q^\infty_+$

[34] McKean and van Moerbeke [18] will serve as a model.

[35] The *spot* means differentiation with respect to λ.

[36] McKean and van Moerbeke [18] misstated this without proof. G is the Green's operator $(Q-\lambda)^{-1}$. DG_{xx} is now the operator $G'_{xx} + G_{xx}D$ *not* the function G'_{xx}.

[37] McKean [14] treats Hill's case; see also [10].

[38] The spectra of Q^∞_- and Q^∞_+ cannot both extend to the left of 0; see art. 2 under *side operators*.

to the left of $0 = \text{bottom spec}\, Q_-^\infty$ to fix ideas and consider the effect of *adding the point* $\mathfrak{p}_0 = (\lambda_0, +)$. Then it is the result of a brief computation that the addition produces a *duality of spectral weights*: with $m = -h'_-(0)/h_-(0)$ evaluated at the ground state, you find

$$\begin{aligned} df_+^\infty &\longrightarrow (\lambda-\lambda_0)^{-1} df_+^0 \\ df^0 &\longrightarrow (\lambda-\lambda_0) df_+^\infty \quad \text{knocking out the weight of the ground state} \\ df_-^\infty &\longrightarrow (\lambda-\lambda_0)^{-1} df_-^0 \quad \text{plus the positive mass } m \text{ at the ground state} \\ df_-^0 &\longrightarrow (\lambda-\lambda_0) df_-^\infty. \end{aligned}$$

The ground state migrates from Q_+^∞ to Q_-^∞. A further addition of $\mathfrak{p}_0 = (\lambda_0, -)$ reproduces Q: in short, *duality is involutive.* A similar duality holds if $\text{bottom spec}\, Q_+^\infty = \text{bottom spec}\, Q_-^\infty = 0$. Then $e_1(x, 0)$ does not vanish and may be used in place of the former ground state. The computation is the same.

12.4. Unimodular Classes

The present article investigates the unimodular spectral class with special attention to the invariant $\sqrt{\det dF}$. *The fundamental matrix* M is written $A + \sqrt{-1}B$ with $A = [a_{ij} : 1 \le i, j \le 2]$ and $B = [b_{ij} : 1 \le i, j \le 2]$.

Item 1. $\sqrt{\det dF}$ is independent of the choice of origin $x = 0$, as noted in art. 2; in particular, *the unimodular class is closed under translation.*

Item 2. $\sqrt{\det B}$ *is a superharmonic function in the open upper half-plane.*

Proof. Let do be the uniform distribution on the perimeter of a circle C in the open half-plane. The interval function

$$D(I) = \sqrt{\det \int_I B\, do}$$

is defined for circular arcs I: it is superadditive, B being symmetric and positive, whence[39]

$$\sqrt{\det B(\text{center})} = D(C) \ge \lim_{N\uparrow\infty} \sum_{n=1}^{N} D\big(I_n\big) = \int_C \sqrt{\det B}\, do. \qquad \square$$

Item 3.

$$H\big(\lambda = a + \sqrt{-1}b\big) = \frac{b}{\pi} \int \big[(\lambda' - a)^2 + b^2\big]^{-1} \sqrt{\det dF(\lambda')}$$

is the greatest harmonic minorant of $\sqrt{\det B}$*; in particular,*

$$\sqrt{\det dF(a)} = \lim_{b\downarrow 0} \sqrt{\det B}\big(a + \sqrt{-1}b\big)\, da.$$

Proof. The superadditivity employed in Item 2 is valid for horizontal lines as well, whence the interval function

$$D(I) = \sqrt{\det \frac{b}{\pi} \int_I \big[(\lambda' - a)^2 + b^2\big]^{-1} dF(\lambda')}$$

[39] B is harmonic. C is divided into N equal arcs $I_n : n \le N$.

formed for fixed $a+\sqrt{-1}b$, satisfies[40]

$$\sqrt{\det B} = D(R) \geq \Sigma D\big(I_n\big) \downarrow \frac{b}{\pi}\int [(\lambda'-a)^2+b^2]^{-1}\sqrt{\det dF}(\lambda') = H,$$

so that H is overestimated by the harmonic minorant m of $\sqrt{\det B}$: $H \leq m$. But also, and for the same reason,

$$\int_I m\,da \leq \int_I \sqrt{\det B}\,da \leq \sqrt{\det \int_I B\,da} = \sqrt{\det F(I)} + o(1) \quad \text{ as } b \downarrow 0$$

for most intervals I, and as the left side in additive while the right side is superadditive, so $m \leq H$. The rest is routine. □

ITEM 4.

$$\det B = \big[h_-^*(0), h_-(0)\big]\big[h_+^*(0), h_+(0)\big] = 1 - \big|\big[h_-^*(0), h_+(0)\big]\big|^2 \leq 1;$$

in particular, $\sqrt{\det dF} \leq d\lambda$, *and the density* $D = \sqrt{\det dF}/d\lambda$ *is the limiting value* $\sqrt{\det B}(\lambda+\sqrt{-1}0+)$ *at almost every point of* $[0,\infty)$. The computation is routine.

ITEM 5. Interprets the density $D = \sqrt{\det dF}/d\lambda$ of Item 4 *as the modulus of a (mean) transmission coefficient*, as advertised in art. 1.

DISCUSSION. $D = \sqrt{\det dF}/d\lambda = |s_{11}|$ in the scattering case; see Example 12.5.1 of art. 5, below. The Jost functions f_- and f_+ and the scattering matrix $[s_{ij} : 1 \leq i,j \leq 2]$ figuring in that computation depend for their definition upon the possibility of standardizing eigenfunctions at $\pm\infty$, as in

$$f_+(x) = s_{11}\exp(\sqrt{-1}kx) + o(1) \quad \text{at } x = +\infty,$$

but this can be side-stepped, even in the most general case. The trick is to standardize the functions h_- and h_+ by $h_-(0) = h_+(0) > 0$ for $\lambda < 0$, keeping $[h_-, h_+] = 1$; then $h_-(0) = h_+(0)$ everywhere off the cut $[0,\infty)$. Now the harmonic functions m_{11}, $-1/m_{11}$, and m_{12} have finite limiting values at almost every point of $[0,\infty)$, so the same is true of $h_-(0) = h_+(0)$, $h'_-(0) + h'_+(0) = m_{12}/h_+(0)$, and $h'_-(0) - h'_+(0) = 1/h_+(0)$, and so also of $h_-(x)$ and $h_+(x)$, independently of $x \in \mathbb{R}$, $[h_-, h_+] = 1$ being maintained on the cut. The values h_- and h_+ from the upper bank of the cut and likewise the values h_-^* and h_+^* from the lower bank provide a base of solutions of $Qh = \lambda h$, so you can patch them across the cut:

$$h_+ = r_{11}h_-^* + r_{12}h_+^*, \qquad h_- = r_{21}h_-^* + r_{22}h_+^*$$

with a matrix $[r_{ij} : 1 \leq i,j \leq 2]$ reminiscent of the scattering matrix in its role of patcher of Jost functions. You read off

$$r_{11} = \big[h_+, h_+^*\big], \qquad r_{12} = \big[h_-^*, h_+\big] = r_{21}^*, \qquad r_{22} = -\big[h_-, h_-^*\big],$$

and $-r_{11}r_{22} + |r_{12}|^2 = 1$ by Item 4, so that the density

$$\sqrt{\det dF}/d\lambda = \lim_{b\downarrow 0}\sqrt{\det B} = \sqrt{-r_{11}r_{22}} = \sqrt{1 - \big|r_{12}\big|^2}$$

is seen as the modulus of a (mean) transmission coefficient.

[40] $\mathbb{R}$ is divided into small intervals I_n. The final step is by routine approximation.

Example. In the scattering case,

$$h_- = c_- f_-, \qquad h_+ = c_+ f_+,$$

and with $c_- = c$ for short, you find $-2\sqrt{-1}ks_{11} = 1/c_- c_+$ and[41]

$$\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} -1/2\sqrt{-1}k|c|^2 & s_{21}^* c^*/c \\ s_{21}c/c^* & 2\sqrt{-1}k\left|s_{11}\right|^2|c|^2 \end{bmatrix};$$

incidentally, $(4k)^{-1} \le |c|^2 \le 4|s_{11}|^2$ on $[0,\infty)$, so that $(1+k^2)^{-1}\lg|c|$ is summable and c extends to a non-vanishing Hardy function off the cut.

Item 6. There is a close but untransparent connection between $\sqrt{\det dF}$ and the vector fields $X : Q \to 2G'_{xx}(\lambda)$ of arts. 1 and 3: with $D = \sqrt{\det dF/d\lambda}$ as before,

$$\frac{\partial D(\lambda)}{\partial q(x)} = -\text{the real part of } G_{xx}(\lambda + \sqrt{-1}0+) \times D \textit{ on the cut};$$

compare Note 4, art. 3.

Proof. D is insensitive to translation so it suffices to compute at $x = 0$. Let $Q^\#$ be the operator $Q + c \times$ the unit mass at $x = 0$, with variable $-1 < c < 1$. This falls outside the class of operators permitted before but never mind. Now if the origin is taken at $x = 0+$, then with $H = 1 + ch_-(0)h_+(0)$,

$$h_-^\#(0) = H^{-1}h_-(0), \qquad h_+^\#(0) = h_+(0),$$

$$h_-^{\#\prime}(0) = H^{-1}[h'_-(0) + ch_-(0)], \qquad h_+^{\#\prime}(0) = h'_+(0),$$

while if it is taken at $x = 0-$, then

$$h_-^\#(0) = h_-(0), \qquad h_+^\#(0) = H^{-1}h_+(0),$$

$$h_-^{\#\prime}(0) = h'_-(0), \qquad h_+^{\#\prime}(0) = H^{-1}\left[h'_+(0) - ch_+(0)\right],$$

and it turns out (as it must) that the 2 determinations of $\det B^\#$ are the same: $\det B^\# = H^{-2}\det B$. It follows that

$$D^\# = \left|1 + (c/2)m_{11}\right|^{-1} D$$

with $m_{11} = m_{11}(\bullet + \sqrt{-1}0+)$ and[42]

$$\frac{\partial D}{\partial q(0)} = \frac{\partial D^\#}{\partial c} \quad \text{evaluated at } c = 0$$

$$= \frac{\partial}{\partial c}\left[\left(1 + \frac{c}{2}a_{11}\right)^2 + \frac{c^2}{4}b_{11}^2\right]^{-1/2} D = (a_{11}/2)D,$$

as promised. The computation is a bit formal but reliable, so I leave it at that. □

Item 7 is a test case. I proposed in art. 1 that the unimodular spectral class of Q and its additive class must be one and the same thing. I prove it now for $Q^0 = -D^2$ whose additive class is the singleton Q^0 itself.

[41] $\lambda = k^2 \ge 0$.
[42] $m_{11} = a_{11} + \sqrt{-1}b_{11}$.

PROOF. The spectral weight of Q^0 is $dF^0 = \operatorname{diag}[\lambda^{-1/2}, \lambda^{+1/2}]d\lambda$. Let dF be the spectral weight of an operator Q from the same unimodular class as Q^0 so that $\operatorname{sp} dF$ belongs to the Lebesgue class on $[0,\infty)$ and $D = \sqrt{\det dF}/d\lambda$ is the indicator thereof. Then[43]

$$p * D = \frac{b}{\pi}\int_0^\infty \left[(\lambda' - a)^2 + b^2\right]^{-1} d\lambda', \quad \text{taken at } \lambda = a + \sqrt{-1}b,$$

is the harmonic minorant of $\sqrt{\det B}$, and for any interval $I \subset [0,\infty)$,

$$\begin{aligned}
&\int_I \left|\left[h_-, h_+^*\right] \text{ taken at } x = 0 \text{ and } a + \sqrt{-1}b\right|^2 da \\
&\quad \le 2\int_I \left(1 - \sqrt{\det B}\right)^{[44]} \le 2\int_I (1 - p * D) \\
&\quad = \frac{2b}{\pi}\int_I da \int_{-\infty}^0 \left[(\lambda' - a)^2 + b^2\right]^{-1} d\lambda' \\
&\quad = \frac{2b}{\pi}\int_I da \int_a^\infty \left(c^2 + b^2\right)^{-1} dc \\
&\quad \le \frac{2b}{\pi}\int_I a^{-1} da,
\end{aligned}$$

whence

$$\begin{aligned}
&\left|\operatorname{imag}\int_I \left[\frac{h_-'}{h_-} - \frac{h_+^{*\prime}}{h_+^*} \text{ taken at } x = 0 \text{ and } a + \sqrt{-1}b\right] da\right|^2 \\
&\le \int_I \left|\left[h_-, h_+^*\right]\right|^2 da \int_I \left|h_- h_+\right|^{-2} da \\
&\le \frac{2}{\pi}\int_I \frac{da}{a}\int_I \frac{4b_{11}}{\left|m_{11}\right|^2}\frac{b}{b_{11}}\, da \\
&= \frac{2}{\pi}\int_I \frac{da}{a} 4\int_I \operatorname{imag}\left(-1/m_{11}\right) \times \left[\frac{1}{\pi}\int_0^\infty \left[(c-a)^2 + b^2\right]^{-1} df_{11}(c)\right]^{-1}.
\end{aligned}$$

Now introduce the representing measures df_-^0 of $-\operatorname{imag} h_-'(0)/h_-(0)$, df_+^0 of $\operatorname{imag} h_+'(0)/h_+(0)$, and df_{00} of $-1/m_{11}$, and make $b \downarrow 0$ in the preceding display to obtain (a) the bound

$$\left|f_+^0(I) - f_-^0(I)\right|^2 \le \frac{2}{\pi}\int_I \frac{da}{a} 4\int_I \left[\frac{1}{\pi}\int_0^\infty (c-a)^{-2} df_{11}(c)\right]^{-1} df_{00}(a),$$

(b) an estimate $|df_+^0 - df_-^0| \le$ a multiple of $a^{-1/2}\sqrt{da\, df_{00}}$ in the small, and (c) the conclusion that df_-^0 and df_+^0 have the same singular part, except perhaps for different jumps at 0. Now df_-^0 and df_+^0 have the same non-singular parts as well: indeed, at almost every point of the cut $[0,\infty)$, the densities $f_-' = df_-^0/da$ and $f_+' = df_+^0/da$ satisfy

$$f_+' - f_-' = \lim_{b\downarrow 0} \operatorname{imag}\left[\frac{h_-'}{h_-} - \frac{h^{*\prime}}{h_+^*}\right] = \lim_{b\downarrow 0} \frac{[h_-, h_+^*]}{h_- h_+^*} = 0$$

[43] p stands for $(b/\pi)[(\lambda' - a)^2 - b^2]^{-1}$. The *star* signifies *convolution*.

[44] $1 - r^2 = (1-r)(1+r) \le 2(1-r)$ if $0 \le r \le 1$.

in view of the fact that $m_{11} = 2h_-h_+$ has a non-vanishing limiting value and the estimate at the start of the proof:

$$\int_I \big|[h_-, h_+^*]\big|^2 \frac{da}{a} \le \frac{2b}{\pi} \int_I \frac{da}{a} = o(1) \quad \text{as } b \downarrow 0$$

if I does not extend to 0. Note, finally, that neither df_-^0 nor df_+^0 can jump at 0; otherwise, the representing measure df_{00} of $(1/2)\,\mathrm{imag}[h_+'/h_+ - h_-'/h_-] = \mathrm{imag}(-1/m_{11})$ has a jump (at 0) so that

$$-1/m_{11} \ge \text{a constant} + (-\lambda)^{-1} f_{00}[0] \uparrow \infty \quad \text{as } \lambda \uparrow 0,$$

contradicting the fact that m_{11} is positive there. The upshot is that $df_-^0 = df_+^0$, which is to say that Q *is symmetrical about* $x = 0$. But the unimodular class is invariant under translation, so the same is true for any choice of origin, whence $Q = -D^2 +$ a constant function, and the constant vanishes because $\mathrm{spec}\, Q$ starts at 0. The proof is finished. □

12.5. Examples

EXAMPLE 12.5.1. Scattering case.[45] $[s_{ij}(k) : 1 \le i, j \le 2]$ is the scattering matrix defined for real $k = \sqrt{\lambda}$, with values in $U(2)$; it is related to the behaviour at infinity of the Jost functions f_- and f_+, as in the table. For $k \ge 0$, f_- and f_+ are the values of h_- and h_+ along the upper bank of the cut $[0, \infty)$, except that $[f_-, f_+] = -2\sqrt{-1}ks_{11}(k)$ is not unity.

	$x \downarrow -\infty$	$x \uparrow +\infty$
f_-	$s_{22}e^{-\sqrt{-1}kx}$	$e^{-\sqrt{-1}kx} + s_{21}e^{\sqrt{-1}kx}$
f_+	$e^{\sqrt{-1}kx} + s_{12}e^{-\sqrt{-1}kx}$	$s_{11}e^{\sqrt{-1}kx}$

The 2×2 spectral weight is comprised of jumps at the bound states and a pure Lebesgue part on $[0, \infty)$ with density

$$\frac{dF}{d\lambda} = \text{the real part of } \frac{1}{ks_{11}} \begin{bmatrix} 2f_-f_+ & (f_-f_+)' \\ (f_-f_+)' & 2f_-'f_+' \end{bmatrix} \quad \text{taken at } x = 0.$$

The additive invariant (second measure class) $\sqrt{\det dF}$ is readily computed: it is unchanged by translation, so the entries of the table may be used in place of f_- and f_+, with the outcome

$$\begin{aligned} \frac{\sqrt{\det dF}}{d\lambda} &= \sqrt{\det} \begin{bmatrix} k\big(1 + \mathrm{real}\, s_{21}e^{2\sqrt{-1}kx}\big) & -\,\mathrm{imag}\, s_{21}e^{2\sqrt{-1}kx} \\ -\,\mathrm{imag}\, s_{21}e^{2\sqrt{-1}kx} & k^{-1}\big(1 - \mathrm{real}\, s_{21}e^{2\sqrt{-1}kx}\big) \end{bmatrix} \\ &= \sqrt{1 - |s_{21}|^2} \\ &= |s_{11}|. \end{aligned}$$

The KDV *invariant manifold* is determined by fixing the transmission coefficient s_{11}, which is to say *by fixing the bound states* $-k_1^2 < \cdots \le -k_g^2$ *and the modulus* $|s_{12}| = \sqrt{1 - |s_{11}|^2}$ because s_{11} encodes just this information:

$$s_{11}(k) = \exp\left[\frac{1}{\pi\sqrt{-1}} \int_{-\infty}^{\infty} \frac{\lg|s_{11}(k')|}{k' - k} dk'\right] \prod_{i=1}^{g} \frac{k + \sqrt{-1}k_i}{k - \sqrt{-1}k_i};$$

[45] Faddeev [8] and/or Deift and Trubowitz [6] are cited for background.

also, it is known that the phase s_{12} together with the logarithms of the norming constants

$$c_i^2 = \int_{-\infty}^{\infty} \big|f_+/s_{11}\big|^2 \quad \text{taken at } k = \sqrt{-1}k_i \ (i \le g)$$

serve as (additive) coordinates on the manifold. Now to elucidate the effect of addition, take $-k_0^2$ to the left of $\operatorname{spec} Q$ and let $\mathfrak{p}_0 = (-k_0^2, +)$ to fix ideas. Then $e_0(x) = e(x, \mathfrak{p}_0) = f_+(x, \sqrt{-1}k_0) = s_{11}(\sqrt{-1}k_0)\exp(-k_0 x)$ at $+\infty$ with a similar exponential behavior at ∞: in fact, $[\lg e_0(x)]''$ vanishes rapidly at $\pm\infty$, so *adding* $\mathfrak{p}_0$ *not only keeps you in the scattering case but also preserves the KDV manifold* in view of note 2, item 7 of art. 3. The effect of addition upon the coordinates is found by comparison of new eigenfunctions $e_0^{-1}[e_0, f_+]$ to old:

$$e_0^{-1}\big[e_0, f_+\big] \simeq e^{k_0 x}\big[e^{-k_0 x}, s_{11}e^{\sqrt{-1}kx}\big] = -\big(k_0 + \sqrt{-1}k\big)s_{11}e^{\sqrt{-1}kx}$$

at $+\infty$; similarly,

$$e_0^{-1}\big[e_0, f_+\big] \simeq \big(-k_0 + \sqrt{-1}k\big)s_{12}e^{-\sqrt{-1}kx} - \big(k_0 + \sqrt{-1}k\big)e^{\sqrt{-1}kx}$$

at $-\infty$, so

$$s_{12} \longrightarrow \frac{k_0 - \sqrt{-1}k}{k_0 + \sqrt{-1}k} = s_{12}.$$

A similar rule applies to the bound states, the ith norming constant being multiplied by $(k_0 + k_i)(k_0 - k_i)^{-1}$ $(i \le g)$; in particular, *repeated additions followed by closure produce the whole invariant manifold*, as you will easily check.

Example 12.5.2. Hill's case. Now h_- and/or h_+ is proportional to the so-called *Baker-Akhiezer function*

$$e(x, \mathfrak{p}) = e_1(x, \lambda) + e_2^{-1}(1, \lambda)\big[m(\mathfrak{p}) - e_1(1, \lambda)\big]e_2(x, \lambda)$$

taken at the point $\mathfrak{p} = (\lambda, \pm)$ of the *multiplier curve* of Q. The latter is the Riemann surface of the 2-valued *multiplier*[46] $m(\mathfrak{p}) = \Delta(\lambda) - \sqrt{\Delta^2(\lambda) - 1}$; the former is the solution of $Qe = \lambda(\mathfrak{p})e$ with $e(x+1) = m(\mathfrak{p})e(x)$ and $e(0) = 1$. The point is that, for λ to the left of $\operatorname{spec} Q$, the number Δ exceeds $+1$, so that the multipliers satisfy $0 < m_+ < 1$ and $1 < m_- < \infty$, with the result that $e(x, \mathfrak{p})$ belongs to $L^2[0, \infty)$ if $m(\mathfrak{p}) = m_+$ and to $L^2(-\infty, 0]$ if $m(\mathfrak{p}) = m_-$. Now the KDV invariant manifold may be described *either* as the class of Hill's operators with fixed periodic/anti-periodic spectrum *or* as (the real part of) the Jacobi variety of the multiplier curve; moreover, $[\lg e(x, \mathfrak{p})]''$ is of period 1, *so addition not only keeps you in the Hill's case but also preserves the KDV invariant manifold*, just as in Example 1. Kodaira [12] computed the 2×2 spectral weight: it is pure Lebesgue with density

$$\frac{dF}{d\lambda} = \frac{\pm 1/2}{\sqrt{1 - \Delta^2}}\begin{bmatrix} 2e_2(1, \lambda) & e_2'(1, \lambda) - e_1(1, \lambda) \\ e_2'(1, \lambda) - e_1(1, \lambda) & -2e_1'(1, \lambda) \end{bmatrix}$$

on the bands of $\operatorname{spec} Q$; the signatures alternate starting with $+1$. The additive invariant $\sqrt{\det dF}$ is easily elicited:

$$\frac{\sqrt{\det dF}}{d\lambda} = \frac{-4e_2e_1 - \big(e_2' - e_1\big)^2}{4\big(1 - \Delta^2\big)} = 1 \quad \text{on } \operatorname{spec} Q;$$

it determines $\operatorname{spec} Q$ and so also the discriminant Δ, the multiplier curve M, and its Jacobian J. The use of the word *addition* can now be fully justified; it will be a by-product of the discussion that *repeated additions* (*followed by closure*) *are transitive on the invariant manifold*, just as in the scattering case.

[46] Δ is the discriminant $(1/2)[e_1(1, \lambda) + e_2'(1, \lambda)]$.

Discussion. Let $\mathfrak{p}_1 + \mathfrak{p}_2 + \cdots$ be the divisor of Q introduced in art. 1, the points being the poles of $e(x, \mathfrak{p})$ determined by the vanishing of $e_2(1, \lambda)$ and the choice of multiplier $m(\mathfrak{p}) = e_2'(1, \lambda)$ instead of the other possibility $e_1(1, \lambda)$, one such to each nontrivial gap $[\lambda_n^-, \lambda_n^+]$ $(n \geq 1)$. Now the operator $Q - 2D^2 \lg e(x, \mathfrak{p}_0)$ produced by addition of a point $\mathfrak{p}_0$ with projection to the left of $\operatorname{spec} Q$ has also such a divisor $\mathfrak{p}_1' + \mathfrak{p}_2' + \cdots$, and *the recipe of addition states that the old (umprimed) divisor with*[47] $-\mathfrak{p}_0$ *adjoined is linearly equivalent to the new (primed) divisor with* ∞ *adjoined*:

$$-\mathfrak{p}_0 + \mathfrak{p}_1 + \mathfrak{p}_2 + \cdots = \infty + \mathfrak{p}_1' + \mathfrak{p}_2' + \cdots \quad \text{in } J.$$

The proof is divided into several steps.

Step 5. The points $\mathfrak{p}_1$, $\mathfrak{p}_2$, *etc.* are the poles of the function

$$f(\mathfrak{p}) = e_1^{-1}(1, \lambda)\big[m(\mathfrak{p}) - e_1(1, \lambda)\big],$$

aside from an extra pole at ∞ which is detected by the development of $f(\mathfrak{p})$ at $\lambda(\mathfrak{p}) = -\infty$:

$$f(\mathfrak{p}) = \frac{e^k - \operatorname{ch} k}{\operatorname{sh} k/k}\big[1 + o(1)\big] = k\big[1 + o(1)\big]$$

if $k = +\sqrt{-\lambda}$.

Step 6. The preceding development of $f(\mathfrak{p})$ is applied to the operator $Q^{\mathfrak{p}_0}$. The corresponding function is[48]

$$f_0(\mathfrak{p}) = \frac{m(\mathfrak{p}) - e_1^{\mathfrak{p}_0}(1, \lambda)}{e_2^{\mathfrak{p}_0}(1, \lambda)} = \frac{m(\mathfrak{p}) - e_0^{-1}\big[e_2, e_0\big] \text{ taken at } x = 1}{e_2^{\mathfrak{p}_0}(1, \lambda)} - c;$$

its poles are $\mathfrak{p}_1'$, $\mathfrak{p}_2'$, *etc.* and ∞, by definition.

Step 7. $f_0(\mathfrak{p})$ takes the value $-c$ at the points $\mathfrak{p} = \mathfrak{p}_1, \mathfrak{p}_2$, *etc.* of the divisor of Q and also at $-\mathfrak{p}_0$ in view of

$$e_0^{-1}\big[e_2, e_0\big] \quad \text{at } x = 1 = e_2'(1) - ce_2(1) = m(\mathfrak{p})$$

in the first case and

$$\frac{\big[e_2, e_0\big] \text{ at } x = 1}{e_0(1)} = \frac{\big[e_2, e_0\big] \text{ at } x = 0}{e_0(1)} = \frac{1}{e_0(1)} = \frac{1}{m(\mathfrak{p}_0)} = m(-\mathfrak{p}_0)$$

in the second.

Step 4 is to notice that $f_0(\mathfrak{p}) = -c$ has no other roots. This follows from the evaluation $(2\pi\sqrt{-1})^{-1} \int d \lg f_0(\mathfrak{p}) = 1$ for small circles about $\mathfrak{p} = \infty$; compare step 1. The upshot is that $-\mathfrak{p}_0 + \mathfrak{p}_1 + \mathfrak{p}_2 +$ *etc.* and $\infty + \mathfrak{p}_1' + \mathfrak{p}_2' +$ *etc.* are the roots and poles of the function $f_0(\mathfrak{p}) + c$, which is what linear equivalence is all about, anyhow.

Example 12.5.3. Bohr's case.[49] I cannot do so much if Q is only almost periodic in the sense of [2]. Fix λ off $\operatorname{spec} Q$. Then $G = G_{xx}(\lambda)$ is an almost periodic function with the same frequency module as Q, and

$$h_{\mp} = \sqrt{G} \exp\left[\pm (1/2) \int_0^x G^{-1} dx'\right],$$

[47] $-\mathfrak{p}_0$ is the point on the sheet opposite to that of $\mathfrak{p}_0$.
[48] $e_1^{\mathfrak{p}_0} = ce_2^{\mathfrak{p}_0} + e_0^{-1}[e_2, e_0]$ with $c = e_0'/e_0$ taken at $x = 0$.
[49] Johnson and Moser [11] and Moser and Pöschel [22] are cited for background.

by elementary computation[50] from which it follows that *addition preserves*:

(1) *the almost periodicity*,
(2) *the frequency module*, and
(3) *the rotation number*

$$r(\lambda) = \lim_{x\uparrow\infty} x^{-1} \operatorname{imag} \lg h_+(x,\lambda)$$
$$= \text{the mean value of imag } -1/2G \text{ off } \operatorname{spec} Q.$$

Proof of (3). Addition of $\mathfrak{p}_0$ changes h_+ into $e_0^{-1}[h_+, e_0]$ so that $\operatorname{imag}\lg h_+$ is changed by the addition of

$$\operatorname{imag}\lg\left[\frac{h'_+}{h_+} - \frac{e'_0}{e_0}\right],$$

and this is bounded off $\operatorname{spec} Q$ because $\operatorname{imag} h'_+/h_+$ is of one signature while e'_0/e_0 is real.

The values of $r(\bullet + \sqrt{-1}0+)$ on the line are known to determine $\operatorname{spec} Q$ and the frequency module,[51] whence it is natural to conjecture that r *and the bound states* (*if any*) *determine the additive class*, but this may be naive. □

Bibliography

[1] V. Bargmann. On the connection between phase shifts and scattering potential. *Rev. Modern Phys.*, 21:488–493, 1949.

[2] H. Bohr. *Fastperiodische Funktionen*. Springer-Verlag, Berlin, 1932.

[3] M. M. Crum. Associated Sturm-Liouville systems. *Quart. J. Math. Oxford Ser. (2)*, 6:121–127, 1955.

[4] G. Darboux. Sur une proposition relative aux équations linéaires. *C. R. Acad. Sci. Paris*, 94:1456–1459, 1882.

[5] P. A. Deift. Applications of a commutation formula. *Duke Math. J.*, 45(2):267–310, 1978.

[6] P. A. Deift and E. Trubowitz. Inverse scattering on the line. *Comm. Pure Appl. Math.*, 32(2):121–251, 1979.

[7] L. D. Faddeev. Properties of the S-matrix of the one-dimensional Schrödinger equation. *Trudy Mat. Inst. Steklov.*, 73:314–336, 1964.

[8] L. D. Faddeev. Properties of the S-matrix of the one-dimensional Schrödinger equation. *Amer. Math. Soc. Transl. (2)*, 65:139–166, 1967.

[9] I. M. Gelfand and B. M. Levitan. On the determination of a differential equation from its spectral function. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 15:309–360, 1951. Translation: American Mathematical Translation, 1 (1955), 253–304.

[10] E. L. Isaacson, H. P. McKean, and E. Trubowitz. The inverse Sturm-Liouville problem. II. *Comm. Pure Appl. Math.*, 37(1):1–11, 1984.

[11] R. Johnson and J. Moser. The rotation number for almost periodic potentials. *Comm. Math. Phys.*, 84(3):403–438, 1982.

[12] K. Kodaira. The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices. *Amer. J. Math.*, 71:921–945, 1949.

[13] H. P. McKean. Integrable systems and algebraic curves. In *Global Analysis (Proc. Biennial Sem. Canad. Math. Congr., Univ. Calgary, Calgary, Alta., 1978)*, volume 755 of *Lecture Notes in Math.*, pages 83–200. Springer, Berlin, 1979.

[14] H. P. McKean. Variations on a theme of Jacobi. *Comm. Pure Appl. Math.*, 38:669–678, 1985.

[50] $KG = 2\lambda G'$ with $K = qD + Dq - (1/2)D^3$, whence $(G')^2 - 2G''G + 4(q-\lambda)G^2 = 1$.

[51] Johnson-Moser [11].

[15] H. P. McKean and E. Trubowitz. Hill's operator and hyperelliptic function theory in the presence of infinitely many branch points. *Comm. Pure Appl. Math.*, 29:143–226, 1976.

[16] H. P. McKean and E. Trubowitz. Hill surfaces and their theta functions. *Bull. Amer. Math. Soc.*, 84:1042–1085, 1978.

[17] H. P. McKean and E. Trubowitz. The spectral class of the quantum-mechanical harmonic oscillator. *Comm. Math. Phys.*, 82(4):471–495, 1981/82.

[18] H. P. McKean and P. van Moerbeke. The spectrum of Hill's equation. *Inv. Math.*, 30:217–274, 1975.

[19] H.P. McKean. Units of Hill curves. In *Geometry and Analysis*, pages 81–94. Indian Acad. Sci., Bangalore, 1980. Papers dedicated to the memory of V. K. Patodi.

[20] R. M. Miura. Korteweg-de Vries equation and generalizations. I. A remarkable explicit nonlinear transformation. *J. Mathematical Phys.*, 9:1202–1204, 1968.

[21] J. Moser. Various aspects of integrable Hamiltonian systems. In *Dynamical Systems (C.I.M.E. Summer School, Bressanone, 1978)*, volume 8 of *Progr. Math.*, pages 233–289. Birkhäuser Boston, Massachusetts, 1980.

[22] J. Moser and J. Pöschel. An extension of a result by Dinaburg and Sinaĭ on quasiperiodic potentials. *Comment. Math. Helv.*, 59(1):39–85, 1984.

[23] D. Mumford. *Tata Lectures on Theta, I*, volume 28 of *Progr. Math.* Birkhäuser, Boston, 1983.

[24] D. Mumford. *Tata Lectures on Theta, II*, volume 43 of *Progr. Math.* Birkhäuser, Boston, 1984.

[25] R. G. Newton. The Marchenko and Gel'fand-Levitan methods in the inverse scattering problem in one and three dimensions. In *Conference on Inverse Scattering: Theory and Application (Tulsa, Okla., 1983)*, pages 1–74. SIAM, Pennsylvania, 1983.

[26] M. H. Stone. *Linear Transformations in Hilbert Space*, volume 15 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Rhode Island, 1932.

[27] H. Weyl. Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen. *Math. Ann.*, 68(2):220–269, 1910.

[28] N. Wiener and P. Masani. The prediction theory of multivariate stochastic processes. I. The regularity condition. *Acta Math.*, 98:111–150, 1957.

[29] N. Wiener and P. Masani. The prediction theory of multivariate stochastic processes. II. The linear predictor. *Acta Math.*, 99:93–137, 1958.

[13] Weighted Trigonometrical Approximation on $\mathbb{R}^1$ with Application to the Germ Field of a Stationary Gaussian Noise

[13] Weighted Trigonometrical Approximation on $\mathbb{R}^1$ with Application to the Germ Field of a Stationary Gaussian Noise. *Acta Math.* **112** (1964), 99–143.

13 Weighted Trigonometrical Approximation on $\mathbb{R}^1$ with Application to the Germ Field of a Stationary Gaussian Noise

N. Levinson and H. P. McKean, Jr.[1,2]

Notation

$f^*(\gamma)$ ($\gamma = a+ib$) denotes the regular extension of $f^*(a) = f(a)^*$ so that $f^*(\gamma) = (f(\gamma^*))^*$, ($\gamma^* = a - ib$).

$\int$ stands for $\int_{-\infty}^{+\infty}$.

$\int_1$ and the like stand for $\int_1^{\infty}$, etc.

$e(\gamma)$ means e^{γ}.

13.1. Introduction

13.1.1. Weighted Trigonometric Approximation. Given a non-trivial, even, non-negative, Lebesgue-measurable weight function $\Delta = \Delta(a)$ with $\int \Delta < \infty$, let Z be the (real) Hilbert space $L^2(R^1, \Delta\, da)$ of Lebesgue-measurable functions f with

$$f^*(-a) = f(a), \qquad \|f\| = \|f\|_\Delta = \left(\int |f|^2\Delta\right)^{\frac{1}{2}} < \infty$$

subject to the usual identifications, and putting Z^{cd} = *the span* (*in* Z) *of* $e(iat)$ ($c \le t \le d$), introduce the following subspaces of Z:

(a) $Z^- = Z^{-\infty 0}$,
(b) $Z^+ = Z^{0\infty}$,
(c) $Z^{+/-}$ = the projection of Z^+ onto Z^-,
(d) $Z^\bullet$ = the class of entire functions $f = f(\gamma)$ ($\gamma = a + ib$) with

$$\overline{\lim_{R\uparrow\infty}}\, R^{-1} \max_{0\le\theta\le 2\pi} \lg\left|f\left(Re^{i\theta}\right)\right| \le 0,$$

which, restricted to the line $b = 0$, belong to Z,
(e) $Z^{0+} = \bigcap_{\delta>0} Z^{0\delta}$,
(f) $Z_\bullet$ = the span of $(ia)^d$, $d = 0, 1, 2,$ etc., $\int a^{2d}\Delta < \infty$,
(g) $Z^{-\infty} = \bigcap_{t<0} Z^{-\infty\, t}$.

$Z^{-\infty\infty} = Z$ since $f \in Z$ implies $f\Delta \in L^1(R^1)$, and in that case $f\Delta = 0$ if $\int f\Delta e(-iat) = 0$ ($t \in R^1$); the functions $f \in Z^\bullet$ are of 0 (minimal) exponential type, so-called.

[1]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[2]Supported in part by the Office of Naval Research and in part by the National Science Foundation, GP-149. Massachusetts Institute of Technology.

$Z^\bullet$ is either dense in Z or a *closed* subspace of Z; the second alternative holds in the case of a *Hardy weight*:

$$\int \frac{\lg \Delta}{1+a^2} > -\infty,$$

and under this condition

$$Z^- \supset Z^{+/-} \supset Z^- \cap Z^+ \supset Z^{0+} = Z^\bullet \supset Z_\bullet.$$

Given a Hardy weight Δ, the problem is to decide if some or all of the above subspaces coincide; for instance, as it turns out, $Z^{+/-} = Z^\bullet$ if and only if $\Delta^{-1} = |f|^2$ with f entire of minimal exponential type, while $Z^\bullet = Z^{0+}$ for the most general Hardy weight.

$Z \neq Z^-$ in the Hardy case, while in the non-Hardy case $Z = Z^- \cap Z^+ = Z^{-\infty}$, and, if $\Delta \in \downarrow$ also, then $Z = Z^{0+}$ too. ($\Delta \in \downarrow$ means that $\Delta(a) \geq \Delta(b)$ for $0 \leq a < b$.)

$Z^{+/-}$ and Z^{0+} receive special attention below for reasons explained in the next part of the introduction.

S. N. Bernstein's problem of finding conditions on a weight $\Delta \leq 1$ so that each continuous function f with $\lim_{|a|\uparrow\infty} |f|\Delta = 0$ should be close to a polynomial p in the sense that $|f-p|\Delta$ be small, is similar to the problem of deciding if $Z_\bullet = Z$ or not, and it turned out that S. N. Mergelyan's solution of Bernstein's problem [10] and I. O. Hačatrjan's amplification of it [5] could be adapted to the present case.

13.1.2. Probabilistic Part. $\Delta\, da$ can be regarded as the spectral weight of a centered Gaussian motion with *sample paths* $t \to x(t) \in R^1$, *universal field* B, *probabilities* $P(\mathrm{B})$, and *expectations* $E(f)$:

$$E\big[x(s)x(t)\big] = \int e^{ia(t-s)}\Delta.$$

Bring in the (real) Hilbert space Q which is the closed span of $x(t)$ ($t \in R^1$) under the norm $\|f\| = [E(f^2)]^{\frac{1}{2}}$ and map $x(t) \to e(iat) \in Z$. Q is mapped 1:1 onto Z, inner products being preserved, and with the notations Q^{cd} = *the span of* $x(t)$ ($c \leq t \leq d$) and B^{cd} = *the smallest Borel subfield of* B *measuring* $x(t)$ ($c \leq t \leq d$), a perfect correspondence is obtained between

(a) Z^-, $Q^- = Q^{-\infty\, 0}$, and $\mathsf{B}^- = \mathsf{B}^{-\infty\, 0}$ = *the past*,
(b) Z^+, $Q^+ = Q^{0\, \infty}$, and $\mathsf{B}^+ = \mathsf{B}^{0\, \infty}$ = *the future*,
(c) $Z^{+/-}$, the projection $Q^{+/-}$ of Q^+ onto Q^-, and $\mathsf{B}^{+/-}$ = *the smallest splitting field of past and future*,
(d) Z^{0+}, $Q^{0+} = \bigcap_{\delta>0} Q^{0\delta}$, and $\mathsf{B}^{0+} = \bigcap_{\delta>0} \mathsf{B}^{0\delta}$ = *the germ*,
(e) $Z_\bullet$, $Q_\bullet$ = the span of $x^{(d)}(0)$, $d = 0, 1, 2$, etc., $E[x^{(d)}(0)^2] < \infty$, and the associated field $\mathsf{B}_\bullet$,
(f) $Z^{-\infty}$, $Q^{-\infty} = \bigcap_{t<0} Q^{-\infty\, t}$, and $\mathsf{B}^{-\infty} = \bigcap_{t<0} \mathsf{B}^{-\infty\, t}$ = *the distant past.*

B^-, B^+, B^{0+}, etc. do not just include the fields of Q^-, Q^+, Q^{0+}, etc., but for instance, if $f \in Q$ is measurable over B^{0+}, then it belongs to Q^{0+}; the proof of this fact and its analogues is facilitated by use of the lemma of Tutubalin-Freĭdlin [11]: if the field A is part of the smallest Borel field containing the fields of B and C and if C is independent of A and B then $\mathsf{A} \subset \mathsf{B}$.

$\mathsf{B}^{+/-}$ (= the splitting field) needs some explanation. Given a pair of fields such as B^- (= the past) and B^+ (= the future), a field $\mathsf{A} \subset \mathsf{B}^-$ is said to be a splitting field of B^- and B^+, if, conditional on A, B^+ is independent of B^-. B^- is a splitting field, and as is not hard to prove, a smallest splitting field exists, coinciding in the present (Gaussian) case with the field of the projection $Q^{+/-}$ (see H. P. McKean, Jr. [9] for the proof). $\mathsf{B}^{+/-}$ and so also $Z^{+/-}$ is a measure of the dependence of the future on the past.

Because $Z^\bullet = Z^{0+}$ for a Hardy weight, the condition $\Delta^{-1} = |f|^2$ (f entire of minimal exponential type) for $Z^{+/-} = Z^\bullet$ is equivalent in the Hardy case to the condition that the motion split over its germ ($\mathsf{B}^{+/-} = \mathsf{B}^{0+}$); this is the principal result of this paper from a probabilistic standpoint. Tutubalin-Freĭdlin's result [11] that if $\Delta \geq |a|^{-d}$ as $|a| \uparrow \infty$ for some $d \geq 2$ then $\mathsf{B}^{0+} = \mathsf{B}_\bullet$, is the sole fact about B^{0+} that has been published to our knowledge.

13.2. Hardy Functions

An even Hardy weight Δ can be expressed as $\Delta = |h|^2$, h belonging to the Hardy class H^{2+} of functions $h = h(\gamma)$ ($\gamma = a + ib$) regular in the half plane ($b > 0$) with $h^*(-a) = h(a)$ and $\int |h(a+ib)|^2\, da$ bounded ($b > 0$); such a Hardy function satisfies

$$\lim_{b\downarrow 0} \int \big|h(a+ib) - h(a)\big|^2\, da = 0 \quad \text{and} \quad \int \big|h(a+ib)\big|^2\, da \leq \int \big|h(a)\big|^2\, da \quad (b > 0).$$

Hardy functions can also be described as the (regular) extensions into $b > 0$ of the Fourier transforms of functions belonging to $L^2(R^1, dt)$ vanishing on the left half line ($t \leq 0$). According to Beurling's nomenclature, each Hardy function comes in two pieces: an *outer* factor o with

$$\lg \big|o(\gamma)\big| = \frac{1}{\pi} \int \frac{b}{(c-a)^2 + b^2} \lg \big|h(c)\big| dc \quad (\gamma = a + ib)$$

and an *inner* factor j with

$$\big|j(\gamma)\big| \leq 1 \quad (b > 0), \qquad \big|j(\gamma)\big| = 1 \quad (b = 0);$$

the complete formula for the outer factor of h is

$$o(\gamma) = e\left[\frac{1}{\pi i} \int \frac{\gamma c - 1}{\gamma + c} \lg \big|h(c)\big| \frac{dc}{1 + c^2}\right].$$

$Z^+ h = H^{2+}$, i.e., $e(i\gamma t)h$ ($t \geq 0$) spans out the whole of H^{2+}, if and only if h is outer. H^{2-} stands for the analogous Hardy class for $b < 0$. $L^2(R^1, da)$ is the (perpendicular) direct sum of H^{2-} and H^{2+}. Hardy classes $H^{1\pm}$ are defined in the same manner except that now it is $\int |h(a+ib)| da$ that is to be bounded. H^{1+} can be described as those functions h belonging to $L^1(R^1, da)$ with $\int e(-iat) h\, da = 0$ ($t \leq 0$); it is characteristic of the moduli of such functions that $\int \lg |h|/(1+a^2) > -\infty$ (see [7] for proofs and additional information).

13.3. Discussion of $Z^- \supset Z^{+/-} \supset Z^- \cap Z^+$

Given Δ as in Sect. 13.1.1, Hardy or not, the inclusions $Z \supset Z^- \supset Z^{+/-} \supset Z^- \cap Z^+$ are obvious, so the problem is to decide in what circumstances some or all of the above subspaces coincide. As it happens,

(a) *either*

$$\int \lg \Delta / (1 + a^2) = -\infty \quad \textit{and} \quad Z = Z^- \cap Z^+ = Z^{-\infty}$$

or

$$\int \lg \Delta / (1 + a^2) > -\infty \quad \textit{and} \quad Z \neq Z^- \neq Z^- \cap Z^+;$$

in the second (Hardy) case, $\int \lg \Delta/(1+a^2) > -\infty$, $\Delta = |h|^2$ with h outer belonging to H^{2+}, and the following statements hold:

(b) $Z^- \neq Z^{+/-}$ *if and only if* $i = h/h^*$, *restricted to the line, coincides with the ratio of* 2 *inner functions*,

(c) $Z^{+/-} = Z^- \cap Z^+$ *if and only if* $i = h/h^*$, *restricted to the line, coincides with an inner function.*

(a) goes back to Szegö; the rest is new.

PROOF OF (a) ADAPTED FROM [7]. $Z \neq Z^-$ implies that for the coprojection f of $e(ias)$ upon Z^-, $f\Delta \neq 0$ for some $s > 0$. Because the projection belongs to Z^-, $e(-ias)fe(iat) \in Z^-$ $(t \leq 0)$ and so is perpendicular (in Z) to f; also, f is perpendicular to $e(iat)$ $(t \leq 0)$, so

$$\int e^{ias}|f|^2\Delta e^{-iat}da = \int f\Delta e^{-tat}da = 0 \quad (t \leq 0).$$

But in view of $\int |f|\Delta \leq \|f\|_\Delta(\int \Delta)^{\frac{1}{2}} < \infty$, it follows that $f\Delta$ belongs to the Hardy class H^{1+}, whence $\int \lg(|f|\Delta)/(1+a^2) > -\infty$. But also $\int \lg(|f|^2\Delta)/(1+a^2) < \infty$ since $f \in Z$, and so $\int \lg \Delta/(1+a^2) > -\infty$, as stated. On the other hand, $\int \lg \Delta/(1+a^2) > -\infty$ implies $\Delta = |h|^2$ with h outer belonging to H^{2+}, and $Z \neq Z^-$ follows: indeed, since Δ is even, $h^*(-a) = h(a)$, and since $h^2 \in H^{1+}$,

$$\int e^{-iat}h^2\,da = \int e^{-iat}i\Delta\,da = 0 \quad (t \leq 0)\ \big(i = h/h^*\big),$$

stating that $i \in Z$ is perpendicular to Z^-. $Z^- \neq Z^- \cap Z^+$ follows, since, in the opposite case, $Z^- \subset Z^+$ so that $Z^+ = Z$ and hence also $Z^- = Z$, against the fact that Δ is a Hardy weight. $Z^{-\infty} = \bigcap_{t<0} Z^{-\infty\, t} = Z$ follows in the non-Hardy case. □

PROOF OF (b). Given $\int \lg \Delta/(1+a^2) > -\infty$, let $\Delta = |h|^2$ with h outer as before and prepare three simple lemmas.

$Z^+h = H^{2+}$ since h is outer as stated in 2.

$Z^-h = iH^{2-}$ because $Z^-h^* = (Z^+h)^* = (H^{2+})^* = H^{2-}$.

$Z^{+/-}h = ipi^{-1}H^{2+}$, p being the projection in $L^2(R^1)$ upon H^{2-}; indeed, ipi^{-1} is a projection and coincides with the identity just on iH^{2-}.

Coming to the actual proof of (b), if the inclusion $Z^- \supset Z^{+/-}$ is proper, then $Z^-h = iH^{2-}$ contains a function $f = i(j_2o_2)^*$ perpendicular to $Z^{+/-}h = ipi^{-1}H^{2+}$, j_2 being an inner and $o_2 \in H^{2+}$ an outer function. Because $ipi^{-1} = 1$ on iH^{2-}, it follows that f is perpendicular in $L^2(R^1)$ to H^{2+}, so $f \in H^{2-}$, i.e., $f = (j_1o_1)^*$, j_1 being an inner and $o_1 \in H^{2+}$ an outer function; in brief, $i(j_2o_2)^* = (j_1o_1)^*$. Because $|o_1| = |o_2|$ on the line $b = 0$, the outer factors can be cancelled, proving that $i = j_2/j_1$. On the other hand, if $i = j_2/j_1$, then $f = i(j_2h)^* \neq 0$ belongs to $iH^{2-} = Z^-h$. Also $f = (j_1h)^* \in H^{2-}$ so that f is perpendicular in $L^2(R^1)$ to H^{2+}, and since $f \in iH^{2-}$, it must be perpendicular to $ipi^-H^{2+} = Z^{+/-}h$ also. $Z^- \neq Z^{+/-}$ follows, completing the proof. □

PROOF OF (c). $Z^- \neq Z^- \cap Z^+$ in the Hardy case, so if $Z^{+/-} = Z^- \cap Z^+$, then $Z^- \neq Z^{+/-}$, and according to (b), $i = h/h^*$ is a ratio j_2/j_1 of inner functions with no common factor. $f \in Z^-h = iH^{2-}$ is perpendicular in $L^2(R^1)$ to $Z^{+/-}h = ipi^{-1}H^{2+}$ if and only if $i^{-1}f \in H^{2-}$ is perpendicular to $pi^{-1}H^{2+}$, or, and this is the same, to $i^{-1}H^{2+}$, and so, computing annihilators in iH^{2-}, $(Z^{+/-}h)^0 = iH^{2-} \cap H^{2-}$. Now $f \in iH^{2-} \cap H^{2-}$ can be expressed as $(j_2/j_i)j_3^*o_3^* = j_4^*o_4^*$ and the outer factors have to match, so $j_2j_4 = j_1j_3$, and since j_1 and j_2 have no common factors, j_1 divides j_4 [1, page 246] and $f \in iH^{2-} \cap (1/j_1)H^{2-}$. Because $j_1^*H^{2-} \subset H^{2-}$, $Z^{+/-}h$ can now be identified as $[iH^{2-} \cap (1/j_1)H^{2-}]^0 = iH^{2-} \cap (1/j_1)H^{2+}$, the annihilator being computed in iH^{2-}; this is because $(1/j_1)H^{2-} = ij_2^*H^{2-} \subset iH^{2-}$ and $(1/j_1)H^{2-} \oplus iH^{2-} \cap (1/j_1)H^{2+}$ is a perpendicular splitting of iH^{2-}. But according to this identification, if $Z^{+/-} = Z^- \cap Z^+$, then $i(j_1h)^* = (1/j_1)h \in Z^{+/-}h \subset Z^+h = H^{2+}$, and h being outer, it follows that j_1 has to be constant, completing half the proof; the opposite implication is obvious using the above identification of $Z^{+/-}h$ in conjunction with $(Z^- \cap Z^+)h = iH^{2-} \cap H^{2+}$. □

EXAMPLE 13.3.1. $h = (1 - i\gamma)^{-3/2}$ is outer belonging to H^{2+} and $Z^- = Z^{+/-}$; indeed, $[(1 + i\gamma)/(1 - i\gamma)]^{3/2} = j_2/j_1$ would mean that $j_1^2[(1 + i\gamma)/(1 - i\gamma)]^3 = j_2^2$, and this would make j_2^2 have a root of *odd* degree at $\gamma = i$.

An outer function h belonging to H^{2+} is determined by its phase factor $i = h/h^*$ if and only if $\dim Z^- \cap Z^+ = 1$; indeed, if $\dim Z^- \cap Z^+ = 1$ and if o is an outer function belonging to H^{2+} with $o/o^* = i$, then $o \in iH^{2-} \cap H^{2+} = Z^- \cap Z^+ h$ and, as such, is a multiple of h. On the other hand, if $o/o^* = i$ implies $o = \text{constant} \times h$, then $\dim Z^- \cap Z^+ = 1$ because if o is the outer factor of $f \in Z^- \cap Z^+ h = iH^{2-} \cap H^{2+}$, then $o/o^* = i/j$ with j an inner multiple of the inner factor of f. $(j + 1)o$ is outer [7, page 76], and since $(j + 1)o/(j + 1)^* o^* = i$, it is a multiple of h. $i(j - 1)o$ is likewise a multiple of h, and so o itself is a multiple of h, $j = 1$, and f too is a multiple of h.

13.4. Discussion of $Z^\bullet$

Before proving the rest of the inclusions $Z^- \cap Z^+ \supset Z^{0+} \supset Z^\bullet \supset Z_\bullet$, Mergelyan's solution of Bernstein's problem and his proof also, is adapted to the present needs.

Given Δ, Hardy or not, let $Z^\bullet = Z^\bullet_\Delta$ be the class of entire functions f of minimal exponential type which, restricted to $b = 0$, belong to Z, let $\Delta^+ = \Delta(1 + a^2)^{-1}$, suppose $\int \Delta^+ = 1$, and putting

$$\sigma^\bullet(\gamma) = \textit{the least upper bound of } \left|f(\gamma)\right| : f \in Z^\bullet_{\Delta^+}, \quad \|f\|_{\Delta^+} \le 1,$$

let us check that the following alternative holds:

either $\sigma^\bullet \equiv \infty (b \neq 0)$,

$$\sup \int \frac{\lg^+ |f|}{1 + a^2} = \int \frac{\lg \sigma^\bullet}{1 + a^2} = \infty, \quad \textit{for } f \in Z^\bullet_{\Delta^+} \textit{ with } \|f\|_{\Delta^+} \le 1,$$

and $Z^\bullet$ *is dense in* Z,

or $\lg \sigma^\bullet$ *is a continuous, non-negative, subharmonic function*,

$$\int \frac{\lg \sigma^\bullet}{1 + a^2} < \infty,$$

$$\lg \sigma^\bullet(\gamma) \le \frac{1}{\pi} \int \frac{b}{(c - a)^2 + b^2} \lg \sigma^\bullet(c) dc \quad (\gamma = a + ib,\ b > 0),$$

$$\varlimsup_{R \uparrow \infty} R^{-1} \max_{0 \le \theta < 2\pi} \lg \sigma^\bullet\left(Re^{i\theta}\right) \le 0,$$

and $Z^\bullet$ *is a closed subspace of* Z;

the second alternative must hold in the case of a Hardy weight as will be proved in Sect. 13.6.2. Because $(f(\gamma^*))^* = f(-\gamma) \in Z^\bullet$ if $f \in Z^\bullet$,

$$\sigma^\bullet(\gamma) = \sigma^\bullet\left(\gamma^*\right) = \sigma^\bullet(-\gamma);$$

this fact is used without additional comment below.

Break up the proof into simple lemmas.

(a) $\sigma^\bullet(\gamma) \equiv \infty$ $(b \neq 0)$ *if and only if* $Z^\bullet$ *is dense in* Z.

PROOF OF (a). $\sigma^\bullet(\beta) = \infty(\beta = a + ib,\ b \neq 0)$ implies that $f \in Z^\bullet_{\Delta^+}$ can be found with $\|f\|_{\Delta^+} \le 1$, $|f(\beta)| > \delta^{-1}$, and hence

$$\left\|\frac{1}{c - \beta} + \frac{f - f(\beta)}{(c - \beta)f(\beta)}\right\|_\Delta = \left\|\frac{f}{(c - \beta)f(\beta)}\right\|_\Delta \le \left|f(\beta)\right|^{-1} \left\|\frac{c - i}{c - \beta}\right\|_\infty \|f\|_{\Delta^+} < \text{constant} \times \delta.$$

Breaking up $[f - f(\beta)]/(\gamma - \beta)f(\beta)$ into the sum of its odd and even parts f_1 and f_2 and then into the sum (with coefficients of modulus 1) of four pieces:

$$f_{11} = \frac{1}{2}(f_1 + f_1^*), \quad f_{12} = \frac{i}{2}(f_1 - f_1^*), \quad f_{21} = \frac{i}{2}(f_2 + f_2^*), \quad f_{22} = \frac{1}{2}(f_2 - f_2^*),$$

each of which belongs to $Z^\bullet_\Delta$, it follows that if $g \in Z$ is perpendicular to $Z^\bullet_\Delta$, then $\int g\Delta/(c-\beta) = 0$ $(\beta = a + ib,\ b \neq 0)$, whence

$$\int \frac{b}{(c-a)^2 + b^2} g\Delta\, dc = 0 \quad (b > 0),$$

and $g\Delta = 0$ as desired. On the other hand, if $Z^\bullet_\Delta$ is dense in Z, then it is possible to find an entire function f of minimal exponential type with $\|1/(c-\beta) - f\|_\Delta < \delta$ $(\beta = a + ib,\ b \neq 0)$. Bring in an entire function g with $[g - g(\beta)]/(\gamma - \beta)g(\beta) = -f$; then

$$\delta > \left\| \frac{g}{(c-\beta)g(\beta)} \right\|_\Delta \geq \textit{a positive constant depending upon } \beta \textit{ alone} \times \frac{\|g\|_{\Delta^+}}{|g(\beta)|},$$

and so

$$|g(\beta)| > \text{constant} \times \delta^{-1} \|g\|_{\Delta^+}.$$

g is now split into the sum (with coefficients of modulus 1) of four members g_{11}, g_{12}, g_{21}, g_{22} of $Z^\bullet_{\Delta^+}$, and it develops that

$$\begin{aligned} \text{constant} \times \delta^{-1}\|g\|_{\Delta^+} &< |g(\beta)| \leq |g_{11}(\beta)| + |g_{12}(\beta)| + |g_{21}(\beta)| + |g_{22}(\beta)| \\ &\leq \sigma^\bullet(\beta)\big(\|g_{11}\|_{\Delta^+} + \|g_{12}\|_{\Delta^+} + \|g_{21}\|_{\Delta^+} + \|g_{22}\|_{\Delta^+}\big) \\ &\leq 2\sigma^\bullet\big(\|g_1\|_{\Delta^+} + \|g_2\|_{\Delta^+}\big) \leq 2\sqrt{2}\sigma^\bullet\big(\|g_1\|^2_{\Delta^+} + \|g_2\|^2_{\Delta^+}\big)^{\frac{1}{2}} \\ &= 2\sqrt{2}\sigma^\bullet\|g\|_{\Delta^+}, \end{aligned}$$

making use of $\int g_1^* g_2 \Delta^+ = 0$. But since δ can be made small, $\sigma^\bullet(\beta)$ is in fact $= \infty$. □

(b) *$Z^\bullet$ dense in Z implies*

$$\sup \int \frac{\lg^+ |f|}{1 + a^2} = \int \frac{\lg \sigma^\bullet}{1 + a^2} = \infty, \quad \textit{for } f \in Z^\bullet_{\Delta^+} \textit{ with } \|f\|_{\Delta^+} \leq 1.$$

Proof of (b). Given $f \in Z^\bullet_{\Delta^+}$, if $\beta = a + ib$ $(b > 0)$, then

$$\lg |f(\beta)| \leq \frac{1}{\pi} \int \frac{b}{(c-a)^2 + b^2} \lg^+ |f(c)| dc$$

as follows from Nevanlinna's theorem [2, 1.2.3] on letting $R \uparrow \infty$ and using

$$\overline{\lim_{R\uparrow\infty}}\, R^{-1} \max_{0 \leq \theta < 2\pi} \lg \big|f\big(Re^{i\theta}\big)\big| \leq 0.$$

Now apply (a). □

(c) $\lg \sigma^\bullet(\beta) \leq \frac{1}{\pi} \int \frac{b}{(c-a)^2 + b^2} \lg \sigma^\bullet(c) dc$ $(\beta = a + ib,\ b > 0)$.

Proof of (c). Obvious from (b). □

(d) *$Z^\bullet$ non-dense implies that $\sigma^\bullet$ is bounded in the neighborhood of each point $\beta = a + ib$ $(b > 0)$; in fact, if $Z^\bullet$ is non-dense $\lg \sigma^\bullet$ is a non-negative continuous subharmonic function $(b \neq 0)$.*

PROOF OF (d). Given $\beta = a + ib$ $(b > 0)$ and a point α near it, take $g \in Z^\bullet_{\Delta+}$ with $\|g\|_{\Delta^+} \leq 1$ and $|g(\alpha)|$ close to $\sigma^\bullet(\alpha)$, and let $f = 1 + [(\gamma - \beta)/(\gamma - \alpha)][(g - g(\alpha))/g(\alpha)]$, observing that f need not belong to $Z^\bullet_{\Delta+}$ since $f^*(-a) = f(a)$ can fail.

$$\|f\|_{\Delta^+} = \left\| \frac{\beta - \alpha}{c - \alpha} - \frac{\beta - \alpha}{c - \alpha}\frac{g}{g(\alpha)} + \frac{g}{g(\alpha)} \right\|_{\Delta^+}$$
$$\leq \left\| \frac{\beta - \alpha}{c - \alpha} \right\|_\infty \left(1 + \left|g(\alpha)\right|^{-1}\right) + \left|g(\alpha)\right|^{-1},$$

and so, as in the second part of the proof of (a),

$$1 = \left|f(\beta)\right| \leq 2\sqrt{2}\sigma^\bullet(\beta)\|f\|_{\Delta^+}$$
$$\leq 2\sqrt{2}\sigma^\bullet(\beta)\left[\left\| \frac{\beta - \alpha}{c - \alpha} \right\|_\infty \left(1 + \left|g(\alpha)\right|^{-1}\right) + \left|g(\alpha)\right|^{-1}\right],$$

proving that $\sigma^\bullet(\alpha)$ is bounded on a neighborhood of β if $\sigma^\bullet(\beta) < \infty$. Because $1 \in Z^\bullet_{\Delta+}$, $\sigma^\bullet \geq 1$ ($\int \Delta^+ = 1$ is used at this place), so $\lg \sigma^\bullet \geq 0$, and since $\lg |f|$ is subharmonic for each $f \in Z^\bullet_{\Delta+}$, $\lg \sigma^\bullet$ is also subharmonic. But now it follows that if $\sigma^\bullet(\beta) = \infty$ at one point $\beta = a + ib$ $(b > 0)$, then it is also ∞ at some point of each punctured neighborhood of β, and arguing as in the first part of the proof of (a) with f perpendicular to $Z^\bullet_\Delta$, $\int f\Delta/(c - \alpha)dc$ is found to vanish at some point of each punctured neighborhood of β and hence to be $\equiv 0$. $Z^\bullet$ dense in Z follows as before, so $Z^\bullet$ non-dense implies the (local) boundedness of $\sigma^\bullet$. It remains to prove that $\sigma^\bullet$ is continuous $(b \neq 0)$. On a small neighborhood of $\alpha = a + ib$, $|f|$ $(f \in Z^\bullet_{\Delta+})$ lies under a universal bound, $\sigma^\bullet$. An application of Cauchy's formula implies that $|f'|$ lies under a universal bound on a smaller neighborhood of α, and so $|f(\beta_2) - f(\beta_1)|$ lies under a universal constant B times $|\beta_2 - \beta_1|$ as β_1 and β_2 range over this smaller neighborhood. But then

$$\left|f\left(\beta_2\right)\right| \leq \left|f\left(\beta_1\right)\right| + B\left|\beta_2 - \beta_1\right| < \sigma^\bullet\left(\beta_1\right) + B\left|\beta_2 - \beta_1\right|,$$

so that

$$\sigma^\bullet\left(\beta_2\right) \leq \sigma^\bullet\left(\beta_1\right) + B\left|\beta_2 - \beta_1\right|,$$

and interchanging the roles of β_1 and β_2 completes the proof of (d). □

(e) *$Z^\bullet$ non-dense implies* $\int \lg^+ |f|/(1 + a^2) \leq \int \lg \sigma^\bullet/(1 + a^2) < \infty$.

PROOF OF (e). $Z^\bullet$ non-dense implies the existence of $g \in Z$ perpendicular to $Z^\bullet_\Delta$, and since, if $f \in Z^\bullet_{\Delta+}$, $(f - f(\beta))/(\gamma - \beta)$ is the sum (with coefficients of modulus 1) of four members of $Z^\bullet_\Delta$,

$$\int \frac{g^* f}{c - \beta}\Delta = \int \frac{g^*\Delta}{c - \beta} f(\beta) \equiv \hat{g} f \quad \left(f \in Z^\bullet_{\Delta+},\ b \neq 0\right).$$

Because $\hat{g}$ is regular and bounded $(b \geq 1)$, $\int \lg |\hat{g}(a + i)|/(1 + a^2) > -\infty$; also

$$\left|\hat{g} f(a + i)\right| \leq \|g\|_\Delta \|f\|_{\Delta^+} \left\| \frac{c - i}{c - a - i} \right\|_\infty,$$

so that $\sigma^\bullet(a + i) \leq \text{constant} \times (1 + a^2)^{\frac{1}{2}} |\hat{g}(a + i)|^{-1}$ and $\int \lg \sigma^\bullet(a + i)/(1 + a^2) < \infty$. But as in the proof of (b),

$$\lg \left|f(a)\right| \leq \frac{1}{\pi} \int \frac{\lg \sigma^\bullet(c + i)}{(c - a)^2 + 1} dc \quad \left(f \in Z^\bullet_{\Delta+}\right),$$

and so

$$\int \frac{\lg \sigma^\bullet(a)}{1+a^2} da \le \int \lg \sigma^\bullet(c+i) dc \frac{1}{\pi} \int \frac{da}{1+a^2} \frac{1}{(c-a)^2+1}$$
$$= 2 \int \frac{\lg \sigma^\bullet(c+i)}{c^2+4} dc < \infty,$$

as stated. □

(f) *If $Z^\bullet$ is non-dense in Z then it is a closed subspace of Z and*

$$\varlimsup_{R\uparrow\infty} R^{-1} \max_{0\le\theta<2\pi} \lg \sigma^\bullet\bigl(Re^{i\theta}\bigr) \le 0.$$

PROOF OF (f).

$$R^{-1} \lg \sigma^\bullet\bigl(Re^{i\theta}\bigr) \le \frac{1}{\pi} \int \frac{\sin\theta(1+c^2)}{(c-R\cos\theta)^2 + R^2\sin^2\theta} \frac{\lg \sigma^\bullet}{1+c^2} dc \quad (0<\theta<\pi)$$

according to (d). A simple estimate, combined with $\sigma^\bullet(\gamma) = \sigma^\bullet(\gamma^*)$ verifies

$$\varlimsup_{R\uparrow\infty} R^{-1} \lg \sigma^\bullet\bigl(Re^{i\theta}\bigr) \le 0 \quad (\theta = \pi/4,\ 3\pi/4,\ 5\pi/4,\ 7\pi/4).$$

Phragmén-Lindelöf is now applied to each of the sectors between $\pi/4$, $3\pi/4$, $5\pi/4$, $7\pi/4$; for instance, in the sector $[\pi/4,\ 3\pi/4]$, each $f \in Z^\bullet_{\Delta+}$ with $\|f\|_{\Delta^+} \le 1$ satisfies

$$\Bigl|f(\gamma)e^{i\gamma 2^{\frac12}\delta}\Bigr| \le \bigl|f\bigl(Re^{i\theta}\bigr)\bigr| e^{-R\delta} \le A \quad (\pi/4 \le \theta \le 3\pi/4)$$

$$\Bigl|f(\gamma)e^{i\gamma 2^{\frac12}\delta}\Bigr| \le \sigma^\bullet\bigl(Re^{i\theta}\bigr) e^{-R\delta} \le B \quad (\theta = \pi/4,\ 3\pi/4)$$

with a constant B not depending upon f, and so

$$\Bigl|f(\gamma)e^{i\gamma(2\delta)^{\frac12}}\Bigr| \le B \quad (\pi/4 \le \theta \le 3\pi/4),$$

or

$$\sigma^\bullet\bigl(Re^{i\theta}\bigr) \le Be^{R(2\delta)^{\frac12}} \quad (\pi/4 \le \theta \le 3\pi/4).$$

$Z^\bullet$ *closed* follows since $|f|$ ($f \in Z^\bullet_\Delta$) lies under a universal bound ($\sigma^\bullet$) on any bounded region of the plane.

Mergelyan's alternative is now proved; several additional comments follow.

Given $f \in Z^\bullet_{\Delta+}$, $(\gamma+i)^{-1} fh \in H^{2+}$ while $(\gamma+i)^{-1} \in H^{2+}$ is an outer function, so that

$$\lg\bigl|fh(i)\bigr|^2 \le \frac{1}{\pi}\int \frac{\lg\bigl|(c+i)^{-1}fh\bigr|^2}{1+c^2} + \frac{1}{\pi}\int \frac{\lg|c+i|^2}{1+c^2}$$
$$= \frac{1}{\pi}\int \frac{\lg|fh|^2}{1+c^2} \le \lg\left(\frac{1}{\pi}\int \frac{|f|^2\Delta}{1+c^2}\right) = \lg\left(\frac{1}{\pi}\|f\|^2_{\Delta+}\right),$$

and so $\pi^{\frac12}\sigma^\bullet(i) \le |h(i)|^{-1}$. Now it is proved that this upper bound is attained if and only if h^{-1} *is entire of minimal exponential type.* Using the compactness that

$$\lim_{R\uparrow\infty} R^{-1} \max_{0\le\theta<2\pi} \lg \sigma^\bullet\bigl(Re^{i\theta}\bigr) \le 0$$

ensures, it is possible to choose $f \in Z^\bullet_{\Delta+}$ with $f(i) = \sigma^\bullet(i)$ and $\|f\|_{\Delta^+} = 1$. As before,

$$\bigl|fh(i)\bigr|^2 \le e\left[\frac{1}{\pi}\int \frac{\lg|f|^2\Delta}{1+a^2}\right] \le \frac{1}{\pi}\int \frac{|f|^2\Delta}{1+a^2} = \frac{1}{\pi},$$

so if $\pi^{\frac{1}{2}}\sigma^\bullet(i) = |h(i)|^{-1}$, then the converse of Jensen's inequality implies that fh is constant; the other implication is trivial.

$\sigma^\bullet(i)$ can also be computed from a Szegö minimum problem:

$$\frac{1}{\sigma^\bullet(i)^2} = \inf \frac{|1-f|^2\Delta}{1+a^2}, \quad \text{for } f \in Z^\bullet_{\Delta^+} \text{ with } f(i) = 0,$$

as the reader can easily check.

Because of the compactness of $Z^\bullet$ used above, it is possible in the non-dense case to find $f = f_\gamma \in Z^\bullet_{\Delta^+}$ with $f(\gamma) = \sigma^\bullet(\gamma)$ and $\|f\|_{\Delta^+} = 1$. f_γ is unique and is perpendicular (in $Z^\bullet_{\Delta^+}$) to each $f \in Z^\bullet_{\Delta^+}$ vanishing at γ. $f_\alpha(\beta)\sigma^\bullet(\alpha)$ acts as a Bergman reproducing kernel for $Z^\bullet_{\Delta^+}$ since

$$\int f^*_\alpha [f - f(\alpha)]\Delta^+ = 0 \quad \big(f \in Z^\bullet_{\Delta^+}\big)$$

implies

$$\int f^*_\alpha f \Delta^+ = f(\alpha)\int f^*_\alpha \Delta^+ = \frac{f(\alpha)}{\sigma^\bullet(\alpha)}\int \big|f_\alpha\big|^2\Delta^+ = \frac{f(\alpha)}{\sigma^\bullet(\alpha)}. \qquad \square$$

13.5. Proof of $Z^\bullet \subset Z^{0+}$ (Δ Hardy or Not)

To begin with, each $f \in Z^\bullet$ can be split into an even part $f_1 = \frac{1}{2}[f(\gamma)+f(-\gamma)] \in Z^\bullet$ and an odd part $f_2 \in Z^\bullet$; the proof is carried out for an even function $f \in Z^\bullet$ with Hadamard factorization

$$f(\gamma) = \gamma^{2m}\prod_{n=1}^{\infty}\left(1 - \frac{\gamma^2}{\gamma_n^2}\right),$$

the odd case being left to the reader. A simple estimate justifies us in ignoring the root of f at $\gamma = 0$; indeed $f_\delta = \delta^{2m}(1-\gamma^2/\delta^2)^m f/\gamma^{2m}$ is an even entire function of minimal exponential type, $|f_\delta/f|$ tends to 1 as $|\gamma| \uparrow \infty$ so that $f_\delta \in Z^\bullet$, and $\|f_\delta - f\|_\Delta$ tends to 0 as $\delta \downarrow 0$ so that if $f_\delta \in Z^{0+}$ then so does f.

Bring in the function

$$g(\gamma) = \prod_{|\gamma_n|<d}\left(1 - \frac{\gamma^2}{\gamma_n^2}\right)\prod_{n>d\delta}\left(1 - \frac{\gamma^2\delta^2}{n^2}\right),$$

depending upon a small positive number δ and a large integral number d. Given $\delta > 0$, $\varepsilon > 0$, and $A < \infty$, it is possible to find $d_1 = d_1(\delta, \varepsilon, A)$ and a universal constant B so that for each $d \geq d_1$,

(a) $|f - g| < \varepsilon \quad (|a| < A)$
(b) $|g| < B|f| \quad (A \leq |a| < d/2)$
(c) $|g| < B \quad (|a| \geq d/2)$
(d) $g \in L^2(R^1)$.

It is best to postpone the proof of (a), (b), (c), (d) and to proceed at once to the

PROOF THAT $f \in Z^{0+}$. Using (a), (b), (c) above,

$$\|f-g\|^2_\Delta < \varepsilon^2\int \Delta + 2(B+1)^2\int_A^{d/2}|f|^2\Delta + 2\int_{d/2}\big(B+|f|\big)^2\Delta$$

tends to 0 as $d \uparrow \infty$, $A \uparrow \infty$, and $\varepsilon \downarrow 0$ in that order. Because the entire function g differs from $\sin\pi\delta\gamma$ by a rational factor and, as such, is of exponential type $\pi\delta$, it follows from (d) in conjunction with the Paley-Wiener theorem that

$$g(\alpha) = \int_{|t|<\pi\delta} e^{iat}\hat{g}(t)dt \quad \text{with} \quad \int_{|t|<\pi\delta} |\hat{g}|^2 dt < \infty.$$

But $\int_{|t|<\pi\delta} e^{iat}\hat{g}\,dt \in Z^{|t|\leq\pi\delta}$, as is obvious upon noting the bound

$$\left\| \int_{|t|<\pi\delta} e^{iat}\hat{g}\,dt \right\|_{\Delta}^{2} \leq 2\pi\delta \int_{|t|<\pi\delta} |\hat{g}|^2 \int \Delta$$

and so $f \in \bigcap_{\delta>0} Z^{|t|<\pi\delta} = Z^{0+}$ (see Sect. 13.6.1). □

Coming to the proof of (a), (b), (c), (d) above, it is convenient to introduce

$$p(\gamma) = p_m(\gamma) = \pi\gamma \prod_{n=1}^{m} \left(1 - \frac{\gamma^2}{n^2}\right)$$

and to check the existence of a universal constant B such that $Q \equiv |\sin(\pi a)/p(a)|$ is bounded as in

$$\text{(e)}\quad Q/B < \begin{cases} e^{-a^2/m} & |a| < m \\ e^{-a/2} & m \leq |a| < 2m \\ e^{-m-2m\lg(a/m)} & |a| \geq 2m. \end{cases}$$

Proof of (e). $Q = \prod_{n>m} (1 - a^2/n^2)$ for $|a| < m$, and since $1 - c \leq e(-c)$, $Q < e(-a^2/(m+1))$. Stirling's approximation is now used to estimate p below for $|a| \geq m$, removing first a factor $a - m$ in case $m \leq |a| < m + \frac{1}{2}$, and then $|\sin\pi\alpha|$ is estimated above by 1 on this range. On the other hand, if m is the biggest integer $< d\delta$ and if $|a| < d/2$, then $\delta|a| < m$ so that the first appraisal listed under (e) supplies us with the bound

$$Q(a\delta) = \prod_{n>d\delta} \left(1 - a^2\delta^2/n^2\right) < Be^{-a^2\delta^2/(m+1)} < Be^{-a^2\delta/2d},$$

and it follows that

$$B\big|f(a)\big| > \prod_{|\gamma_n|<d} \left|1 - \frac{a^2}{\gamma_n^2}\right| \prod_{n>d\delta} \left(1 - \frac{a^2\delta^2}{n^2}\right) = |g|,$$

as desired. □

Proof of (c) and (d). Let $\#(R)$ be the number of n's with $|\gamma_n| < R$. On the range $|a| \geq d/2$,

$$\begin{aligned} \lg \prod_{|\gamma_n|<d} \left|1 - \frac{a^2}{\gamma_n^2}\right| &\leq \int_0^d \lg\left(1 + \frac{a^2}{R^2}\right) d\#(R) \\ &= \#(d)\lg\left(1 + \frac{a^2}{d^2}\right) + \int_0^d \frac{2a^2}{a^2 + R^2}\frac{\#}{R}dR \\ &\leq 2\#(d)\lg\left(3|a|/d\right) + 2\int_0^d \frac{\#}{R}dR \\ &= o\big[d + d\lg\left(|a|/d\right)\big] \end{aligned}$$

for large d, while according to (e), if $|a| \geq d/2$ and if m is the biggest integer $< d\delta$, then

$$Q(a\delta) < Be\Big[-\frac{1}{2}d\delta\big(1 + \lg(a/d)\big)\Big].$$

But then $|g| < B$ for large d as stated in (c), while for $d > 8/\delta$

$$|g| < Be\Big[-\frac{1}{4}d\delta\big(1+\lg(a/d)\big)\Big] \quad \big(|a| > d/2\big).$$

But for still larger d, $d\delta(1+\lg(a/d)) - 8\lg\ a > 0$ for $a > d/2$, since the left side is positive at $a = d/2$ and increasing for $a > d/2$. Thus

$$|g| < B/a^2 \quad \big(|a| > d/2\big)$$

so that $g \in L^2(R^1)$ as stated in (d). □

13.6.

13.6.1. Proof of $Z^- \cap Z^+ \supset Z^{0+}$ (Δ Hardy or Not). Given $f \in Z^{0+} \subset Z^+$, then $e(-ia\delta)f \in Z^{-\delta 0} \subset Z^-$, and

$$\big\|\big(e^{-ia\delta}-1\big)f\big\|_\Delta \le \max_{|a|\le n}\big|e^{-ia\delta}-1\big|\|f\|_\Delta + 2\Big(\int_{|a|>n}|f|^2\Delta\Big)^{\frac{1}{2}}$$

$$\le n\delta\|f\|_\Delta + 2\Big(\int_{|a|>n}|f|^2\Delta\Big)^{\frac{1}{2}}$$

is small for $\delta = n^{-2}$ and $n \uparrow \infty$, so that $f \in Z^-$ also. Our proof justifies

$$Z^{0+} = \bigcap_{\delta<0} Z^{\delta 0} = \bigcap_{\delta>0} Z^{|t|<\delta};$$

this fact will be used without additional comment below.

13.6.2. Proof of $Z^{0+} = Z^\bullet$ (Δ Hardy). $Z^{0+} \subset Z^\bullet$ is proved next for a Hardy weight Δ. Combined with the previous result $Z^{0+} \supset Z^\bullet$, this gives $Z^{0+} = Z^\bullet$.

Given $f \in Z^{0+}$, it is possible to find a finite sum

$$f_n = \sum c_k^n e\big(i\gamma\, t_k^n\big) \quad \text{with} \quad 0 \le t_k^n < 1/n,\ \big\|f - f_n\big\|_\Delta < 1/n,$$

and hence

$$\big\|f_n\big\|_\Delta < 1/n + \|f\|_\Delta \le 1 + \|f\|_\Delta.$$

Phragmén-Lindelöf is now applied to obtain bounds on $|f_n|$. Because $|f_n|$ is bounded ($b \ge 0$) and f_n is entire, $f_n h \in H^{2+}$, so

$$\int \big|f_n h(a+ib)\big|^2\, da \le \int \big|f_n\big|^2\Delta$$

is bounded ($b > 0,\ n \ge 1$), and an application of Cauchy's formula to a ring supplies us with the bound

$$\big|f_n h\big| \le B_1 \quad (b \ge 1,\ n \ge 1).$$

Also, $|e(-i\gamma/n)f_n|$ is bounded ($b < 0$), so

$$\big|e^{-i\gamma/n} f_n h^*\big| \le B_2 \quad (b \le -1,\ n \ge 1)$$

with a similar proof. Next, the underestimate

$$\begin{aligned}\pi \lg \big|h(a+ib)\big| &= \pi \lg \big|h^*(a-ib)\big| \\ &\geq \int \frac{b \lg^- |h|}{(c-a)^2+b^2} dc \geq B_3(1+a^2) \int \frac{\lg^- |h|}{1+c^2} dc \\ &\geq B_4 \big| e^{-B_5 \gamma^2} \big| \quad \big(1 \leq b \leq 2,\ B_4 > 0\big)\end{aligned}$$

justifies the bound

$$\big|g_n\big| \leq B_6 \quad \text{for } 1 \leq b \leq 2,\ n \geq 1 \text{ with } g_n \equiv e\big(-B_5\gamma^2\big) f_n.$$

Because $|g_n|$ tends to 0 at the ends of the strip $|b| \leq 2$, it is bounded ($\leq B_6$) in the whole strip according to the maximum modulus principle. In particular, $|f_n| \leq B_7$ on the disc $|\gamma| \leq 2$. A second underestimate of $|h|$ is obtained from the Poisson integral for $\lg |h|$: $\lim_{R\uparrow\infty} R^{-1} \lg |h(Re^{i\theta})| = 0$ ($\theta = \pi/4,\ 3\pi/4$), and it follows from the resulting bound

$$\big|f_n\big| \leq B_8 e^{\delta R} \quad (R \geq 1,\ \theta = \pi/4,\ 3\pi/4)$$

and its companion

$$\big|e^{-i\gamma/n} f_n\big| \leq B_9 e^{\delta R} \quad (R \geq 1,\ \theta = 5\pi/4,\ 7\pi/4)$$

combined with an application of Phragmén-Lindelöf to each of the four sectors between $\pi/4,\ 3\pi/4,\ 5\pi/4,\ 7\pi/4$, that

$$\big|f_n\big| \leq B_{10} e^{(\delta+1/n)R}.$$

But now it is legitimate to suppose that as $n \uparrow \infty$, f_n tends on the whole plane to an entire function f_∞; moreover, this function is specified on the line $b=0$ since $\|f_n - f\|_\Delta$ tends to 0 as $n \uparrow \infty$. Accordingly, the entire function f_∞ is an extension of f, and since $|f_\infty| \leq B_{10} e(\delta R)$, it is clear that $f \in Z^\bullet_\Delta$ as desired.

If Δ is non-Hardy then it is possible for Z^{0+} to contain $Z^\bullet$ properly. Indeed let $\Delta(a)$ be even, non-increasing for $a > 0$, and non-Hardy. Then, as will be proved in Sect. 13.8, $Z^{0+} = Z \neq Z^\bullet$.

Δ non-Hardy does not ensure that $Z^\bullet$ is dense in Z; in fact if $\int_{-1}^{1} \lg \Delta / (1+a^2) = -\infty$ while $\Delta \geq 1/a^2$ ($|a| \geq 1$), then $f \in Z^\bullet$ satisfies $\int |f|^2/(1+a^2) < \infty$, and a simple application of Phragmén-Lindelöf implies that f is constant; in short, $\dim Z^\bullet = 1$.

13.6.3. A Condition That $Z^- \cap Z^+ = Z^\bullet$ (Δ Hardy). *$Z^- \cap Z^+ = Z^\bullet$ if Δ is a Hardy weight and if $\int_{-d}^{+d} \Delta^{-1} < \infty$ ($d < \infty$).*

PROOF. The idea is that $f \in Z^- \cap Z^+$ is regular for $b \neq 0$ and can be continued across $b = 0$ if Δ is not too small (see T. Carleman [3] for a similar argument).

Given $f \in Z^- \cap Z^+$, then $fh \in H^{+2}$, $\lim_{b\downarrow 0} f(a+ib) = f(a)$ except at a set of points of Lebesgue measure 0 [7, page 123], and so the Lebesgue measure of

$$A \equiv \Big(a : \sup_{0 \leq b < \delta} \big|f(a+ib)\big| > \varepsilon^{-1},\ |a| < d\Big)$$

tends to 0 as δ and $\varepsilon \downarrow 0$; it is to be proved that

$$\sup_{0 \leq b < \delta} \int_A \big|f(a+ib)\big| da$$

is small for small δ and ε for each $d < \infty$. Bring in the summable weight

$$B = \Delta^{-1} \qquad \bigl(|c| \le 2d\bigr)$$
$$= \bigl(1 + c^2\bigr)^{-1} \quad \bigl(|c| > 2d\bigr).$$

Then for large d,

$$\begin{aligned}
&\left(\int_A \bigl|f(a+ib)\bigr| da\right)^2 \\
&\le \int \bigl|fh(a+ib)\bigr|^2 da \int_A \bigl[\Delta(a+ib)\bigr]^{-1} da \\
&\le \|f\|_\Delta^2 \int_A da\, e\left[\frac{1}{\pi}\int_{|c|\le 2d} \frac{b}{(c-a)^2+b^2} \lg \Delta^{-1}\, dc\right] \\
&\qquad \times e\left[\frac{1}{\pi}\int_{|c|>2d} \frac{b}{(c-a)^2+b^2} \lg \Delta^{-1}\, dc\right] \\
&\le 2\|f\|_\Delta^2 \int_A da\, e\left[\frac{1}{\pi}\int \frac{b}{(c-a)^2+b^2} \lg B\, dc\right]
\end{aligned}$$

and an application of Jensen's inequality implies

$$\begin{aligned}
\sup_{0\le b<\delta} \left[\int_A \bigl|f(a+ib)\bigr| da\right]^2 &\le 2\|f\|_\Delta^2 \int B\, dc \sup_{0\le b<\delta} \int_A \frac{b}{(c-a)^2+b^2} \frac{da}{\pi} \\
&\downarrow 2\|f\|_\Delta^2 \int_{\substack{\cap A \\ \delta,\,\varepsilon>0}} B\, dc = 0 \quad \bigl(\delta, \varepsilon \downarrow 0\bigr).
\end{aligned}$$

Using this appraisal, it follows that

$$\lim_{b\downarrow 0} \int_{-d}^{+d} \bigl|f(a+ib) - f(a)\bigr| da = 0;$$

the analogous result for $b < 0$ follows from a similar appraisal. Choose c so that $f(c+ib)$ tends boundedly to $f(c)$ as $b \downarrow 0$ and define

$$g(\gamma) = \int_c^a f(\xi + ib) d\xi + i \int_0^b f(c + i\eta) d\eta \quad (\gamma = a + ib).$$

g is regular ($b \ne 0$) since $f \in Z^- \cap Z^+$ is such, it is continuous across $b = 0$ and hence entire, so $f = g'$ is likewise entire, and all that remains to be proved is that f is of minimal exponential type.

Because $fh \in H^{2+}$, $\int |\lg|fh||/(1+a^2) < \infty$, and since $\lg^+|f| \le \lg^+|fh| - \lg^-|h|$, the integral $\int |\lg|f||/(1+a^2)$ is also convergent; also, $\lg|fh|$ is smaller than its Poisson integral, so

$$\lg^+ \bigl|f\bigl(Re^{i\theta}\bigr)\bigr| \le \frac{1}{\pi}\int \frac{R\sin\theta \lg^+ \bigl|f(c)\bigr| dc}{R^2 - 2Rc\cos\theta + c^2} \quad (0 < \theta < \pi),$$

$\lg|h|$ being expressible by its Poisson integral since h is an outer function. According to this bound,

$$\int_0^\pi \lg^+ \bigl|f\bigl(Re^{i\theta}\bigr)\bigr| d\theta \le \frac{2}{\pi}\int_0 \lg^+ \bigl|f(c)\bigr| \lg\left|\frac{R+c}{R-c}\right| \frac{dc}{c}$$

and

$$\int_R^{2R} dR \int_0^{\pi} d\theta \lg^+ \left|f\big(Re^{i\theta}\big)\right| \le \frac{2}{\pi}\int_0 \lg^+ \left|f(c)\right| dc \int_{R/c}^{2R/c} \lg\left|\frac{t+1}{t-1}\right| dt$$
$$< B_1\big(1+R^2\big)\int_0 \frac{\lg^+ \left|f(c)\right|}{1+c^2},$$

as a simple appraisal justifies. A similar bound holds for $\lg^+|f|$ in the lower half plane $b<0$, so that

$$\int_R^{2R} dR \int_0^{2\pi} d\theta \lg^+ \left|f\big(Re^{i\theta}\big)\right| < B_2\big(1+R^2\big),$$

and it follows that between each large R and its double $2\,R$ can be found an R_1 with

$$\int_0^{2\pi} \lg^+ \left|f\big(R_1 e^{i\theta}\big)\right| d\theta < 2B_2R_1.$$

An application of the Poisson-Jensen formula now supplies us with the bound

$$\lg^+|f| < B_3 R \quad \big(R\uparrow\infty\big),$$

and a second application of the fact that $\lg^+|f|$ is smaller than its Poisson integral supplies the additional information that

$$\varlimsup_{R\uparrow\infty} R^{-1}\lg^+ \left|f\big(Re^{i\theta}\big)\right| \le 0 \quad (\theta=\pi/4,\ 3\pi/4,\ 5\pi/4,\ 7\pi/4).$$

Phragmén-Lindelöf is now applied to each of the four sectors between, with the result that

$$\varlimsup_{R\uparrow\infty} R^{-1}\max_{0\le\theta<2\pi}\lg^+ \left|f\big(Re^{i\theta}\big)\right| \le 0,$$

and the proof is complete. □

A second proof of $Z^{0+}\subset Z^{\bullet}$ can be based on the above; indeed, if $f\in Z^{0+}$ and if f_n is chosen as in Sect. 13.6.2, then

$$\int \left|\big(f-f_n\big)h(a+i)\right|^2 da \le \left\|f-f_n\right\|_\Delta^2 < 1/n^2,$$

and so $f(a+i)\in Z^{0+}_{\Delta(a+i)}$ with $\Delta(a+i)=|h(a+i)|^2$. But $\Delta(a+i)$ is positive and continuous, so

$$Z^{0+}_{\Delta(a+i)} \subset Z^{-}_{\Delta(a+i)}\cap Z^{+}_{\Delta(a+i)} = Z^{\bullet}_{\Delta(a+i)},$$

proving that $f(a+i)$ is entire of minimal exponential type.

$Z^{\bullet}\neq Z^-\cap Z^+$ if, for instance, $\int_{-1}^{+1}\Delta/a^2<\infty$; indeed in this case,

$$\frac{1}{\pm ia+\delta} = \int_0^{\infty} e^{-\delta t}e^{\pm iat}dt \in Z^{\pm} \quad (\delta>0),$$

while

$$\left\|\frac{1}{ia\pm\delta}-\frac{1}{ia}\right\|_\Delta^2 \le \delta^2\int_{|a|>1}\Delta + \int_{|a|\le 1}\frac{\Delta}{a^2}\frac{\delta^2}{a^2+\delta^2}$$

tends to 0 as $\delta\downarrow 0$, so that $1/ia\in Z^-\cap Z^+$.

The Hardy weight $\Delta = a^2 e(-2|a|^{-\frac{1}{2}})/(1+a^4)$ illustrates the point that $f\in Z^-\cap Z^+$ can be regular in the punctured plane but have an essential singular point at $\gamma=0$. Define

$f = \gamma^{-1}\cos(1/\gamma^{\frac{1}{2}})$; then $f_\delta = f(\gamma + i\delta)$ $(\delta > 0)$ is of modulus $\le |a|^{-1}e(1/|a|^{\frac{1}{2}})$ on the line so that $\|f - f_\delta\|_\Delta$ tends to 0 as $\delta \downarrow 0$, while, as an application of the Paley-Wiener theorem justifies, $f_\delta = \int_0^\infty e(iat)\hat{f}_\delta(t)dt$ with $\hat{f}_\delta$ and $t\hat{f}_\delta \in L^2[0,\infty)$. $f_{0+} = f \in Z^+$ follows and a similar argument with $\delta < 0$ proves that $f \in Z^-$ also.

13.6.4. A Condition That Genus $Z^\bullet = 0$ (Δ Hardy).

Each $f \in Z^\bullet$ is of genus 0 and $\int_1 \lg\max_{0\le\theta<2\pi}|f(Re^{i\theta})|/R^2 < \infty$ if $\int_1 \lg^-\Delta(ib)/b^2 > -\infty$ or, and this is the same, if $\int_1 \lg^-\Delta\lg a/a^2 > -\infty$.

Proof. To begin with, $\int_1 \lg^-\Delta(ib)/b^2$ and $\int_1 \lg^-\Delta(a)\lg a/a^2$ converge and diverge together; indeed, since $\int_1 \lg^+\Delta(a)\lg a/a^2 \le \int_1 \Delta < \infty$, the convergence of $\int_1 \lg^-\Delta(a)\lg a/a^2$ combined with the Poisson formula

$$\lg\Delta(ib) = \frac{1}{\pi}\int \frac{b}{a^2+b^2}\lg\Delta(a)da,$$

leads at once to the bound

$$\int_1 \frac{\big|\lg\Delta(ib)\big|}{b^2} \le \frac{1}{\pi}\int \big|\lg\Delta(a)\big|da\int_1 \frac{db}{b\big(b^2+a^2\big)},$$

the second integral converging, since

$$\int_1 \frac{db}{b\big(b^2+a^2\big)} \sim \frac{\lg a}{a^2} \quad (a\uparrow\infty).$$

On the other hand, if $\int_1 \lg^-\Delta(ib)/b^2 > -\infty$, then $\int_1 \lg^-\Delta(a)\lg a/a^2$ is not smaller than a positive multiple of

$$\begin{aligned}\int_1 \lg^-\Delta(a)da\frac{1}{\pi}\int_1 \frac{db}{b\big(b^2+a^2\big)} &\ge \int_1 \frac{db}{b^2}\frac{1}{\pi}\int\frac{b}{a^2+b^2}\lg^-\Delta(a)da\\ &= \int_1 \frac{db}{b^2}\left(\lg\Delta(ib) - \frac{1}{\pi}\int\frac{b}{a^2+b^2}\lg^+\Delta(a)da\right)\\ &\ge \int_1 \lg^-\Delta(ib)/b^2 - \text{constant}\times\int\lg^+\Delta(a) > -\infty.\end{aligned}$$

Given $\int_1 \lg^-\Delta(ib)/b^2 > -\infty$, if $f \in Z^\bullet_\Delta$, then f is of genus 0 and

$$\int_1 \lg\max_{0\le\theta<2\pi}\big|f\big(Re^{i\theta}\big)\big|/R^2 < \infty;$$

indeed, since $\Delta(ib)$ is bounded $(b \ge 1)$,

$$\begin{aligned}\Delta^o(b) = \Delta^o(-b) &= \Delta(ib)/b^2 \quad (b > 1)\\ &= 1 \qquad\qquad (0 \le b \le 1)\end{aligned}$$

is a Hardy weight, and if $f \in Z^\bullet_\Delta$, then $|fh|$ is bounded $(b \ge 1)$, $|fh^*|$ is bounded $(b \le -1)$, and $\int|f(ib)|^2\Delta^o db < \infty$, i.e., $f(i\gamma) \in Z^\bullet_{\Delta^o}$. But then $\int_1 |\lg|f(ib)||/b^2 < \infty$, and combining this with $\int_1 |\lg|f(a)||/a^2 < \infty$ and an application of Carleman's theorem, one finds that the sum of the reciprocals of the moduli of the roots of f has to converge [2, 2.3.14], i.e., that the genus of f is 0. Because $f^+ = f + f^* \in Z^\bullet_\Delta$ satisfies

$$\int_1 \lg^+\big|f^+(ib)\big|b^2 < \infty \quad\text{and}\quad \int_1 \lg^+\big|f^+(a)\big|/a^2 < \infty,$$

it is of genus 0. It is also even, so $\int_1 \lg \max_{0\le\theta<2\pi} |f_+(Re^{i\theta})|/R^2 < \infty$ [2, 2.12.5]; the same holds for $f_- = f - f^* \in Z^\bullet_\Delta$ since γf_- is entire, even, and of genus 0, so

$$\int_1 \lg \max_{0\le\theta<2\pi} \left|f\left(Re^{i\theta}\right)\right|/R^2 < \infty,$$

as stated.

$\int \lg^- \Delta(ib)/b^2$ can diverge even though each $f \in Z^\bullet_\Delta$ is of genus 0, as can be seen from the Hardy weight Δ:

$$\begin{aligned} e^{a^{\frac12}}\Delta &= 1 && \text{on } [0,1)+[2,3)+\text{etc.} \\ &= e\big[-a/\lg^2(a+1)\big] && \text{on } [1,2)+[3,4)+\text{etc.} \end{aligned}$$

Δ is Hardy since $(a \lg^2(a+1))^{-1}$ is summable, while

$$\int_1 \lg^- \Delta \lg a/a^2 \le \sum_{\substack{d\,\text{odd} \\ d\ge1}} \int_d^{d+1} \big(a\lg(a+1)\big)^{-1} = -\infty,$$

so that $\int_1 \lg \Delta(ib)/b^2 = -\infty$. Given $f \in Z^\bullet_\Delta$,

$$B_1 = \|f\|^2_\Delta > \int_{2d}^{2d+1} |f|^2 e^{-2a^{\frac12}} > |f|^2 e^{-2a^{\frac12}}$$

at some point $2d \le a < 2d+1$ $(d \ge 0)$, so an application of the Duffin-Schaeffer theorem [2, 10.5.1] applied to $fe(-\gamma^{\frac12})$ on the half plane $a \ge 0$ supplies us with the bound $|f|e(-a^{\frac12}) < B_2$ on the half line $a \ge 0$. $|f|e(|a|^{\frac12}) < B_3$ on the left half line for similar reasons. Phragmén-Lindelöf applied to $fe(-(2\gamma)^{\frac12}e^{-i\pi/4})$ on the half plane $b \ge 0$ together with an analogous argument on $b > 0$ supplies the bound $|f| < B_4 e[(2R)^{\frac12}]$ on the whole plane, and it follows that f is of genus 0. □

13.6.5. Rational Weights.

$\dim Z^{+/-} = d < \infty$ if and only if Δ is a rational function of degree $2d$.

See, for example, Hida [6] from whom the following proof is adapted.

PROOF. $\dim Z^{+/-} = d < \infty$ implies $Z^{+/-} \ne Z$, so Δ is a Hardy weight and can be expressed as $|h|^2$ with h outer. Define the Fourier transform $\hat f(t) = (1/2\pi)\int e(-iat) f(a)da$ and note that if $i = h/h^*$ and if p is the projection upon H^{2-}, then $Z^{+/-}h = ipi^{-1}H^{2+}$ is of the same dimension d as

$$\begin{aligned} \big[pi^{-1}H^{2+}\big]^\wedge &= \text{span}\,\big[pi^{-1}e^{iat}h : t > 0\big]^\wedge \\ &= \text{span}\,\big[pe^{iat}h^* : t > 0\big]^\wedge \\ &= \text{span}\,\big[\big(e^{iat}h^*\big)^\wedge k(s) : t > 0\big] \\ &= \text{span}\,\big[\hat h(t-s)k(s) : t > 0\big], \end{aligned}$$

where $k(s)$ is the indicator of $s \le 0$. $[\pi^{-1}H^{2+}]$ has a unit perpendicular basis $f_1, \ldots, f_d$, and $\hat{h}(t-s) = c_1(t)f_1(s) + \cdots + c_d(t)f_d(s)$ $(s \le 0)$ with (real) coefficients $c_1, \ldots, c_d$. Choose $g_1, \ldots, g_d \in C^\infty(-\infty, 0]$ vanishing near $-\infty$ and 0 with $\det[\int_{-\infty}^0 f_i g_j] \ne 0$; then

$$\sum_{i \le d} c_i \int_{-\infty}^0 f_i g_j \, ds = \int_{-\infty}^0 \hat{h}(t-s) g_j \, ds \quad (j \le d,\ t > 0),$$

so that $c_1, \ldots, c_d \in C^\infty(0, \infty)$, and it follows that $\hat{h} \in C^\infty(0,\infty)$ also. Given $0 < t_0 < \cdots < t_d$, a dependence with non-trivial (real) coefficients must prevail between $\hat{h}(t_0 - s), \ldots, \hat{h}(t_d - s)$ $(s \le 0)$, and since $\hat{h} \in C^\infty$ $(0,\infty)$, it is possible to find a differential operator D with constant (real) coefficients and degree $\le d$ annihilating $\hat{h}$ on the half line $t > 0$. But this means that $\hat{h}$ is a sum of $\le d$ terms $t^a e^{bt} \overset{\cos}{\underset{\sin}{\cdot}} ct$, the permissible a filling out a series $0, 1, 2,$ etc., $b < 0$, and the trigonometrical factors either absent or both permissible. Δ rational of degree $\le 2d$ follows at once upon taking the inverse Fourier transform. On the other hand, if Δ is rational of degree $2d$, then it is a Hardy weight $|h|^2$ with h outer, h is also rational (of degree d), $\hat{h}$ is a sum of terms $t^a e^{bt} \overset{\cos}{\underset{\sin}{\cdot}} ct$, as above, the number of them coinciding with $\deg h$ and the trigonometrical factors either absent or present in pairs, and $\dim Z^{+/-} = d$ follows from $\dim \operatorname{span}[\hat{h}(t-s)i(s) : t > 0] = d$. □

Δ rational of degree $2d$ implies that

(a) *$h = p_0 p_1 / p_2$, p_0, p_1, p_2 being polynomials in $i\gamma$ with roots on the line in the case of p_0 and in the open half plane $b < 0$ in the case of p_1 and p_2, and of degrees d_0, d_1, d_2 $(= d)$ with $d_0 + d_1 < d_2$,*
(b) *$Z^\bullet = Z^{0+} = Z_\bullet =$ polynomials in $i\gamma$ of degree $< d_2 - d_1 - d_0$,*
(c) *$Z^- \cap Z^+ = 1/p_0 \times$ polynomials in $i\gamma$ of degree $< d_2 - d_1$,*
(d) *$Z^{+/-} = 1/p_0 p_1^* \times$ polynomials in $i\gamma$ of degree $< d_2 (= d)$,*

esp.,

(e) *$Z^\bullet = Z^- \cap Z^+$ if and only if h has no roots on $b = 0$,*
(f) *$Z^- \cap Z^+ = Z^{+/-}$ if and only if h has no roots in $b < 0$,*
(g) *$Z^{+/-} = Z^- \cap Z^+ = Z^\bullet - Z^{0+} = Z_\bullet$ if and only if h has no roots at all.*

Proof of (a). Obvious. □

Proof of (b). $f \in Z^- \cap Z^\bullet_\Delta$ implies $\int |f|^2/(1+a^2)^d < \infty$, and a simple application of Phragmén-Lindelöf implies that f is a polynomial; the bound on its degree is obvious. □

Proof of (c). $f \in Z^- \cap Z^+$ implies $p_0 f \in Z^-_{\Delta^o} \cap Z^+_{\Delta^o}$ $(\Delta^o = |p_1/p_2|^2)$, and since Δ^o is bounded from 0 on bounded intervals, $p_0 f \in Z^\bullet_{\Delta^o}$ (Sect. 13.6.3). But then $p_0 f$ has to be a polynomial as in the proof of (b) above, the bound on the degree of this polynomial is obvious, and the rest of the proof is a routine application of $Z^- \cap Z^+ h = iH^{2-} \cap H^{2+}$ $(i = h/h^*)$. □

Proof of (d). Use the formula $Z^{+/-}h = iH^{2-} \cap (1/j_1)H^{2+}$ $(i = j_2/j_1)$ of Sect. 13.3 and match dimensions. □

Proof of (e), (f), (g). Obvious. □

13.7. A Condition That $Z^{+/-} = Z^\bullet$ (Δ Hardy)

Given a Hardy weight $\Delta = |h|^2$ (h outer), $Z^{+/-} = Z^\bullet$ if and only if h is the reciprocal of an entire function of minimal exponential type.

PROOF. Suppose h is the reciprocal of an entire function f of minimal exponential type; then $h = 1/f$ implies $\int_{|a|<d} \Delta^{-1} < \infty$ ($d < \infty$), so $Z^\bullet = Z^- \cap Z^+$ (Sect. 13.6.3), and to complete the proof of $Z^{+/-} = Z^\bullet$, it is enough to check that $i = h/h^* = f^*/f$ is an inner function (Sect. 13.3(c)). But $1/f = h$ being outer, it is root-free ($b \geq 0$), and

$$\lg |f| = \frac{1}{\pi} \int \frac{b}{(c-a)^2 + b^2} \lg |f| dc \quad (b > 0),$$

while f^*, as an entire function of minimal exponential type with $\int \lg |f^*|/(1+a^2) < \infty$, satisfies

$$\lg \left|f^*\right| \leq \frac{1}{\pi} \int \frac{b}{(c-a)^2 + b^2} \lg \left|f^*\right| dc \quad (b > 0),$$

so f^*/f is regular ($b > 0$) with

$$\left|f^*/f\right| = 1 \qquad (b = 0)$$

$$\left|f^*/f\right| \leq e\left[\frac{1}{\pi} \int \frac{b}{(c-a)^2 + b^2} \lg \left|f^*/f\right| dc\right] = 1 \quad (b > 0),$$

i.e., f^*/f is inner.

On the other hand, if $Z^{+/-} = Z^\bullet$ and if p is the projection upon H^{2-}, then the projection of $e(iat)$ ($t > 0$) upon Z^-:

$$\begin{aligned}
& h^{-1} i p i^{-1} e^{iat} h \quad \left(i = h/h^*\right) \\
&= h^{-1} i p e^{iat} h^* \\
&= h^{-1} i \frac{1}{2\pi} \int_{-\infty}^{0} e^{ias} ds \int e^{-ics} e^{ict} h^* \, dc \\
&= h^{-1} i \frac{1}{2\pi} \int_{-\infty}^{0} e^{ias} ds \left(\int e^{-ic(t-s)} h \, dc \right)^* \\
&= h^{-1} i \frac{1}{2\pi} \int_{-\infty}^{0} e^{ias} ds \; \hat{h}(t-s) \quad \left(\hat{h} = \frac{1}{2\pi} \int e^{-iat} h \, dt = \hat{h}^*\right),
\end{aligned}$$

belongs to $Z^\bullet$, and since its conjugate also belongs to $Z^\bullet$,

$$\frac{e^{-iat}}{2\pi\, h} \int_t^\infty e^{ias} \hat{h} \, ds \equiv f_t(a) \in Z^\bullet_\Delta \quad (t > 0).$$

Choose $t > 0$ belonging to the Lebesgue set of $\hat{h}$ so that $\lim_{\delta \downarrow 0} \delta^{-1} \int_t^{t+\delta} \hat{h} \, ds = \hat{h}(t) \neq 0$ and $\delta^{-1} \int_t^{t+\delta} |\hat{h}| ds$ is bounded as $\delta \downarrow 0$.

$$\begin{aligned}
& 2\pi \left\| f_{t+\delta} - f_t \right\|_{\Delta+} \\
&\leq \left\| \left(e^{-ia(t+\delta)} - e^{-iat}\right) \int_{t+\delta}^\infty e^{ias} \hat{h} \, ds \right\|_{1/(1+a^2)} + \left\| e^{-iat} \int_t^{t+\delta} e^{ias} \hat{h} \, ds \right\|_{1/(1+a^2)}
\end{aligned}$$

$$= \left\| \left(e^{ia\delta} - 1\right) \int_{t+\delta}^{\infty} e^{ias} \hat{h}\, ds \right\|_{1/(1+a^2)} + \left\| \int_t^{t+\delta} e^{ias} \hat{h}\, ds \right\|_{1/(1+a^2)}$$

$$\le \text{constant} \times \delta \left\| \int_{t+\delta}^{\infty} e^{ias} \hat{h}\, ds \right\|_1 + \int_t^{t+\delta} |\hat{h}| ds \left(\int \frac{da}{1+a^2} \right)^{1/2}$$

$$< \text{constant} \times \delta,$$

and it follows, thanks to the $\overline{\lim}_{R\uparrow\infty} R^{-1} \max_{0\le\theta<2\pi} \lg \sigma^\bullet(Re^{i\theta}) \le 0$, that $\delta^{-1}(f_{t+\delta} - f_t)$ can be made to tend on the whole plane to some $f^\bullet \in Z^\bullet_{\Delta+}$ as $\delta \downarrow 0$ via some series $\delta_1 > \delta_2 >$ etc. Going back to the definition of $f_t \equiv f$, it develops that

$$-\hat{h}(t)/2\pi\, h(a) = \left[iaf + f^\bullet\right] \in Z^\bullet_{\Delta+},$$

and the proof is complete. □

13.8. A Condition That $Z^{0+} = Z$

$Z^{0+} = Z$ if $\int_1 da/a^2 \lg \int_a \Delta e^{-2B} = -\infty$ with $0 \le B \in\uparrow$, $\int_1 e^{-2B} < \infty$, and $\int_1 B/a^2 < \infty$. Δ has to be non-Hardy for this integral to diverge since

$$\begin{aligned}
\int_1 \frac{da}{a^2} \lg \int_a \Delta e^{-2B} &\ge \int_1 \frac{da}{a^2} \lg \left[a^3 \int_a \frac{\Delta e^{-2B}}{c^3} \right] \\
&= \int_1 \frac{\lg(2a)}{a^2} da + \int_1 \frac{da}{a^2} \lg \left[\frac{a^2}{2} \int_a \frac{\Delta e^{-2B}}{c^3} \right] \\
&\ge \int_1 \frac{\lg(2a)}{a^2} da + \int_1 \frac{da}{a^2} \left[\frac{a^2}{2} \int_a \frac{\lg \Delta e^{-2B}}{c^3} \right] \\
&\ge \int_1 \frac{\lg(2a)}{a^2} da + \frac{1}{2} \int_1 da \int_a \frac{\lg^- \Delta}{c^3} - \int_1 da \int_a \frac{B}{c^3} \\
&> \text{constant} + \frac{1}{2} \int_1 \lg^- \Delta/a^2;
\end{aligned}$$

also, if $\Delta \in\downarrow$, then $\int_1 da\, a^{-2} \lg \int_a \Delta e^{-2B}$ and $\int_1 \lg^- \Delta/a^2$ converge or diverge together, since under this condition,

$$\begin{aligned}
\int_1 \frac{da}{a^2} \lg \int_a \Delta e^{-2B} &\le \int_1 \frac{da}{a^2} \left(\lg \Delta + \lg \int_a e^{-2B} \right) \\
&\le \int_1 \lg \Delta/a^2 + \lg \int_1 \frac{da}{a^2} \int_a e^{-2B} \\
&< \int_1 \lg \Delta/a^2 + \text{constant};
\end{aligned}$$

esp., if $\Delta \in\downarrow$, then $Z^{0+} = Z$ if and only if $\int_1 \lg \Delta/a^2 = -\infty$.

As to the proof of the original statement, if $\int_1 da\, a^{-2} \lg \int_a \Delta e^{-2B} = -\infty$ with B as above and if $Z^{0+} \ne Z$, then $Z^{|t|<\delta} \ne Z$ for small δ, and it is possible to find $f \in Z$ with $\int f e(iat) \Delta\, da = 0$ ($|t| < \delta$). But

$$\int_a |f| \Delta e^{-B} \le \|f\|_\Delta \left(\int_a \Delta e^{-2B} \right)^{\frac{1}{2}} \quad (a \ge 1),$$

so that

$$\int_1 \frac{da}{a^2} \lg \int_a |f|\Delta e^{-B} = -\infty,$$

and according to Levinson [8, page 81], this cannot happen unless $f = 0$.

13.9. Discussion of $Z_\bullet$

I. O. Hačatrjan's contribution to the Bernstein problem [5] is adapted as follows.

Consider the span $Z_\bullet = Z_{\bullet\Delta}$ of (real) polynomials p of $i\gamma$ belonging to Z, let $\int a^{2d}\Delta < \infty$ ($d \geq 1$), let $\sigma_\bullet(\gamma)$ be the least upper bound of $|p(\gamma)|$ for $p \in Z_{\bullet\Delta^+}$ with $\|p\|_{\Delta^+} \leq 1$, and let us prove that the following alternative holds:

either

$$\sigma_\bullet \equiv \infty \quad (b \neq 0),$$

$$\sup \int \frac{\lg^+ |p|}{1+a^2} = \int \frac{\lg \sigma_\bullet}{1+a^2} = \infty, \quad \textit{for } p \in Z_{\bullet\Delta^+} \textit{ with } \|p\|_{\Delta^+} \leq 1,$$

and $Z_\bullet = Z$,

or $\lg \sigma_\bullet$ *is a continuous, non-negative, subharmonic function*,

$$\int \frac{\lg \sigma_\bullet}{1+a^2} < \infty,$$

$$\lg \sigma_\bullet(\gamma) \leq \frac{1}{\pi} \int \frac{b}{(c-a)^2+b^2} \lg \sigma_\bullet(c)dc \quad (\gamma = a+ib,\ b > 0),$$

$$\overline{\lim_{R\uparrow\infty}}\ R^{-1} \max_{0\leq\theta<2\pi} \lg \sigma_\bullet\big(Re^{i\theta}\big) \leq 0,$$

and $Z_\bullet \neq Z$;

in the second case, $Z_\bullet \subset Z^\bullet$, *the two coinciding if and only if* $\sigma_\bullet \equiv \sigma^\bullet$ ($b \neq 0$).

Proof. The proof is identical to the discussion of $Z^\bullet$ (Sect. 13.4), excepting the final statement to which attention is now directed.

Given $\sigma_\bullet = \sigma^\bullet < \infty$ while $Z_\bullet \neq Z^\bullet$, then it would be possible to find $f \in Z^\bullet_\Delta$, $f \not\equiv 0$, with $\int f^* a^d \Delta = 0$ ($d \geq 0$). This implies

$$\int f^* \frac{p - p(\beta)}{c - \beta}\Delta = 0 \quad (\beta = a+ib,\ b \neq 0),$$

and it follows that

$$\left|\int \frac{f^*\Delta}{c-\beta}\right| = \left|\int \frac{f^*\Delta p}{(c-\beta)p(\beta)}\right| \leq \left\|\frac{c-i}{c-\beta} f\right\|_\Delta \big|p(\beta)\big|^{-1} \|p\|_{\Delta^+} \quad (\beta = a+ib,\ b \neq 0),$$

esp.,

$$\left|\int \frac{f^*\Delta}{c-ib}\right| = o\big(\sigma_\bullet(ib)^{-1}\big) \quad \text{as } |b| \uparrow \infty.$$

Chose $g \in Z^\bullet_{\Delta^+}$; then $\int f^* g\Delta(c-\beta)^{-1}$ tends to 0 at both ends of $a = 0$ so that

$$\hat{g} \equiv \int f^* \frac{g - g(\beta)}{c-\beta}\Delta$$

satisfies

$$\begin{aligned}\big|\hat g(ib)\big| &\le o(1) + \big|g(ib)\big|\left|\int \frac{f^*\Delta}{c-ib}\right| \\ &= o(1) + \big|g(ib)\big|o\big(\sigma_\bullet(ib)^{-1}\big) \\ &= o(1) + \big|g(ib)\big|o\big(\sigma^\bullet(ib)^{-1}\big) \\ &= o(1) \quad \big(|b|\uparrow\infty\big),\end{aligned}$$

and since $\hat g$ is entire of minimal exponential type, Phragmén-Lindelöf implies $\hat g \equiv 0$. But then $\int f^* g(c-\beta)^{-1}\Delta = g(\beta)\int f^*\Delta(c-\beta)^{-1} = 0$ if β is a root of $g \in Z^\bullet_{\Delta+}$ $(b \ne 0)$, so taking $g = (\gamma - i)f \in Z^\bullet_{\Delta+}$ and $\beta = i$, $\|f\|^2 = \int f^* g(c-i)^{-1}\Delta = 0$, and the proof is complete. □

13.10.

13.10.1. Special Case ($1/\Delta = 1 + c_1 a^2 +$ **etc.**)**.** Hačatrjan [5] states the analogue for the Bernstein problem of the following result:

> If $1/\Delta = 1 + c_1a^2 + c_2a^4 + etc.$ $(c_1, c_2, etc. \ge 0)$ and if $\int a^{2d}\Delta < \infty$ $(d \ge 0)$, then either Δ is non-Hardy and $Z_\bullet = Z$ or Δ is Hardy and $Z_\bullet = Z^\bullet$.

Proof. $p_d = \sum_{n\le d} c_n\gamma^{2n}$ can be expressed as $|q_d|^2$, q_d being a polynomial in $i\gamma$ of degree d with no roots in the closed half plane $b \ge 0$. As $d \uparrow \infty$,

$$\lg\big|q_d(i)\big|^2 = \frac{1}{\pi}\int \frac{\lg\big|q_d\big|^2}{1+c^2} \uparrow \frac{1}{\pi}\int\frac{\lg\Delta^{-1}}{1+c^2}$$

while

$$\big\|q_d\big\|^2_{\Delta+} = \frac{1}{\pi}\int\frac{p_d\Delta}{1+c^2} \uparrow 1,$$

so *either* $\int \lg\Delta/(1+c^2) = -\infty$, $\sigma_\bullet(i) = \infty$, and $Z_\bullet = Z$ *or* Δ is Hardy ($\Delta = |h|^2$ with h outer). Because $|q_d|^2 = p_d \le \Delta^{-1}$, an application of Lebesgue's dominated convergence test shows that $h^{-1} = \lim_{d\uparrow\infty} q_d$ $(b \ge 0)$ in the second case.

Now in the second case, if $f \in Z^\bullet_\Delta$ is perpendicular to $Z_{\bullet\Delta}$, if $g \in Z^\bullet_{\Delta+}$, and if

$$\hat g(\beta) \equiv \int f^*\frac{g - g(\beta)}{c-\beta}\Delta$$

as before, then

$$\begin{aligned}\left|q_d(ib)\int\frac{f^*\Delta\,dc}{c-ib}\right| &= \left|\int\frac{f^*q_d\Delta\,dc}{c-ib}\right| \le \|f\|_\Delta\left(\int\frac{\big|q_d\big|^2\Delta\,dc}{c^2+b^2}\right)^{1/2} \\ &\le \|f\|_\Delta\left(\int\frac{dc}{c^2+b^2}\right)^{1/2} = \|f\|_\Delta(\pi/b)^{1/2},\end{aligned}$$

and so

$$\begin{aligned}\big|\hat g(ib)\big| &\le \left|\int\frac{f^*g\Delta\,dc}{c-ib}\right| + \big|g(ib)\big|\left|\int\frac{f^*\Delta\,dc}{c-ib}\right| \\ &\le \|f\|_\Delta\left(\int\frac{c^2+1}{c^2+b^2}|g|^2\Delta^+dc\right)^{1/2} + \inf_{d>0}\left|\frac{g(ib)}{q_d(ib)}\right|\left|\int\frac{f^*q_d\Delta\,dc}{c-ib}\right| \\ &= o(1) + \big|gh(ib)\big|\|f\|_\Delta(\pi/b)^{1/2}.\end{aligned}$$

Since the Poisson integral applies as an inequality to $\lg|(\gamma+i)^{-1}gh|$ and as an equality to $\lg|\gamma+i|$,

$$\big|gh(ib)\big|^2 \le e\left[\frac{1}{\pi}\int \frac{b}{b^2+c^2}\lg|gh|^2\right] \le \frac{1}{\pi}\int \frac{b\big(c^2+1\big)}{b^2+c^2}|g|^2\Delta^+ = o(b),$$

and so $\lim_{b\uparrow\infty}|\hat{g}(ib)| = 0$. Repeating the proof as $b\downarrow -\infty$ justifies $\lim_{b\downarrow-\infty}|\hat{g}(ib)| = 0$, and now $\hat{g}=f=0$ follows as in Sect. 13.9.

A special case of the above is the fact that if h is the reciprocal of an entire function and if the roots of h^{-1} fall in the sector $-3\pi/4 \le \theta \le -\pi/4$, then $Z_\bullet = Z^\bullet$; obvious improvements can be made, but $Z_\bullet = Z^\bullet$ does not hold without some condition on the roots of h^{-1} as the example of Sect. 13.11 proves. □

As a second application, it will be proved that

$$Z_\bullet = Z^\bullet \textit{ in case } \Delta(a) = e\big(-2|a|^p\big)\ (0<p<1);$$

similar but more complicated cases can be treated in the same fashion (see below).

Proof. It suffices to construct a weight $\Delta^o = (1+c_1a^2+\text{etc.})^{-1}$ with non-negative coefficients, positive multiples of which bound Δ above and below. Define $\#(R) = [\theta R^p + 1/2]$ with an adjustable $\theta>0$, the bracket denoting the integral part, and let

$$\begin{aligned} -\lg\Delta^o(a) &= \int_0 \lg\left(1+\frac{a^2}{R^2}\right)d\#(R) = 2a^2\int_0 \frac{\#(R)dR}{\big(a^2+R^2\big)R} \\ &= \frac{2a^2}{p}\int_0 \frac{\big[\theta c+1/2\big]dc}{\big(a^2+c^{2/p}\big)c} \quad \big(c=R^p\big) \\ &= J_1+J_2 \end{aligned}$$

with

$$J_1 = \frac{2a^2}{p}\int_0 \frac{\big[\theta c+1/2\big]+1/2-\big(\theta c+1/2\big)}{\big(a^2+c^{2/p}\big)c}dc$$

and

$$J_2 = \frac{2a^2\theta}{p}\int_0 \big(a^2+c^{2/p}\big)^{-1}dc.$$

In J_2, substitute $c=|a|^pt$ and let $\theta^{-1} = (2/p)\int_0(1+t^{2/p})^{-1}$, obtaining $J_2 = 2|a|^p$. Coming to J_1, note that the numerator under the integral sign is periodic and that its average over a period is 0, so that J_1 tends to a constant as $|a|\uparrow\infty$. J_1 is then bounded, so Δ is bounded above and below by positive multiples of Δ^o, and the proof is complete. □

$Z_\bullet = Z^\bullet$ also holds in the more general case of a Hardy weight.

$$\Delta = \Delta(0)e\left(-\int_0^{|a|}\frac{\omega(c)}{c}dc\right)$$

provided $\omega\in\uparrow$ and $\omega(c)\lg c$ tends to ∞ as $c\uparrow\infty$.

Proof. Under the above condition it is possible, according to Y. Domar [4], to find a reciprocal weight $1/\Delta^o = 1+c_1a^2+$etc. with non-negative coefficients such that Δ is bounded above by a positive multiple of Δ^o and below by a positive multiple of $\Delta^\theta = \Delta^o(\theta a)$ with a constant depending upon $\theta>1$ alone. Because

$$Z_{\bullet\Delta^\theta} = Z^\bullet_{\Delta^\theta} \supset Z^\bullet_\Delta,$$

each $f \in Z_\Delta^\bullet$ can be approximated in Z_{Δ^θ} by a polynomial p so as to have

$$\int \left|f(a/\theta) - p(a/\theta)\right|^2 \Delta \le \text{constant} \times \theta \|f - p\|_{\Delta^\theta}^2$$

small, and to complete the proof it suffices to check that $f_\theta(a) = f(a/\theta)$ tends to f in Z_Δ as $\theta \downarrow 1$. But this is obvious from the fact that

$$\|f_\theta\|_\Delta^2 = \theta \int |f|^2 \Delta(\theta a) \sim \|f\|_\Delta^2 \quad (\theta \downarrow 1)$$

while f_θ tends to f pointwise under a local bound.

By the same method it is easy to prove that if Δ has the above form with $\omega \in\uparrow$ and $\int_1 \omega/c^2 = \infty$ (non-Hardy case), then $Z_\bullet = Z$. □

Domar's paper was brought to our notice through the kindness of Professor L. Carleson.

13.10.2. A Special Case ($\Delta = e^{-2|a|^{\frac{1}{2}}}$)**.** $\Delta = \exp(-2|a|^{\frac{1}{2}})$ falls under the discussion of Sect. 13.10.1, but it is entertaining to check $Z^\bullet = Z_\bullet$ from scratch using the following special proof.

$\Delta = |h|^2$ with

$$h = e\Big[-(2\gamma)^{\frac{1}{2}} e^{-i\pi/4}\Big] = \int_0^\infty e^{i\gamma t} \frac{e^{-1/2t}}{\left(2\pi t^3\right)^{\frac{1}{2}}} dt,$$

and h is outer since

$$\lg\left|h(i)\right| = -2^{\frac{1}{2}} = \frac{1}{\pi}\int \frac{\lg|h|}{1+a^2}$$

(see [7, page 62]).

Given $f \in Z_\Delta^\bullet$, a simple application of Phragmén-Lindelöf supplies us with the bound

$$f(\gamma) \le Be\Big[\big(\sqrt{2}+\delta\big)\sqrt{R}\Big] \quad (\delta > 0);$$

hence, $|f(\gamma^2)| \le Be[(\sqrt{2}+\delta)R]$, and according to Pólya's theorem [2, 5.3.5],

$$f(\gamma^2) = \int e^{\gamma w} g = \int e^{-\gamma w} g = \int \cosh(\gamma w) g\, dw,$$

i.e.,

$$f(\gamma) = \int \cosh\left(\sqrt{\gamma} w\right) g\, dw,$$

the integral being extended over $|w| = 2^{\frac{1}{2}} + \delta$ and g being regular outside $|w| = 2^{\frac{1}{2}}$ and at ∞. Accordingly, if $f \in Z^\bullet$ is perpendicular to $Z_\bullet$, then

$$\begin{aligned}
0 &= \int f a^d \Delta\, da = \int g\, dw \int \cosh\left(\sqrt{a} w\right) a^d \Delta\, da \\
&= \int g \Big[\int_0^\infty \cosh\left(\sqrt{a} w\right) a^d e^{-2a^{\frac{1}{2}}} + \int_0^\infty \cos\left(\sqrt{a} w\right)(-a)^d e^{-2a^{\frac{1}{2}}}\Big] \\
&= \int g D^{2d} \Big[\int_0^\infty \cosh\left(\sqrt{a} w\right) e^{-2a^{\frac{1}{2}}} + \int_0^\infty \cos\left(\sqrt{a} w\right) e^{-2a^{\frac{1}{2}}}\Big] \\
&= 2\int g D^{2d+1}\Big[\int_0^\infty \sinh(aw) e^{-2a} + \int_0^\infty \sin(aw) e^{-2a}\Big]
\end{aligned}$$

$$= \int gD^{2d+1}\left[\frac{1}{2-w} - \frac{1}{2+w} + \frac{1}{2i+w} - \frac{1}{2i-w}\right]$$

$$= \int gD^{2d+1}\frac{16w}{16-w^4}.$$

Because $\int e^{\gamma w}g = f(\gamma^2)$ is an even function, $\int gw^d = 0$ (d odd) and since $w/(16-w^4)$ is a sum of powers w^d ($d \equiv 1(4)$), it follows that

$$0 = \int gD^d\left[\frac{1}{2-w} - \frac{1}{2+w} + \frac{1}{2i+w} - \frac{1}{2i-w}\right] \quad (d \geq 0),$$

and so

$$0 = \int g\left[\frac{1}{2-w+t} - \frac{1}{2+w-t} + \frac{1}{2i+w-t} - \frac{1}{2i-w+t}\right]dw$$

$$= g(t+2) + g(t-2) - g(t-2i) - g(t+2i)$$

for small $|t|$.

Draw four circles, each of radius $2^{\frac{1}{2}}$, having centers at 2, $2i$, -2 and $-2i$, respectively. The circles with centers at 2 and $2i$ are tangent at A, which is $1+i$. The circles with centers at 2 and $-2i$ are tangent at B, which is $1-i$. The point C is $-3+i$ and lies on the circle with center at -2. Using this diagram depicting four discs on each of which just one of the summands can be singular, it follows that $g(t-2) = -g(t+2) + g(t-2i) + g(t+2i)$ can be singular only at A and B since the second member is non-singular on the rest of $|t-2| \leq 2^{\frac{1}{2}}$. Now if $g(t-2)$ is singular at A, then $g(t+2)$ is singular at $C = A-4$ and that is impossible, so $g(t-2)$ cannot be singular at A, nor, for similar reasons, at B. But then g is entire, and by Cauchy's theorem, $f(\gamma^2) = \int \cosh(\gamma w)g = 0$, completing the proof.

$Z^- = Z^{+/-} \neq Z^- \cap Z^+ = Z^\bullet = Z^{0+} = Z_\bullet$ can be proved at little extra cost. $Z^- \cap Z^+ = Z^\bullet$ is obvious from Sect. 13.6, and so it suffices to prove that $i = h/h^* = e[2i\,\mathrm{sgn}(a)|a|^{\frac{1}{2}}]$ is not a ratio j_2/j_1 of inner functions (Sect. 13.3). But in the opposite case, $if \in H^{2+}$ ($f = j_1 h$), so that

$$0 = \frac{1}{2}\int e^{-iat} if\, da \quad (t < 0)$$

$$= \mathrm{Re}\left[\int_0^\infty e^{-iat}e^{2ia^{\frac{1}{2}}} f\, da\right] = \mathrm{Im}\left[\int_0^\infty e^{bt}e^{(2b)^{\frac{1}{2}}(i-1)} f(ib)db\right],$$

since

$$\left|\int_0^{\pi/2} e^{-iRe^{i\theta}t}e^{2iR^{\frac{1}{2}}e^{i\theta/2}} f\big(Re^{i\theta}\big)Re^{i\theta}i\, d\theta\right|$$

$$\leq \int_0^{\pi/2} e^{R\sin\theta t}e^{-2R^{\frac{1}{2}}\sin\theta/2}e^{-(2R)^{\frac{1}{2}}\cos(\theta/2-\pi/4)}R\, d\theta$$

tends to 0 as $R \uparrow \infty$. Because $f = f^*$ ($a = 0$),

$$0 = \mathrm{Im}\left[e^{(2b)^{\frac{1}{2}}(i-1)} f(ib)\right] = \sin(2b)^{\frac{1}{2}}e^{-(2b)^{\frac{1}{2}}} f(ib) \quad (b \geq 0),$$

and that is absurd.

An entertaining illustration of the delicacy of the projection $Z^{+/-}$ is thus obtained. $Z^{+/-} \neq Z^\bullet$ as was just proved, so naturally the condition that $Z^{+/-} = Z^\bullet$, to wit, that $\Delta = |f|^{-2}$ with f entire of minimal exponential type, does not hold. But as proved in Sect. 13.10.1, $e(-2|a|^{\frac{1}{2}})$ is bounded above and below by positive multiples of such a weight.

13.11. An Example (Δ Hardy, $\dim Z_\bullet = \infty$, $Z^\bullet = Z^{0+} \neq Z_\bullet$)

A weight Δ exists with the following properties:

(a) $\int \lg \Delta/(1+a^2) > -\infty$, i.e., Δ *is a Hardy weight*,
(b) $\int a^{2d}\Delta < \infty$ $(d \geq 0)$, i.e., $\dim Z_\bullet = \infty$,
(c) $Z_\bullet \neq Z^\bullet = Z^{0+}$.

Consider for the proof

$$\delta_n = 1/\sinh \pi n, \qquad \gamma_{+n} = n^2 - i\delta_n, \qquad \gamma_{-n} = -n^2 - i\delta_n,$$

$$1/h(\gamma) = \prod_{|n|>0} \left(1 - \frac{\gamma}{\gamma_n}\right), \qquad \Delta = |h|^2,$$

$$f = \frac{\sin \pi\sqrt{\gamma}\sinh \pi\sqrt{\gamma}}{\pi^2\gamma} = \prod_{n\geq 1}\left(1 - \frac{\gamma^2}{n^4}\right), \text{ and } g = f/(1-\gamma^2) = \prod_{n\geq 2}\left(1-\frac{\gamma^2}{n^4}\right),$$

and break up the proof into a series of simple lemmas.

(a) $0 < B_1 < |fh| < B_2$ *if* $|\gamma \pm n^2| \geq \frac{1}{2}$ $(n \geq 1)$, *while* $0 < B_3 < |fh||(\gamma - \gamma_{\pm n})/(\gamma \mp n^2)| < B_4$ *if* $|\gamma \pm n^2| < \frac{1}{2}$; *a similar appraisal holds with* h^* *in place of* h.
(b) $g \in Z^\bullet_\Delta$.
(c) Δ *is a Hardy weight and* $\int a^{2d}\Delta < \infty$ $(d \geq 0)$.
(d) $g \notin Z_{\bullet\Delta}$.

Proof of (a). Obvious. □

Proof of (b). g is entire of minimal exponential type with $g^*(-a) = g(a)$, so it is enough to check that $\|g\|_\Delta < \infty$. But (a) supplies us with the bound $|fh| < B_5$, so $|gh| < B_5/(1-a^2)$, and since $|gh| < B_6$ for small $|a|$, $\|g\|_\Delta < \infty$. □

Proof of (c). h^{-1} is entire and free of roots in the closed half-plane $b \geq 0$, and $\Delta(a + ib) \in\downarrow$ as a function of $b > 0$, so it suffices to check

$$\int_8 a^{2d}\Delta \leq \sum_{n=3}^{\infty} \int_{n^2-n+\frac{1}{4}}^{n^2+n+\frac{1}{4}} a^{2d}\Delta < \infty \quad (d \geq 0).$$

But on $|a - n^2| < \frac{1}{2}$,

$$\Delta < B_4^2|f|^{-2}\frac{\left(a-n^2\right)^2}{\left(a-n^2\right)^2+\delta_n^2}, \qquad \frac{|a-n^2|}{|f|} = -\frac{\pi^2 a}{\sinh \pi\sqrt{a}}\left|\frac{a-n^2}{\sin \pi\sqrt{a}}\right| < B_7 n^3 e^{-\pi n},$$

and hence

$$a^{2d}\Delta < B_8 \frac{n^{2d+6}e^{-2\pi n}}{\left(a-n^2\right)^2 + \delta_n^2}$$

on this range, while on the rest of $n^2 - n + \frac{1}{4} \leq a < n^2 + n + \frac{1}{4}$,

$$a^{2d}\Delta < (n+1)^{2d}B_2^2|f|^{-2} < B_9 n^{2d+6}e^{-2\pi n},$$

so that

$$\int_{n^2-n+\frac{1}{4}}^{n^2+n+\frac{1}{4}} a^{2d}\Delta < B_{10}\left[n^{2d+6}e^{-2\pi n}\int \frac{da}{a^2+\delta_n^2} + n^{2d+7}e^{-2\pi n}\right] < B_{11}n^{2d+7}e^{-\pi n},$$

which is the general term of a convergent sum. □

PROOF OF (d). $g \in Z_{\bullet\Delta}$ implies the existence of polynomials $p_\delta \in Z_{\bullet\Delta}$ with $\|g - p_\delta\|_\Delta < \delta$. p_δ can be supposed even since g is such; also, as $\delta \downarrow 0$, p_δ tends to g on the whole plane under a local bound ($\sigma_\bullet < \infty$), so that $p_{0+}(0) = g(0) = 1$, and according to Hurwitz's theorem, the roots of p_δ tend to the roots $\pm 2^2$, $\pm 3^2$, etc. of g. Rotate the roots of p_δ onto the line $b = 0$ and put its bottom coefficient $= 1$, defining a new polynomial q_δ with $|q_\delta| \le |p_\delta / p_\delta(0)|$ $(b = 0)$ and $\|q_\delta\|_\Delta \le \|p_\delta\|_\Delta / |p_\delta(0)|$ bounded as $\delta \downarrow 0$; it is this boundedness of $\|q_\delta\|_\Delta$ that leads to a contradiction. □

Evaluate $\int q_\delta^2 h^*$, integrating about the semicircle $Re^{i\theta}(-\pi/2 \le \theta \le \pi/2)$ and then down along the segment joining iR to $-iR$ with R half an odd integer. Bound the integral on the arc with the aid of $|fh^*| < B_2$ and let $R \uparrow \infty$, obtaining

$$\frac{1}{2\pi} \int q_\delta^2 h^*(ib)db = \sum_{n=1}^{\infty} \frac{q_\delta^2(\gamma_n^*)}{(1/h^*)'(\gamma_n^*)} \equiv Q_\delta.$$

Because $h^*(ib) > 0$ and $|q_\delta(ib)| \ge |p_\delta(ib)/p_\delta(0)|$, an application of Fatou's lemma combined with $|fh^*| > B_1 > 0$ justifies the under-estimate:

$$Q_{0+} \ge \frac{1}{2\pi} \int g^2 h^*(ib) > B_{13} \int_1 f(ib)/b^4 > B_{14} \int_1 e^{\pi(2b)^{\frac{1}{2}}}/b^5 = \infty.$$

Q_δ is now estimated again with the contradictory result that it is bounded as $\delta \downarrow 0$.

$\int (q_\delta h)^2 = 0$, the integral being taken around the arc $Re^{i\theta}(0 \le \theta \le \pi/2)$, down the segment joining iR to 0, and thence out along the segment joining 0 to R with R half an odd integer. Bound the integral along the arc as before and let $R \uparrow \infty$, obtaining

$$\int_0^\infty (q_\delta h)^2(ib) = -i \int_0^\infty (q_\delta h)^2(a) \le \|q_\delta\|_\Delta^2 < B_{15},$$

the first integrand being positive.

$\int (q_\delta h)^2 (\gamma - \gamma_n)/(\gamma - \gamma_n^*)$ is now evaluated along the same curve, giving

$$-\int_0^\infty (q_\delta h)^2(ib) \frac{ib - \gamma_n}{ib - \gamma_n^*} - i \int_0^\infty (q_\delta h)^2(a) \frac{a - \gamma_n}{a - \gamma_n^*} = 4\pi i \delta_n (q_\delta h)^2(\gamma_n^*);$$

this supplies the bound

$$4\pi\delta_n |q_\delta h(\gamma_n^*)|^2 \le \int_0^\infty (q_\delta h)^2(ib) + \int_0^\infty |q_\delta h|^2(a) < 2B_{15} = B_{16},$$

and it follows that

$$Q_{0+} < B_{16} \sum_{n=1}^{\infty} e^{\pi n} \left| \frac{h^{-2}(\gamma_n^*)}{(1/h^*)'(\gamma_n^*)} \right|.$$

But, since

$$|(\gamma - \gamma_n^*)h^*| < B_4 \frac{|\gamma - n^2|}{|f|} \quad \text{near } \gamma = \gamma_n^*,$$

$$|(1/h^*)'(\gamma_n^*)|^{-1} \le 2B_4 e^{-\pi n}/|f(\gamma_n^*)|,$$

while

$$|h(\gamma_n^*)|^{-2} < 4B_3^{-2}|f(\gamma_n^*)|^2,$$

and combining these bounds leads at once to the desired contradiction:

$$Q_{0+} < B_{17} \sum_{n=1}^{\infty} |f(\gamma_n^*)| < B_{18} \sum_{n=1}^{\infty} n^{-3} < \infty.$$

$Z^\bullet$ is sometimes closed under $f \to {}'f = if'$, but this can fail; indeed in the above case,

$$\Delta > \frac{B_3^2}{|f|^2} \frac{(a-n^2)^2}{(a-n^2)^2+\delta_n^2} > B_{19} \frac{n^6\delta_n^2}{(a-n^2)^2+\delta_n^2} \quad \left(|a-n^2| < \sqrt{\delta_n}\right),$$

while on the same range,

$$|'g| > B_{20} e^{\pi n} n^{-7}$$

so that $\|'g\|_\Delta = \infty$ because

$$\int_{n^2-\delta_n^{\frac{1}{2}}}^{n^2+\delta_n^{\frac{1}{2}}} \frac{n^6\delta_n^2 e^{2\pi n-14}}{(a-n^2)^2+\delta_n^2} > B_{21} n^{-8} \int_{-\delta_n^{\frac{1}{2}}}^{+\delta_n^{\frac{1}{2}}} \frac{1}{a^2+\delta_n^2} > B_{22} n^{-8} e^{\pi n} \quad (n \uparrow \infty)$$

is the general term of a divergent sum.

13.12. Hardy Weights with Arithmetical Gaps

Consider a weight Δ that bounds above a decreasing Hardy weight $|h|^2$ (h outer) on an arithmetical series of intervals:

$$\big|a-(2n-1)c\big| < d \quad (0<d<c,\ n=0,\ \pm 1,\ etc.)$$

but is otherwise unspecified. Then
(a) $Z^\bullet$ is a closed subspace of Z,
(b) $Z^\bullet \supset Z^{0+}$, and hence in accordance with Section 13.5, $Z^\bullet = Z^{0+}$.

As an application, it is easy to derive the lemma of Tutubalin-Freĭdlin [11]: that if $\Delta \geq |a|^{-2m}$ ($m>0$) far out, then $Z^{0+} = Z_\bullet$; indeed, according to (b), $f \in Z^{0+}$ is an entire function of minimal exponential type, and since $\infty > \int |f|^2/(1+a^2)^m$, a simple application of Phragmén-Lindelöf implies that f is a polynomial (of degree $< m$). Actually, it is enough to have $\Delta \geq |a|^{-2m}$ on an arithmetical series of intervals: as the reader can easily check using (b) and the Duffin-Schaeffer theorem [2, 10.5.1].

Proof of (a). Similar to that of (b). □

Proof of (b). $f \in Z^{0+}$ implies the existence of a sum f_δ of trigonometrical functions $e(iat)$ with $|t|<\delta$, real coefficients, and $\|f-f_\delta\|_\Delta < \delta$, and it follows that

$$B_1 > \big\|f_\delta\big\|_\Delta^2 \geq \int_{(2n-1)c-d}^{(2n-1)c+d} \big|f_\delta h\big|^2 \geq 2d\big|f_\delta h(a_n)\big|^2$$

for some $|a_n-(2n-1)c| < d$ with a constant B_1 not depending upon δ. Bring in an entire function g of exponential type $\leq \varepsilon$ with $|g|<|h|$ far out on $b=0$ and $|g| \geq \frac{1}{2}$ on the two 45° lines: to be explicit, let

$$g(\gamma) = e^{-\frac{2}{3}} \prod_{n=n_1}^{\infty} \cos\big(\gamma/\gamma_n\big)$$

with

$$1 < \gamma_1 < \gamma_2 < \text{etc.}$$

and

$$\begin{aligned} \#(R) = \sum_{\gamma_n<R} 1 &= 0 && (R<1) \\ &= \left[3\int_1^R \frac{\lg|h|^{-1}}{A} dA\right] && (R \geq 1), \end{aligned}$$

the bracket denoting the integral part and $|h(1)|$ being supposed ≤ 1, choose n_1 so that

$$\begin{aligned}|g(\gamma)| &\le \prod_{n=n_1}^{\infty} e^{R/\gamma n} = e\left[R\int_C \frac{d\#(B)}{B}\right] \quad (C=\gamma_{n_1},\ |\gamma|=R)\\ &\le e\left[R\int_C \frac{\#(B)}{B^2}\right] < e\left[3R\int_C \frac{dB}{B^2}\int_1^B \frac{\lg|h|^{-1}}{A}dA\right]\\ &= e\left[3R\frac{1}{C}\int_1^C \frac{\lg|h|^{-1}}{A}dA + 3R\int_C^\infty \frac{\lg|h|^{-1}}{B^2}dB\right]\\ &< e^{\varepsilon R},\end{aligned}$$

and use the obvious $|\cos a| < e(-a^2/3)$ $(|a|\le 1)$ to bound $|g(a)|$ for large $|a|$ as follows:

$$\begin{aligned}e^{\frac{2}{3}} \le |g(a)| \prod_{\gamma n\ge|a|} e^{-a^2/3\gamma_n^2} &= e\left[-\frac{a^2}{3}\int_{|a|}\frac{d\#(R)}{R^2}\right]\\ &= e\left[\frac{a^2}{3}\int_{|a|}\frac{\#(R)-\#\big(|a|\big)}{R^3}\right]\\ &< e\left[\frac{a^2}{2}\int_{|a|}\int_{|a|}^R \frac{-\lg|h|^{-1}}{A}dA\frac{dR}{R^3} + \frac{a^2}{3}\int_{|a|}\frac{dR}{R^3}\right]\\ &= e\left[-\frac{a^2}{2}\int_{|a|}\frac{\lg|h|^{-1}}{R^3}dR + \frac{2}{3}\right]\\ &\le \frac{a^2}{2}\int_{|a|}|h|\frac{dR}{R^3}e^{\frac{2}{3}}\\ &\le |h|e^{\frac{2}{3}}.\end{aligned}$$

$f_\delta g$ is then entire of exponential type $\delta+\varepsilon$ and $|f_\delta g(a_n)| < B_2$ with a constant B_2 not depending upon δ. An application of the Duffin-Schaeffer theorem [2, 10.5.3] implies $|f_\delta g| < B_3$ on the whole line $b=0$ if $\delta+\varepsilon$ is small enough, B_3 being likewise independent of δ. Phragmén-Lindelöf now implies that $|f_\delta g| < B_3 e[(\delta+\varepsilon)R]$, and since $|g|\ge\frac{1}{2}$ on the two 45° lines, $|f_\delta| < 2B_3 e[(\delta+\varepsilon)R]$ there. Phragmén-Lindelöf is now applied to each of the four sectors between the 45° lines; this supplies us with the bound $|f_\delta| < 2B_3 e[2(\delta+\varepsilon)R]$, establishing the compactness of f_δ as $\delta\downarrow 0$, and it follows that each limit function f_{0+} is entire of exponential type $\le 2\varepsilon$ with $\|f - f_{0+}\|_\Delta = 0$. But this means that f is the restriction to $b=0$ of an entire function of exponential type $\le 2\varepsilon$, and since ε can be made as small as desired, $f\in Z_\Delta^\bullet$, and the proof is complete. □

13.13. Entire Functions of Positive Type

Given a Hardy weight $\Delta = |h|^2$ and a positive number ϱ, let $Z^{\bullet\varrho}$ be the class of entire functions $f = f(\gamma)$ of exponential type $\le\varrho$:

$$\varlimsup_{R\uparrow\infty} R^{-1}\max_{0\le\theta<2\pi}\lg\left|f\big(Re^{i\theta}\big)\right| \le \varrho,$$

which, restricted to the line $b=0$, belong to Z. Then

$$Z^{\bullet\varrho} = Z^{|t|\le\varrho+} = \bigcap_{\varrho'>\varrho} Z^{|t|\le\varrho'}.$$

PROOF. We first prove the inclusion

$$Z^{\bullet\varrho} \supset Z^{|t|\le\varrho+}.$$

If $f \in Z^{|t|\le\varrho+}$, then it is possible to find (real) sums of trigonometrical functions:

$$f_n(\gamma) = \sum_{k\le n} c_k^n e\bigl(i\gamma t_k^n\bigr)$$

with $|t_k^n| < \varrho + 1/n$ and $\|f - f_n\|_\Delta < 1/n$. Given $\delta > 1/n$, $f_n e[i\gamma(\varrho+\delta)]h$ belongs to H^{2+}, and much as in Sect. 13.6.2,

$$\bigl|f_n h\bigr| < B_1 e^{(\varrho+\delta)R} \quad (b \ge 1), \qquad \bigl|f_n h^*\bigr| < B_2 e^{(\varrho+\delta)R} \quad (b \le -1),$$

and

$$\bigl|f_n\bigr| < B_3 \quad \bigl(|\gamma| \le 2\bigr)$$

with constants B_1, B_2, B_3 not depending upon n. An appraisal of h on $\theta = \pi/4$, $3\pi/4$ and of h^* on $\theta = 5\pi/4$, $7\pi/4$ leads to

$$\bigl|f_n\bigr| < B_4 e^{(\varrho+2\delta)R}$$

much as in Sect. 13.6.2, B_4 being likewise independent of n, and since $\|f - f_n\|_\Delta < 1/n$, it follows that as $n \uparrow \infty$, f_n tends on the whole plane to an entire function f_∞ of exponential type $\le \varrho$, coinciding with f on $b = 0$. But then $f \in Z^{\bullet\varrho}$, and the inclusion is proved.

As in Sect. 13.5, it suffices for the proof of the opposite inclusion:

$$Z^{\bullet\varrho} \subset Z^{|t|\le\varrho+}$$

to consider *even* functions $f \in Z^{\bullet\varrho}$ with Hadamard factorization

$$f(\gamma) = \prod_{n=1}^{\infty}\left(1 - \frac{\gamma^2}{\gamma_n^2}\right).$$

Because

$$\lg^+ \bigl|f(a)\bigr|^2 \le \lg^+\Bigl(\bigl|f(a)\bigr|^2\Delta\Bigr) - \lg^-\Delta \le \bigl|f(a)\bigr|^2\Delta - \lg^-\Delta,$$

f satisfies

$$\int \frac{\lg^+\bigl|f(a)\bigr|}{1+a^2} < \infty;$$

it follows that

$$\varlimsup_{R\uparrow\infty} R^{-1}\lg\bigl|f\bigl(Re^{i\theta}\bigr)\bigr| \le \varrho|\sin\theta|$$

[8, page 27] and that the roots of f in the half-plane $a > 0$ have a density $D \le \varrho/\pi$:

$$\lim_{n\to\infty} n/\bigl|\gamma_n\bigr| = D$$

[8, Theorem VIII]. Also, it is permissible to assume that the roots of f are real: indeed, if

$$f_1(\gamma) = \prod_{n=1}^{d}\left(1 - \frac{\gamma^2}{\gamma_n^2}\right) f_2(\gamma) \quad \text{with} \quad f_2(\gamma) = \prod_{n>d}\left(1 - \frac{\gamma^2}{|\gamma_n^2|}\right)$$

then $|f_1(a)| \le |f(a)|$ and the roots of $f_2(\gamma)$ have the same density D; this implies [2, 8.2.1] that f_2 is of type πD. Hence f_1 is also of type πD and so $f_1 \in Z^{\bullet\varrho}$. But then $(\gamma^2-1)^d f_2 \in Z^{\bullet\varrho}$, so $(\gamma^2-1)^n f_2 \in Z^{\bullet\varrho}$ $(n \le d)$. All these functions have real zeros and hence we may assume them in $Z^{|t|\le\varrho+}$. f_1 is a sum of these, so $f_1 \in Z^{|t|\le\varrho+}$, and since $\|f - f_1\|_\Delta$ is small for large d it follows that $f \in Z^{|t|\le\varrho+}$ also. From here on the roots of f are real: $0 < \gamma_1 \le \gamma_2 \le$ etc.

Given $\varrho' > \varrho$, let us grant the existence of an entire function g of exponential type $\le \varrho'$ with $\|f-g\|_\Delta$ as small as desired and $g \in L^2(R^1)$. As in Sect. 13.5, an application of the Paley-Wiener theorem implies $f \in Z^{|t|\le\varrho'}$, and $f \in Z^{|t|\le\varrho+}$ follows. Accordingly, it suffices to produce such an entire function g.

Given a small positive number $\varepsilon < 1$, define

$$\delta = (\varepsilon/8)^2, \qquad D_* = D - \delta/2, \qquad D^* = D + \delta/2,$$

$$g_1(\gamma) = \prod_{\gamma_n \le d}\left(1 - \frac{\gamma^2}{\gamma_n^2}\right), \qquad g_2(\gamma) = \prod_{n > D^* d}\left(1 - \frac{D^{*2}\gamma^2}{n^2}\right),$$

$$g_3(\gamma) = \prod_{n > \varepsilon d}\left(1 - \frac{\varepsilon^2\gamma^2}{n^2}\right),$$

and let us check the following lemmas leading to the properties of $g = g_1 g_2 g_3$ needed for the proof of $f \in Z^{|t|\le\varrho+}$ indicated above; in the lemmas, c_1, c_2, etc. denote positive constants depending upon ε alone, and it is understood that if ε and/or d is unspecified, then ε has to be small enough and d large enough, the smallest admissible d depending in general upon ε. At a first reading, just note the statements of lemmas (a)–(g) and then turn to (h). □

(a) *g is an entire function of exponential type $\pi(D^* + \varepsilon) \le \varrho + \pi(\delta/2 + \varepsilon)$.*

Proof of (a). Obvious. □

(b) *$|f - g|$ tends to 0 as $d \uparrow \infty$ independently of $\varepsilon(< 1)$ and of $|a| \le A$ for each $A > 0$.*

Proof of (b).

$$e\big(-2A^2\varepsilon^2/n^2\big) \le 1 - a^2\varepsilon^2/n^2 \le 1 \quad \big(|a| \le A\big)$$

for $n > \varepsilon d$ and $d > 2A$, so that as $d \uparrow \infty$

$$e\left(-2A^2 \sum_{n > \varepsilon d} \varepsilon^2 n^{-2}\right) \le g_3(a) \le 1$$

is close to 1 independently of $\varepsilon(< 1)$ and of $|a| \le A$. □

(c) *$|g| \le B|f|$ for $|a| \le d/2$, B being the universal constant involved in the appraisal* (e) *of Sect. 13.5.*

Proof of (c). Because the roots of f have density D,

$$n/D^* < \gamma_n < n/D_* \quad \big(n \ge n_0\big)$$

with n_0 depending only upon D_* and D^* and so only upon ε. Given $d > n_0$ and $0 \le a \le d/2$, if δ is so small that $D^*/D_* < 2$, then

$$\big|f/g_1\big| = \prod_{\gamma_n > d}\left(1 - \frac{a^2}{\gamma_n^2}\right) > \prod_{n > D_* d}\left(1 - \frac{D^{*2}a^2}{n^2}\right)$$

so that

$$\big|f/g_1 g_2\big| > \prod_{D_* d < n \le D^* a}\left(1 - \frac{D^{*2}a^2}{n^2}\right),$$

and since, in this product,

$$D^{*2}a^2/n^2 < \frac{(D+\delta/2)^2}{4(D-\delta/2)^2} < \frac{1}{2}$$

for small δ, the bound $1-c > e(-2c)$ $(0 < c \le \frac{1}{2})$ implies

$$\big|f/g_1g_2\big| > e\left[-2a^2D^{*2}\sum_{D_*d<n\le D^*d} n^{-2}\right]$$

$$> e\big[-3a^2\big(D^*-D_*\big)/d\big] = e\big(-3a^2\delta/d\big).$$

On the other hand, the appraisal (e) of Sect. 13.5 implies

$$g_3 < Be\big(-a^2\varepsilon/d\big) \quad (0 \le a \le d/2),$$

and since $3\delta < \varepsilon$ for small ε, the desired bound follows. □

(d) $|g| < c_1$ $(d/2 < |a| \le D_*d/D^*)$.

Proof of (d). Given $d > 2n_0$ with n_0 as in the proof of (c), it is possible to find c_2 and c_3 depending upon $n_0 = n_0(\varepsilon)$ (and so upon ε) such that

$$\big|g_1\big| < c_1a^{c_2}\prod_{n<D_*a}\left(\frac{D^{*2}a^2}{n^2}-1\right)\prod_{D^*a<n<D_*d}\left(1-\frac{D_*^2a^2}{n^2}\right)$$

for $d/2 < a \le D_*d/D^*$. Define $c_3 = c_1/(\pi D^*)$; then

$$\big|g_1g_2\big| < c_3a^{c_2-1}\big|\sin\pi D^*a\big|J_1/J_2J_3,$$

$$J_1 = \prod_{D^*a<n<D_*d} = \frac{n^2-D_*^2a^2}{n^2-D^{*2}a^2}, \qquad J_2 = \prod_{D_*a\le n\le D^*a}\left(\frac{a^2D^{*2}}{n^2}-1\right),$$

$$J_3 = \prod_{D_*d\le n\le D^*a}\left(1-\frac{D^{*2}a^2}{n^2}\right).$$

J_1 is supposed non-void since the proof simplifies in the opposite case; also, it is supposed below that the smallest integer $n_1 > D^*a$ does not exceed $D^*a+\frac{1}{2}$, the discussion of J_1 being simpler and that of J_2 just a little more complicated if $n_1 > D^*a+\frac{1}{2}$. Bring out the leading factor of J_1:

$$\frac{n_1^2-D_*^2a^2}{n_1^2-D^*a^2} < \frac{n_1-D_*a}{n_1-D^*a} \le \frac{1+a\delta}{n_1-D^*a} < \frac{e^{a\delta}}{n_1-D^*a};$$

the product of other factors of J_1 does not exceed

$$\prod_{D^*a+\frac{1}{2}<n<D_*d}\frac{n-D_*a}{n-D^*a} = e\left[\sum_{D^*a+\frac{1}{2}<n<D_*d}\lg\left(1+\frac{a\delta}{n-D^*a}\right)\right]$$

$$< e\left[2\int_0^{D_*d-D^*a}\lg\big(1+a\delta/c\big)dc\right]$$

$$< e\left[2a\delta\int_0^{D^*/\delta}\lg(1+1/c)dc\right]$$

since $D_*d < 2D^*a$, and using the bound $\lg(1+1/c) < 1/c$, it follows that

$$J_1 < e\left[2a\delta\left(\int_0^1\lg(1+1/c)dc+\lg D^*/\delta\right)\right]\frac{e^{a\delta}}{n_1-D^*a} < \frac{e^{a\delta^{\frac{1}{2}}}}{n_1-D^*a}$$

for small δ. Stirling's approximation is now applied to obtain an underestimate of J_2 for small δ, using $D^*a-(n_1-1)>\frac{1}{2}$:

$$\begin{aligned}
J_2 &> \prod_{D_*a\le n\le D^*a}\frac{D^*a-n}{n} > \frac{\Gamma(a\delta)}{\bigl(D^*a\bigr)^{a\delta+1}}\\
&> c_4(a\delta)^{a\delta-\frac{1}{2}}e^{-a\delta}\bigl(D^*a\bigr)^{-a\delta-1}\\
&> c_4\bigl(D^*a\bigr)^{-\frac{3}{2}}(\delta/eD^*)^{a\delta}\\
&= c_4\bigl(D^*a\bigr)^{-\frac{3}{2}}e\Bigl[-a\delta\Bigl(\lg\frac{D^*}{\delta}+1\Bigr)\Bigr]\\
&> c_4\bigl(D^*a\bigr)^{-\frac{3}{2}}e^{-a\delta^{\frac{1}{2}}}
\end{aligned}$$

with a universal constant c_4. Similarly

$$\begin{aligned}
J_3 &\ge \prod_{D_*d\le n\le D^*d}\Bigl(\frac{n-aD^*}{n}\Bigr) \ge \frac{\Gamma\bigl(D^*(d-a)\bigr)}{\Gamma\bigl(D_*d-aD^*+1\bigr)\bigl(D^*d\bigr)^{\delta d+1}}\\
&\ge C_5\frac{\bigl[D^*(d-a)\bigr]^{D^*(d-a)-\frac{1}{2}}e^{-D^*(d-a)}}{\bigl(D_*d-aD^*\bigr)^{D_*d-aD^*+\frac{1}{2}}e^{-D_*d+aD^*}\bigl(D^*d\bigr)^{\delta d+1}}\\
&\ge C_5\frac{e^{-\delta d}}{\bigl(D_*d-aD^*\bigr)D^*d}\Bigl[\frac{D^*(d-a)}{D_*d-aD^*}\Bigr]^{D_*d-aD^*-\frac{1}{2}}\Bigl(\frac{d-a}{d}\Bigr)^{\delta d}\\
&\ge c_5\frac{e^{-\delta d}}{D^*D_*d^2}\Bigl(1-\frac{a}{d}\Bigr)^{\delta d} \ge c_5e^{-2\delta a}a^{-2}\Bigl(1-\frac{D_*}{D^*}\Bigr)^{\delta d}/\bigl(4D^*D_*\bigr)\\
&\ge c_5a^{-2}e\bigl[-2\delta a-\delta d\lg\bigl(D^*/\delta\bigr)\bigr]/\bigl(4D^*D_*\bigr) \ge c_5a^{-2}e\Bigl(-\sqrt{\delta}a\Bigr)/\bigl(4D^*D_*\bigr)
\end{aligned}$$

with a universal constant c_5. Combining the bounds for J_1, J_2, J_3 and using $0<n_1-D^*a\le\frac{1}{2}$, it follows that

$$\bigl|g_1g_2\bigr| < c_6a^{c_2+3}\left|\frac{\sin\pi D^*a}{n_1-D^*a}\right|e^{3a\delta^{\frac{1}{2}}} < c_7a^{c_2+3}e^{3a\delta^{\frac{1}{2}}} < c_7e^{4a\delta^{\frac{1}{2}}}$$

with c_7 depending upon ε alone, d being increased if need be so as to achieve $a^{c_2+3}<e(a\delta^{\frac{1}{2}})$. But now the familiar appraisal (e) of Sect. 13.5 implies

$$\bigl|g_3\bigr| < Be^{-4a\delta^{\frac{1}{2}}},$$

and so

$$|g| = \bigl|g_1g_2g_3\bigr| < Bc_7 \equiv c_1,$$

completing the proof of (d). □

(e) $|g|<c_8\ (D_*d/D^*<|a|\le d)$.

PROOF OF (e).

$$\bigl|g_1\bigr| < c_9a^{c_{10}}\prod_{n<D_*a}\Bigl(\frac{D^{*2}a^2}{n^2}-1\Bigr)$$

for $D_*d/D^* < a \le d$ with constants c_9 and c_{10} depending upon $n_0 = n_0(\varepsilon)$ alone, so

$$|g_1 g_2| < c_{11} a^{c_{10}} |\sin \pi D^* a| / J_4$$

with

$$J_4 = \prod_{D_* a \le n \le D^* d} \left| 1 - \frac{D^{*2} a^2}{n^2} \right| \ge \prod_{D_* a \le n \le D^* d} \left| 1 - \frac{D^* a}{n} \right|$$

$$\ge \left| \frac{n_2 - D^* a}{n^2} \right| \frac{\Gamma\big(D^* d - D_* a\big)\Gamma(a\delta)}{\big(D^* d\big)^{D^* d - D_* a + 3}},$$

n_2 being determined from $-\frac{1}{2} < n_2 - D^* a \le \frac{1}{2}$. Both gamma functions contribute to this underestimate if, as is supposed below, $D^* a$ is not too close to $D_* a$ or to $D^* d$; the appraisal of J_4 is similar in the opposite case. Stirling's approximation is now applied to obtain

$$J_4 > c_{12} |n_2 - D^* a| \big(D^* d\big)^{-5} J_5 J_6$$

with

$$J_5 = e\left[- D^* d \left(\frac{d-a}{d} \right) \lg \left(\frac{d}{d-a} \right) \right]$$

and

$$J_6 = e\left[- D^* d \left(\frac{a\delta}{D^* d} \right) \lg \left(\frac{D^* d}{a\delta} \right) \right].$$

Because $d - a \le d(1 - D_*/D^*) = d\delta/D^*$ and $a\delta \le d\delta$, both J_5 and J_6 are bigger than $e(-a\delta^{\frac{1}{2}})$ for small δ, so

$$J_4 > c_{13} |n_2 - D^* a| a^{-5} e\Big(- 3a\sqrt{\delta} \Big),$$

and the proof is completed as in (d) above. □

(f) $|g| < c_{14}$ $(d < |a| \le 2d)$.

Proof of (f).

$$|g_1| < c_{15} a^{c_{16}} \prod_{n < D_* d} \left(\frac{D^{*2} a^2}{n^2} - 1 \right) e^{2n\delta}$$

for $d < a \le 2d$, the exponential accounting for the factors of

$$\prod_{D_* d \le n \le D^* d} \left(\frac{D^{*2} a^2}{n^2} - 1 \right)$$

that exceed 1; the rest of the proof is similar to but simpler than that of (e). □

(g) $|g| < c_{17}$ $(|a| > 2d)$, *and* $g \in L^2(R^1)$.

Proof of (g).

$$|g_1| < c_{18} a^{c_{19}} \prod_{n \le D^* d} \left(\frac{D^{*2} a^2}{n^2} - 1 \right)$$

for $a > 2d$, so

$$|g_1 g_2| < c_{20} a^{c_{19}} |\sin \pi D^* a| \le c_{20} a^{c_{19}},$$

and using the familiar appraisal (e) of Sect. 13.5 to bound g_3, it develops that

$$|g| < B c_{21} a^{c_{22}} e^{-\varepsilon d (1 + 2(\lg a/d))}.$$

But

$$d\lg(a/d) > \frac{d\lg 2}{\lg(2d)}\lg a \quad (a > 2d),$$

and so

$$|g| < c_{23}a^{c_{22}-2\varepsilon d\lg 2/\lg(2d)}$$

is bounded ($a > 2d$) and belongs to $L^2(R^1)$ if d is large enough. □

(h) $\|f-g\|_\Delta$ *can be made as small as desired by appropriate choice of* ε *and* d.

PROOF OF (h).

$$\frac{1}{2}\|f-g\|_\Delta^2 \le \int_0^A |f-g|^2\Delta + (2B+1)^2\int_A^{d/2}|f|^2\Delta + \int_{d/2}^\infty \big(c_{24}+|f|\big)^2\Delta$$

with an adjustable number A, a universal constant B, and c_{24} (= the greatest of c_1, c_8, c_{14}, c_{17}) depending upon ε alone, *provided* ε is small enough and $d(> 2A)$ is large enough, the smallest admissible d depending upon ε. A is now chosen so large that $(2B+1)^2\int_A^\infty |f|^2\Delta < 1/n$ and then ε is chosen so small that $c_{24} = c_{24}(\varepsilon) < \infty$ and d is made so big that neither $\int_0^A |f-g|^2\Delta$ nor $\int_{d/2}^\infty (c_{24}+|f|)^2\Delta$ exceeds $1/n$, with the result that $\|f-g\|_\Delta^2 < 6/n$. □

13.14. Another Condition for $Z^{+/-} = Z^{0+}$ (Δ Hardy)

Because $Z^{|t|\le\varrho+}$ is closed so is $Z^{\bullet\varrho}$, but it is possible to go another step and prove that, *if*

$$\sigma^{\bullet\varrho}(\gamma) = \sup\big|f(\gamma)\big| \quad f\in Z^{\bullet\varrho}_{\Delta^+},\ \|f\|_{\Delta^+}\le 1,$$

then $\lg\sigma^{\bullet\varrho}$ *is a non-negative, continuous subharmonic function such that*

$$\overline{\lim_{R\uparrow\infty}}\, R^{-1}\max_{0\le\theta<2n}\lg\sigma^{\bullet\varrho}\big(Re^{i\theta}\big) = \varrho.$$

PROOF. Only the last statement needs a proof. Given $f\in Z^{\bullet\varrho}_{\Delta^+}$, $(\gamma+i)^{-1}e^{i\gamma\varrho}fh\in H^{2+}$, and so

$$\lg\left|\frac{e^{i\gamma\varrho}fh}{\gamma+i}\right| \le \frac{1}{\pi}\int\frac{b\,dc}{(c-a)^2+b^2}\lg\frac{|fh|}{|a+i|} \quad (\gamma = a+ib,\ b>0);$$

this leads at once to

$$\lg\big[e^{-b\varrho}\sigma^{\bullet\varrho}(\gamma)\big] \le \frac{1}{\pi}\int\frac{b\,dc}{(c-a)^2+b^2}\lg\sigma^{\bullet\varrho}$$

since $h(\gamma)/(\gamma+i)$ is outer. $\int\lg\sigma^{\bullet\varrho}/(1+a^2) < \infty$ is now proved as in Sect. 13.4(e), and it follows that

$$\overline{\lim_{R\uparrow\infty}}\, R^{-1}\lg\sigma^{\bullet\varrho}\big(Re^{i\theta}\big) \le \varrho|\sin\theta|$$

for $\theta = \pi/4, 3\pi/4$; the same holds by a similar argument for $\theta = 5\pi/4, 7\pi/4$. An application of Phragmén-Lindelöf as in Sect. 13.4(f) completes the proof that $\sigma^{\bullet\varrho}$ is of type $\le\varrho$, and that the equality must hold follows since $e(-i\gamma\varrho)\in Z^{\bullet\varrho}_{\Delta^+}$.

As an application of the bound for $\sigma^{\bullet\varrho}$, it will be proved that if $Z^{|t|\le\varrho+}\supset Z^{+/-}$, and indeed if the projection of $e(ias)$ upon Z^- belongs to $Z^{|t|\le\varrho+}$ for *single* $s>0$, then $Z^{+/-} =$

Z^{0+}. Indeed, if the projection belongs to $Z^{|t|\le\varrho+}$ for a single $s>0$, then it does so for a whole (bounded) interval of s with a larger ϱ, and selecting such an s from the Lebesgue set of

$$h = \frac{1}{2\pi}\int_0 e^{-iat}\hat{h}(a)da$$

and arguing as in Sect. 13.7 with $\sigma^{\bullet\varrho}$ in place of $\sigma^{\bullet}$, it is found that h^{-1} is an entire function of exponential type $\le\varrho$. But then $i = h/h^*$ is inner as in Sect. 13.7 so that $Z^{+/-} = Z^- \cap Z^+$; also $Z^- \cap Z^+ = Z^{\bullet}$ since $1/\Delta$ is locally summable (Sect. 13.6.3), and so $Z^{+/-} = Z^{\bullet} = Z^{0+}$ as stated. □

Bibliography

[1] A. Beurling. On two problems concerning linear transformations in Hilbert space. *Acta Math.*, 81:239–255, 1949.

[2] R. P. Boas. *Entire functions*. Academic Press, New York, 1954.

[3] T. Carleman. *L'Intégrale de Fourier et Questions que s'y Rattachent*, volume 1 of *Publications Scientifiques de l'Institut Mittag-Leffler*. Almqvist & Wiksells, Uppsala, 1944.

[4] Y. Domar. Closed primary ideals in a class of Banach algebras. *Math. Scand.*, 7:109–125, 1959.

[5] I. O. Hačatrjan. Weighted approximation of entire functions of zero degree by polynomials on the real axis. *Dokl. Akad. Nauk SSSR*, 145:744–747, 1962.

[6] T. Hida. Canonical representations of Gaussian processes and their applications. *Mem. Coll. Sci. Univ. Kyoto. Ser. A. Math.*, 33:109–155, 1960.

[7] K. Hoffman. *Banach Spaces of Analytic Functions*. Prentice-Hall Series in Modern Analysis. Prentice-Hall Inc., New Jersey, 1962.

[8] N. Levinson. *Gap and Density Theorems*. Number 26 in American Mathematical Society Colloquium Publications. American Mathematical Society, New York, 1940.

[9] H. P. McKean. Brownian motion with a several-dimensional time. *Teor. Verojatnost. i Primenen.*, 8:357–378, 1963.

[10] S. N. Mergelyan. Weighted approximations by polynomials. *Uspehi Mat. Nauk (N.S.)*, 11:107–152, 1956. *Amer. Math. Soc. Transl. Ser. 2*, 10:59–106, 1958.

[11] V. N. Tutubalin and M. I. Freĭdlin. On the structure of the infinitesimal σ-algebra of a Gaussian process. *Teor. Verojatnost. i Primenen.*, 7:204–208, 1962. Translation in "*Theor. Probab. Appl.* 7:196–199, 1962".

[14] Brownian Local Times

[14] Brownian Local Times. *Adv. Math.* **16** (1975), 91–111.

14 Brownian Local Times

Henry P. McKean Jr.[1]

Dedicated To Norman Levinson

14.1. Simple Brownian Motions

The purpose of this section is to review the elementary facts. The proofs are indicated. Itô-McKean [3] and Lévy [5] can be consulted for additional information.

14.1.1. Standard Brownian Motion. Fix a standard one-dimensional Brownian motion BM with sample paths $\mathfrak{x} : t \to \mathfrak{x}(t)$ and probabilities $P_a(B)$ depending upon the starting point $\mathfrak{x}(0) = a$ and the event B. The infinitesimal operator is $\mathbf{G} = \frac{1}{2}\mathbf{D}^2$ acting upon $\mathsf{C}^2(R^1)$, and the transition density is

$$p(t, x, y) = (2\pi t)^{-1/2} e^{-(x-y)^2/2t}.$$

The Brownian traveller begins afresh at stopping times, the most important of which are the passage times

$$\mathfrak{m}_x = \min(t : \mathfrak{x}(t) = x)$$

for $x \geqslant 0$, say. $\mathbf{M} = [\mathfrak{m}, P_0]$ is additive: in fact, it is the one-sided stable process of exponent $\frac{1}{2}$, so-called, and you have the formula

$$E_0\left[e^{-\alpha \mathfrak{m}_x}\right] = e^{-(2\alpha)^{1/2} x}$$

for $\alpha > 0$ and $x \geqslant 0$, or what is the same,

$$P_0\left[\mathfrak{m}_x \in dt\right] = \left(2\pi t^3\right)^{-1/2} x e^{-x^2/2t} dt.$$

Proof. $\exp[\alpha\mathfrak{x}(t) - \alpha^2 t/2]$ is a martingale over BM, and for paths starting at $x = 0$, it is bounded up to the stopping time $\mathfrak{m} = \mathfrak{m}_x$ for fixed $x > 0$. Therefore, you have

$$1 = E_0\left[e^{\alpha\mathfrak{x}(\mathfrak{m}) - \alpha^2\mathfrak{m}/2}\right] = e^{\alpha x} E_0\left[e^{-\alpha^2\mathfrak{m}/2}\right]$$

if you grant $P_0[\mathfrak{m} < \infty] = 1$. The proof is finished by putting $(2\alpha)^{1/2}$ in place of α and inverting the transform. □

[1] The Courant Institute of Mathematical Sciences, New York, NY 10012, USA, mckean@cims.nyu.edu.

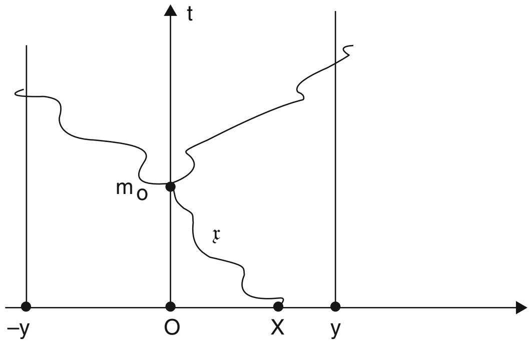

FIGURE 14.1.

14.1.2. Absorbing Brownian Motion. The absorbing Brownian motion BM^∞ is the diffusion on the half-line $x \geqslant 0$ with infinitesimal operator $\mathbf{G}^\infty = \frac{1}{2}\mathbf{D}^2$ acting upon functions f of class $\mathsf{C}^2[0,\infty)$ with $f(0) = 0$. It can be presented in terms of the standard Brownian motion $\mathfrak{x}$ starting at $x > 0$ as

$$\mathfrak{x}^\infty(t) = \begin{cases} \mathfrak{x}(t) & \text{if } t < \mathfrak{m}_0 \\ \infty & \text{if } t \geqslant \mathfrak{m}_0. \end{cases}$$

The jump to ∞ is spoken of as "killing." The absorbing transition density is

$$p^\infty(t,x,y) = (2\pi t)^{-1/2}\left[e^{-(x-y)^2/2t} - e^{-(x+y)^2/2t}\right],$$

as you can verify either from the known law of $\mathfrak{m}_0$ or from Fig. 14.1 taking advantage of the fact that the standard Brownian traveller begins afresh at time $\mathfrak{m}_0$ and cannot come from $x > 0$ to $-y < 0$ without hitting the origin. By the second way,

$$\begin{aligned} P_x\left[\mathfrak{x}^\infty(t) > y\right] &= P_x\left[\mathfrak{x}(t) > y,\ t < \mathfrak{m}_0\right] \\ &= P_x\left[\mathfrak{x}(t) > y\right] - P_x\left[\mathfrak{x}(t) < -y\right], \end{aligned}$$

and the formula follows upon differentiating by y. The formula for p^∞ leads at once to the joint law of $\mathfrak{x}(t)$ and the maximum function $\mathfrak{t}^-(t) = \max_{s \leqslant t} \mathfrak{x}(s)$, namely, for $-\infty < x < y$ and $y > 0$,

$$\begin{aligned} P_0\left[\mathfrak{x}(t) < x,\ \mathfrak{t}^-(t) < y\right] &= P_0\left[\mathfrak{x}(t) < x,\ \mathfrak{m}_y > t\right] \\ &= P_y\left[\mathfrak{x}(t) > y - x,\ \mathfrak{m}_0 > t\right], \end{aligned}$$

so

$$P_0\left[\mathfrak{x}(t) \in dx,\ \mathfrak{t}^-(t) \in dy\right] = \left(2/\pi t^3\right)^{1/2}(2y-x)e^{-(2y-x)^2/2t}dx\,dy.$$

14.1.3. Reflecting Brownian Motion. The reflecting Brownian motion BM^+ attached to the infinitesimal operator $\mathbf{G}^+ = \frac{1}{2}\mathbf{D}^2$ acting upon functions f of class $\mathsf{C}^2[0,\infty)$ with $f^+(0) = 0$ can be similarly presented as

$$\mathfrak{x}^+(t) = |\mathfrak{x}(t)| \quad \text{for } 0 \leqslant t < \infty.$$

Because $-\mathfrak{x}$ is likewise a standard Brownian motion, the motion $\mathfrak{x}^+$ so presented begins afresh at its stopping times, and the identification of $\mathfrak{x}^+$ with BM^+ is easily made by computing its transition density

$$p^+(t,x,y) = (2\pi t)^{-1/2}\left[e^{-(x-y)^2/2t} + e^{-(x+y)^2/2t}\right]$$

from the self-evident formula

$$P_x\left[\mathfrak{x}^+(t) > y\right] = P_x\left[\mathfrak{x}(t) > y\right] + P_x\left[\mathfrak{x}(t) < -y\right].$$

14.1.4. Elastic Brownian Motion. The elastic Brownian motion BM^γ is the diffusion on the half-line with infinitesimal operator $\mathbf{G} = \frac{1}{2}\mathbf{D}^2$ acting upon functions f of class $\mathsf{C}^2[0,\infty)$ with $\gamma f(0) = f^+(0)$. The elastic transition density is

$$p^\gamma(t,x,y) = p^\infty(t,x,y) + \int_0^t \left(2\pi s^3\right)^{-1/2}(x+y)e^{-(x+y)^2/2s}p^\gamma(t-s,0,0)ds,$$

in which

$$p^\gamma(t,0,0) = 2\int_0^\infty e^{-\gamma x}\left(2\pi t^3\right)^{-1/2} x e^{-x^2/2t}dx.$$

BM^∞ is the old absorbing Brownian motion, while BM^0 is the reflecting Brownian motion BM^+, and for intermediate $0 < \gamma < \infty$, you have something between the two, with a little killing $[\int_0^\infty p^\gamma dy < 1]$ but only while the particle is *at* the origin $[\mathbf{G}^\gamma = \frac{1}{2}\mathbf{D}^2$ for $x > 0]$. The precise mechanism of this killing will be described in Sect. 14.2.8 by presenting the elastic Brownian motion in terms of the standard Brownian motion and its local time.

14.2. Brownian Local Time

14.2.1. P. Lévy's Presentation of BM^+. Lévy [5, p. 234] presented BM^+ in terms of the standard Brownian motion $\mathfrak{x}$ starting at $x \geqslant 0$ by means of the recipe

$$\mathfrak{x}^-(t) = \begin{cases} \mathfrak{x}(t) & \text{if } t < \mathfrak{m}_0 \\ \mathfrak{t}^-(t) - \mathfrak{x}(t) & \text{if } t \geqslant \mathfrak{m}_0, \end{cases}$$

in which $\mathfrak{t}^-(t)$ is the maximum of $\mathfrak{x}(s)$ for $\mathfrak{m}_0 \leqslant s \leqslant t$; see Fig. 14.2. The proof that $\mathfrak{x}^-$ is indeed a reflecting Brownian motion consists in checking first that the motion $\mathfrak{x}^-$, so presented, begins afresh at its stopping times, and second that $\mathfrak{x}^-$ has transition density $p^- = p^+$. A close examination of the picture will convince you of the former, and the latter is easily deduced from the joint law of $\mathfrak{x}$ and $\mathfrak{t}^-$ of Sect. 14.1.2. The computation is simplified if $x = 0$, which suffices for the proof; see Itô-McKean [3, pp. 40–42] for the details.

14.2.2. Mesure du voisinage. The presentation $\mathfrak{x}^-$ of Sect. 14.2.1 is now used as follows: From the picture, it is plain that $\mathfrak{t}^-$ is a continuous increasing function of $t > 0$ which is flat off the roots $\mathfrak{Z}^- = (t : \mathfrak{x}^-(t) = 0)$ of $\mathfrak{x}^-$. What is not so plain is that $\mathfrak{t}^-$ is measurable over $\mathfrak{x}^-$, as follows from the remarkable formula of Lévy [5, pp. 239–241]:

$$\mathfrak{t}^-(t) = \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\,\text{measure}\left(s \leqslant t : \mathfrak{x}^-(s) < \epsilon\right).$$

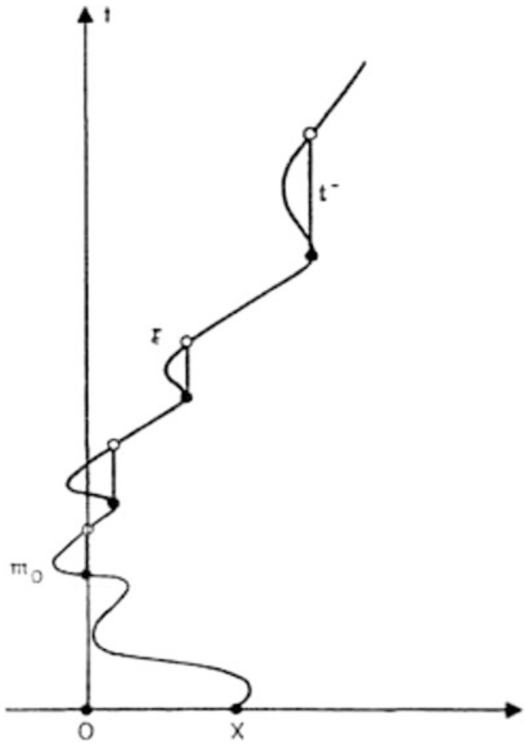

FIGURE 14.2.

Because $\mathfrak{x}^-$ is a representation of BM^+, it follows that the original presentation $\mathfrak{x}^+$ admits a similar functional $\mathfrak{t}^+$, flat off the roots $\mathfrak{Z}^+ = (t : \mathfrak{x}^+(t) = 0)$ and expressible by a similar formula:

$$\mathfrak{t}^+(t) = \lim_{\epsilon \downarrow 0} (2\epsilon)^{-1} \text{ measure} \left(s \leqslant t : \mathfrak{x}^+(s) < \epsilon \right).$$

The latter is the *mesure du voisinage* of Lévy [5, p. 228], alias the reflecting Brownian local time. The above is taken for granted now: suffice it to say that the (pathwise) existence of the mesure du voisinage $\mathfrak{t}^+$ is proved in Sect. 14.3 and that the identification of the mesure du voisinage of $\mathfrak{x}^-$ as $\mathfrak{t}^- = \max_{s \leqslant t} \mathfrak{x}(s)$ is easily proved by checking that

$$\mathfrak{t}^-(t) = \lim_{\epsilon \downarrow 0} (2\epsilon)^{-1} \text{ measure} \left(s \leqslant t : \mathfrak{x}^-(s) < \epsilon \right)$$

in mean-square; see Itô-McKean [3, pp. 63–64] for the details. The numerous properties and uses of the local time $\mathfrak{t}^+$ occupy the rest of this section.

14.2.3. Inverse Local Time. The inverse function of the reflecting Brownian local time,

$$\mathfrak{t}^{-1}(t) = \max \left(s : \mathfrak{t}^+(s) = t \right),$$

considered for paths starting at $\mathfrak{x}^+(0) = 0$, is identical in law to the inverse function of $\mathfrak{t}^-$, alias the standard Brownian passage times $\mathfrak{m}$. As such, it is a copy of the one-sided stable process of exponent $\frac{1}{2}$, and you easily deduce the results of P. Lévy [5, pp. 224–225]:

$$\mathfrak{t}^+(t) = \lim_{\epsilon \downarrow 0} \left(\frac{\pi\epsilon}{2} \right)^{1/2} \times \left(\text{the number of intervals of } [0,t] - \mathfrak{Z}^+ \text{ of length} \geqslant \epsilon \right)$$

and

$$\mathfrak{t}^+(t) = \lim_{\epsilon\downarrow 0} \left(\frac{\pi}{2\epsilon}\right)^{1/2} \times \big(\text{the measure of intervals of } [0,t] - \mathfrak{Z}^+ \text{ of length} < \epsilon\big).$$

PROOF. The first formula stems from the fact that the number of cited intervals is the same as the number of jumps of $\mathfrak{t}^{-1}$ of magnitude $\geqslant \epsilon$, up to time $\mathfrak{t}^+(t)$ (give or take one), as you can see from a picture. The proof is finished by an application of the strong law of large numbers: for fixed $t \geqslant 0$, the number $\mathfrak{n}(\epsilon)$ of jumps of $\mathfrak{t}^{-1}(s) : s \leqslant t$ of magnitude exceeding ϵ is a homogeneous Poisson process relative to the parameter $(2/\pi\epsilon)^{1/2}$, so

$$\lim_{\epsilon\downarrow 0} \left(\frac{\pi\epsilon}{2}\right)^{1/2} \mathfrak{n}(\epsilon)$$

is its rate, namely t, and plainly this holds for all $t \geqslant 0$, simultaneously, permitting you to substitute $\mathfrak{t}^+(t)$ in place of t. The proof of the second formula is just as easy; see Itô-McKean [3, pp. 31–33, 42–43] for more details. □

14.2.4. Downcrossings. Another entertaining way of computing the local time is embodied in the formula of Itô-McKean [3, p. 48]:

$$\mathfrak{t}^+(t) = \lim_{\epsilon\downarrow 0} \frac{\epsilon}{2} \times \begin{array}{l}\text{the number of times } \mathfrak{x}^+(s) : s \leqslant t \\ \text{crosses down from } x = \epsilon \text{ to } x = 0.\end{array}$$

PROOF. The idea of the proof is that, for fixed $\epsilon > 0$, the quantity measure $(s \leqslant t : \mathfrak{x}^+(s) < \epsilon)$ is (approximately) the sum of independent copies of measure $(s \leqslant \mathfrak{m} : \mathfrak{x}^+(s) < \epsilon)$, in which $\mathfrak{m}$ is the time it takes $\mathfrak{x}^+$ to reach ϵ and come back to 0, one such copy to each downcrossing. This suggests that you should have

$$\begin{aligned} 2\epsilon\mathfrak{t}^+(t) = \big[1 + o(1)\big] &\times \text{the number of downcrossings} \\ &\times E_0\big[\,\text{measure}(s \leqslant \mathfrak{m} : \mathfrak{x}^+(s) < \epsilon)\big], \end{aligned}$$

or approximately so, and the proof is finished by evaluating

$$\begin{aligned} &E_0\big[\,\text{measure}(s \leqslant \mathfrak{m} : \mathfrak{x}^+(s) < \epsilon)\big] \\ &\qquad = E_0\big[\mathfrak{m}_\epsilon^+\big] + E_\epsilon\big[\,\text{measure}(s \leqslant \mathfrak{m}_0 : \mathfrak{x}(s) < \epsilon)\big] \\ &\qquad = 2\epsilon^2. \end{aligned}$$

The first expectation is computed by stopping the martingale $\mathfrak{x}^2 - t$ at time $\mathfrak{m}_\epsilon^+ = \min(t : |\mathfrak{x}(t)| = \epsilon)$, while the second comes from the evaluation

$$\int_0^\infty p^\infty(t,x,y)dt = 2y$$

for $x > y$; for more details, see Itô-McKean [3, pp. 48–50]. □

14.2.5. Hausdorff Measure. The fact that $\mathfrak{t}^+$ is flat off $\mathfrak{Z}^+$ and is a copy of the Brownian maximum function leads at once to the fact that $\mathfrak{Z}^+$ is of Hausdorff dimension $\geqslant \frac{1}{2}$.

PROOF. Fix a number $\delta < \frac{1}{2}$ and cover $\mathfrak{Z}^+ \cap [0,1]$ by nonoverlapping intervals I of such small lengths that $\mathfrak{t}^+$ increases by $\mathfrak{t}^+(I) < |I|^\delta$ on each of them. Then

$$0 < \mathfrak{t}^+(1) \leqslant \sum \mathfrak{t}^+(I) \leqslant \sum |I|^\delta,$$

whence the outer δ-dimensional measure of $\mathfrak{Z}^+ \cap [0,1]$ is positive. □

The fact that the dimension of $\mathfrak{Z}^+$ is precisely $\frac{1}{2}$ now follows from the formula of Itô-McKean [3, pp. 50–51]:

$$\mathfrak{t}^+(t) = \lim_{n\uparrow\infty}\left(\frac{\pi}{2}\right)^{1/2}\sum_{k\cdot 2^{-n}\leqslant t}\left({}_k\mathfrak{z}_n{}^+ - {}_k\mathfrak{z}_n{}^-\right)^{1/2},$$

in which ${}_k\mathfrak{z}_n{}^-[{}_k\mathfrak{z}_n{}^+]$ is the smallest (biggest) root of $\mathfrak{x}^+ = 0$ in the interval $(k-1)2^{-n} \leqslant t \leqslant k2^{-n}$, with the understanding that ${}_k\mathfrak{z}_n{}^- = {}_k\mathfrak{z}_n{}^+ = 0$ if no such root exists.

Proof. Bring in the field $\mathfrak{Z}_n$ of $\mathfrak{x}^+(k2^{-n})$, ${}_k\mathfrak{z}_n{}^-$, and ${}_k\mathfrak{z}_n{}^+$ for $k \geqslant 1$, and notice that $\mathfrak{Z}_n$ increases to the full field of $\mathfrak{x}^+$ as $n \uparrow \infty$. The formula is proved (for $t = 1$, say) by a self-evident application of the martingale theorem to the conditional expectation

$$E_0\left[\mathfrak{t}^+(1) \mid \mathfrak{Z}_n\right] = \left(\frac{\pi}{2}\right)^{1/2}\sum_{k\leqslant 2^n}\left({}_k\mathfrak{z}_n{}^+ - {}_k\mathfrak{z}_n{}^-\right)^{1/2}.$$

The actual computation of the latter is made as follows. Pick $0 \leqslant t_1 < t_1^* < t_2^* < t_2 < \infty$ and look at

$$E_0\left[\mathfrak{t}^+(t_2) - \mathfrak{t}^+(t_1) \mid \mathfrak{x}^+(s) : s \leqslant t_1,\ \mathfrak{z}^- = t_1^*,\ \mathfrak{z}^+ = t_2^*,\ \mathfrak{x}^+(s) : s \geqslant t_2\right],$$

in which $\mathfrak{z}^-[\mathfrak{z}^+]$ is the smallest [biggest] root of $\mathfrak{x}^+(s) = 0$ in $t_1 \leqslant s \leqslant t_2$. Because $\mathfrak{t}^+(t_2) - \mathfrak{t}^+(t_1)$ does not depend upon what happens before the stopping time $\mathfrak{z}^-$ or after the fixed time t_2, the expectation simplifies to

$$E_0\left[\mathfrak{t}^+(t_2 - t_1^*) \mid \mathfrak{z}^+ = t_2^* - t_1^*,\ \mathfrak{x}^+(t_2 - t_1^*) = x\right]$$

with a self-evident abuse of notation. Besides, $\mathfrak{t}^+(t_2 - t_1^*)$ depends in a *reversible* way upon the tied reflecting motion $\mathfrak{x}^+(s) : s \leqslant t_2 - t_1^*$ with law $P_0[B \mid \mathfrak{x}^+(t_2 - t_1^*) = x]$, so the expectation is reduced to

$$E_x\left[\mathfrak{t}^+(t_2 - t_1^*) \mid \mathfrak{m}_0 = t_2 - t_2^*,\ x^+(t_2 - t_1^*) = 0\right]$$

$$= E_0\left[\mathfrak{t}^+(t_2^* - t_1^*) \mid \mathfrak{x}^+(t_2^* - t_1^*) = 0\right]$$

by a second application of the passage time trick. Now you have simply to compute. You have

$$P_0\left[\mathfrak{x}^+(t) \in dx,\ \mathfrak{t}^+(t) \in dy\right] = \left(\frac{2}{\pi t^3}\right)^{1/2}(x+y)e^{-(x+y)^2/2t}dx\,dy$$

from Sect. 14.1.2 via the presentation $\mathfrak{x}^- = \mathfrak{t}^- - \mathfrak{x}$, and the evaluation

$$E_0\left[\mathfrak{t}^+(t) \mid \mathfrak{x}^+(t) = 0\right] = (\pi t/2)^{1/2}$$

follows. □

14.2.6. Germ Field. As you may easily believe by now, the germ field, defined by intersecting over $T > 0$ the field of $\mathfrak{x}^+(t) : t \leqslant T$, is precisely the field of $\mathfrak{t}^+$. The proof can be found in Itô-McKean [3, p. 79]. The moral is that $\mathfrak{t}^+$ accounts for the whole of the fine structure of $\mathfrak{x}^+$ in the vicinity of $x = 0$.

14.2.7. Reflecting Brownian Motion. Skorokhod [10] noticed a very amusing variant of the presentation $\mathfrak{x}^- = \mathfrak{t}^- - \mathfrak{x}$. You may say that the reflecting Brownian motion $\mathfrak{x}^-$ is the sum of a standard Brownian motion $(-\mathfrak{x})$ plus the effect of a singular force $(\mathfrak{t}^-)$, acting while the traveller is at $x = 0$ so as to keep him in the half-line $x \geqslant 0$. To put the matter a little differently, the pair $\mathfrak{x}^+$ and $\mathfrak{t}^+$ is the only solution of $\mathfrak{x}^+ = \mathfrak{x} + \mathfrak{t}^+$ subject to (1) $\mathfrak{x}^+$ is non-negative and continuous, (2) $\mathfrak{t}^+$ is increasing and flat off $\mathfrak{Z}^+ = (t : \mathfrak{x}^+(t) = 0)$, (3) both $\mathfrak{x}^+$ and $\mathfrak{t}^+$ depend upon the standard Brownian motion $\mathfrak{x}$ in a non-anticipating way. McKean [7, pp. 71–77] can be consulted for additional information.

Proof. Pick solutions $\mathfrak{x}_1^+$, $\mathfrak{t}_1^+$ and $\mathfrak{x}_2^+$, $\mathfrak{t}_2^+$. Then $\mathfrak{x}_1^+ - \mathfrak{x}_2^+ = \mathfrak{t}_1^+ - \mathfrak{t}_2^+$, and if ever $\mathfrak{x}_1^+ > \mathfrak{x}_2^+$, then $\mathfrak{x}_1^+ > 0$, $\mathfrak{t}_1^+$ is flat nearby, $\mathfrak{t}_2^+$ increases, and the difference $\mathfrak{x}_1^+ - \mathfrak{x}_2^+$ goes down. But this means that $\mathfrak{x}_1^+ \leqslant \mathfrak{x}_2^+$, and a self-evident reprise finishes the proof. □

Remark 14.2.1. M. Motoo [private communication] informs me that you do *not* obtain the reflecting Brownian motion $\mathfrak{x}^+ = \mathfrak{x} + \mathfrak{t}^+$ by making $\delta \downarrow 0$ in the diffusion with infinitesimal operator

$$\mathbf{G} = \frac{1}{2}\mathbf{D}^2 + (k/2\delta) \times (\text{the indicator function of } 0 \leqslant x < \delta)\mathbf{D},$$

though you would expect to do so for $k = 1$: in fact, you get a motion of the form

$$\mathfrak{x}^* = \mathfrak{x} + \big[(e^k - 1)/(e^k + 1)\big]\mathfrak{t}^*$$

not confined to the half-line. The moral is that the multiplier $(e^k - 1)(e^k + 1)^{-1} < 1$ is not large enough to prevent the traveller from entering $x < 0$.

14.2.8. Elastic Brownian Motion. The presentation of the elastic Brownian motion BM^γ advertised in Sect. 14.1.4 will now be explained. You take the reflecting Brownian path $\mathfrak{x}^+$ and kill it at time $\mathfrak{m}_\infty$, subject to the conditional law

$$P\big[\mathfrak{m}_\infty > t \mid \mathfrak{x}^+\big] = e^{-\gamma \mathfrak{t}^+(t)},$$

and you prove that the killed motion $\mathfrak{x}^\gamma$ is a presentation of the elastic Brownian motion associated with $\gamma f(0) + f^+(0) = 0$.

Amplification 14.2.2. $\mathfrak{m}_\infty$ is called an exponential local holding time because it is distributed like $\mathfrak{t}^{-1}(\mathfrak{e})$, in which $\mathfrak{e}$ is an exponential holding time with rate γ, independent of $\mathfrak{x}^+$.

Proof. The first task is to verify that $\mathfrak{x}^\gamma$, so presented, begins afresh at stopping times, which is easy; see Itô-McKean [3, pp. 45–46] for details. Then you compute the transition probabilities

$$\begin{aligned} P_x\big[\mathfrak{x}^\gamma(t) \leqslant y\big] &= P_x\big[\mathfrak{x}^+(t) \leqslant y,\ t < \mathfrak{m}_\infty\big] \\ &= E_x\big[\mathfrak{x}^+(t) \leqslant y,\ e^{-\gamma \mathfrak{t}^+(t)}\big] \\ &= P_x\big[\mathfrak{x}(t) < y,\ t < \mathfrak{m}_0\big] + \int_0^t P_x\big[\mathfrak{m}_0 \in ds\big] \\ &= P_0\big[\mathfrak{x}^+(t-s) \leqslant y,\ e^{-\gamma \mathfrak{t}^+(t-s)}\big]. \end{aligned}$$

Now the identification follows from the formula of Sect. 14.1.4 and the joint law of $\mathfrak{x}^+$ and $\mathfrak{t}^+$, employed at the end of Sect. 14.2.5; for additional information on this subject, see Itô-McKean [2], in which the local time is used in a similar style to make the general Brownian motion on the half-line attached to the infinitesimal operator $\mathbf{G} = \frac{1}{2}\mathbf{D}^2$ acting upon functions f of class $\mathsf{C}^2[0,\infty)$ with

$$p_1 f(0) + p_2 f^+(0) + p_3 \mathbf{G} f(0) = \int_0^\infty \big[f(x) - f(0)\big]\, dp_4(x)$$

for fixed non-negative p_1, p_2, p_3, and increasing $p_4(x)$, subject to $\int_0^1 x\, dp_4(x) < \infty$, $\int_1^\infty dp_4(x) < \infty$, and p_1, p_2, p_3, p_4 not all trivial. □

14.3. Existence of Standard Brownian Local Times

The purpose of this section is to prove that for the standard Brownian motion $\mathfrak{x}$,

$$\text{measure}\big(s \leqslant t : a \leqslant \mathfrak{x}(s) < b\big) = 2\int_a^b \mathfrak{t}(t,x)dx,$$

simultaneously for all $t \geqslant 0$ and $-\infty < a < b < \infty$, with a density function $\mathfrak{t}$, alias the standard Brownian local time, which is continuous in the pair $(t,x) \in [0,\infty) \times R^1$. The fact is due to Trotter [11]; it was also stated by Lévy [5, p. 239]. The existence of the reflecting Brownian local time

$$\mathfrak{t}^+(t) = \lim_{\epsilon \downarrow 0}(2\epsilon)^{-1}\,\text{measure}\big(s \leqslant t : \mathfrak{x}^+(s) < \epsilon\big) = 2\mathfrak{t}(t,0)$$

of Sect. 14.2.3 is a self-evident consequence. Deeper properties of $\mathfrak{t}$ are studied in Sect. 14.4 from which a different existence proof could be extracted, but it seems that the method employed below is simpler if you want the present theorem, only. The proof is based upon the formula of H. Tanaka [private communication]:

$$\big[\mathfrak{x}(t) - x\big]^+ - \big[\mathfrak{x}(0) - x\big]^+ = \int_0^t e_{x\infty}\big[\mathfrak{x}(s)\big]\,d\mathfrak{x}^+(s) + \mathfrak{t}(t,x),$$

in which y^+ stands for the larger of y and 0 and $e_{ab}(x)$ is the indicator of the interval $a \leqslant x < b$. The reader is assumed to be familiar with such Brownian integrals and with Itô's lemma which states that

$$f\big[\mathfrak{x}(t)\big] - f\big[\mathfrak{x}(0)\big] = \int_0^t f'\big[\mathfrak{x}(s)\big]\,d\mathfrak{x}(s) + \frac{1}{2}\int_0^t f''\big[\mathfrak{x}(s)\big]\,ds$$

for functions f of class $\mathsf{C}^2(R^1)$; see, for example, McKean [7]. Tanaka's formula is suggested by applying Itô's lemma to the function $f(y) = [y - x]^+$. This is *not* a proof because f'' is not even a bonafide function, but it suggests that you should *declare*

$$\mathfrak{t}(t,x) = \big[\mathfrak{x}(t) - x\big]^+ - \big[\mathfrak{x}(0) - x\big]^+ - \int_0^t e_{x\infty}\big[\mathfrak{x}(s)\big]\,d\mathfrak{x}(s)$$

and try to prove that $\mathfrak{t}$, so presented, actually *is* the local time for $\mathfrak{x}$. The proof is indicated below; for more details, see McKean [7, pp. 68–71].

Proof. A preliminary difficulty to be overcome is that the Brownian integrals are defined for each $x \in R^1$, separately, and you want to look at them for all $(t,x) \in [0,\infty) \times R^1$, simultaneously. Now for any nonanticipating Brownian functional e,

$$E\left[\left(\int e\,d\mathfrak{x}\right)^4\right] \leqslant 36E\left[\left(\int e^2\,dt\right)^2\right],$$

whence the local bound

$$E_0\left|\int_0^{t_1} e_{a\infty}(\mathfrak{x})d\mathfrak{x} - \int_0^{t_2} e_{b\infty}(\mathfrak{x})d\mathfrak{x}\right|^4 \leqslant \text{a constant multiple of } \big(t_2 - t_1\big)^2 + (b-a)^2,$$

and with this in hand, the familiar lemma of Čentsov-Kolmogorov supplies you with a version $\mathfrak{e}(t,x)$ of the Brownian integral which is almost surely continuous in the pair $(t,x) \in [0,\infty) \times R^1$. To identify

$$\mathfrak{t}(t,x) = \big[\mathfrak{x}(t) - x\big]^+ - \big[\mathfrak{x}(0) - x\big]^+ - \mathfrak{e}(t,x)$$

as the local time of $\mathfrak{x}$, fix $t \geqslant 0$ and $-\infty < a < b < \infty$. Then

$$\int_a^b \big[\mathfrak{x}(t) - x\big]^+ dx - \int_a^b \big[\mathfrak{x}(0) - x\big]^+ dx$$
$$= \int_a^b dx \int_{\mathfrak{x}(0)}^{\mathfrak{x}(t)} e_{x\infty}(y)dy = \int_{\mathfrak{x}(0)}^{\mathfrak{x}(t)} dy \int_{-\infty}^{y} e_{ab}(x)dx,$$

as you will easily see from a picture. A small extension of Itô's lemma plus a little fooling about with Riemann sums permits you to re-express this as

$$\int_0^t \left[\int_0^{\mathfrak{x}} e_{ab}(x)dx\right] d\mathfrak{x} + \frac{1}{2}\int_0^t e_{ab}(\mathfrak{x})ds$$
$$= \int_a^b dx \int_0^t e_{x\infty}(\mathfrak{x})d\mathfrak{x} + \frac{1}{2}\int_0^t e_{ab}(\mathfrak{x})ds$$
$$= \int_a^b \mathfrak{e}(t,x)dx + \frac{1}{2}\,\text{measure}\,\big(s \leqslant t : a \leqslant \mathfrak{x}(s) < b\big).$$

But now what sits at the beginning and at the end is continuous in $(t, a, b) \in [0, \infty) \times R^2$, so the formula holds without exception, for all $t \geqslant 0$ and $a \leqslant b$, simultaneously. The proof is finished. □

AMPLIFICATION 14.3.1. A little extra computation produces a modulus of continuity for $\mathfrak{t}$ due to Ray [8, p. 615]:

$$\big|\mathfrak{t}(t,y) - \mathfrak{t}(t,x)\big| \leqslant \big\|\mathfrak{t}(t,\cdot)\big\|_\infty |2\delta \lg \delta|^{1/2}$$

for $\delta = |x - y| \downarrow 0$ and fixed $t > 0$; see McKean [7, p. 70]. The modulus is actually sharp, as may be confirmed by means of the deeper results of Ray [8] explained in Sect. 14.4.5.

14.4. Deeper Properties of Brownian Local Time

The standard Brownian local time $\mathfrak{t} = \mathfrak{t}(t, x)$ has an even more recondite property than any described above. I allude to the remarkable fact that if the stopping time $\mathfrak{m}$ is either a passage time or an independent exponential holding time, then $\mathfrak{t}(\mathfrak{m}, x)$ is a diffusion relative to the *spatial* parameter $x \in R^1$, closely related to the two- and four-dimensional Bessel motions. This development is associated with the names of Knight [4], Ray [8], Silverstein [9], and Williams [12, 13, 14]. I will follow the methods of Williams, as I find them most appealing, but all of these papers will replay close study. Here, I will explain less and assume more; in particular, I count upon familiarity with the time substitution recipe of Itô-McKean [3]; see also Breiman [1, pp. 370–375] for a nicer account.

14.4.1. D. Williams: Markovian Properties of Brownian Local Time. Define $\mathfrak{f}(t) = \text{measure}(s \leqslant t : \mathfrak{x}(s) \geqslant 0)$ for a standard Brownian motion $\mathfrak{x}$ starting at $x \geqslant 0$. Itô-McKean [3, pp. 81–82] prove that $\mathfrak{x}^*(t) = \mathfrak{x}[\mathfrak{f}^{-1}(t)]$ is still another presentation of the reflecting Brownian motion BM^+. By substituting the "clock" $\mathfrak{f}^{-1}$ into $\mathfrak{x}$, you are throwing out the left-hand (shaded) excursions in Fig. 14.3 and pushing down the survivors to close up the gaps so produced. The local time of $\mathfrak{x}^*$ at $x = 0$ is

$$\mathfrak{t}^*(t) = \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\,\text{measure}\,\big(s \leqslant t : \mathfrak{x}^*(s) < \epsilon\big) = \mathfrak{t}_0\big(\mathfrak{f}^{-1}\big),$$

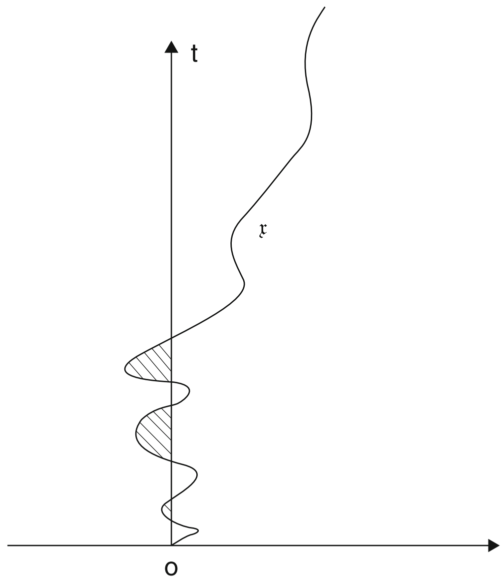

FIGURE 14.3.

$\mathfrak{t}_0$ being the local time for $\mathfrak{x}$ at $x = 0$, and for $\gamma > 0$ you have, what is not so easy to see, the formula of Williams [7]:

$$E_0\big[e^{-\gamma \mathfrak{f}^{-1}(t)} \mid \mathfrak{x}^*\big] = e^{-\gamma t - (2\gamma)^{1/2}\mathfrak{t}^*(t)}.$$

This is not needed below, but it is simple enough (if you know how) and will give you the flavor of what is to follow.

PROOF. To begin with, $\mathfrak{f}^{-1}(t) = \mathfrak{t}_0^{-1}[\mathfrak{t}^*(t)]$, and as you can see from the picture, $\mathfrak{t}_0^{-1}$ is identical in law to $\mathfrak{t}^{*-1}$ plus an independent copy thereof, namely, an independent copy $\mathfrak{m}$ of standard Brownian passage times. But now $\mathfrak{f}^{-1}(t)$ is presented as the sum of t and an independent copy of $\mathfrak{m}$ with parameter $x = \mathfrak{t}^*(t)$, so

$$E_0\big[e^{-\gamma \mathfrak{f}^{-1}(t)} \mid \mathfrak{x}^*\big] = E_0\big[e^{-\gamma t}e^{-\gamma \mathfrak{m}_x} \mid \mathfrak{x}^*\big] = e^{-\gamma t - (2\gamma)^{1/2}\mathfrak{t}^*(t)}.$$

The proof is finished. □

BONUS 1. By taking expectations on both sides, you can easily deduce the arcsine law of Lévy [5, p. 323]:

$$P_0\big[\mathfrak{f}(t) \leqslant T\big] = \frac{2}{\pi}\sin^{-1}(t/T)^{1/2}$$

for $t \leqslant T$.

A deeper formula of Williams [12] is

$$E_0\big[e^{-\gamma \mathfrak{t}_b(\mathfrak{f}^{-1}(t))} \mid \mathfrak{x}^*\big] = \exp\left[-\frac{\gamma \mathfrak{t}^*(t)}{1-\gamma b}\right],$$

for $b < 0$, in which $\mathfrak{t}_b$ is standard Brownian local time at $x = b$.

Proof. $\mathfrak{f}^{-1}(t)$ is identical in law to t plus an independent copy of $\mathfrak{m}$ with parameter $\mathfrak{t}^*(t)$, as above. Because $\mathfrak{t}_b(\mathfrak{f}^{-1})$ is flat on the right-hand excursions of $\mathfrak{x}$, you can identify it with $\mathfrak{t}_b[\mathfrak{t}_0^{-1}(\mathfrak{t}^*)]$, in which $\mathfrak{t}^*$ is fixed while $\mathfrak{t}_0$ and $\mathfrak{t}_b$ are based upon an independent standard Brownian motion. A careful look at Fig. 14.3 will confirm this: the chief point is that $\mathfrak{x}^*$ is independent of the left-hand excursions of $\mathfrak{x}$, while it is these, only, that account for $\mathfrak{t}_b$. The formula now follows from the evaluation

$$E_0\left[e^{-\gamma\mathfrak{t}_b(\mathfrak{t}_0^{-1}(t))}\right] = \exp\left[\frac{-\gamma t}{1-\gamma b}\right]$$

of Itô-McKean [3, Problem 2.8.4]. □

An immediate consequence is the theorem of Ray [8] and Knight [4] that the process $[\mathfrak{t}(\mathfrak{m}_1, 1-x) : 0 \leqslant x \leqslant 1, P_0]$ is a copy of $[\frac{1}{2}\mathfrak{r}_2^2(x) : 0 \leqslant x \leqslant 1]$, in which $\mathfrak{r}_2$ is the two-dimensional Bessel process, i.e., the radial part of a two-dimensional Brownian motion; see Itô-McKean [3, Problems 2.8.5 and 2.8.6] for another proof.

Proof. Williams' second formula is obviously still correct if you replace t by a stopping time of $\mathfrak{x}^*$ such as $\mathfrak{m}_1^* = \min(t : \mathfrak{x}^*(t) = 1) = \mathfrak{f}(\mathfrak{m}_1)$, so you have

$$E_0\left[e^{-\gamma\mathfrak{t}_b(\mathfrak{m}_1)} \mid \mathfrak{x}^*\right] = \exp\left[-\frac{\gamma\mathfrak{t}_0(\mathfrak{m}_1)}{1-\gamma b}\right].$$

The fact that $\mathfrak{t}(\mathfrak{m}_1, \cdot)$ is Markovian is now self-evident, $\mathfrak{t}(\mathfrak{m}_1, x) : 0 \leqslant x \leqslant 1$ being measurable over $\mathfrak{x}^*$, and now it is a simple exercise to identify the latter as $\frac{1}{2}\mathfrak{r}_2^2$; for example, it is easy to pick off the infinitesimal operator $\mathbf{G} = t\mathbf{D}^2 + \mathbf{D}$. □

14.4.2. Aside on Cameron-Martin's Formula. An amusing side product of the presentation $\mathfrak{t}(\mathfrak{m}_1, 1-x) = \frac{1}{2}\mathfrak{r}_2^2(x)$ is the formula of Cameron-Martin

$$E_0\left[e^{-\gamma\int_0^1 \mathfrak{x}^2(t)dt}\right] = \left(\cosh(2\gamma)^{1/2}\right)^{-1/2}.$$

Proof. $\mathfrak{r}_2^2$ is the sum of two independent copies of $\mathfrak{x}^2$, so

$$\begin{aligned}
\left(E_0\left[e^{-\gamma\int_0^1 \mathfrak{x}^2(t)dt}\right]\right)^2 &= E_0\left[e^{-\gamma\int_0^1 \mathfrak{r}_2^2(x)dx}\right] \\
&= E_0\left[e^{-\gamma\int_0^1 2\mathfrak{t}(\mathfrak{m}_1,x)dx}\right] \\
&= E_0\left[e^{-\gamma\mathfrak{f}(m_1)}\right] \\
&= E_0\left[e^{-\gamma\mathfrak{m}_1^*}\right],
\end{aligned}$$

in which $\mathfrak{m}_1^*$ is the passage time of the reflecting Brownian motion $\mathfrak{x}^*$ to 1, alias the passage time $\mathfrak{m}$ of a standard Brownian motion $\mathfrak{x}$ to ± 1. The evaluation now follows from the fact that $\exp[(2\gamma)^{1/2}\mathfrak{x} - \gamma t]$ is a martingale over $\mathfrak{x}$ and from the independence of $\mathfrak{x}(\mathfrak{m})$ and $\mathfrak{m}$:

$$1 = E_0\left[e^{(2\gamma)^{1/2}\mathfrak{x}(\mathfrak{m})-\gamma\mathfrak{m}}\right] = \cosh(2\gamma)^{1/2}E_0\left[e^{-\gamma\mathfrak{m}}\right].$$

□

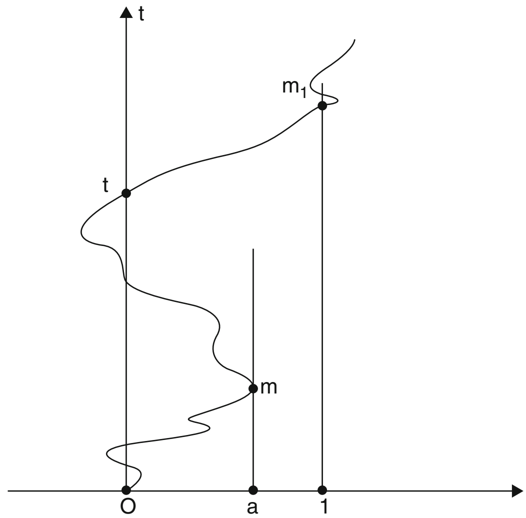

FIGURE 14.4.

14.4.3. D. Williams Continued. The present article looks at $[\mathfrak{t}(\mathfrak{m}_1, x) : 0 \leqslant x \leqslant 1, P_0]$ in more detail and then at $[\mathfrak{t}(\mathfrak{m}_1, x) : x \leqslant 0, P_0]$, following Williams [13] in part. Bring in the last leaving time $\mathfrak{l} = \max(t < \mathfrak{m}_1 : \mathfrak{x}(t) = 0)$ and the maximum $\mathfrak{a}$ of $\mathfrak{x}(t)$ for $t \leqslant \mathfrak{l}$, as in Fig. 14.4, and let $\mathfrak{m}$ be *the* root $t < \mathfrak{l}$ of $\mathfrak{x}(t) = \mathfrak{a}$; see Itô-McKean [3, Problem 2.2.2] for a proof that there is no other such root. The number $0 < \mathfrak{a} < 1$ is uniformly distributed, and conditional upon its value $\mathfrak{a} = a$, the excursions $\mathfrak{x}(t) : 0 \leqslant t \leqslant \mathfrak{m}$ and $\mathfrak{x}(\mathfrak{l} - t) : 0 \leqslant t \leqslant \mathfrak{l} - \mathfrak{m}$ look like independent copies of $\mathfrak{x}(t) : t \leqslant \mathfrak{m}_a$, as you can easily believe but less easily prove.[2] Therefore, by Sect. 14.4.1, $\mathfrak{t}(\mathfrak{m}, x) : 0 \leqslant x \leqslant 1$ and $\mathfrak{t}(\mathfrak{l}, x) - \mathfrak{t}(\mathfrak{m}, x) : 0 \leqslant x \leqslant 1$ look like independent copies of $\frac{1}{2}\mathfrak{r}_2^2[(a-x)^+]$, while their sum looks like $\frac{1}{2}\mathfrak{r}_4^2[(a-x)^+]$, in which $\mathfrak{r}_4$ is a four-dimensional Bessel process, i.e., the radial part of a four-dimensional Brownian motion.

A similar presentation of $\mathfrak{t}(\mathfrak{m}_1, x) - \mathfrak{t}(\mathfrak{l}, x) : 0 \leqslant x \leqslant 1$ can be obtained, but it is not so cheap. The reversed excursion $1 - \mathfrak{x}(\mathfrak{m}_1 - t) : 0 \leqslant t \leqslant m_1 - \mathfrak{l}$ looks like a standard Brownian motion conditioned so as not to come back to $x = 0$, and stopped at its passage time to $x = 1$; as such, it is identical in law to the three-dimensional Bessel process $\mathfrak{r}_3$ similarly stopped; see McKean [6] for a full proof. An indication is provided by the evaluation

$$\lim_{T\uparrow\infty} P_x\big[\mathfrak{x}(t) \in dy \mid \mathfrak{m}_0 > T\big]$$
$$= P_x\big[\mathfrak{x}(t) \in dy,\ \mathfrak{m}_0 > t\big] \lim_{T\uparrow\infty} \frac{P_y\big[\mathfrak{m}_0 > T - t\big]}{P_x\big[\mathfrak{m}_0 > T\big]}$$
$$= P_x\big[\mathfrak{x}(t) \in dy,\ \mathfrak{m}_0 > t\big](y/x),$$

[2]D. Williams [May 8, 1972] writes: "I still cannot find a *sane* proof of [this.]—To get a sane proof, I formulated a Splitting Time Theorem that good Markov processes start afresh at 'splitting times' (but with new laws). You of all people should recognize the terminology! M. Jacobsen (visiting here from Copenhagen) has a nice Galmarino-type definition of splitting times, but we cannot get near the theorem. All the $\mathfrak{l}$'s and $\mathfrak{m}$'s in your paper *are* splitting times so the result would be useful."

which suggests that the infinitesimal operator of the Brownian motion, so conditioned, is, or ought to be,

$$\mathbf{G}_3 = \frac{1}{2}x^{-1}\mathbf{D}^2 x = \frac{1}{2}\mathbf{D}^2 + x^{-1}\mathbf{D},$$

and that is precisely the generator of $\mathfrak{r}_3$. By symmetry, the excursion $\mathfrak{x}(t) : \mathfrak{l} \leqslant t \leqslant \mathfrak{m}_1$ may now be presented as $\mathfrak{r}_3(t) : 0 \leqslant t \leqslant \mathfrak{m}_3$, in which $\mathfrak{m}_3$ is the passage time $\min(t : \mathfrak{r}_3(t) = 1)$, so to finish the story of $\mathfrak{t}(\mathfrak{m}_1, x) : 0 \leqslant x \leqslant 1$, you have only to find the proper presentation of the Bessel local time

$$\mathfrak{t}_3(x) = \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\,\text{measure}\big(t \leqslant \mathfrak{m}_3 : x \leqslant \mathfrak{r}_3(t) < x+\epsilon\big),$$

alias $\mathfrak{t}(\mathfrak{m}_1, x) - \mathfrak{t}(\mathfrak{l}, x)$, for $0 \leqslant x \leqslant 1$. The result of Ray [8] and Knight [4] is that this process is a copy of $\frac{1}{2}x^2\mathfrak{r}_2^2[x^{-1}(1-x)] : 0 \leqslant x \leqslant 1$. Here you may recognize $x\mathfrak{r}_2[x^{-1}(1-x)]$ as a representation of the two-dimensional Bessel process starting at $\mathfrak{r}_2 = 0$ and conditioned so as to come back to $\mathfrak{r}_2 = 0$ at time 1, as you infer from the fact that $x\mathfrak{x}[x^{-1}(1-x)]$ is a representation of the standard Brownian motion, so conditioned.

Proof of $\mathfrak{t}_3(x) = \frac{1}{2}x^2\mathfrak{r}_2^2[x^{-1}(1-x)]$. The recipe of time substitutions of Itô-McKean [3, pp. 167–170] shows that in the scale $y = -1/r_3$, the three-dimensional Bessel process looks like a standard Brownian motion $\mathfrak{n}$ run with the clock $\mathfrak{f}^{-1}$ inverse to $\mathfrak{f}(t) = \int_0^l \mathfrak{n}^{-4}(s)ds$. The passage time $\mathfrak{m}_3$ is now presented as $\mathfrak{f}(\mathfrak{m}_{-1})$ with $\mathfrak{m}_{-1} = \min(t : \mathfrak{n}(t) = -1)$, so that if e is the indicator function of $-1/x \leqslant y < -1/(x+\epsilon)$, then you have

$$\begin{aligned}
\mathfrak{t}_3(x) &= \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\,\text{measure}\big(t \leqslant \mathfrak{m}_3 : x \leqslant \mathfrak{r}_3(t) < x+\epsilon\big)\\
&= \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\,\text{measure}\left(t \leqslant \mathfrak{f}(\mathfrak{m}_{-1}) : \frac{-1}{x} \leqslant \mathfrak{n}\big[\mathfrak{f}^{-1}(t)\big] < \frac{-1}{x+\epsilon}\right)\\
&= \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\int_0^{\mathfrak{f}(\mathfrak{m}_{-1})} e\big[\mathfrak{n}\big(\mathfrak{f}^{-1}\big)\big]\,dt\\
&= \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\int_0^{\mathfrak{m}_{-1}} e\big[\mathfrak{n}(t)\big]\,d\mathfrak{f}(t)\\
&= \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\int_0^{\mathfrak{m}_{-1}} e\big[\mathfrak{n}(t)\big]\,\mathfrak{n}^{-4}(t)dt\\
&= \lim_{\epsilon\downarrow 0}(2\epsilon)^{-1}\int_{-1/x}^{-1/(x+\epsilon)} \mathfrak{t}\big(\mathfrak{m}_{-1}, y\big)y^{-4}2\,dy\\
&= x^2\mathfrak{t}\big(\mathfrak{m}_{-1}, -1/x\big),
\end{aligned}$$

in which $\mathfrak{t}$ is local time for $\mathfrak{n}$. But, by Sect. 14.4.1, this means that $\mathfrak{t}_3$ may be presented in the advertised form:

$$x^2 \times \frac{1}{2}\mathfrak{r}_2^2\left[-1-\left(-\frac{1}{x}\right)\right] = \frac{1}{2}x^2\mathfrak{r}_2^2\big[x^{-1}(1-x)\big].$$

The proof is finished. □

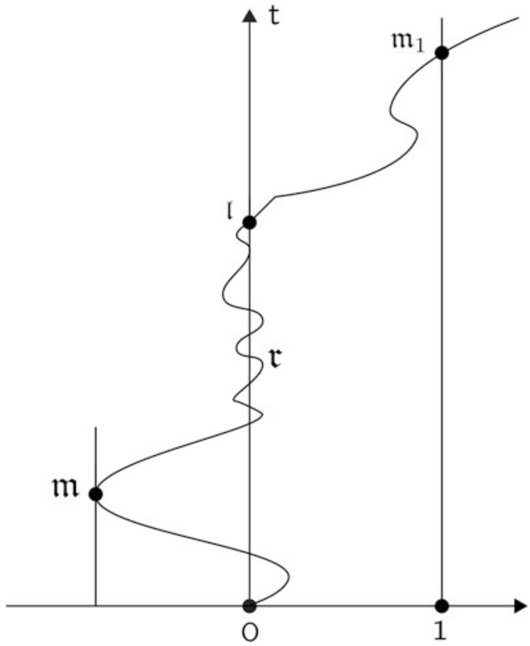

FIGURE 14.5.

The same line of reasoning may be applied to $\mathfrak{t}(\mathfrak{m}_1, x)$ for $x \leqslant 0$. The minimum $\mathfrak{b}$ of $\mathfrak{x}(t)$ for $t \leqslant \mathfrak{l}$ is distributed according to the law $P_0[\mathfrak{b} < x] = (1-x)^{-1}$ for $x \leqslant 0$, and conditional upon $\mathfrak{b} = b$, the excursion $1 - \mathfrak{x}(\mathfrak{m}_1 - t) : 0 \leqslant t \leqslant \mathfrak{m}_1 - \mathfrak{m}$ looks like $\mathfrak{r}_3$ stopped at its passage time to $1 - b$, $\mathfrak{m}$ being *the* root of $\mathfrak{x} = b$; see Fig. 14.5. Therefore, $\mathfrak{t}(\mathfrak{m}_1, x) - \mathfrak{t}(\mathfrak{m}, x) : x \leqslant 0$ can be presented as

$$(1-x)^2 \mathfrak{t}\big[\mathfrak{m}_{-(1-b)^{-1}},\ -(1-x)^{-1}\big],$$

alias

$$\frac{1}{2}(1-x)^2 \mathfrak{r}_2^2\big[(1-x)^{-1} - (1-b)^{-1}\big].$$

Besides, it is only the excursion $\mathfrak{x}(\mathfrak{l} - t) : 0 \leqslant t \leqslant \mathfrak{l} - \mathfrak{m}$ that contributes, and the excursion $\mathfrak{x}(t) : 0 \leqslant t \leqslant \mathfrak{m}$ is an independent copy thereof, *so the total local time* $\mathfrak{t}(\mathfrak{m}_1, x) : x \leqslant 0$ *can be presented as*

$$\frac{1}{2}(1-x)^2 \mathfrak{r}_4^2\big[(1-x)^{-1} - (1-b)^{-1}\big],$$

a fact which is also due to Ray [8] and Knight [4].

14.4.4. A Time Reversal. To prepare for the next article, it is convenient to prove at this place that if $\mathfrak{e}$ is an exponential holding time with rate $\frac{1}{2}$ and if $\mathfrak{x}$ is an independent standard Brownian motion starting at $x = 0$, then conditional upon the place $\mathfrak{a} = \mathfrak{x}(\mathfrak{e}) > 0$, the reversed process $\mathfrak{x}^{-1}(t) = \mathfrak{x}(\mathfrak{e} - t)\ 0 \leqslant t \leqslant \mathfrak{e}$ is identical in law to the motion $\mathfrak{a} + \mathfrak{x}(t) - t$ stopped at its last leaving time from $x = 0$.

PROOF. The main step is to compute

$$\begin{aligned}
&P_0\big[\mathfrak{x}^{-1}(t_1)\in dx_1,\dots,\mathfrak{x}^{-1}(t_n)\in dx_n,\ t_n<\mathfrak{e}\mid \mathfrak{x}(\mathfrak{e})=\mathfrak{a}\big]\\
&=\frac{\int_{t_n}^{\infty}P[\mathfrak{e}\in dt]P_0\big[\mathfrak{x}(t-t_1)\in dx_1,\dots,\mathfrak{x}(t-t_n)\in dx_n,\ \mathfrak{x}(t)\in da\big]}{P_0\big[\mathfrak{x}(\mathfrak{e})\in da\big]}\\
&=e^a\int_{t_n}^{\infty}e^{-t/2}dtP_0\big[\mathfrak{x}(t-t_1)\in dx_1,\dots,\mathfrak{x}(t-t_n)\in dx_n\mid \mathfrak{x}(t)=a\big]\frac{e^{-a^2/2t}}{(2\pi t)^{1/2}}\\
&=e^a\int_{t_n}^{\infty}e^{-t/2}dtP_a\big[\mathfrak{x}(t_1)\in dx_1,\dots,\mathfrak{x}(t_n)\in dx_n\mid \mathfrak{x}(t)=0\big]\frac{e^{-a^2/2t}}{(2\pi t)^{1/2}}\\
&=e^a\int_{t_n}^{\infty}e^{-t/2}dtP_a\big[\mathfrak{x}(t_1)\in dx_1,\dots,\mathfrak{x}(t_n)\in dx_n\big]\frac{e^{-x_n{}^2/2(t-t_n)}}{\big(2\pi(t-t_n)\big)^{1/2}}\\
&=P_a\big[\mathfrak{x}(t_1)\in dx_1,\dots,\mathfrak{x}(t_n)\in dx_n\big]\times e^{a-x_n-t_n/2},
\end{aligned}$$

for $x_n>0$. The factor $\exp(a-x_n-t_n/2)$ is now brought underneath the expectation sign in the form $\exp(-[\mathfrak{x}(t_n)-x(0)]-t_n/2)$, in which you recognize the Cameron-Martin factor for the Brownian motion with drift -1. The rest of the proof is easy. □

14.4.5. D. Williams Continued. The purpose of this article is to prove, after the manner of Williams [14], the results of Ray [8] concerning the standard Brownian local times $\mathfrak{t}(\mathfrak{e},x):0\leqslant x\leqslant\infty$ for $\mathfrak{e}$ an exponential holding time with rate $\frac{1}{2}$ which is independent of $\mathfrak{x}$, everything being considered *conditional upon the place* $\mathfrak{a}=\mathfrak{x}(\mathfrak{e})>0$. The local time is unchanged if you go over to the reversed process $\mathfrak{x}^{-1}(t)=\mathfrak{x}(\mathfrak{e}-t):0\leqslant t\leqslant\mathfrak{e}$ of Sect. 14.4.4, so you may as well look at the local time of $\mathfrak{n}=\mathfrak{a}+\mathfrak{x}(t)-t$ up to its last leaving time from $x=0$. The situation is depicted in Fig. 14.6, in which $\mathfrak{b}=\max_{t\geqslant 0}\mathfrak{n}(t)$, $\mathfrak{m}$ is *the* root of $\mathfrak{n}(t)=b$, and $\mathfrak{l}$ is the last leaving time of $\mathfrak{n}$ from $0<\mathfrak{a}<\mathfrak{b}$. The recipe of time substitutions is now used much as in Sect. 14.4.3: In the scale $x=\frac{1}{2}e^{2y}$, $\mathfrak{n}$ looks like a new standard Brownian motion $\mathfrak{x}$ run with the clock $\mathfrak{f}^{-1}$ inverse to $\mathfrak{f}(t)=\int_0[4\mathfrak{x}^2(s)]^{-1}\,ds$, and you deduce that *the original*

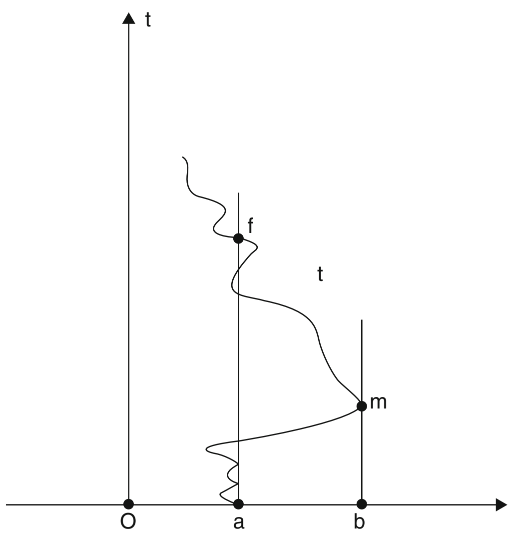

FIGURE 14.6.

Brownian local times $\mathfrak{t}(\mathfrak{e}, x)$ can be presented as

$$\frac{1}{2}e^{-2x}\mathfrak{r}_2^2\big(e^{2x}/2\big) \quad \text{if } 0 \leqslant x \leqslant a$$

and as

$$\frac{1}{8}e^{2x}\mathfrak{r}_4^2\big[2\big(e^{-2x} - e^{-2b}\big)\big] \quad \text{if } a \leqslant x \leqslant b.$$

The details are left to you as an exercise.

REMARK 14.4.1. The fact that Ray's modulus for $\mathfrak{t}$, described in Sect. 14.3, is exact is an immediate consequence of these presentations.

Bibliography

[1] L. Breiman. *Probability*. Addison-Wesley, Massachussetts, 1968.

[2] K. Itô and H. P. McKean. Brownian motion on a half line. *Ill. J. Math.*, 7:181–231, 1963.

[3] K. Itô and H. P. McKean. *Diffusion Processes and Their Sample Paths*. Academic Press, 1965.

[4] F. B. Knight. Random walks and a sojourn density process for Brownian motion. *Trans. AMS*, 167:56–86, 1963.

[5] P. Lévy. *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, Paris, 1948.

[6] H. P. McKean. Excursions of a non-singular diffusion. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 1:230–239, 1963.

[7] H. P. McKean. *Stochastic Integrals*, volume 5 of *Probability and Mathematical Statistics*. Academic Press, 1969.

[8] D. B. Ray. Sojourn times of a diffusion process. *Illinois J. Math.*, 7:615–630, 1963.

[9] M. L. Silverstein. A new approach to local times. *J. Math. Mech.*, 17:1023–1054, 1968.

[10] A. V. Skorohod. Stochastic equations for diffusion processes with a boundary. *Teor. Verojatnost. i Primenen*, 6:287–298, 1961.

[11] H. F. Trotter. A property of Brownian motion paths. *Illinois J. Math.*, 2:425–433, 1958.

[12] D. Williams. Markov properties of Brownian local time. *Bull. Amer. Math. Soc.*, 75:1035–1036, 1964.

[13] D. Williams. Decomposition of the Brownian path. *Bull. Amer. Math. Soc.*, 70:871–973, 1970.

[14] D. Williams. Markov properties of Brownian local time (2). *Unpublished*, 1971.

[15] Brownian Motions on a Half Line

[15] Brownian Motions on a Half Line. *Ill. Jour.* **7** (1963), 181–231.

15 Brownian Motions on a Half Line

K. Itô and H. P. McKean, Jr.[1]

Dedicated to W. Feller

'Numbering (in italics). (1) means formula 1 in the present section; (2.1) means formula (1) of Section 15.2, etc.'

15.1. The Classical Brownian Motions

Consider the space of all (continuous) *sample paths* $w : [0, +\infty) \to R^1$ with coordinates $\mathfrak{x}(t, w) = \mathfrak{x}(t)$ $(t \geqq 0)$, the field $\mathbf{A}$ of events

(1) $$B = w^{-1}_{t_1 t_2 \cdots t_n}(A) = \big(w : \big(\mathfrak{x}(t_1), \mathfrak{x}(t_2), \ldots, \mathfrak{x}(t_n)\big) \in A\big)$$
$$0 < t_1 < t_2 < \cdots < t_n, \quad A \in \mathbf{B}\big(R^n\big),\ n \geqq 1,$$ [2]

and the Gauss kernel

(2) $$g(t, a, b) = e^{-(b-a)^2/2t}/(2\pi t)^{1/2}, \quad (t, a, b) \in (0, +\infty) \times R^2.$$

Because of

(3a) $$g(t, a, b) > 0,$$

(3b) $$\int g(t, a, b) db = 1,$$

(3c) $$g(t, a, b) = \int g(t - s, a, c) g(s, c, b) dc \quad (t > s),$$

the function

(4) $$P_a(B) = \int_A g\big(t_1, a, b_1\big) g\big(t_2 - t_1, b_1, b_2\big) \cdots g\big(t_n - t_{n-1}, b_{n-1}, b_n\big) \cdot db_1 db_2 \cdots db_n$$

of $B = w^{-1}_{t_1 t_2 \cdots t_n}(A) \in \mathbf{A}$ is well-defined, nonnegative, additive, and of total mass +1 for each $a \in R^1$, and, as N. Wiener [20] discovered, the estimate

(5) $$\int_{|a-b|>\varepsilon} g(t, a, b)\, db < \text{constant} \times \varepsilon^{-1} t^{1/2} e^{-\varepsilon^2/2t}, \quad t \downarrow 0$$

permits us to extend it to a nonnegative Borel measure $P_a(B)$ of total mass +1 on the Borel extension $\mathbf{B}$ of $\mathbf{A}$ (see P. Lévy [13] for an alternative proof).

---

[1]Fulbright grantee 1957–1958 during which time the major part of this material was obtained; the support of the Office of Naval Research, U.S. Govt. during the summer of 1961 is gratefully acknowledged also.

[2]$B(R^n)$ is the usual topological Borel field of the n-dimensional euclidean space R^n.

Granting this, it is apparent that $P_a(\mathfrak{x}(0) \in db)$ is the unit mass at $b = a$. $P_a(B)$ is now interpreted as *the chance of the event B for paths starting at the point a* and the sample path $w : t \to \mathfrak{x}(t)$ with these probabilities imposed is called *standard Brownian motion starting at a*.

Given $t \geqq 0$, if $B \in \mathbf{B}$ and if w_t^+ denotes the shifted path $w_t^+ : s \to \mathfrak{x}(t+s, w)$, then (4) implies

(6) $$P_a\big(w_t^+ \in B \mid \mathfrak{x}(s) : s \leqq t\big) = P_b(B), \quad b = \mathfrak{x}(t),$$

i.e., the *law* of the *future* $\mathfrak{x}(s) : s > t$ *conditional on the past* $\mathfrak{x}(s) : s \leqq t$ depends upon the *present* $b = \mathfrak{x}(t)$ alone (in short, *the Brownian traveller starts afresh at each constant time* $t \geqq 0$).

Because the Gauss kernel $g(t, a, b)$ is the fundamental solution of the heat flow problem

(7) $$\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial a^2}, \quad (t, a) \in (0, +\infty) \times R^1,$$

the operator $\mathfrak{G} = D^2/2$ acting on[3] $C^2(R^1)$ is said to *generate* the standard Brownian motion, and it is natural to seek other differential operators $\mathfrak{G}^\bullet$ giving rise via the fundamental solution of $\partial u/\partial t = \mathfrak{G}^\bullet u$ and the rule (4) to similar (stochastic) motions.

Consider, for example, the operator[4]

(8) $$\mathfrak{G}^+ = \mathfrak{G} \mid C^2[0, +\infty) \cap \big(u : u^+(0) = 0\big) :$$

the fundamental solution of $\partial u/\partial t = \mathfrak{G}^+ u$ is

(9) $$g^+(t, a, b) = e^{-(b-a)^2/2t}/(2\pi t)^{1/2} + e^{-(b+a)^2/2t}/(2\pi t)^{1/2}, \quad t > 0 \leqq a, b,$$

which satisfies (3a), (3b), and (3c), and the corresponding (*reflecting Brownian*) *motion* is identical in law to

(10) $$\mathfrak{x}^+ = |\mathfrak{x}|,$$

where $\mathfrak{x}$ is a standard Brownian motion.

Consider next the operator

(11) $$\mathfrak{G}^- = \mathfrak{G} \mid C^2[0, +\infty) \cap \big(u : u(0) = 0\big) :$$

the fundamental solution of $\partial u/\partial t = \mathfrak{G}^- u$ is

(12) $$g^-(t, a, b) = e^{-(b-a)^2/2t}/(2\pi t)^{1/2} - e^{-(b+a)^2/2t}/(2\pi t)^{1/2}, \quad t > 0 \leqq a, b,$$

which satisfies (3) with

(1.3bbis) $$\int g^-(t, a, b)db < 1$$

in place of (3b), and the corresponding (*absorbing Brownian*) *motion* is identical in law to

(13) $$\begin{aligned} \mathfrak{x}^-(t) &= \mathfrak{x}^+(t) \quad \text{if } t < \mathfrak{m}_0, \\ &= \infty \qquad\ \text{if } t \geqq \mathfrak{m}_0, \end{aligned}$$

where $\mathfrak{x}^+$ is the reflecting Brownian motion described above, $\mathfrak{m}_0$ is its passage time $\mathfrak{m}_0 = \min(t : \mathfrak{x}^+(t) = 0)$, and ∞ is an extra state adjoined to R^1.

Given $0 < \gamma < +\infty$, the operator

(14) $$\mathfrak{G}^\gamma = \mathfrak{G} \mid C^2[0, +\infty) \cap \big(u : \gamma u(0) = u^+(0)\big)$$

---

[3] $C^d(R^1)$ is the space of bounded continuous functions $f : R^1 \to R^1$ with d bounded continuous derivatives.

[4] $C^2[0, +\infty)$ is the space of functions $u \in C[0, +\infty)$ with $D^2 u \in C(0, +\infty)$ and $(D^2u)(0) \equiv (D^2u)(0+)$ existing. $u^+(0) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1}[u(\varepsilon) - u(0)]$.

is also possible: the fundamental solution of $\partial u/\partial t = \mathfrak{G}^\gamma u$ is

$$g^\gamma(t,a,b) = g^\gamma(t,b,a)$$

(15a)
$$= g^-(t,a,b) + \int_0^t \frac{a}{(2\pi s^3)^{1/2}} e^{-a^2/2s} g^\gamma(t-s,0,b)ds, \quad t > 0 < a,b,$$

(15b)
$$g^\gamma(t,0,0) = 2\int_0^{+\infty} e^{-\gamma c} \frac{c}{(2\pi t^3)^{1/2}} e^{-c^2/2t} dc, \quad t > 0,$$

which satisfies (3) with (1.3bbis) in place of (3b), and the corresponding (*elastic Brownian*) *motion* is identical in law to

(16a)
$$\mathfrak{x}^\gamma(t) = \mathfrak{x}^+(t) \quad \text{if } t < \mathfrak{m}_\infty,$$
$$= \infty \quad \text{if } t \geqq \mathfrak{m}_\infty,$$

(16b)
$$\mathfrak{m}_\infty = \mathfrak{t}^{-1}(\mathfrak{e}/\gamma),$$

where $\mathfrak{e}$ is an exponential holding time independent of the reflecting Brownian motion $\mathfrak{x}^+$ with law $P.(\mathfrak{e} > t) = e^{-t}$ and $\mathfrak{t}^{-1}$ is the inverse function of the *reflecting Brownian local time*:

(17)
$$\mathfrak{t}^+(t) = \lim_{\varepsilon \downarrow 0} (2\varepsilon)^{-1} \text{measure}\big(s : \mathfrak{x}^+(s) < \varepsilon, \ s \leqq t\big)$$

(see Sects. 15.3, 15.4, 15.14 for additional information about local times).

15.2. Feller's Brownian Motions

W. Feller [3] discovered that the classical Brownian generators $\mathfrak{G}^\pm$ and $\mathfrak{G}^\gamma$ $(0 < \gamma < +\infty)$ of Sect. 15.1 are the simplest members of a wide class of restrictions $\mathfrak{G}^\bullet$ of $\mathfrak{G} \mid C^2[0,+\infty)$ which generate what could be called Brownian motions on $[0,+\infty)$. Feller found that the domain $D(\mathfrak{G}^\bullet) \subset C^2[0,+\infty)$ of such a generator could be described in terms of three nonnegative numbers p_1, p_2, p_3, and a nonnegative mass distribution $p_4(dl)$ $(l > 0)$ subject to[5]

(1)
$$p_1 + p_2 + p_3 + \int_{0+} (l \wedge 1) p_4(dl) = 1$$

as follows:

(2)
$$D\big(\mathfrak{G}^\bullet\big) = C^2[0,+\infty) \cap \bigg(u : p_1 u(0) - p_2 u^+(0) + p_3(\mathfrak{G}u)(0) = \int_{0+} \big[u(l) - u(0)\big] p_4(dl)\bigg).$$

M. Kac [10] cited the problem of describing the sample paths of the elastic Brownian motion $(p_3 = p_4 = 0 < p_1 p_2)$, and it was W. Feller's (private) suggestion that these should be the reflecting Brownian sample paths, killed at the instant some increasing function $\mathfrak{t}^+(\mathfrak{Z}^+ \cap [0,t])$ of the visiting set $\mathfrak{Z}^+ \equiv (t : \mathfrak{x}^+(t) = 0)$ hits a certain level, that was the starting point of this paper.

P. Lévy's profound studies [13] had clarified the fine structure of the standard and reflecting Brownian motions and their local times, the papers of E. B. Dynkin [2] and G. Hunt [7] on Markov times provided an indispensable tool, H. Trotter [17] proved a deep result about local times, and W. Feller [4] had presented a (partial) description of the sample paths of the Brownian motion associated with $\mathfrak{G}^\bullet$ in the special case $p_4(0,+\infty) < +\infty$ (the case $p_4(0,+\infty) = +\infty$ was not discovered in Feller's original proof of (2), but this error was corrected by W. Feller [5] and A. D. Ventcel' [18]).

[5] $a \wedge b$ is the smaller of a and b. $\int_{0+}$ means $\int_{0<l<+\infty}$.

It was left to use these ideas (and some new ones) to build up the sample paths of Feller's Brownian motions from the reflecting Brownian motion and its local time and (independent) exponential holding times and differential processes; that is the aim of the present paper.

15.3. Outline

Brownian motions on $[0,+\infty)$ are defined from a probabilistic point of view in Sect. 15.5, and a special case is disposed of in Sect. 15.6. Green operators

$$G^\bullet_\alpha : f \longrightarrow E^\bullet_\cdot\left(\int_0 e^{-\alpha t} f(\mathfrak{x}^\bullet)\,dt\right)$$

and the generator $\mathfrak{G}^\bullet$ $(= \alpha - G^{\bullet -1}_\alpha)$ are introduced in Sect. 15.7 and computed in Sect. 15.8 using a method of E. B. Dynkin [2]. $\mathfrak{G}^\bullet$ turns out to be the restriction of $\mathfrak{G} \mid C^2[0,+\infty)$ to a domain $D(\mathfrak{G}^\bullet)$ as described in (2.2); it is the *simplest complete invariant* of the motion, i.e., the associated sample paths can be built up from

(a) *a reflecting Brownian motion* $\mathfrak{x}^+$,
(b) *a differential process* $\mathfrak{p}$ *with increasing sample paths based on* p_2 *and* p_4,
(c) *a stochastic clock* $\mathfrak{f}^{-1}$ *based on* $\mathfrak{x}^+$, $\mathfrak{p}$, *and* p_3,
(d) *a killing time based on* $\mathfrak{x}^+$, $\mathfrak{p}$, $\mathfrak{f}^{-1}$, *and* p_1

(see Sects. 15.9, 15.10, 15.11, 15.12, 15.13, 15.14, and 15.15).

Consider, for the sake of conversation, the case:

$$p_4(0,+\infty) = +\infty \quad \text{if } p_2 = 0, \tag{1}$$

introduce the reflecting Brownian motion $\mathfrak{x}^+$ as described in Sect. 15.1 $(u^+(0) = 0)$, and let $\mathfrak{t}^+$ be P. Lévy's *mesure du voisinage* (*local time*)

$$\mathfrak{t}^+(t) = \lim_{\varepsilon\downarrow 0}(2\varepsilon)^{-1}\,\text{measure}\left(s : \mathfrak{x}^+(s) < \varepsilon,\ s \leqq t\right) \tag{2}$$

as described in Sect. 15.4.

Given $p_1 = p_3 = 0$, if $\mathfrak{p}(dt \times dl)$ is a Poisson measure as described in Sect. 15.11 with mean $dt \times p_4(dl)$ independent of $\mathfrak{x}^+$, if $\mathfrak{p}$ is the (increasing) differential process

$$\mathfrak{p}(t) = p_2 t + \int_{0+} l\mathfrak{p}\left([0,t]\times dl\right), \quad t \geqq 0, \tag{3}$$

and if $\mathfrak{p}^{-1}$ is its inverse function, then the desired motion is identical in law to[6]

$$\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+, \tag{4}$$

which could be described as *a reflecting Brownian motion jumping out from* $l = 0$ *like the germ of the differential process* $\mathfrak{p}$ *run with the clock* $\mathfrak{p}^{-1}\mathfrak{t}^+$ (see Sect. 15.12 for pictures).

$p^{-1}\mathfrak{t}^+$ can be interpreted as a *local time* for the new sample path $\mathfrak{x}^\bullet$ (see Sect. 15.14), and, with its help, the description of the sample paths can be completed as follows: in case $p_1 = 0$, the desired motion is identical in law to

$$\mathfrak{x}^\bullet(\mathfrak{f}^{-1}), \quad \mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+, \tag{5a}$$

where the stochastic clock $\mathfrak{f}^{-1}$ is the inverse function of

$$\mathfrak{f} = t + p_3\mathfrak{p}^{-1}\left(\mathfrak{t}^+(t)\right), \tag{5b}$$

while, in case $p_1 > 0$, it is identical in law to $\mathfrak{x}^\bullet(\mathfrak{f}^{-1})$ *killed* (i.e., sent off to an extra state ∞) at a time $\mathfrak{m}^\bullet_\infty$ $(< +\infty)$ with conditional distribution

$$P_\cdot\left(\mathfrak{m}^\bullet_\infty > t \mid \mathfrak{x}^\bullet(\mathfrak{f}^{-1})\right) = e^{-p_1\mathfrak{p}^{-1}\mathfrak{t}^+\mathfrak{f}^{-1}}. \tag{6}$$

[6] $\mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+$ means $\mathfrak{p}(\mathfrak{p}^{-1}(\mathfrak{t}^+))$.

Here are two simple cases to be treated in Sect. 15.10.

Given $p_1 = p_4 = 0 < p_2p_3$ (i.e., $u^+(0) = (p_3/p_2)(\mathfrak{G}^\bullet u)(0)$), the desired motion is identical in law to

$$\mathfrak{x}^\bullet = \mathfrak{x}^+\big(\mathfrak{f}^{-1}\big), \tag{7a}$$

$$\mathfrak{f} = t + \big(p_3/p_2\big)\mathfrak{t}^+. \tag{7b}$$

$\mathfrak{f}^{-1}$ *counts standard time while* $\mathfrak{x}^\bullet(t) > 0$ *but runs slow on the barrier, and hence, compared to the reflecting Brownian motion,* $\mathfrak{x}^\bullet$ *lingers at* $l = 0$ *a little longer than it should*; as a matter of fact,

$$\text{measure}\,\big(s : \mathfrak{x}^\bullet(s) = 0,\ s \leqq t\big) = p_3\mathfrak{t}^+\big(\mathfrak{f}^{-1}(t)\big) > 0 \tag{8}$$

if $t > \min(s : \mathfrak{x}^\bullet(s) = 0)$.

Given $p_3 = p_4 = 0 < p_1p_2$ (i.e., $(p_1/p_2)u(0) = u^+(0)$), the desired (elastic Brownian) motion is identical in law to a reflecting Brownian motion, killed at time $\mathfrak{m}^\bullet_\infty$ with conditional distribution

$$P_\cdot\big(\mathfrak{m}^\bullet_\infty > t \mid \mathfrak{x}^+\big) = e^{-(p_1/p_2)\mathfrak{t}^+(t)}, \tag{9}$$

i.e., *killed on the barrier* $l = 0$ *at a rate* $(p_1/p_2)\mathfrak{t}^+(dt) : dt$ *proportional to the local time.*

Brownian motions with similar barriers at both ends of $[-1, +1]$ or with a two-sided barrier on the line or the unit circle are studied in Sects. 15.16 and 15.17, Sect. 15.18 treats a wider class of Brownian motions on $[0, +\infty)$, substantiating a conjecture of N. Ikeda, Sect. 15.19 describes the sample paths in case a diffusion operator $\mathfrak{G}u = u^+(dl)/e(dl)$ is used in place of the reflecting Brownian generator $\mathfrak{G}^+$, and Sect. 15.20 indicates how to adapt the method to birth and death processes.

15.4. Standard Brownian Motion: Stopping Times and Local Times

Before coming to Brownian motions on a half line, it is convenient to collect in one place some facts about the standard Brownian motion on the line (see K. Itô and H. P. McKean, Jr. [9] for the proofs and additional information).

Consider a standard Brownian motion with sample paths $w : t \to \mathfrak{x}(t)$, universal field $\mathbf{B}$, and probabilities $P_a(B)$ as described in Sect. 15.1, define[7] $\mathbf{B}_t = \mathbf{B}[\mathfrak{x}(s) : s \leqq t]$, and, if $\mathfrak{m} = \mathfrak{m}(w)$ is a stopping time, i.e., if

$$0 \leqq \mathfrak{m} \leqq +\infty, \tag{1a}$$

$$(\mathfrak{m} < t) \in \mathbf{B}_t, \quad t \geqq 0, \tag{1b}$$ [8]

then introduce the associated field

$$\mathbf{B}_{\mathfrak{m}+} = \mathbf{B} \cap \big(B : (\mathfrak{m} < t) \cap B \in \mathbf{B}_t,\ t \geqq 0\big). \tag{2}$$

$\mathbf{B}_{\mathfrak{m}+} = \bigcap_{s>t} \mathbf{B}_s$ in case $\mathfrak{m} \equiv t$; in general, $(\mathfrak{m} < t) \in \mathbf{B}_{\mathfrak{m}+}$ $(t \geqq 0)$, and, with the aid of

$$\mathbf{B}_{\mathfrak{a}+} \subset \mathbf{B}_{\mathfrak{b}+}, \quad \mathfrak{a} \leqq \mathfrak{b}, \tag{3a}$$

$$\mathbf{B}_{\mathfrak{a}+} = \bigcap_{\varepsilon>0} \mathbf{B}_{\mathfrak{b}+}, \quad \mathfrak{b} = \mathfrak{a} + \varepsilon, \tag{3b}$$

it is not hard to check that $\mathbf{B}_{\mathfrak{m}+}$ measures the *past* $\mathfrak{x}(t) : t \leqq \mathfrak{m}+$, i.e.,

$$\mathbf{B}_{\mathfrak{m}+} \supset \bigcap_{\varepsilon>0} \mathbf{B}\big[\mathfrak{x}\big(t \wedge (\mathfrak{m} + \varepsilon)\big) : t \geqq 0\big]. \tag{4}$$

---

[7] $\mathbf{B}[\mathfrak{q}(t) : a \leqq t < b]$ means the smallest Borel subfield of $\mathbf{B}$ measuring the motion indicated inside the brackets.

[8] $(\mathfrak{m} < t)$ is short for $(w : \mathfrak{m} < t)$.

E. B. Dynkin [2] and G. Hunt [7] discovered that *the Brownian traveller starts afresh at a stopping time*; this means that for each stopping time $\mathfrak{m}$, each $a \in R^1$, and each $B \in \mathbf{B}$,

$$P_a\big(w^+_{\mathfrak{m}} \in B \mid \mathbf{B}_{\mathfrak{m}+}\big) = P_b(B), \quad b = \mathfrak{x}(\mathfrak{m}) \tag{5}$$

where $w^+_{\mathfrak{m}}$ denotes the shifted path $w^+_{\mathfrak{m}} : t \to \mathfrak{x}(t + \mathfrak{m})$, $\mathfrak{x}(+\infty) \equiv \infty$[9], and $P_\infty(\mathfrak{x}(t) \equiv \infty,\ t \geqq 0) = 1$. Because $\mathfrak{m} \equiv t$ is a stopping time, (5) includes the *simple Markovian evolution* noted in (1.6); an alternative statement is that *conditional on* $\mathfrak{m} < +\infty$ *and on the present state* $b = \mathfrak{x}(\mathfrak{m})$, *the future* $\mathfrak{x}(t + \mathfrak{m}) : t \geqq 0$ *is a standard Brownian motion, independent of* $\mathfrak{m}$ *and of the past* $\mathfrak{x}(t) : t \leqq \mathfrak{m}+$.

Given $l > 0$, the *passage time* $\mathfrak{m}_l = \min(t : \mathfrak{x}(t) = l)$ is a stopping time, and the motion $[\mathfrak{m}_l : l \geqq 0, P_0]$ is a differential process, homogeneous in the parameter l; it is, in fact, *the one-sided stable process* with exponent $\frac{1}{2}$, rate $\sqrt{2}$, and law

$$P_0\big(\mathfrak{m}_l \in dt\big) = \frac{l}{\big(2\pi t^3\big)^{1/2}} e^{-l^2/2t} dt \tag{6}$$

as P. Lévy [12] discovered.

$\mathfrak{m}$, itself, is a sum of positive jumps (see Sect. 15.11 for information on this point), and its inverse function $\mathfrak{t}^-(t) = \max_{s \leqq t} \mathfrak{x}(s)$ is continuous and flat outside a (Cantor-like) set of times of Hausdorff-Besicovitch dimension number $\frac{1}{2}$; the joint law

$$P_0\big[\mathfrak{x}(t) \in da,\ \mathfrak{t}^-(t) \in db\big] = 2\frac{2b - a}{\big(2\pi t^3\big)^{1/2}} e^{-(2b-a)^2/2t} da\, db, \quad b \geqq 0,\ a \leqq b \tag{7}$$

is cited for future use.

Consider, next, the reflecting Brownian motion $\mathfrak{x}^+ = |\mathfrak{x}|$.

Given a reflecting Brownian stopping time $\mathfrak{m}$, i.e., a time $0 \leqq \mathfrak{m} \leqq +\infty$ with $(\mathfrak{m} < t) \in \mathbf{B}[\mathfrak{x}^+(s) : s \leqq t]$ $(t \geqq 0)$, $\mathfrak{m}$ is likewise a standard Brownian stopping time, and it follows that, conditional on $\mathfrak{m} < +\infty$ and $b = \mathfrak{x}^+(\mathfrak{m})$, the shifted path $\mathfrak{x}^+(l + \mathfrak{m}) : t \geqq 0$ is a reflecting Brownian motion, independent of $\mathfrak{m}$ and of the past $\mathfrak{x}^+(t) : t \leqq \mathfrak{m}$; *in brief, the reflecting Brownian motion starts afresh at its stopping times* (Fig. 15.1).

P. Lévy [13] observed that if $\mathfrak{x}$ is a standard Brownian motion starting at 0, then $\mathfrak{x}^- = \mathfrak{t}^- - \mathfrak{x}$ $(\mathfrak{t}^- = \max_{s \leqq t} \mathfrak{x}(s))$ is identical in law to the reflecting Brownian motion $\mathfrak{x}^+$ starting at 0. Figure 15.2 is a mere caricature of the path, the actual visiting set $(t : \mathfrak{x} = 0)$ being a closed Cantor-like set of Lebesgue measure 0.

P. Lévy also indicated a proof of

$$P_0\Big[\lim_{\varepsilon \downarrow 0}(2\varepsilon)^{-1} \text{measure}\big(s : \mathfrak{x}^-(s) < \varepsilon,\ s \leqq t\big) = \mathfrak{t}^-(t), t \geqq 0\Big] = 1, \tag{8}$$

which implies that $\mathfrak{t}^-$ is a function of $\mathfrak{x}^-$ alone, and deduced the existence of *the reflecting Brownian local time (mesure du voisinage)*:

$$\mathfrak{t}^+(t) = \lim_{\varepsilon \downarrow 0}(2\varepsilon)^{-1} \text{measure}\big(s : \mathfrak{x}^+(s) < \varepsilon,\ s \leqq t\big) \tag{9}$$

(see H. Trotter [17] for a complete proof). $\mathfrak{t}^+$ grows on the visiting set $\mathfrak{Z}^+ = (t : \mathfrak{x}^+(t) = 0)$; it is identical in law to $\mathfrak{t}^-$, and its inverse function $\mathfrak{t}^{-1}$ is identical in law to the standard Brownian passage times; especially, the joint law

$$P_0\big[\mathfrak{x}^+(t) \in da,\ \mathfrak{t}^+(t) \in db\big] = 2\frac{b + a}{\big(2\pi t^3\big)^{1/2}} e^{-(b+a)^2/2t} da\, db, \quad a, b \geqq 0, \tag{10}$$

is deduced from the joint law of $\mathfrak{x}$ and $\mathfrak{t}^-$ above.

[9] ∞ is an extra state $\notin R^1$.

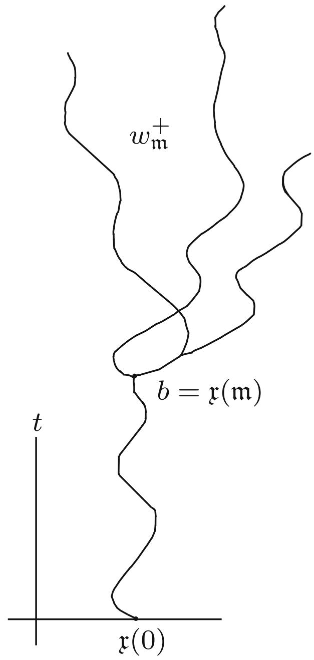

Figure 15.1

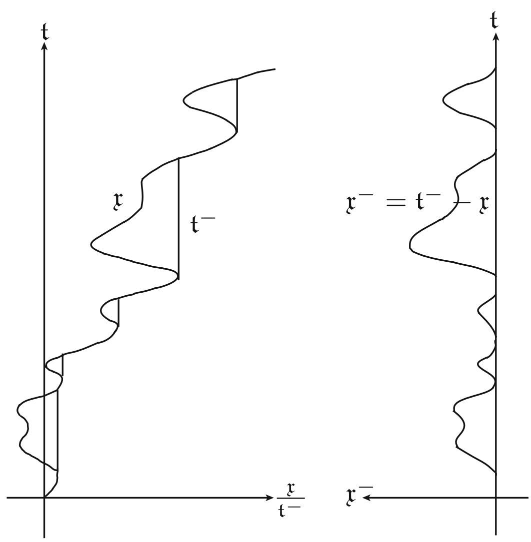

Figure 15.2

Skorohod [15, 16] has made the point that if $\mathfrak{x}$ is a standard Brownian motion, if $0 \leqq \mathfrak{x}^\bullet$ is continuous, if $0 \leqq \mathfrak{t}^\bullet$ is continuous, increasing, and flat outside $\mathfrak{Z}^\bullet = (t : \mathfrak{x}^\bullet = 0)$, and if $\mathfrak{x}^\bullet = \mathfrak{t}^\bullet - \mathfrak{x}$, then $\mathfrak{x}^\bullet = \mathfrak{x}^-$ and $\mathfrak{t}^\bullet = \mathfrak{t}^-$.

15.5. Brownian Motions on $[0, +\infty)$

Given *probabilities* $P_a^\bullet(B)$ ($a \in [0, +\infty) \cup \infty$) defined on the natural *universal field* $\mathbf{B}^\bullet$ of the *path space* comprising all *sample paths*

(1a) $$w^\bullet : t \longrightarrow \mathfrak{x}^\bullet(t) \equiv \mathfrak{x}^\bullet(t+) \in [0, +\infty) \cup \infty,$$

(1b) $$\mathfrak{x}^\bullet(t) \equiv \infty, \quad t \geqq \mathfrak{m}_\infty^\bullet \equiv \inf\bigl(t : \mathfrak{x}^\bullet = \infty\bigr)$$

and subject to

(2a) $$P_a^\bullet(B) \textit{ is a Borel function of } a,$$

(2b) $$P_a^\bullet\bigl[\mathfrak{x}^\bullet(0) \in db\bigr] \textit{ is the unit mass at } b = a \quad (a \neq 0),$$

let us speak of the associated motion as

(a) *simple Markov* if it starts afresh at constant times:

(3a) $$P_\cdot^\bullet\bigl(w_s^{\bullet +} \in B \mid \mathbf{B}_s^\bullet\bigr) = P_a^\bullet(B), \quad s \geqq 0,\ B \in \mathbf{B}^\bullet,\ a = \mathfrak{x}^\bullet(s),$$

where $w_s^{\bullet +}$ is the shifted path $t \to \mathfrak{x}^\bullet(t+s)$ and $\mathbf{B}_s^\bullet$ is the field of $\mathfrak{x}^\bullet(t) : t \leqq s$,

(b) *strict Markov* if it starts afresh at its stopping times:

(3b) $$P_\cdot^\bullet\bigl(w_{\mathfrak{m}^\bullet}^{\bullet +} \in B \mid \mathbf{B}_{\mathfrak{m}^\bullet +}^\bullet\bigr) = P_a^\bullet(B), \quad B \in \mathbf{B}^\bullet,\ a = \mathfrak{x}^\bullet\bigl(\mathfrak{m}^\bullet\bigr),$$

for each stopping time

(4a) $$0 \leqq \mathfrak{m}^\bullet \leqq +\infty,$$

(4b) $$\bigl(\mathfrak{m}^\bullet < t\bigr) \in \mathbf{B}_t^\bullet \quad (t \geqq 0),$$

where $\mathfrak{x}^\bullet(+\infty) \equiv \infty$ and $\mathbf{B}_{\mathfrak{m}^\bullet +}^\bullet$ is the field of events

(5a) $$B \in \mathbf{B}^\bullet,$$

(5b) $$B \cap \bigl(\mathfrak{m}^\bullet < t\bigr) \in \mathbf{B}_t^\bullet \quad (t \geqq 0),$$

(c) *a Brownian motion* if, in addition to (b), the stopped path

(6a) $$\mathfrak{x}^\bullet(t) : t < \mathfrak{m}_{0+} = \lim_{\varepsilon \downarrow 0} \inf\bigl(t : \mathfrak{x}^\bullet < \varepsilon\bigr), \quad \mathfrak{x}^\bullet(0) = l > 0,$$

is identical in law to the stopped standard Brownian motion

(6b) $$\mathfrak{x}(t) : t < \mathfrak{m}_0 = \min(t : \mathfrak{x} = 0), \quad \mathfrak{x}(0) = l.$$

$E_\cdot^\bullet$ denotes the integral (expectation) based upon $P_\cdot^\bullet$, and $E_\cdot^\bullet(e, B) = E_\cdot^\bullet(B, e)$ denotes the integral of $e = e(w^\bullet)$ extended over B; the subscript $\cdot$ as in (3a) and (3b) stands for an unspecified point of $[0, +\infty) \cup \infty$ with the understanding that if several dots appear in a *single* formula, then it is the *same* point that is meant each time.

15.6. Special Case: $p_+(0) < 1$

Given a Brownian motion as described above and a sample path $\mathfrak{x}^\bullet$ starting at $\mathfrak{x}^\bullet(0) = l > 0$, the *crossing time*

(1) $$\mathfrak{m}^\bullet = \mathfrak{m}^\bullet_\varepsilon = \inf\big(t : \mathfrak{x}^\bullet(t) < \varepsilon\big), \quad 0 < \varepsilon < l,$$

is a stopping time, $P^\bullet_l[\mathfrak{x}^\bullet(\mathfrak{m}^\bullet_\varepsilon) = \varepsilon] = 1$, $\mathfrak{m}^\bullet_{0+} = \lim_{\delta\downarrow 0} \mathfrak{m}^\bullet_\delta = \mathfrak{m}^\bullet_\varepsilon + \mathfrak{m}^\bullet_{0+}(w^{\bullet +}_{\mathfrak{m}^\bullet})$, and, since the stopped path $\mathfrak{x}^\bullet(t) : t < \mathfrak{m}^\bullet_{0+}$ is standard Brownian,

(2) $$\begin{aligned} &E^\bullet_l\big[e^{-\alpha\mathfrak{m}^\bullet_{0+}}, \mathfrak{x}^\bullet\big(\mathfrak{m}^\bullet_{0+}\big) \in B\big] \\ &= E^\bullet_l\big(e^{-\alpha\mathfrak{m}^\bullet_\varepsilon} E^\bullet_l\big[\exp\big(-\alpha\mathfrak{m}^\bullet_{0+}\big(w^{\bullet +}_{\mathfrak{m}^\bullet}\big)\big), \mathfrak{x}^\bullet\big(\mathfrak{m}^\bullet_{0+}\big(w^{\bullet +}_{\mathfrak{m}^\bullet}\big), w^{\bullet +}_{\mathfrak{m}^\bullet}\big) \in B \mid \mathbf{B}^\bullet_{\mathfrak{m}^\bullet +}\big]\big) \\ &= E^\bullet_l\big(e^{-\alpha\mathfrak{m}^\bullet_\varepsilon}\big) E^\bullet_\varepsilon\big[e^{-\alpha\mathfrak{m}^\bullet_{0+}}, \mathfrak{x}^\bullet\big(\mathfrak{m}^\bullet_{0+}\big) \in B\big] \\ &\longrightarrow E^\bullet_l\big(e^{-\alpha\mathfrak{m}^\bullet_{0+}}\big) P^\bullet_\varepsilon\big[\mathfrak{x}^\bullet\big(\mathfrak{m}^\bullet_{0+}\big) \in B\big] \quad (\varepsilon \downarrow 0) \\ &= e^{-(2\alpha)^{1/2l}} P^\bullet_l\big[\mathfrak{x}^\bullet\big(\mathfrak{m}^\bullet_{0+}\big) \in B\big],^{10} \end{aligned}$$

i.e., $\mathfrak{x}^\bullet(\mathfrak{m}^\bullet_{0+})$ *is independent of* $\mathfrak{m}^\bullet_{0+}$, *and its law* $p_+(B) = P^\bullet_l[\mathfrak{x}^\bullet(\mathfrak{m}^\bullet_{0+}) \in B]$ *does not depend on* $l > 0$.

Consider the law $p(dl) \equiv P^\bullet_0[\mathfrak{x}^\bullet(0) \in dl]$, and, in case $p(0) = 1$, let $\mathfrak{e}$ be the *exit time* $\inf(t : \mathfrak{x}^\bullet(t) \neq 0)$.

Because

(3a) $$p_+(0) = P^\bullet_l\big[\mathfrak{x}^\bullet\big(\mathfrak{m}^\bullet_{0+}\big) = 0,\ \mathfrak{x}^\bullet\big(0, w^{\bullet +}_{\mathfrak{m}^\bullet_{0+}}\big) = 0\big] = p_+(0)p(0), \quad l > 0,$$

and

(3b) $$p(0) = P^\bullet_0\big[\mathfrak{x}^\bullet(0) = 0,\ \mathfrak{x}^\bullet\big(0, w^{\bullet +}_0\big) = 0\big] = p(0)^2,$$

the possibilities are

(4a) $$p(0) = p_+(0) = 0,$$

(4b) $$p(0) = 1 > p_+(0),$$

(4c) $$p(0) = p_+(0) = 1.$$

(4a) is the simplest case. Figure 15.3 shows the motion $[\mathfrak{x}^\bullet, P^\bullet_0]$: the jumps l_1, l_2, etc. are independent with common law $p_+(dl)$, the initial position l_0 is independent of l_1, l_2, etc. with law $p(dl)$, and the excursions leading back to $l = 0+$ are standard Brownian.

(4b) is more interesting. $\mathfrak{e}$ *is an exponential holding time independent of* $\mathfrak{x}^\bullet(\mathfrak{e})$ *with law* e^{-t/p_3} $(0 \leqq p_3 \leqq +\infty)$; indeed, if $s \geqq 0$, then $(\mathfrak{e} > s) \in \mathbf{B}^\bullet_{s+} = \bigcap_{t>s} \mathbf{B}^\bullet_t$, whence

(5) $$P^\bullet_0(\mathfrak{e} > t + s) = P^\bullet_0\big(\mathfrak{e} > s,\ \mathfrak{e}\big(w^{\bullet +}_s\big) > t\big) = P^\bullet_0(\mathfrak{e} > s)P^\bullet_0(\mathfrak{e} > t)$$

and

(6) $$\begin{aligned} P^\bullet_0\big[\mathfrak{e} > s,\ \mathfrak{x}^\bullet(\mathfrak{e}) \in dl\big] &= P^\bullet_0\big[\mathfrak{e} > s,\ \mathfrak{x}^\bullet\big(\mathfrak{e}\big(w^{\bullet +}_s\big) + s\big) \in dl\big] \\ &= P^\bullet_0(\mathfrak{e} > s)P^\bullet_0\big[\mathfrak{x}^\bullet(\mathfrak{e}) \in dl\big], \end{aligned}$$

completing the proof.

p_3 *has to be positive*; in the opposite case,

$$P^\bullet_0(\mathfrak{e} = 0) = p(0) = P^\bullet_0\Big(\lim_{\varepsilon\downarrow 0} \mathfrak{m}^\bullet_\varepsilon = 0\Big) = 1,$$

[10] $\mathfrak{m}^\bullet_{0+}$ is identical in law to the standard Brownian passage time $\mathfrak{m}_0 = \min(t : \mathfrak{x}(t) = 0)$, and hence $E^\bullet_l(\exp(-\alpha\mathfrak{m}^\bullet_{0+})) = \exp(-(2\alpha)^{1/2}l)$ (see (6)).

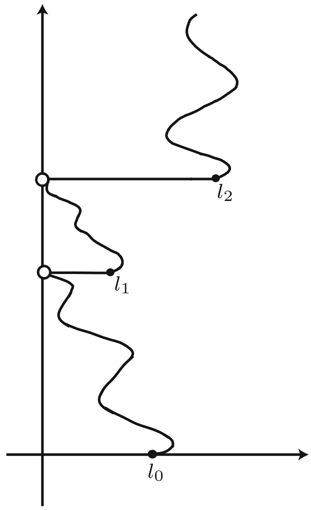

Figure 15.3

where now $\mathfrak{m}_\varepsilon^\bullet$ is the sum of the crossing time $\mathfrak{m}^\bullet = \inf(t : \mathfrak{x}^\bullet(t) > \varepsilon)$ and $\mathfrak{m}_{0+}^\bullet(w_{\mathfrak{m}^\bullet}^{\bullet +})$, and hence

(7)
$$\begin{aligned} 1 = p(0) &= P_0^\bullet\Big(\lim_{\varepsilon\downarrow 0}\mathfrak{x}^\bullet\big(\mathfrak{m}_\varepsilon^\bullet\big) = 0\Big) \\ &= \lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0} P_0^\bullet\big(\mathfrak{x}^\bullet\big(\mathfrak{m}_\varepsilon^\bullet\big) < \delta\big) \\ &= \lim_{\delta\downarrow 0} p_+[0,\delta) \\ &= p_+(0), \end{aligned}$$

contradicting $p_+(0) < 1$.

$p_-(dl) \equiv P_0^\bullet[\mathfrak{x}^\bullet(\mathfrak{e}) \in dl, \mathfrak{e} < +\infty]$ *attributes no mass to* $l = 0$ as is clear from

(8a)
$$P_0^\bullet(\mathfrak{e} > 0) = \lim_{t\downarrow 0} e^{-t/p_3} = 1$$

and

(8b)
$$p_-(0) = P_0^\bullet\big[\mathfrak{x}^\bullet(\mathfrak{e}) = 0,\ \mathfrak{e} < +\infty,\ \mathfrak{e}\big(w_{\mathfrak{e}}^{\bullet +}\big) = 0\big] \leqq P_0^\bullet(\mathfrak{e} = 0).$$

Figure 15.4 is now evident; the jumps l_1^-, l_2^-, etc., l_1^+, l_2^+, etc., and the holding times $\mathfrak{e}_1$, $\mathfrak{e}_2$, etc. are independent with common laws $P(l_1^- \in dl) = p_-(dl)$, $P(l_1^+ \in dl) = p_+(dl)$, $P(\mathfrak{e}_1 > t) = e^{-t/p_3}$, and the excursions leading back to $l = 0+$ are standard Brownian.

(4c) occupies us in Sects. 15.7 to 15.15; a further class of ramified simple Markov motions is studied in Sect. 15.18.

15.7. Green Operators and Generators: $p_+(0) = 1$

Consider the case $p_+(0) = 1$ (4c), and introduce the Green operators

(1)
$$G_\alpha^\bullet : f \in C[0,+\infty) \longrightarrow E_\cdot^\bullet\Big(\int_0^{\mathfrak{m}_\infty^\bullet} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big)\,dt\Big), \quad \alpha > 0.$$

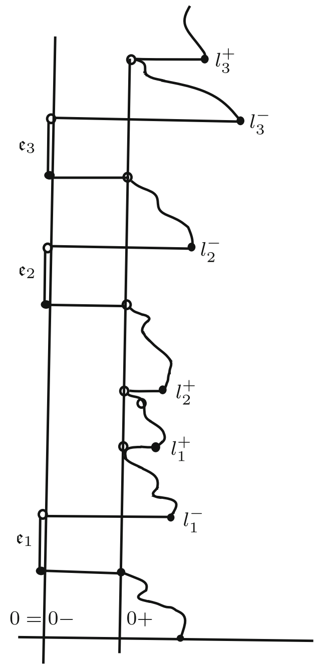

Figure 15.4

Because $\mathfrak{m}^\bullet \equiv \mathfrak{m}^\bullet_{0+} = \lim_{\varepsilon\downarrow 0} \inf(t : \mathfrak{x}^\bullet(t) < \varepsilon)$ is a stopping time and $P^\bullet_\cdot(\mathfrak{x}^\bullet(\mathfrak{m}^\bullet) = 0) \equiv 1$,

$$(2)\qquad \begin{aligned} \big(G^\bullet_\alpha f\big)(l) &= E^\bullet_l\bigg(\int_0^{\mathfrak{m}^\bullet_{0+}} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big)\,dt\bigg) \\ &\quad + E^\bullet_l\bigg(e^{-\alpha m^\bullet_{0+}} E^\bullet_l\bigg(\int_0^{\mathfrak{m}^\bullet_\infty(w^{\bullet +}_{\mathfrak{m}^\bullet})} e^{-\alpha t} f\big[\mathfrak{x}^\bullet\big(t+\mathfrak{m}^\bullet\big)\big]\,dt \mid B^\bullet_{\mathfrak{m}^\bullet +}\bigg)\bigg) \\ &= \big(G^-_\alpha f\big)(l) + E^\bullet_l\big(e^{-\alpha m^\bullet_{0+}}\big) E_0\bigg(\int_0^{\mathfrak{m}^\bullet_\infty} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big)\,dt\bigg) \\ &= \big(G^-_\alpha f\big)(l) + e^{-(2\alpha)^{1/2} l}\big(G^\bullet_\alpha f\big)(0), \end{aligned}$$

where G^-_α is the Green operator for the (absorbing) Brownian motion with instant killing at $l = 0$:

$$(3)\qquad \begin{aligned} \big(G^-_\alpha f\big)(a) &= E_a\bigg(\int_0^{\mathfrak{m}_0} e^{-\alpha t} f(\mathfrak{x})\,dt\bigg) \\ &= \int_0^{+\infty} \frac{e^{-(2\alpha)^{1/2}|b-a|} - e^{-(2\alpha)^{1/2}|b+a|}}{(2\alpha)^{1/2}} f\,db, \qquad a \geqq 0; \end{aligned}$$

especially, $G^\bullet_\alpha$ maps $C[0,+\infty)$ *into* $C^2[0,+\infty)$.

Given $\alpha, \beta > 0$ and $f \in C[0,+\infty)$,

$$\begin{aligned}(\alpha-\beta)G^{\bullet}_{\alpha}G^{\bullet}_{\beta}f &= (\alpha-\beta)E^{\bullet}_{\cdot}\left(\int_0^{\mathfrak{m}^{\bullet}_{\infty}} e^{-\alpha t}\big(G^{\bullet}_{\beta}f\big)\big(\mathfrak{x}^{\bullet}\big)dt\right)\\ &= (\alpha-\beta)E^{\bullet}_{\cdot}\left(\int_0^{\mathfrak{m}^{\bullet}_{\infty}} e^{-\alpha t}dtE^{\bullet}_{\mathfrak{x}^{\bullet}(t)}\left(\int_0^{\mathfrak{m}^{\bullet}_{\infty}} e^{-\beta s}f\big(\mathfrak{x}^{\bullet}\big)ds\right)\right)\\ &= (\alpha-\beta)E^{\bullet}_{\cdot}\left(\int_0^{\mathfrak{m}^{\bullet}_{\infty}} e^{-(\alpha-\beta)t}dt\int_t^{\mathfrak{m}^{\bullet}_{\infty}} e^{-\beta s}f\big(\mathfrak{x}^{\bullet}\big)ds\right)\\ &= E^{\bullet}_{\cdot}\left(\int_0^{\mathfrak{m}^{\bullet}_{\infty}} e^{-\beta s}f\big(\mathfrak{x}^{\bullet}\big)ds(\alpha-\beta)\int_0^{s} e^{-(\alpha-\beta)t}dt\right)\\ &= G^{\bullet}_{\beta}f - G^{\bullet}_{\alpha}f,\end{aligned} \tag{4}$$

i.e.,

$$G^{\bullet}_{\alpha} - G^{\bullet}_{\beta} + (\alpha-\beta)G^{\bullet}_{\alpha}G^{\bullet}_{\beta} = 0, \quad \alpha,\beta > 0, \tag{5}$$

proving that the *range* $G^{\bullet}_{\alpha}C[0,+\infty) \equiv D(\mathfrak{G}^{\bullet})$ and the *null-space* $G^{\bullet -1}_{\alpha}(0)$ are both independent of $\alpha > 0$; in fact, $G^{-1}_{\beta}(0) = \bigcap_{\alpha>0} G^{\bullet -1}_{\alpha}(0) = 0$ because if f belongs to it, then

$$0 = \lim_{\alpha\uparrow+\infty} \alpha\big(G^{\bullet}_{\alpha}f\big)(l) = \lim_{\alpha\uparrow+\infty} E^{\bullet}_{l}\left(\alpha\int_0^{\mathfrak{m}^{\bullet}_{\infty}} e^{-\alpha t}f\big(\mathfrak{x}^{\bullet}\big)dt\right) = f(l), \quad l \geqq 0, \tag{6}$$

thanks to $P^{\bullet}_{l}(\mathfrak{x}^{\bullet}(0+) = l) \equiv 1$ $(l \geqq 0)$.

$G^{\bullet}_{\alpha}$ is now seen to be invertible, and another application of (5) implies that

$$\mathfrak{G}^{\bullet} \equiv \alpha - G^{\bullet -1}_{\alpha} : D\big(\mathfrak{G}^{\bullet}\big) \longrightarrow C[0,+\infty) \tag{7}$$

is likewise independent of $\alpha > 0$.

$\mathfrak{G}^{\bullet}$ *is the generator cited in the section title; it is a contraction of* $\mathfrak{G} = D^2/2$ *acting on* $C^2[0,+\infty)$ because

$$D\big(\mathfrak{G}^{\bullet}\big) = G^{\bullet}_{1}C[0,+\infty) \subset C^2[0,+\infty) \tag{8a}$$

and

$$(\alpha - \mathfrak{G}^{\bullet})G^{\bullet}_{\alpha} = 1, \quad \alpha > 0. \tag{8b}$$

Given two Brownian motions with the *same generator*, their Green operators and hence their transition probabilities and laws in function space are the same, i.e., $\mathfrak{G}^{\bullet}$ *is a complete invariant of the Brownian motion.*

15.8. Generator and Green Operators Computed: $p_+(0) = 1$

$D(\mathfrak{G}^{\bullet})$ can be described in terms of three nonnegative numbers p_1, p_2, p_3 and a nonnegative mass distribution $p_4(dl)$ $(l > 0)$ subject to

$$p_1 + p_2 + p_3 + \int_{0+} (l \wedge 1)p_4(dl) = 1 \tag{1a}$$

and

$$p_4(0,+\infty) = +\infty \quad \text{in case } p_2 = p_3 = 0, \tag{1b}$$

namely, $D(\mathfrak{G}^\bullet)$ is *the class of functions* $u \in C^2[0,+\infty)$ *subject to*[11]

$$p_1 u(0) + p_3(\mathfrak{G}u)(0) = p_2 u^+(0) + \int_{0+} \big[u(l) - u(0)\big] p_4(dl), \tag{2a}$$

as will now be proved.

(1b) is automatic from the rest because if $p_2 = p_3 = 0$ and $p_4(0,+\infty) < +\infty$, then an application of (2a) to $u = \alpha G^\bullet_\alpha f \in D(\mathfrak{G}^\bullet)$ implies, on letting $\alpha \uparrow +\infty$, that

$$\big[p_1 + p_4(0,+\infty)\big] f(0) = \int_{0+} f p_4(dl) \quad \text{for each } f \in C[0,+\infty),$$

which is absurd in view of (1a). Besides, it is enough to prove that

$$\begin{aligned} &D\big(\mathfrak{G}^\bullet\big) \subset C^2[0,+\infty) \\ &\qquad \cap \left(u : p_1 u(0) + p_3(\mathfrak{G}u)(0) = p_2 u^+(0) + \int_{0+} \big[u(l) - u(0)\big] p_4(dl)\right) \end{aligned} \tag{2b}$$

for some choice of p_1, p_2, p_3, p_4 subject to (1a), because, if u is a member of the second line, then so is the bounded solution $u^\bullet = G^\bullet_1(1-\mathfrak{G})u - u$ of $\mathfrak{G}u^\bullet = u^\bullet$, and, expressing $u^\bullet$ as $c_1 e^{2^{1/2}l} + c_2 e^{-2^{1/2}l}$, it is found that $c_1 = c_2 = u^\bullet \equiv 0$, i.e., $u = G^\bullet_1(1-\mathfrak{G})u \in D(\mathfrak{G}^\bullet)$.

Consider, for the proof of (2b), the *exit time*

$$\mathfrak{e} = \inf\big(t : \mathfrak{x}^\bullet(t) \neq 0\big) \tag{3}$$

and its law

$$P^\bullet_0(\mathfrak{e} > t) = \mathfrak{e}^{-t/k} \quad (0 \leqq k \leqq +\infty), \tag{4}$$

and bear in mind that $\mathfrak{x}^\bullet(\mathfrak{e})$ is independent of $\mathfrak{e}$:

$$P^\bullet_0\big[\mathfrak{e} > t,\ \mathfrak{x}^\bullet(\mathfrak{e}) \in dl\big] = \mathfrak{e}^{-t/k} p(dl). \tag{5}$$

If $k = +\infty$ ($\mathfrak{e} \equiv +\infty$), then $(\mathfrak{G}^\bullet u)(0) = 0$ for each $u \in D(\mathfrak{G}^\bullet)$, and (2b) holds with $p_1 = p_2 = p_4 = 0$ and $p_3 = 1$.

If $0 < k < +\infty$, then

$$p(0) = P^\bullet_0\big[\mathfrak{x}^\bullet(\mathfrak{e}) = 0,\ \mathfrak{e}\big(w^{\bullet +}_{\mathfrak{e}}\big) = 0\big] \leqq P^\bullet_0(\mathfrak{e} = 0) = 0, \tag{6}$$

and choosing $u = G^\bullet_\alpha f \in D(\mathfrak{G}^\bullet)$, it appears that

$$\begin{aligned} u(0) &= f(0) E^\bullet_0\left(\int_0^{\mathfrak{e}} e^{-\alpha t} dt\right) + E^\bullet_0\big[e^{-\alpha t} u\big(\mathfrak{x}^\bullet(\mathfrak{e})\big),\ \mathfrak{e} < \mathfrak{m}^\bullet_\infty\big] \\ &= \frac{\alpha u(0) - \big(\mathfrak{G}^\bullet u\big)(0)}{\alpha + k} + \frac{k}{\alpha + k} \int_{0+} u p(dl), \end{aligned} \tag{7a}$$

[12]

or, what is the same,

$$u(0) + k^{-1}\big(\mathfrak{G}^\bullet u\big)(0) = \int_{0+} u p(dl), \tag{7b}$$

i.e., (2b) holds with $p_1 : p_2 : p_3 : p_4 = 1 : 0 : k^{-1} : p$.

But, if $k = 0$ ($\mathfrak{e} \equiv 0$), the proof is less simple; the method used below is due to E. B. Dynkin [2].

$(\mathfrak{G}^\bullet u)(0) < -1$ for some $u \in D(\mathfrak{G}^\bullet)$ (if not, then $(\mathfrak{G}^\bullet u)(0) \equiv 0$, $f(0) = (1 - \mathfrak{G}^\bullet)G^\bullet_1 f(0) = (G^\bullet_1 f)(0)$ for each $f \in C[0,+\infty)$, and $P^\bullet_0(\mathfrak{e} = +\infty) = 1$), so, choosing $\varepsilon > 0$ so small that

---

[11] $\mathfrak{G} = D^2/2$.

[12] $(\alpha - \mathfrak{G}^\bullet)G^\bullet_\alpha = 1$.

$(\mathfrak{G}^\bullet u)(l) < -1$ $(l \leqq \varepsilon)$ and introducing the *crossing time* $\mathfrak{m}_\varepsilon^\bullet = \inf(t : \mathfrak{x}^\bullet(t) > \varepsilon)$, it is clear from

$$
\begin{aligned}
u(0) &= E_0^\bullet\Big(\int_0^{\mathfrak{m}_\infty^\bullet} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big)\,dt\Big), \quad f = \big(\alpha - \mathfrak{G}^\bullet\big)u, \\
&= E_0^\bullet\Big(\int_0^{\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet} e^{-\alpha t}\big(\alpha - \mathfrak{G}^\bullet\big)u\big(\mathfrak{x}^\bullet\big)\,dt\Big) \\
&\quad + E_0^\bullet\big[e^{-\alpha \mathfrak{m}_\varepsilon^\bullet} u\big(\mathfrak{x}^\bullet\big(\mathfrak{m}_\varepsilon^\bullet\big)\big),\ \mathfrak{m}_\varepsilon^\bullet < \mathfrak{m}_\infty^\bullet\big]
\end{aligned}
\tag{8}
$$

that

$$
E_0^\bullet\big(\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet\big) \leqq \lim_{\varepsilon\downarrow 0} E_0^\bullet\Big(\int_0^{\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet} e^{-\alpha t}\big(\alpha - \mathfrak{G}^\bullet\big)u\big(\mathfrak{x}^\bullet\big)\,dt\Big) < +\infty. \tag{9}
$$

$(\mathfrak{G}^\bullet u)(0) < -1$ has no special advantage for the derivation of (8) which holds for each $u \in D(\mathfrak{G}^\bullet)$ and $\varepsilon > 0$; thus, keeping $\varepsilon > 0$ so small that $E_0^\bullet(\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet) < +\infty$ and letting $\alpha \downarrow 0$ in (8) implies

$$
\begin{aligned}
u(0) &= -E_0^\bullet\Big(\int_0^{\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet} \big(\mathfrak{G}^\bullet u\big)\big(\mathfrak{x}^\bullet\big)\,dt\Big) \\
&\quad + E_0^\bullet\big[u\big(\mathfrak{x}^\bullet\big(\mathfrak{m}_\varepsilon^\bullet\big)\big),\ \mathfrak{m}_\varepsilon^\bullet < +\infty\big], \quad u \in D\big(\mathfrak{G}^\bullet\big),
\end{aligned}
\tag{10}
$$

and letting $\varepsilon\downarrow 0$ in (10) establishes E. B. Dynkin's *formula for the generator*:

$$
\big(\mathfrak{G}^\bullet u\big)(0) = \lim_{\varepsilon\downarrow 0} \int_{[\varepsilon,+\infty)\cup\infty} \big[u(l) - u(0)\big] p_\varepsilon(dl), \quad u \in D\big(\mathfrak{G}^\bullet\big),\ u(\infty) \equiv 0, \tag{11a}
$$

$$
p_\varepsilon(dl) = E_0^\bullet\big(\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet\big)^{-1} P_0^\bullet\big[\mathfrak{x}^\bullet\big(\mathfrak{m}_\varepsilon^\bullet \wedge \mathfrak{m}_\infty^\bullet\big) \in dl\big], \tag{11b}
$$

or, what is better for the present purpose,

$$
\lim_{\varepsilon\downarrow 0}\left[\frac{p_\varepsilon(\infty)}{D} u(0) + \frac{(\mathfrak{G}^\bullet u)(0)}{D} - \int_{[\varepsilon,+\infty)} u^\bullet(l)(l \wedge 1)\frac{p_\varepsilon(dl)}{D}\right] = 0, \tag{12a}
$$

$$
D = p_\varepsilon(\infty) + 1 + \int_{0+} (l \wedge 1) p_\varepsilon(dl), \tag{12b}
$$

$$
\begin{aligned}
u^\bullet(l) &= \frac{u(l) - u(0)}{l \wedge 1} \quad \text{if } l > 0, \\
&= u^+(0) \quad \text{if } l = 0.
\end{aligned}
\tag{12c}
$$

Because $D(\mathfrak{G}^\bullet) \subset C^2[0,+\infty)$, $u^\bullet \in C[0,+\infty)$, and selecting $\varepsilon = \varepsilon_1 > \varepsilon_2 >$ etc. $\downarrow 0$ so as to have

$$
\lim_{\varepsilon\downarrow 0} p_\varepsilon(\infty)/D = p_1, \tag{13a}
$$

$$
\lim_{\varepsilon\downarrow 0} 1/D = p_3, \tag{13b}
$$

$$
\lim_{\varepsilon\downarrow 0} (l \wedge 1) p_\varepsilon(dl)/D = p_*(dl), \text{[13]} \tag{13c}
$$

[13] $\int_0 f(l \wedge 1) D^{-1} p_\varepsilon(dl)$ converges as $\varepsilon \downarrow 0$ to $\int f p_*(dl)$ extended over $[0,+\infty]$ for each $f \in C[0,+\infty]$.

existing, it is clear from (12) that

$$p_1 u(0) + p_3(\mathfrak{G}^\bullet u)(0) = p_2 u^+(0) + \int_{(0,+\infty]} [u(l) - u(0)] p_4(dl), \tag{14a}$$

$$p_2 = p_*(0), \qquad p_4(dl) = p_*(dl)/(l \wedge 1) \quad (l > 0), \tag{14b}$$

$$p_1 + p_2 + p_3 + \int_{(0,+\infty]} (l \wedge 1) p_4(dl) = 1 \tag{14c}$$

for each $u \in D(\mathfrak{G}^\bullet)$ having a limit $u(+\infty)$ at $l = +\infty$.

But $p_4(+\infty) = 0$ because, if $f = e^{-n/l}$, then $u = G_1^\bullet f \in D(\mathfrak{G}^\bullet)$, $u(+\infty) = 1$, and at the same time $u(0)$, $u^+(0)$, $\mathfrak{G}^\bullet u(0)$, and $\int_{l<+\infty} [u(l) - u(0)] p_4(dl)$ are all small for large n, and this permits us to derive (14a) anew for *each* $u \in D(\mathfrak{G}^\bullet)$, completing the proof of (2b).

Given $u \in D(\mathfrak{G}^\bullet)$ and inserting (7.2) into (2a), a little algebra justifies

$$(G_\alpha^\bullet f)(0) = \frac{p_2 2 \int_{0+} e^{-(2\alpha)^{1/2} l} f(l) dl + p_3 f(0) + \int_{0+} (G_\alpha^- f)(l) p_4(dl)}{p_1 + (2\alpha)^{1/2} p_2 + \alpha p_3 + \int_{0+} [1 - e^{-(2\alpha)^{1/2} l}] p_4(dl)}, \tag{15}$$

which finishes the computation of the Green operators.

15.9. Special Case: $p_2 = 0 < p_3$ and $p_4 < +\infty$

Consider the special case

$$p_2 = 0 < p_3, \tag{1a}$$

$$p_4 = p_4(0, \infty) < +\infty, \tag{1b}$$

and introduce a motion $\mathfrak{x}^\bullet$ based on a reflecting Brownian motion with sample paths $t \to \mathfrak{x}^+(t)$ and probabilities $P_a(B)$ $(a \geqq 0)$ as follows.

Given a sample path $\mathfrak{x}^+$ starting at a point of $[0, +\infty)$, let $\mathfrak{x}^\bullet = \mathfrak{x}^+$ up to the passage time $\mathfrak{m}_0 = \min(t : \mathfrak{x}^+(t) = 0)$; then make $\mathfrak{x}^\bullet$ wait at 0 for an exponential holding time $\mathfrak{e}_1$ with conditional law

$$P_\cdot(\mathfrak{e}_1 > t \mid \mathfrak{x}^+) = e^{-((p_1 + p_4)/p_2)t}; \tag{1}$$

at the end of that time let it jump to a point $l_1 \in (0, +\infty) \cup \infty$ with conditional law

$$\begin{aligned} P_\cdot(l_1 \in dl \mid \mathfrak{e}_1, \mathfrak{x}^+) &= p_4(dl)/(p_1 + p_4) \quad &\text{if } l > 0, \\ &= p_1/(p_1 + p_4) &\text{if } l = 0, \end{aligned} \tag{2}$$

and, if $+\infty > l_1 > 0$, let it start afresh, while, if $l_1 = \infty$, let $\mathfrak{x}^\bullet = \infty$ at all later times (see Fig. 15.5).

Because $\mathfrak{x}^\bullet$ starts afresh at the passage time $\mathfrak{m}_0$,

$$\begin{aligned} (G_\alpha^\bullet f)(l) &= E_l\left(\int_0^{\mathfrak{m}_\infty^\bullet} e^{-\alpha t} f(\mathfrak{x}^\bullet) dt\right) \quad \left(\mathfrak{m}_\infty^\bullet = \min\left(t : \mathfrak{x}^\bullet(t) = \infty\right)\right) \\ &= E_l\left(\int_0^{\mathfrak{m}_0} e^{-\alpha t} f(\mathfrak{x}^+) dt\right) + E_l(e^{-\alpha \mathfrak{m}_0}) E_0\left(\int_0^{\mathfrak{m}_\infty^\bullet} e^{-\alpha t} f(\mathfrak{x}^\bullet) dt\right) \\ &= (G_\alpha^- f)(l) + e^{-(2\alpha)^{1/2} l} (G_\alpha^\bullet f)(0) \end{aligned} \tag{3}$$

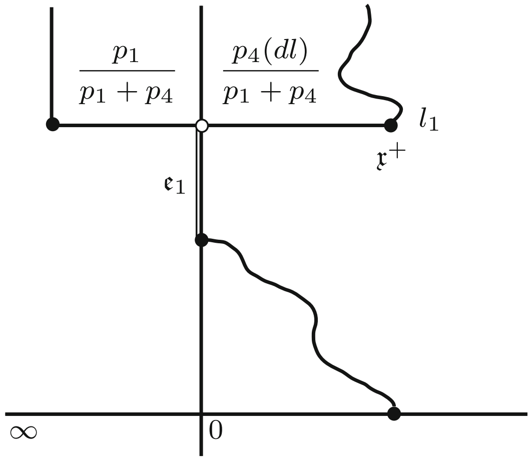

Figure 15.5

as in (7.2), whence

$$\begin{aligned}\big(G^\bullet_\alpha f\big)(0) &= f(0)E_0\bigg(\int_0^{\mathfrak{e}_1} e^{-\alpha t}dt\bigg)\\ &\quad + E_0\big(e^{-\alpha\mathfrak{e}_1}\big)E_0\big[\big(G^\bullet_\alpha f\big)(l_1),\ \mathfrak{e}_1 < \mathfrak{m}^\bullet_\infty\big]\\ &= \frac{p_3 f(0)}{p_1+\alpha p_3+p_4}\\ &\quad + \frac{1}{p_1+\alpha p_3+p_4}\bigg[\int_{0+}\big(G^-_\alpha f\big)(l)p_4(dl)\\ &\qquad + \int_{0+} e^{-(2\alpha)^{1/2}l}p_4(dl)\big(G^\bullet_\alpha f\big)(0)\bigg],\end{aligned} \tag{4}$$

and, solving for $(G^\bullet_\alpha f)(0)$, one finds

$$\big(G^\bullet_\alpha f\big)(0) = \frac{p_3 f(0) + \int_{0+}\big(G^-_\alpha f\big)(l)p_4(dl)}{p_1+\alpha p_3+\int_{0+}\big[1-e^{-(2\alpha)^{1/2}l}\big]p_4(dl)}. \tag{5}$$

Granting that the dot motion starts afresh at constant times (the reader will fill this gap), a comparison of (5) and (8.15) permits its identification as the Brownian motion associated with the operator $\mathfrak{G}^\bullet$ with domain

$$D\big(\mathfrak{G}^\bullet\big) = C^2[0,+\infty)\cap\bigg(u : p_1u(0)+p_3(\mathfrak{G}u)(0) = \int_{0+}\big[u(l)-u(0)\big]p_4(dl)\bigg); \tag{6}$$

the proof that $\mathfrak{x}^\bullet$ is a Brownian motion can be based on the fact, used several times below, that if a motion is *simple* Markov and if its Green operators map $C[0,+\infty)$ into itself, then it is also *strict* Markov (see, for example K. Itô and H. P. McKean, Jr. [9]).

15.10. Special Case: $p_2 > 0 = p_4$

Given a reflecting Brownian motion with sample paths $t\to\mathfrak{x}^+(t)$, probabilities $P_a(B)$, and local time

$$\mathfrak{t}^+(t) = \lim_{\varepsilon\downarrow 0}(2\varepsilon)^{-1}\,\text{measure}\big(s:\mathfrak{x}^+(s)<\varepsilon,\ s\leqq t\big), \tag{1}$$

it is possible to build up all the Brownian motions attached to the generators

(2) $$\mathfrak{G}^\bullet = \mathfrak{G} \mid C^2[0,+\infty] \cap \big(u : p_1 u(0) - p_2 u^+(0) + p_3(\mathfrak{G}u)(0) = 0\big), \quad p_2 > 0$$

with the aid of an extra exponential holding time $\mathfrak{e}$ with conditional law

(3) $$P.\big(\mathfrak{e} > t \mid \mathfrak{x}^+\big) = e^{-t}.$$

Beginning with the elastic Brownian case ($p_1 > 0 = p_3$), the desired motion is

(4a) $$\begin{aligned} \mathfrak{x}^\bullet(t) &= \mathfrak{x}^+(t) \quad \text{if } t < \mathfrak{m}_\infty^\bullet, \\ &= \infty \qquad \text{if } t \geqq \mathfrak{m}_\infty^\bullet, \end{aligned}$$

(4b) $$\mathfrak{m}_\infty^\bullet = \mathfrak{t}^{-1}\big((p_2/p_1)\mathfrak{e}\big) = \min\big(t : \mathfrak{t}^+(t) = (p_2/p_1)\mathfrak{e}\big)$$

as stated in Sects. 15.1 and 15.3.

With the aid of the conditional law

(5) $$P.\big(\mathfrak{m}_\infty^\bullet > t \mid \mathfrak{x}^+\big) = P.\big(\mathfrak{e} > (p_1/p_2)\mathfrak{t}^+(t) \mid \mathfrak{x}^+\big) = e^{-(p_1/p_2)\mathfrak{t}^+(t)}$$

and the *addition rule*

(6) $$\mathfrak{t}^+\big(t_2\big) = \mathfrak{t}^+\big(t_1\big) + \mathfrak{t}^+\big(t_2 - t_1, w_{t_1}^+\big), \quad t_2 \geqq t_1,$$

it is clear that, if $db \subset [0,+\infty)$ and if $\mathfrak{m}_\infty^\bullet > t_1 \leqq t_2$, then

(7a) $$\begin{aligned} &P.\big[\mathfrak{x}^\bullet\big(t_2\big) \in db \mid \mathfrak{x}^+(s) : s \leqq t_1,\ \mathfrak{m}_\infty^\bullet \wedge t_1,\ \mathfrak{m}_\infty^\bullet > t_1\big] \\ &= \frac{P.\big[\mathfrak{x}^+\big(t_2\big) \in db,\ \mathfrak{m}_\infty^\bullet > t_2 \mid \mathfrak{x}^+(s) : s \leqq t_1\big]}{P.\big(\mathfrak{m}_\infty^\bullet > t_1\big)} \\ &= E.\big[\mathfrak{x}^+\big(t_2\big) \in db,\ e^{-(p_1/p_2)\mathfrak{t}^+(t_2)} \mid \mathfrak{x}^+(s) : s \leqq t_1\big] e^{+(p_1/p_2)\mathfrak{t}^+(t_1)} \\ &= E.\big[\mathfrak{x}^+\big(t_2\big) \in db,\ e^{-(p_1/p_2)\mathfrak{t}^+(t_2 - t_1, w_{t_1}^+)} \mid \mathfrak{x}^+(s) : s \leqq t_1\big] \\ &= E_a\big[\mathfrak{x}^+\big(t_2 - t_1\big) \in db,\ e^{-(p_1/p_2)\mathfrak{t}^+(t_2 - t_1)}\big], \quad a = \mathfrak{x}^+\big(t_1\big), \\ &= P_a\big[\mathfrak{x}^\bullet\big(t_2 - t_1\big) \in db\big], \quad a = \mathfrak{x}^\bullet\big(t_1\big), \end{aligned}$$

while, if $\mathfrak{m}_\infty^\bullet \leqq t_1$, then $\mathfrak{x}^\bullet(t_1) = \infty$, and

(7b) $$\begin{aligned} &P.\big[\mathfrak{x}^\bullet\big(t_2\big) \in db \mid \mathfrak{x}^+(s) : s \leqq t_1,\ \mathfrak{m}_\infty^\bullet \wedge t_1,\ \mathfrak{m}_\infty^\bullet \leqq t_1\big] \\ &\qquad = 0 = P_\infty\big[\mathfrak{x}^\bullet\big(t_2 - t_1\big) \in db\big].\end{aligned}$$ [14]

Since $\mathfrak{x}^\bullet(s) : s \leqq t_1$ is a Borel function of $\mathfrak{x}^+(s) : s \leqq t_1$, $\mathfrak{m}_\infty^\bullet \wedge t_1$, and of the indicator of $(\mathfrak{m}_\infty^\bullet < t_1)$, it follows that

(8) $$P.\big[\mathfrak{x}^\bullet\big(t_2\big) \in db \mid \mathfrak{x}^\bullet(s) : s \leqq t_1\big] = P_a\big[\mathfrak{x}^\bullet\big(t_2 - t_1\big) \in db\big], \quad a = \mathfrak{x}^\bullet\big(t_1\big),$$

establishing the simple Markovian nature of the dot motion.

Consider for the next step, its Green operators

$$G_\alpha^\bullet f = E.\bigg(\int_0^{\mathfrak{m}_\infty^\bullet} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big) dt\bigg),$$

[14] $P_\infty[\mathfrak{x}^\bullet \equiv \infty] = 1$ as usual.

and use the conditional law of $\mathfrak{m}^\bullet_\infty$ to check

$$
\begin{aligned}
G^\bullet_\alpha f &= E_\cdot\left(\int_0^{+\infty} e^{-(p_1/p_2)\mathfrak{t}^+(s)}\frac{p_1}{p_2}\mathfrak{t}^+(ds)\int_0^s e^{-\alpha t} f\big(\mathfrak{x}^+\big)\,dt\right)\\
&= E_\cdot\left(\int_0^{+\infty} e^{-\alpha t} f\big(\mathfrak{x}^+\big)\,dt\int_t^{+\infty} e^{-(p_1/p_2)\mathfrak{t}^+}\mathfrak{t}^+(ds)\right) \qquad (9)\\
&= E_\cdot\left(\int_0^{+\infty} e^{-\alpha t} e^{-(p_1/p_2)\mathfrak{t}^+} f\big(\mathfrak{x}^+\big)\,dt\right).
\end{aligned}
$$

Because $\mathfrak{m}_0 = \min(t : \mathfrak{x}^+(t) = 0)$ is a stopping time and $\mathfrak{t}^+(t) = 0$ $(t \leqq \mathfrak{m}_0)$,

$$
\begin{aligned}
\big(G^\bullet_\alpha f\big)(l) &= E_l\left(\int_0^{\mathfrak{m}_0} e^{-\alpha t} f\big(\mathfrak{x}^+\big)\,dt\right)\\
&\quad + E_l\left(e^{-\alpha\mathfrak{m}_0}\int_0^{+\infty} e^{-\alpha t}\exp\big\{-\big(p_1/p_2\big)\mathfrak{t}^+\big(t, w^+_{\mathfrak{m}_0}\big)\big\} f\big[\mathfrak{x}^+\big(t+\mathfrak{m}_0\big)\big]\,dt\right) \qquad (10)\\
&= \big(G^-_\alpha f\big)(l) + E_l\big(e^{-\alpha\mathfrak{m}_0}\big)E_0\left(\int_0^{+\infty} e^{-\alpha t} e^{-(p_1/p_2)\mathfrak{t}^+} f\big(\mathfrak{x}^+\big)\,dt\right)\\
&= \big(G^-_\alpha f\big)(l) + e^{-(2\alpha)^{1/2} l}\big(G^\bullet_\alpha f\big)(0), \quad l \geqq 0,
\end{aligned}
$$

and now the identification of the dot motion as the elastic Brownian motion will be complete as soon as it is verified that

$$
\big(G^\bullet_\alpha f\big)(0) = \frac{p_2 2\int_0^{+\infty} e^{-(2\alpha)^{1/2} l} f(l)\,dl}{p_1 + (2\alpha)^{1/2} p_2}; \qquad (11)
$$

in fact, this will prove that the dot motion is simple Markov with the correct (elastic Brownian) Green operators, and the proof can be completed as at the end of Section 9.

But (11) is trivial; in fact, using the joint law (4.10),

$$
\begin{aligned}
\big(G^\bullet_\alpha f\big)(0) &= E_0\left(\int_0^{+\infty} e^{-\alpha t} e^{-(p_1/p_2)\mathfrak{t}^+} f\big(\mathfrak{x}^+\big)\,dt\right)\\
&= \int_0^{+\infty} e^{-\alpha t}dt\int_0^{+\infty} db\int_0^{+\infty} da\, 2\frac{b+a}{\big(2\pi t^3\big)^{1/2}} e^{-(b+a)^2/2t} e^{-(p_1/p_2)b} f(a) \qquad (12)\\
&= 2\int_0^{+\infty} db\int_0^{+\infty} da\, e^{-(2\alpha)^{1/2}(b+a)} e^{-(p_1/p_2)b} f(a)\\
&= \frac{p_2 2\int_{0+} e^{-(2\alpha)^{1/2} l} f(l)\,dl}{p_1 + (2\alpha)^{1/2} p_2}
\end{aligned}
$$

as stated.

Consider next, the case $p_3 > 0 = p_1$, and let us prove the desired motion to be[15]

$$
\mathfrak{x}^\bullet = \mathfrak{x}^+\big(\mathfrak{f}^{-1}\big), \quad \mathfrak{f} = t + \big(p_3/p_2\big)\mathfrak{t}^+. \qquad (13)
$$

Beginning, as before, with the proof that the dot motion is *simple Markov*, if $t_2 \geqq t_1$ and if $\mathfrak{m} = \mathfrak{f}^{-1}(t_1)$, then

(a) $(\mathfrak{m} < t) = (t_1 < \mathfrak{f}(t)) \in \mathbf{B}[\mathfrak{x}^+(s) : s \leqq t]$, i.e., $\mathfrak{m}$ *is a stopping time*;

(b) $\mathfrak{f}(\mathfrak{m} + s) = \mathfrak{f}(\mathfrak{m}) + \mathfrak{f}(s, w^+_{\mathfrak{m}}) = t_1 + (t_2 - t_1)$ *if* $s = \mathfrak{f}^{-1}(t_2 - t_1, w^+_{\mathfrak{m}})$ *and so* $\mathfrak{f}^{-1}(t_2) = \mathfrak{m} + s = \mathfrak{m} + \mathfrak{f}^{-1}(t_2 - t_1, w^+_{\mathfrak{m}})$;

[15] $\mathfrak{f}^{-1}$ is the inverse function of $\mathfrak{f}$.

(c) $\mathfrak{x}^\bullet(t_2) = \mathfrak{x}^+[\mathfrak{f}^{-1}(t_2 - t_1, w^+_{\mathfrak{m}}) + \mathfrak{m}]$;
(d) $\mathfrak{x}^\bullet(s) : s \leqq t_1$ *is a Borel function of the stopped path* $t \to \mathfrak{x}^+(t \wedge \mathfrak{m})$ *and of* $\mathfrak{f}^{-1}(s) : s \leqq t_1$;
(e) $\mathfrak{f}^{-1}(s)$ *is the solution* r *of* $f(r) = s$ $(\leqq t_1 = f(\mathfrak{m}))$ *and, as such, it is likewise a Borel function of the stopped path*;

and now, using the strict Markovian nature of $\mathfrak{x}^+$, the law of $\mathfrak{x}^\bullet(t_2)$ conditional on $\mathbf{B}_{\mathfrak{m}+} \supset \mathbf{B}[\mathfrak{x}^\bullet(s) : s \leqq t_1]$ is found to be

$$
\begin{aligned}
&P_\cdot\big(\mathfrak{x}^+\big[\mathfrak{f}^{-1}\big(t_2 - t_1, w^+_{\mathfrak{m}}\big) + \mathfrak{m}\big] \in db \mid \mathbf{B}_{\mathfrak{m}+}\big) \\
&\qquad = P_a\big(\mathfrak{x}^+\big[\mathfrak{f}^{-1}\big(t_2 - t_1\big)\big] \in db\big), \quad a = \mathfrak{x}^+(\mathfrak{m}), \\
&\qquad = P_a\big(\mathfrak{x}^\bullet\big(t_2 - t_1\big) \in db\big), \quad a = \mathfrak{x}^\bullet\big(t_1\big),
\end{aligned} \tag{14a}
$$

whence, taking the expectation of both sides conditional on $\mathbf{B}[\mathfrak{x}^\bullet(s) : s \leqq t_1]$,

$$
P_\cdot\big(\mathfrak{x}^\bullet\big(t_2\big) \in db \mid \mathfrak{x}^\bullet(s) : s \leqq t_1\big) = P_a\big(\mathfrak{x}^\bullet\big(t_2 - t_1\big) \in db\big), \quad a = \mathfrak{x}^\bullet\big(t_1\big), \tag{14b}
$$

i.e., $\mathfrak{x}^\bullet = \mathfrak{x}^+(\mathfrak{f}^{-1})$ starts afresh at time t_1, as was to be proved.

Coming to the Green operators

$$
G^\bullet_\alpha f = E_\cdot\bigg(\int_0^{+\infty} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big)\,dt\bigg),
$$

since $\mathfrak{m}_0 = \min(t : \mathfrak{x}^+(t) = 0)$ is a stopping time and $\mathfrak{f}^{-1} \equiv t$ $(t \leqq \mathfrak{m}_0)$,

$$
\begin{aligned}
\big(G^\bullet_\alpha f\big)(l) &= E_l\bigg(\int_0^{\mathfrak{m}_0} e^{-\alpha t} f\big(\mathfrak{x}^+\big)\,dt\bigg) \\
&\quad + E_l\bigg(e^{-\alpha\mathfrak{m}_0}\int_0^{+\infty} e^{-\alpha t} f\big[\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big(t, w^+_{\mathfrak{m}_0}\big) + \mathfrak{m}_0\big)\big]\,dt\bigg) \\
&= \big(G^-_\alpha f\big)(l) + e^{-(2\alpha)^{1/2} l}\big(G^\bullet_\alpha f\big)(0)
\end{aligned} \tag{15}
$$

as in the elastic Brownian case, and to complete the identification of $\mathfrak{x}^\bullet$ it is sufficient to check that

$$
\begin{aligned}
\big(G^\bullet_\alpha f\big)(0) &= E_0\bigg(\int_0^{+\infty} e^{-\alpha t} f\big[\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big)\big]\,dt\bigg) \\
&= E_0\bigg(\int_0^{+\infty} e^{-\alpha\mathfrak{f}} f\big(\mathfrak{x}^+\big)\,\mathfrak{f}(dt)\bigg) \\
&= E_0\bigg(\int_0^{+\infty} e^{-\alpha[t + (p_3/p_2)\mathfrak{t}^+]} f\big(\mathfrak{x}^+\big)\,dt\bigg) \\
&\quad + f(0)E_0\bigg(\int_0^{+\infty} e^{-\alpha[t + (p_3/p_2)\mathfrak{t}^+]}\frac{p_3}{p_2}\mathfrak{t}^+(dt)\bigg)^{16} \\
&= \frac{p_2 2\int_{0+} e^{-(2\alpha)^{1/2} l} f(l)\,dl}{(2\alpha)^{1/2} p_2 + \alpha p_3}
\end{aligned} \tag{16}
$$

[16] $\mathfrak{t}^+(dt) = 0$ off $\mathfrak{Z}^+ = (t : \mathfrak{x}^+(t) = 0)$.

$$+\frac{f(0)}{\alpha}\left[1-E_0\left(\int_0^{+\infty} e^{-\alpha[t+(p_3/p_2)\mathfrak{t}^+]}dt\right)\right]^{17,\ 18}$$
$$=\frac{p_2 2\int_{0+} e^{-(2\alpha)^{1/2}l} f(l)dl + p_3 f(0)}{(2\alpha)^{1/2}p_2+\alpha p_3}$$

as it should be.

Consider now the case $0<p_1p_2p_3$; this time the motion is

(17a)
$$\begin{aligned}\mathfrak{x}^\bullet(t) &= \mathfrak{x}^+(\mathfrak{f}^{-1}) \quad \text{if } t<\mathfrak{m}^\bullet_\infty,\\ &=\infty \qquad\qquad \text{if } t\geqq \mathfrak{m}^\bullet_\infty,\end{aligned}$$

(17b)
$$\mathfrak{m}^\bullet_\infty=\mathfrak{f}\left[\mathfrak{t}^{-1}\big((p_2/p_1)\mathfrak{e}\big)\right]=\left[\mathfrak{t}^+\big(\mathfrak{f}^{-1}\big)\right]^{-1}\big((p_2/p_1)\mathfrak{e}\big),$$

as will still be proved.

$\mathfrak{x}^+(\mathfrak{f}^{-1})$ is a Brownian motion, its *local time*

(18)
$$\begin{aligned}\mathfrak{t}^\bullet(t) &= \text{measure }\big(s:\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big)=0,\ s\leqq t\big)\\ &=\text{measure }\big(s:\mathfrak{f}^{-1}(s)\in\mathfrak{Z}^+,\ s\leqq t\big)\\ &=\text{measure }\mathfrak{f}\big(\mathfrak{Z}^+\big)\cap[0,t]\\ &=\int_{\mathfrak{Z}^+\cap[0,\mathfrak{f}^{-1}(t)]}\mathfrak{f}(ds)\\ &=\big(p_3/p_2\big)\mathfrak{t}^+\left[\mathfrak{f}^{-1}(t)\right]^{19}\end{aligned}$$

satisfies the addition rule (6), and substituting them in place of $\mathfrak{x}^+$ and $\mathfrak{t}^+$ in the derivation of the simple Markovian nature of the elastic Brownian motion, it is found that the present motion is likewise simple Markov.

$G^\bullet_\alpha f=G^-_\alpha f+e^{-(2\alpha)^{1/2}l}(G^\bullet_\alpha f)(0)$ is derived as before, that the dot motion is Brownian follows, and now, using the evaluation (12) with $p_1+\alpha p_3$ in place of p_1 in conjunction with the conditional law

(19)
$$\begin{aligned}P_\cdot\big(\mathfrak{m}^\bullet_\infty>t\mid\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big)\big) &= P_\cdot\big(\mathfrak{e}>\big(p_1/p_2\big)\mathfrak{t}^+\big(\mathfrak{f}^{-1}\big)\mid\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big)\big)\\ &=e^{-(p_1/p_2)\mathfrak{t}^+(\mathfrak{f}^{-1})}=e^{-(p_1/p_3)\mathfrak{t}^\bullet(t)},\end{aligned}$$

it develops that

(20)
$$\begin{aligned}\big(G^\bullet_\alpha f\big)(0) &= E_0\left(\int_0^{\mathfrak{m}^\bullet_\infty} e^{-\alpha t} f\left[\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big)\right]dt\right)\\ &=E_0\left(\int_0^{+\infty} e^{-\alpha t}e^{-(p_1/p_2)\mathfrak{t}^+(\mathfrak{f}^{-1})} f\left[\mathfrak{x}^+\big(\mathfrak{f}^{-1}\big)\right]dt\right)\\ &=E_0\left(\int_0^{+\infty} e^{-\alpha t}e^{-((p_1+\alpha p_3)/p_2)\mathfrak{t}^+} f\big(\mathfrak{x}^+\big)dt\right)\\ &\quad+f(0)E_0\left(\int_0^{+\infty} e^{-\alpha t}e^{-((p_1+\alpha p_3)/p_2)\mathfrak{t}^+}\frac{p_3}{p_2}\mathfrak{t}^+(dt)\right)\\ &=\frac{p_2 2\int_{0+} e^{-(2\alpha)^{1/2}l}f(l)dl+p_3 f(0)}{p_1+(2\alpha)^{1/2}p_2+\alpha p_3},\end{aligned}$$

completing the proof.

---

[17] Use (12) with αp_3 in place of p_1.
[18] Do a partial integration under the expectation sign.
[19] $\text{measure}(\mathfrak{Z}^+)=0$. $\mathfrak{t}^+(dt)=0$ outside $\mathfrak{Z}^+$.

A second description of the present motion is available: *it is the elastic Brownian motion* $\mathfrak{x}^\bullet$ *described in* (4) *run with the new stochastic clock* $\mathfrak{f}^{-1}$ *which is the inverse function of*

(21a) $$\mathfrak{f} = t + \big(p_3/p_2\big) \times \text{the elastic Brownian local time } \mathfrak{t}^\bullet,$$

(21b) $$\begin{aligned}\mathfrak{t}^\bullet(t) &= \lim_{\varepsilon\downarrow 0}(2\varepsilon)^{-1}\,\text{measure}\,\big(s : \mathfrak{x}^\bullet(s) < \varepsilon, s \leqq t\big)\\ &= \mathfrak{t}^+\big(t \wedge \mathfrak{m}^\bullet_\infty\big), \quad \mathfrak{m}^\bullet_\infty = \min\big(t : \mathfrak{x}^\bullet = \infty\big).\end{aligned}$$

15.11. Increasing Differential Processes

Before describing the sample paths in the case $p_4 = p_4(0,+\infty) = +\infty$, it will be helpful to list some properties of differential processes with increasing sample paths.

Given a stochastic process with universal field $\mathbf{B}$, probabilities P, and sample paths $t \to \mathfrak{p}(t)$ with

(1a) $$\mathfrak{p}(0) = 0,$$

(1b) $$\mathfrak{p}(s) \leqq \mathfrak{p}(t), \quad s \leqq t,$$

(1c) $$\mathfrak{p}(t+) = \mathfrak{p}(t) < +\infty, \quad t \geqq 0,$$

which is *differential* in the sense that *shifted path* $\mathfrak{p}_+(t) \equiv \mathfrak{p}(t+s) - \mathfrak{p}(s)$ *is independent of its past* $\mathfrak{p}(t) : t \leqq s$ *and identical in law to* $\mathfrak{p}$, P. Lévy [11][20] proved that

(2a) $$E\big(e^{-\alpha\mathfrak{p}(t)}\big) = \exp\left\{-t\left[p_2\alpha + \int_{0+}(1-e^{-\alpha l})p(dl)\right]\right\}, \quad \alpha > 0,$$

(2b) $$p_2 \geqq 0, \quad p(dl) \geqq 0, \quad \int_{0+}(l \wedge 1)p(dl) < +\infty$$

and expressed $\mathfrak{p}$ as

(3) $$\mathfrak{p}(t) = p_2 t + \int_{0+} l\,\mathfrak{p}\big([0,t] \times dl\big), \quad t \geqq 0,$$

in which $\mathfrak{p}(dt \times dl) =$ *the number of jumps of* $\mathfrak{p}$ *of magnitude* $\in dl$ *occurring in time* dt is differential in the pair $(t,l) \in [0,+\infty) \times (0,+\infty)$ and Poisson distributed with mean $dt\,p(dl)$, i.e., if Q_1, Q_2, etc. are disjoint figures of $[0,+\infty) \times (0,+\infty)$, then $\mathfrak{p}(Q_1)$, $\mathfrak{p}(Q_2)$, etc., are independent, and

(4) $$P\big(\mathfrak{p}(Q) = n\big) = \big(|Q|^n/n!\big)e^{-|Q|}, \quad n \geqq 0, \ |Q| = \int_Q dt\,p(dl);$$

in short, $\mathfrak{p}(t)$ is *the (direct) integral* $\int_{0+} l\mathfrak{p}([0,t] \times dl)$ *of the differential Poisson processes* $\mathfrak{p}([0,t] \times dl)$ *with rates* $p(dl)$ *plus a linear part* $p_2 t$.

Given nonnegative p_2 and $p(dl)$ with $\int_{0+}(l \wedge 1)p(dl) < +\infty$ as in (2b), it is possible to make a Poisson measure $\mathfrak{p}(dt \times dl)$ with mean $dt\,p(dl)$ as described above; the associated $\mathfrak{p}(t) = p_2 t + \int_{0+} l\mathfrak{p}([0,t] \times dl)$ is a differential process having (2a) as its Lévy formula.

G. Hunt [7] discovered that if $\mathfrak{m}$ is a *stopping time*, i.e., if

(5) $$(\mathfrak{m} < t) \in \mathbf{B}\big[\mathfrak{p}(s) : s \leqq t\big] \times \mathbf{B}^\bullet, \quad t \geqq 0,$$

for some field $\mathbf{B}^\bullet$ independent of $\mathfrak{p}$, then $\mathfrak{p}$ *starts afresh at time* $t = \mathfrak{m}$, *i.e., the shifted path* $\mathfrak{p}_+(t) \equiv \mathfrak{p}(t+\mathfrak{m}) - \mathfrak{p}(\mathfrak{m})$ *is independent of the past* $\mathfrak{p}(t) : t \leqq \mathfrak{m}$ *and identical in law to* $\mathfrak{p}$ *itself.*

[20] See also K. Itô [8].

Given $a \geqq 0$, if P_a is the law that P induces on the space of sample paths $\mathfrak{q} = \mathfrak{p} + a$, then

$$P_\cdot\bigl(\mathfrak{q}(t_2) \in db \mid \mathfrak{q}(s) : s \leqq t_1\bigr) = P_a\bigl(\mathfrak{q}(t_2 - t_1) \in db\bigr), \quad t_2 \geqq t_1,\ a = \mathfrak{q}(t_1), \tag{6}$$

the associated Green operators $f \to E(\int_0^{+\infty} e^{-\alpha t} f(\mathfrak{q})dt)$ map $C[0, +\infty)$ into itself, and the associated generator $\mathfrak{Q}$ is

$$(\mathfrak{Q}f)(a) = p_2 f^+(a) + \int_{0+} \bigl[f(b+a) - f(a)\bigr] p(db), \quad f \in C^1[0, +\infty). \tag{7}$$

Given $t \geqq 0$, $\mathfrak{p}([0,t] \times [\varepsilon, +\infty))$ is Poisson distributed and differential in ε with mean $tp[\varepsilon, +\infty)$; as such, it is identical in law to a standard Poisson process $\mathfrak{q}$ with unit jumps and unit rate run with the clock $tp[\varepsilon, +\infty)$, and, using the strong law of large numbers, it follows that

$$\lim_{\varepsilon \downarrow 0} \frac{\mathfrak{p}\bigl([0,t] \times [\varepsilon, +\infty)\bigr)}{p[\varepsilon, +\infty)} = \lim_{\varepsilon \downarrow 0} \frac{\mathfrak{q}\bigl(tp[\varepsilon, +\infty)\bigr)}{p[\varepsilon, +\infty)} = t, \tag{8}$$

which will be helpful to us in Sect. 15.14.

Consider the special case $p(0, +\infty) < +\infty$ pictured in Figure 15.6: the exponential holding times $\mathfrak{e}_1$, $\mathfrak{e}_2$, etc. between jumps are independent with common law $P(\mathfrak{e}_1 > t) = e^{-p(0,+\infty)t}$, the jumps l_1, l_2, etc. are likewise independent with common law $P(l_1 \in dl) = p(0, +\infty)^{-1} p(dl)$, and the slope of the slanting lines is $1/p_2$.

Consider, as a second example, the standard Brownian passage times $\mathfrak{m}_a = \min(t : \mathfrak{x} = a)$ $(a \geqq 0)$ under the law $P = P_0$. Because the Brownian traveller starts afresh at its passage times, the shifted path $\mathfrak{m}_{b+a} - \mathfrak{m}_a = \mathfrak{m}_{b+a}(w^+_{\mathfrak{m}_a})$ is independent of $\mathfrak{m}_b : b \leqq a$ and identical in law to $\mathfrak{m}_\cdot$, i.e., $\mathfrak{m}_\cdot$ is differential (it is the *one-sided stable process with exponent* $\frac{1}{2}$ *and rate* $\sqrt{2}$ as noted in Sect. 15.4);

$$p_2 = 0, \tag{9a}$$

$$p(dl) = dl/\bigl(2\pi l^3\bigr)^{1/2} \tag{9b}$$

can be read off

$$E_0\bigl(e^{-\alpha \mathfrak{m}_a}\bigr) = e^{-(2\alpha)^{1/2} a} = \exp\Bigl\{ -a \int_{0+} (1 - e^{-\alpha l}) \frac{dl}{(2\pi l^3)^{1/2}} \Bigr\}. \tag{10}$$

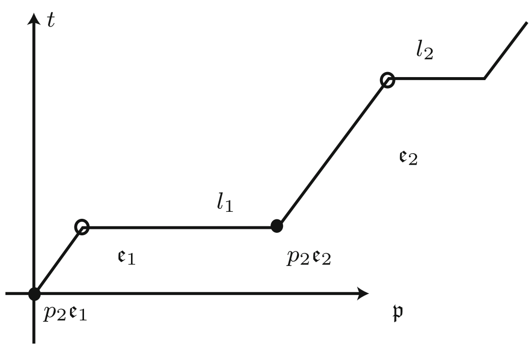

Figure 15.6

$\mathfrak{m}_a$ is left-continuous, so in the direct integral $[0,a)$ must be used in place of $[0,a]$:

$$\mathfrak{m}_a = \int_{0+} lp\big([0,a)\times dl\big).$$

15.12. Sample Paths: $p_1 = p_3 = 0 < p_4$ $(p_2 > 0/p_4 = +\infty)$

Given a reflecting Brownian motion with local time $\mathfrak{t}^+$, a nonnegative number p_2, and a nonnegative mass distribution $p_4(dl)$ $(l > 0)$ with $p_4 = p_4(0,+\infty) = +\infty$ in case $p_2 = 0$, introduce the Poisson measure $\mathfrak{p}(dt\times dl)$ with mean $dt\ p_4(dl)$, make up the associated differential process

(1) $$\mathfrak{p}(t) = p_2 t + \int_{0+} l\mathfrak{p}\big([0,t]\times dl\big),$$

and consider the sample path[21]

(2a) $$\mathfrak{x}^\bullet(t) = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+(t) - \mathfrak{t}^+(t) + \mathfrak{x}^+(t), \quad t \geqq 0,$$

(2b) $$\mathfrak{p}^{-1}(l) = \inf\big(t : \mathfrak{p}(t) > l\big)$$

and its alternative description

(3) $$\mathfrak{x}^\bullet(t) = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^-(t) + \mathfrak{x}^-(t), \quad t \geqq 0,$$

in terms of the *standard Brownian motion* $\mathfrak{x}^- = -\mathfrak{t}^+ + \mathfrak{x}^+$ and its *minimum function* $\mathfrak{t}^-(t) = \mathfrak{t}^+(t) = -(\min_{s\leqq t}\mathfrak{x}^-(s)\wedge 0)$; it is to be proved that $\mathfrak{x}^\bullet$ is the Brownian motion associated with

(4) $$p_2 u^+(0) + \int_{0+}\big[u(l) - u(0)\big]p_4(dl) = 0,$$

but before doing that let us look at some pictures of the sample path.

Consider the case $p_4 < +\infty$: the jumps l_1, l_2, etc. of $\mathfrak{p}$ are finite in number per unit time and can be labelled in their correct temporal order. $\mathfrak{p}$ and $\mathfrak{p}^{-1}$ are seen in Fig. 15.6, $\mathfrak{p}\mathfrak{p}^{-1}$ in Fig. 15.7 of the present section, and the path $\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+$ in Fig. 15.8, in which $\mathfrak{t}^{-1}$

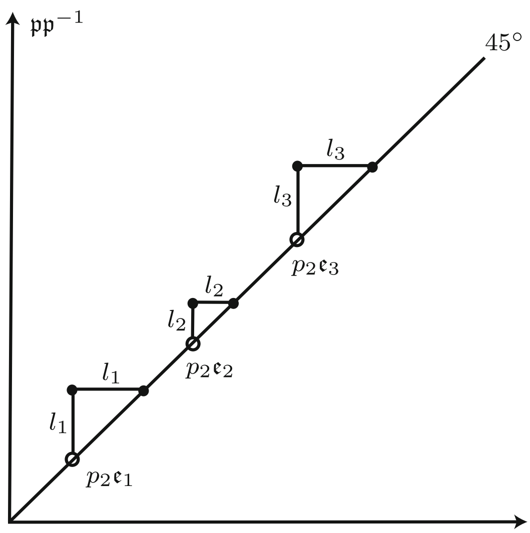

Figure 15.7

[21] $\mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+(t)$ is short for $\mathfrak{p}(\mathfrak{p}^{-1}(\mathfrak{t}^+(t)))$.

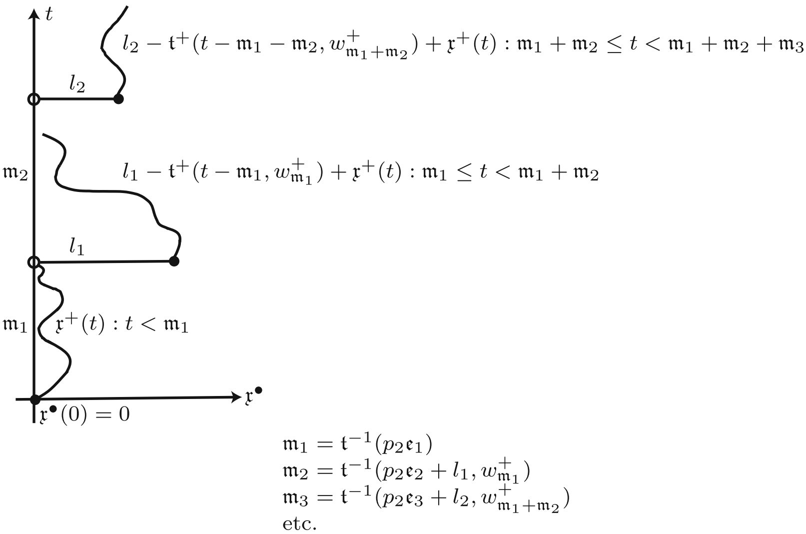

Figure 15.8

is left-continuous as usual and $\mathfrak{e}_1$, $\mathfrak{e}_2$, etc. are the exponential holding times between jumps of $\mathfrak{p}$.

Coming to the case $p_4 = +\infty$, $\mathfrak{p}(t)$ experiences an infinite number of jumps during each time interval $[t_1, t_2)$ $(t_1 < t_2)$, but

$$\mathfrak{p}\big([t_1, t_2) \times [\varepsilon, +\infty)\big) < \infty \quad \big(t_2 < +\infty,\ \varepsilon > 0\big),$$

and so it is legitimate to label the jumps as follows:

(a) arrange in separate rows the jumps occurring in $(0, 1]$, $(1, 2]$, etc.;
(b) in each row, arrange the jumps in order of magnitude beginning with the largest one;
(c) if several jumps of the same magnitude occur in a single row, arrange them in correct temporal order;
(d) number the rows jumps row by row as indicated below:

$$\begin{aligned} &l_1 \geqq l_3 \geqq l_6 \geqq l_{10}, \\ &l_2 \geqq l_5 \geqq l_9, \\ &l_4 \geqq l_8, \\ &l_7 \quad \text{etc.} \end{aligned}$$

Figure 15.8 gives an approximate idea of the sample path in the case $p_2 = 0$. Figure 15.9 $(p_2 = 0,\ \mathfrak{x}(0) = 0)$ is based on the alternative description 3: the standard Brownian path $\mathfrak{x}^-$ has been slanted off to the left for the purposes of the picture, and the rule is to translate the excursions of $\mathfrak{x}^-$ between the endpoints of the flat stretches of $\mathfrak{p}^{-1}$ until the left legs of the hatched curvilinear triangles abut on the time axis and then to fill up the gaps with $\mathfrak{x}^\bullet = 0$.

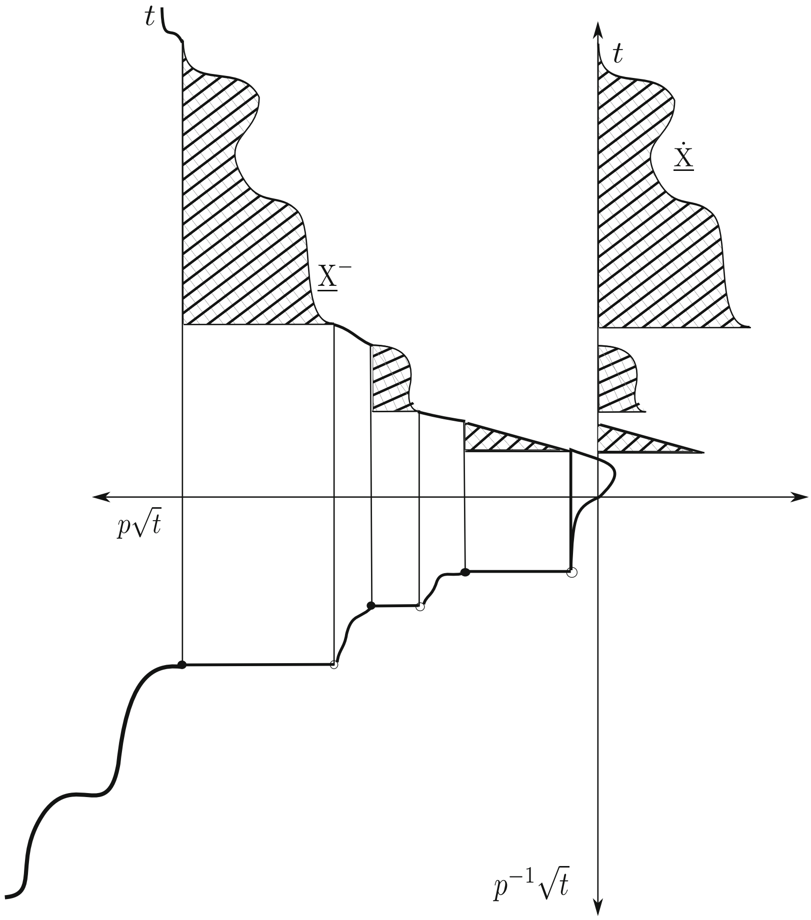

Figure 15.9

The picture is not so simple in case $p_2 > 0$: then $\mathfrak{Q} = (l : \mathfrak{p}\mathfrak{p}^{-1}(l) = l)$ has positive measure, and, on $\mathfrak{Q}^- = (t : t^{-1}(t) \in \mathfrak{Q})$, $\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}t^- - \mathfrak{x}^-$ reduces to the reflecting Brownian motion $\mathfrak{t}^- - \mathfrak{x}^- = \mathfrak{x}^+$.

15.13. Simple Markovian Character: $p_1 = p_3 = 0$ ($p_2 > 0/p_4 = +\infty$)

Consider the sample path

(1) $$\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+ = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^- + \mathfrak{x}^-$$

described in Sect. 15.12.

Given $t_2 \geqq t_1 \geqq 0$, if $\mathfrak{m} = \mathfrak{p}^{-1}\mathfrak{t}^-(t_1)$, if $\mathfrak{p}_+(t) = \mathfrak{p}(t+\mathfrak{m}) - \mathfrak{p}(\mathfrak{m})$, and if $\mathfrak{t}_+^- = -\min_{s\leqq t}[\mathfrak{x}^-(s+t_1) - \mathfrak{x}^-(t_1)]$, then, as the reader will check,

$$
\begin{aligned}
&\mathfrak{p}^{-1}\mathfrak{t}^-(t_2) - \mathfrak{p}^{-1}\mathfrak{t}^-(t_1) \\
&\quad = \inf\big(s : \mathfrak{p}(s) > \mathfrak{t}^-(t_2) - \mathfrak{p}^{-1}\mathfrak{t}^-(t_1)\big) \\
&\quad = \inf\big(s : \mathfrak{p}(s+\mathfrak{m}) > \mathfrak{t}^-(t_2)\big) \\
&\quad = \inf\big(s : \mathfrak{p}_+(s) + \mathfrak{p}(\mathfrak{m}) > \big[\mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^-(t_1)\big] \vee \mathfrak{t}^-(t_1)\big)^{22} \\
&\quad = \inf\big(s : \mathfrak{p}_+(s) > \big[\mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^\bullet(t_1)\big] \vee \big[\mathfrak{t}^-(t_1) - \mathfrak{p}(\mathfrak{m})\big]\big) \\
&\quad = \inf\big(s : \mathfrak{p}_+(s) > \big[\mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^\bullet(t_1)\big] \vee 0\big),
\end{aligned}
\tag{2}
$$

where the last step is justified as follows: $\mathfrak{a} = \mathfrak{t}^-(t_1) - \mathfrak{p}(\mathfrak{m}) \leqq 0$ since *either* $p_2 > 0$ *or* $p_4(0,+\infty) = +\infty$, $\mathfrak{p}^{-1}(0) = 0$, and it follows that *either* $\mathfrak{b} = \mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^\bullet(t_1) < 0$ and $\inf(s : \mathfrak{p}_+(s) > \mathfrak{a}\vee\mathfrak{b}) = \inf(s : \mathfrak{p}_+ > 0) = 0$ *or else* $\mathfrak{b} \geqq 0$ and $\mathfrak{a}\vee\mathfrak{b} = \mathfrak{b}$.

Coming to the sample path itself, an application of (2) implies

$$
\begin{aligned}
\mathfrak{x}^\bullet(t_2) &= \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^-(t_2) + \mathfrak{x}^-(t_2) \\
&= \mathfrak{p}\big(\mathfrak{p}_+^{-1}\big(\big[\mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^\bullet(t_1)\big]\vee 0\big) + \mathfrak{m}\big) + \mathfrak{x}^-(t_2) \\
&= \mathfrak{p}_+\mathfrak{p}_+^{-1}\big(\big[\mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^\bullet(t_1)\big]\vee 0\big) + \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^-(t_1) + \mathfrak{x}^-(t_2) \\
&= \mathfrak{p}_+\mathfrak{p}_+^{-1}\big(\big[\mathfrak{t}_+^-(t_2-t_1) - \mathfrak{x}^\bullet(t_1)\big]\vee 0\big) + \big[\mathfrak{x}^-(t_2) - \mathfrak{x}^-(t_1)\big] + \mathfrak{x}^\bullet(t_1) \\
&= \mathfrak{p}_+\mathfrak{p}_+^{-1}\mathring{\mathfrak{t}}(t_2-t_1) + \mathring{\mathfrak{x}}(t_2-t_1)
\end{aligned}
\tag{3}
$$

with

$$
t \longrightarrow \mathring{\mathfrak{x}}(t) = \big[\mathfrak{x}^-(t+t_1) - \mathfrak{x}^-(t_1)\big] + \mathfrak{x}^\bullet(t_1)
\tag{4a}
$$

Consider this conditional on $\mathfrak{x}^\bullet(t_1) = a \geqq 0$.

Because of the differential character of the standard Brownian motion $\mathfrak{x}^-$, $\mathring{\mathfrak{x}}$ is likewise a standard Brownian motion starting at $\mathring{\mathfrak{x}}(0) = \mathfrak{x}^\bullet(t_1) = a$, independent of $\mathfrak{x}^-(s) : s \leqq t_1$ and of $\mathfrak{p}$ (and hence independent of $\mathfrak{x}^\bullet(s) : s \leqq t$, and of $\mathfrak{p}_+$ also) with minimum function

$$
\begin{aligned}
-\Big(\min_{s\leqq t}\mathring{\mathfrak{x}}(s)\wedge 0\Big) &= -\Big(\min_{s\leqq t}\big[\mathfrak{x}^-(s+t_1) - \mathfrak{x}^-(t_1)\big] + \mathfrak{x}^\bullet(t_1)\wedge 0\Big) \\
&= \Big[-\min_{s\leqq t}\big[\mathfrak{x}^-(s+t_1) - \mathfrak{x}^-(t_1)\big] - \mathfrak{x}^\bullet(t_1)\Big]\vee 0 \\
&= \big[\mathfrak{t}_+^-(t) - \mathfrak{x}^\bullet(t_1)\big]\vee 0 \\
&= \mathring{\mathfrak{t}}(t).
\end{aligned}
\tag{4b}
$$

Given $t \geqq 0$, the indicator of the event

$$
(\mathfrak{m} > t) = \big(\mathfrak{p}^{-1}\mathfrak{t}^-(t_1) > t\big) = \big(\mathfrak{t}^-(t_1) > \mathfrak{p}(t)\big)
\tag{5}
$$

is a Borel function of $\mathfrak{p}(s) : s \leqq t$ and $\mathfrak{x}^-(s) : s \leqq t_1$, and since $\mathfrak{x}^-$ and $\mathfrak{p}$ are independent, $\mathfrak{m}$ is a *stopping time* for $\mathfrak{p}$, i.e., $\mathfrak{p}_+$ is identical in law to $\mathfrak{p}$ and independent of $\mathfrak{x}^-$ and of $\mathfrak{p}(s) : s \leqq \mathfrak{m}$ and hence independent of $\mathfrak{x}^\bullet(s) : s \leqq t_1$ and of $\mathring{\mathfrak{x}}$.

[22] $a \vee b$ is the larger of a and b.

But now it is clear that, conditional on $\mathfrak{x}^\bullet(t_1) = a$, $\mathfrak{x}^\bullet(t_2)$ is independent of the past $\mathfrak{x}^\bullet(s) : s \leqq t_1$ with law

$$P_a\big[\mathfrak{x}^\bullet(t) \in db\big], \quad a = \mathfrak{x}^\bullet\big(t_1\big),\ t = t_2 - t_1 \tag{6}$$

as was to be proved.

15.14. Local Times: $p_1 = p_3 = 0$ $(p_2 > 0/p_4 = +\infty)$

Because the reflecting Brownian local time $\mathfrak{t}^+$ was central to the construction of the Brownian motions in the case $p_4 = 0$ treated in Sect. 15.10, one expects that a similar local time $\mathfrak{t}^\bullet$ based upon the path $\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+$ should figure in the general case; the purpose of this section is to prove its existence (Fig. 15.10).

Given $p_2 > 0$, the contention is that the *local time*

$$\mathfrak{t}^\bullet(t) = \lim_{\varepsilon \downarrow 0} \big(2\varepsilon p_2\big)^{-1} \text{measure}\,\big(s : \mathfrak{x}^\bullet(s) < \varepsilon,\ s \leqq t\big), \quad t \geqq 0 \tag{1a}$$

exists and can be expressed as

$$\begin{aligned} \mathfrak{t}^\bullet(t) &= p_2^{-1}\mathfrak{t}^+\big(\mathfrak{Q}^+ \cap [0,t]\big) \\ &= p_2^{-1}\big|\mathfrak{Q} \cap \big[0, \mathfrak{t}^+(t)\big]\big| \\ &= \mathfrak{p}^{-1}\mathfrak{t}^+(t), \end{aligned} \tag{1b}$$

in which

$$\mathfrak{Q} = \big(t : \mathfrak{p}\mathfrak{p}^{-1}(t) = t\big), \tag{2a}$$

$$\mathfrak{Q}^+ = \big(t : \mathfrak{t}^+(t) \in \mathfrak{Q}\big). \tag{2b}$$

Consider, for the proof, the intervals $[l_1^-, l_1^+)$, $[l_2^-, l_2^+)$, etc. of the complement of $\mathfrak{Q}$, and note that the complement of $\mathfrak{Q}^+$ is the union of the intervals $[\mathfrak{t}^{-1}(l_1^-), \mathfrak{t}^{-1}(l_1^+))$, $[\mathfrak{t}^{-1}(l_2^-), \mathfrak{t}^{-1}(l_2^+))$, etc., whence[23] $\partial\mathfrak{Q}^+$ is *countable*.

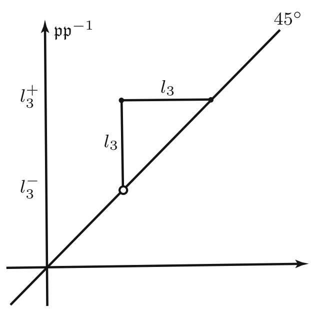

Figure 15.10

[23] $\partial\mathfrak{Q}^+$ denotes the boundary of $\mathfrak{Q}^+$.

Because $\mathfrak{t}^+$ is continuous and $\mathfrak{x}^\bullet = \mathfrak{x}^+$ on $\mathfrak{Q}^+$,

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0}(2\varepsilon)^{-1}\,\text{measure}\,\big(s : \mathfrak{x}^\bullet(s) < \varepsilon,\ s \in \mathfrak{Q}^+ \cap [0,t]\big)\\
&\qquad = \lim_{\varepsilon\downarrow 0}(2\varepsilon)^{-1}\,\text{measure}\,\big(s : \mathfrak{x}^+(s) < \varepsilon,\ s \in \mathfrak{Q}^+ \cap [0,t]\big)\\
&\qquad = \mathfrak{t}^+\big(\mathfrak{Q}^+ \cap [0,t]\big).
\end{aligned}\tag{3}$$

Consider, next,

$$\begin{aligned}
\mathfrak{Q}^-_\varepsilon &= \big(t : t \notin \mathfrak{Q}^+,\ \mathfrak{x}^\bullet < \varepsilon\big)\\
&= \bigcup_{n\geqq 1}\big[\mathfrak{t}^{-1}(l^-_n), \mathfrak{t}^{-1}(l^+_n)\big) \cap \big(t : l^+_n - \mathfrak{t}^+ < \varepsilon\big)\\
&= \bigcup_{n\geqq 1}\big[\mathfrak{t}^{-1}(l^-_n), \mathfrak{t}^{-1}(l^+_n)\big) \cap \big[\mathfrak{t}^{-1}(l^+_n - \varepsilon), +\infty\big)\\
&= \bigcup_{n\geqq 1}\big[\mathfrak{t}^{-1}\big(l^-_n \vee (l^+_n - \varepsilon)\big), \mathfrak{t}^{-1}(l^+_n)\big).
\end{aligned}\tag{4}$$

Because $\mathfrak{t}^{-1}$ is left-continuous, $\bigcap_{\varepsilon>0}\mathfrak{Q}^-_\varepsilon = \emptyset$, and seeing as $\partial\mathfrak{Q}^-_\varepsilon$ is countable and $\mathfrak{t}^+$ is continuous, it develops, much as in (3), that

$$\begin{aligned}
&\overline{\lim_{\varepsilon\downarrow 0}}(2\varepsilon)^{-1}\,\text{measure}\,\big(s : \mathfrak{x}^\bullet(s) < \varepsilon,\ s \in [0,t] - \mathfrak{Q}^+\big)\\
&\qquad = \overline{\lim_{\varepsilon\downarrow 0}}(2\varepsilon)^{-1}\,\text{measure}\,\big(s : \mathfrak{x}^+(s) < \varepsilon,\ s \in \mathfrak{Q}^-_\delta \cap [0,t]\big)\\
&\qquad = \mathfrak{t}^+\big(\mathfrak{Q}^-_\delta \cap [0,t]\big)\\
&\qquad \downarrow 0 \quad (\delta \downarrow 0),
\end{aligned}\tag{5}$$

which justifies the definition (1a) and the first line of (1b); the second line of (1b) is immediate from the definition of $\mathfrak{Q}^+$; as to the third line,

$$\begin{aligned}
\mathfrak{p}\mathfrak{p}^{-1} &= t, \quad t \in \mathfrak{Q},\\
&= l^+_n, \quad t \in \big[l^-_n, l^+_n\big)\ (n \geqq 1),\\
&= p_2\mathfrak{p}^{-1} + \int_{0+} l_{\mathfrak{p}}\big(\big[0, \mathfrak{p}^{-1}\big] \times dl\big),
\end{aligned}\tag{6}$$

and, picking out the continuous part on both sides, it is clear that

$$\begin{aligned}
\mathfrak{p}^{-1}(dt) &= p_2^{-1}dt \quad \text{on } \mathfrak{Q},\\
&= 0 \qquad\quad \text{off } \mathfrak{Q},
\end{aligned}\tag{7}$$

completing the proof.

$\mathfrak{p}^{-1}\mathfrak{t}^+$ can still be interpreted as a local time in case $p_2 = 0$ ($p_4 = +\infty$):

$$\mathfrak{p}^{-1}\mathfrak{t}^+(t) = \lim_{\varepsilon\downarrow 0}\frac{\sum_{l_n>\varepsilon}\text{measure}\,\big(s : \mathfrak{x}^\bullet(s) < \varepsilon, s \in \big[\mathfrak{t}^{-1}(l^-_n), \mathfrak{t}^{-1}(l^+_n)\big) \cap [0,t]\big)}{\varepsilon^2 p_4[\varepsilon, +\infty)}.\tag{8}$$

Consider, for the proof, the scaled *visiting times*:

$$\mathfrak{d}_n = \varepsilon^{-2}\,\text{measure}\,\big(s : \mathfrak{x}^\bullet(s) < \varepsilon,\ s \in \big[\mathfrak{t}^{-1}(l^-_n), \mathfrak{t}^{-1}(l^+_n)\big)\big).\tag{9}$$

Conditional on $\mathfrak{p}$ (i.e., conditional on $l_1^{\pm}$, $l_2^{\pm}$, etc.), the visiting times $\mathfrak{d}_n$ are independent because $\mathfrak{x}^+$ starts from scratch at the place $\mathfrak{x}^+(\mathfrak{m}) = 0$ at time $\mathfrak{m} = \mathfrak{t}^{-1}(l_n^-)$ $(n \geqq 1)$; in addition, if $l_n > \varepsilon$, then $\mathfrak{d}_n$ is identical in law to measure $(s : \mathfrak{x}(s) > 0,\ s < \mathfrak{m}_1)$, where $\mathfrak{x}$ is a standard Brownian motion starting at 0 and $\mathfrak{m}_1$ is its passage time to 1, as will now be verified.

Given $\sigma > 0$, the *scaling*

(10) $$\mathfrak{x}(t) \longrightarrow \sigma\mathfrak{x}\big(t/\sigma^2\big)$$

preserves the Wiener measure for standard Brownian paths starting at 0 and sends

(11a) $$\mathfrak{x}^+(t) \longrightarrow \sigma\mathfrak{x}^+\big(t/\sigma^2\big),$$

(11b) $$\mathfrak{t}^+(t) \longrightarrow \sigma\mathfrak{t}^+\big(t/\sigma^2\big),$$

(11c) $$\mathfrak{t}^{-1}(t) \longrightarrow \sigma^2\mathfrak{t}^{-1}\big(t/\sigma\big),$$

(12a) $$\mathfrak{x}^-(t) \longrightarrow \sigma\mathfrak{x}^-\big(t/\sigma^2\big),$$

(12b) $$\mathfrak{t}^-(t) \longrightarrow \sigma\mathfrak{t}^-\big(t/\sigma^2\big),$$

(12c) $$\mathfrak{m}_l \longrightarrow \sigma^2\mathfrak{m}_{l/\sigma},$$

where $\mathfrak{x}^-$ is the standard Brownian motion $\mathfrak{t}^+ - \mathfrak{x}^+$, $\mathfrak{t}^- = \mathfrak{t}^+ = \max_{s \leqq t}\mathfrak{x}^-(s)$, and $\mathfrak{m}_l = \min(t : \mathfrak{x}^- = l)$, and, using $\mathfrak{a} \equiv \mathfrak{b}$ to indicate that $\mathfrak{a}$ and $\mathfrak{b}$ are identical in law, it follows from the rules (11) and (12) that in case $l_n > \varepsilon$,

(13) $$\begin{aligned}
\mathfrak{d}_n &\equiv \varepsilon^{-2}\,\text{measure}\,\big(s : l_n - \mathfrak{t}^+(s) + \mathfrak{x}^+(s) < \varepsilon,\ s < \mathfrak{t}^{-1}(l_n)\big), \quad \mathfrak{x}^+(0) = 0,\\
&\equiv \varepsilon^{-2}\,\text{measure}\,\big(s : 1 - \mathfrak{t}^+\big(s/\sigma^2\big) + \mathfrak{x}^+\big(s/\sigma^2\big) < \varepsilon/\sigma,\ s/\sigma^2 < \mathfrak{t}^{-1}(1)\big), \quad \sigma = l_n,\\
&\equiv \varepsilon^{-2}l_n^2\,\text{measure}\,\big(s : 1 - \mathfrak{t}^+\big(s\big) + \mathfrak{x}^+(s) < \varepsilon/l_n,\ s < \mathfrak{t}^{-1}(1)\big)\\
&\equiv \varepsilon^{-2}l_n^2\,\text{measure}\,\big(s : \mathfrak{x}^-(s) > 1 - \varepsilon/l_n,\ s < \mathfrak{m}_1\big)\\
&\equiv \varepsilon^{-2}l_n^2\,\text{measure}\,\big(s : \mathfrak{x}^-(s) > 1 - \varepsilon/l_n,\ \mathfrak{m}_{1-\varepsilon/l_n} \leqq s < \mathfrak{m}_1\big)\\
&\equiv \varepsilon^{-2}l_n^2\,\text{measure}\,\big(s : \mathfrak{x}(s) > 0,\ s < \mathfrak{m}_{\varepsilon/l_n}\big), \quad \mathfrak{x}(0) = 0,\\
&\equiv \text{measure}\,\big(s : \mathfrak{x}(s) > 0,\ s < \mathfrak{m}_1\big),
\end{aligned}$$

where the scaling (10) was used in step 2 $(\sigma = l_n)$ and in step 7 $(\sigma = \varepsilon/l_n)$.

Coming back to (8), the strong law of large numbers combined with the rule

(14) $$\lim_{\varepsilon \downarrow 0} p_4[\varepsilon, +\infty)^{-1}\mathfrak{p}\big([0,t] \times [\varepsilon, +\infty)\big) = t \quad (t \geqq 0)$$

(see Sect. 15.11) and the simple evaluation

(15) $$\begin{aligned}
E_0\big(\mathfrak{d}_1\big) &= \int_0^{+\infty} dt P_1\big[\mathfrak{x}(t) < 1,\ \mathfrak{m}_0 > t\big]\\
&= \int_0^{+\infty} dt \int_0^1 \frac{e^{-(a-1)^2/2t} - e^{-(a+1)^2/2t} da}{(2\pi t)^{1/2}}\\
&= \int_0^1 2a da = 1,
\end{aligned}$$

justifies

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0}\frac{\sum_{l_n>\varepsilon}\text{measure}\big(s:\mathfrak{x}^\bullet(s)<\varepsilon,\ s\in\big[\mathfrak{t}^{-1}\big(l_n^-\big),\mathfrak{t}^{-1}\big(l_n^+\big)\big)\cap[0,t]\big)}{\varepsilon^2 p_4[\varepsilon,+\infty)}\\
&\qquad=\lim_{\varepsilon\downarrow 0}\sum_{\substack{l_n>\varepsilon\\ \mathfrak{t}^{-1}(l_n^+)\leqq t}}\mathfrak{d}_n/p_4[\varepsilon,+\infty)\\
(16)\qquad&\qquad=E_0\big(\mathfrak{d}_1\big)\lim_{\varepsilon\downarrow 0}\frac{\#\big(l_n:l_n>\varepsilon,\ \mathfrak{t}^{-1}\big(l_n^+\big)\leqq t\big)}{p_4[\varepsilon,+\infty)}{}^{24}\\
&\qquad=\lim_{\varepsilon\downarrow 0}\frac{\#\big(l_n:l_n>\varepsilon,\ l_n^+<\mathfrak{t}^+(t)\big)}{p_4[\varepsilon,+\infty)}\\
&\qquad=\lim_{\varepsilon\downarrow 0}\frac{\mathfrak{p}\big(\big[0,\mathfrak{p}^{-1}\mathfrak{t}^+(t)\big]\times[\varepsilon,+\infty)\big)}{p_4[\varepsilon,+\infty)}\\
&\qquad=\mathfrak{p}^{-1}\mathfrak{t}^+(t),
\end{aligned}$$

where the use of $l_n^+<\mathfrak{t}^+(t)$ in place of $\mathfrak{t}^{-1}(l_n^+)\leqq t$ in step 3 is justified because both describe the same class of jumps plus or minus a single jump and $p_4[\varepsilon,+\infty)\uparrow+\infty$ as $\varepsilon\downarrow 0$; a picture helps to see that $l_n^+<\mathfrak{t}^+$ and $\mathfrak{p}^{-1}(l_n^+)<\mathfrak{p}^{-1}(\mathfrak{t}^+)$ are identical as needed in step 4.

$\mathfrak{p}^{-1}\mathfrak{t}^+$ *cannot be computed from the sample path if* $p_2=0$ *and* $p_4<+\infty$, as is clear from Fig. 15.11 in which

$$\begin{aligned}
(17)\qquad \mathfrak{p}^{-1}\mathfrak{t}^+(t)&=\mathfrak{e}_1+\cdots+\mathfrak{e}_n,\\
\mathfrak{t}^{-1}\big(l_1+\cdots+l_{n-1}\big)&\leqq t<\mathfrak{t}^{-1}\big(l_1+\cdots+l_n\big),
\end{aligned}$$

and $\mathfrak{x}^\bullet$ is independent of the holding times $\mathfrak{e}_1$, $\mathfrak{e}_2$, etc. But it still has some features of a local time: it is the sum of n independent holding times $\mathfrak{e}$ with common conditional law $P_\cdot(\mathfrak{e}_1>t\mid\mathfrak{x}^\bullet)=e^{-p_4t}$, where n is the number of times that the sample path approaches 0 before time t (see Fig. 15.12).

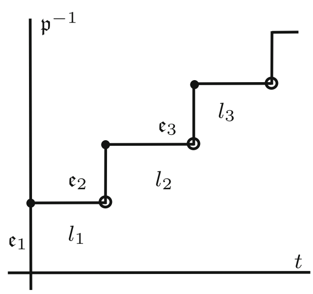

Figure 15.11

[24] $\#(l_n$: etc.) denotes the number of jumps l_n with the properties described inside.

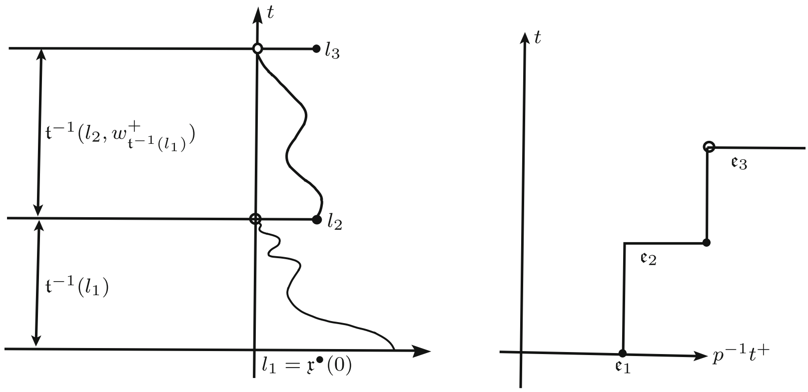

Figure 15.12

15.15. Sample Paths and Green Operators: $p_1 u(0) + p_3(\mathbf{G}u)(0) = p_2 u^+(0) + \int_{0+} [u(l) - u(0)] p_4(dl)$ $(p_2 > 0/p_4 = +\infty)$

Consider the motion $\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+$ and its local time $\mathfrak{t}^\bullet = \mathfrak{p}^{-1}\mathfrak{t}^+$, and let us use them to build up the sample paths in the general case $(p_2 > 0/p_4 = +\infty)$ imitating the prescription of Sect. 15.10:

(1a)
$$\mathfrak{y}^\bullet(t) = \mathfrak{x}^\bullet\left[\mathfrak{f}^{-1}(t)\right] \quad \text{if } t < \mathfrak{m}^\bullet_\infty, \qquad = \infty \quad \text{if } t \geqq \mathfrak{m}^\bullet_\infty,$$

(1b)
$$\mathfrak{f}(t) = t + p_2 \mathfrak{t}^\bullet(t),$$

(1c)
$$P_\cdot\left(\mathfrak{m}^\bullet_\infty > t \mid \mathfrak{x}^\bullet\right) = e^{-p_1 t^\bullet[\mathfrak{f}^{-1}(t)]}.$$

Given $l \geqq 0$,

(2)
$$\begin{aligned}
\left(G^\bullet_\alpha f\right)(l) &= E_l\left(\int_0^{m^\bullet_\infty} e^{-\alpha t} f\left(\mathfrak{y}^\bullet\right) dt\right) \\
&= E_l\left(\int_0^{+\infty} e^{-\alpha t} e^{-p_1 \mathfrak{t}^\bullet(\mathfrak{f}^{-1})} f\left[\mathfrak{x}^\bullet\left(\mathfrak{f}^{-1}\right)\right] dt\right) \\
&= E_l\left[\int_0^{+\infty} e^{-\alpha \mathfrak{f}} e^{-p_1 \mathfrak{t}^\bullet} f\left(\mathfrak{x}^\bullet\right) f(dt)\right] \\
&= E_l\left[\int_0^{\mathfrak{m}_0} e^{-\alpha t} f\left(\mathfrak{x}^+\right) dt\right] \\
&\quad + E_l\left(e^{-\alpha \mathfrak{m}_0}\right) E_0\left[\int_0^{+\infty} e^{-\alpha \mathfrak{t}} e^{-p_1 \mathfrak{t}^\bullet} f\left(\mathfrak{x}^\bullet\right) \mathfrak{f}(dt)\right]^{25} \\
&= \left(G^-_\alpha f\right)(l) + e^{-(2\alpha)^{1/2} l}\left(G^\bullet_\alpha f\right)(0),
\end{aligned}$$

especially, the Green operators map $C[0, +\infty)$ into itself in the special case $p_1 = p_2 = 0$ $(\mathfrak{y}^\bullet = \mathfrak{x}^\bullet)$, and since $\mathfrak{x}^\bullet$ starts afresh at constant times, it follows that it must be a Brownian motion. $\mathfrak{y}^\bullet$ is likewise a Brownian motion as is clear on arguing as in Sect. 15.10 with $\mathfrak{x}^\bullet$ and

[25] $\mathfrak{x}^\bullet = \mathfrak{x}^+$ and $\mathfrak{t}^\bullet = 0$ up to time $\mathfrak{m}_0 = \min(t : \mathfrak{x}^+ = 0)$, and $\mathfrak{x}^\bullet$ starts afresh at that moment.

$\mathfrak{t}^\bullet$ in place of $\mathfrak{x}^+$ and $\mathfrak{t}^+$, and now, for the identification of its generator as the contraction of $\mathbf{G} = D^2/2$ to

$$D(\mathbf{G}^\bullet) = C^2[0,+\infty) \cap \left(u : p_1 u(0) + p_3(\mathbf{G}u)(0) = p_2 u^+(0) + \int_{0+} [u(l) - u(0)] p_4(dl)\right), \tag{3}$$

it suffices to make the evaluation

$$\begin{aligned} e = (G^\bullet_\alpha f)(0) &= E_0\left[\int_0^{+\infty} e^{-\alpha \mathfrak{f}} e^{-p_1 \mathfrak{t}^\bullet} f(\mathfrak{x}^\bullet)\mathfrak{f}(dt)\right] \\ &= \frac{p_2 2\int_{0+} e^{-(2\alpha)^{1/2} l} f(l)dl + p_3 f(0) + \int_{0+}(G^-_\alpha f)(l) p_4(dl)}{p_1 + (2\alpha)^{1/2} p_2 + \alpha p_3 + \int_{0+}[1 - e^{-(2\alpha)^{1/2} l}] p_4(dl)} \end{aligned} \tag{4}$$

e is decomposed into simpler integrals in several steps (see the explanation below):

$$\begin{aligned} e &= E_0\left[\int_0^{+\infty} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{t}^\bullet} f(\mathfrak{x}^\bullet) dt\right] \\ &\quad + p_3 f(0) E_0\left[\int_0^{+\infty} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{t}^\bullet} \mathfrak{t}^\bullet(dt)\right] \\ &= \sum_{n\geqq 1} E_0\left[\int_{[\mathfrak{t}^{-1}(l_n^-),\, \mathfrak{t}^{-1}(l_n^+))} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}\mathfrak{t}^+} f(l_n^+ - \mathfrak{t}^+ + \mathfrak{x}^+) dt\right] \\ &\quad + E_0\left[\int_0^{+\infty} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}\mathfrak{t}^+} f(\mathfrak{x}^+) dt\right] \\ &\quad - \sum_{n\geqq 1} E_0\left[\int_{[\mathfrak{t}^{-1}(l_n^-),\mathfrak{t}^{-1}(l_n^+))} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}\mathfrak{t}^+} f(\mathfrak{x}^+) dt\right] \\ &\quad + p_3 f(0) E_0\left[\int_0^{+\infty} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}\mathfrak{t}^+} \mathfrak{p}^{-1}\mathfrak{t}^+(dt)\right] \\ &= \sum_{n\geqq 1} E_0\left[e^{-\alpha \mathfrak{t}^{-1}(l_n^-)} e^{-(p_1+\alpha p_3)p^{-1}(l_n^+)} \cdot E_0\left(\int_0^{\mathfrak{t}^{-1}(l_n)} e^{-\alpha t} f[l_n - \mathfrak{t}^+ + \mathfrak{x}^+] dt \mid l_n\right)\right] \\ &\quad + E_0\left[\int_0^{+\infty} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}\mathfrak{t}^+} f(\mathfrak{x}^+) dt\right] \\ &\quad - \sum_{n\geqq 1} E_0\left[e^{-\alpha \mathfrak{t}^{-1}(l_n^-)} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(l_n^+)} E_0\left(\int_0^{\mathfrak{t}^{-1}(l_n)} e^{-\alpha t} f(\mathfrak{x}^+) dt \mid l_n\right)\right] \\ &\quad + \frac{p_3 f(0)}{p_1 + \alpha p_3}\left[1 - \alpha E_0\left(\int_0^{+\infty} e^{-\alpha t} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}\mathfrak{t}^+} dt\right)\right]^{26} \\ &\equiv e_1 + e_2 - e_3 + e_4, \end{aligned} \tag{5}$$

where $\mathfrak{t}^\bullet(dt) = 0$ *outside* $\mathfrak{Z}^\bullet \equiv (t : \mathfrak{x}^\bullet = 0)$ was used in step 1, $[0,+\infty)$ was split into $\mathfrak{Q}^+ + \bigcup_{n\geqq 1}[\mathfrak{t}^{-1}(l_n^-), \mathfrak{t}^{-1}(l_n^+))$ in step 2, and $\mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+$ was evaluated as $\mathfrak{t}^+$ or l_n^+ according as $t \in \mathfrak{Q}^+$ or $\mathfrak{t}^{-1}(l_n^-) \leqq t < \mathfrak{t}^{-1}(l_n^+)$, and, in step 3, it was noted that, conditional on $\mathfrak{p}$, the

[26] $p_3/(p_1 + \alpha p_3) = 0$ if $p_3 = 0$.

standard Brownian traveller starts afresh at time $\mathfrak{m} = \mathfrak{t}^{-1}(l_n^-)$ at the place $l = 0$; the addition rule

$$\mathfrak{t}^{-1}\big(l_n^+\big) = \mathfrak{t}^{-1}\big(l_n^-\big) + \mathfrak{t}^{-1}\big(l_n, w^+_{\mathfrak{t}^{-1}(l_n^-)}\big)$$

was also used in step 3, and a partial (time) integration was performed under the expectation sign in the expression for e_4.

To compute e_1, substitute the standard Brownian motion $\mathfrak{x}^- = \mathfrak{t}^+ - \mathfrak{x}^+$ and its passage times $\mathfrak{m}_l = \mathfrak{t}^{-1}(l)$ into the conditional expectation and integrate them out, next integrate out $\mathfrak{t}^{-1}(l_n^-)$ conditional on $\mathfrak{p}$, express the integral in terms of the Poisson measure $\mathfrak{p}(dt \times dl)$, and use the differential character of the latter to integrate it out also:

$$\begin{aligned}
e_1 &= \sum_{n\geqq 1} E_0\left[e^{-\alpha\mathfrak{t}^{-1}(l_n^-)} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(l_n^+)} E_0\left(\int_0^{\mathfrak{m}_{l_n}} e^{-\alpha t} f\big(l_n - \mathfrak{x}^-\big)dt \mid l_n\right)\right] \\
&= \sum_{n\geqq 1} E_0\big[e^{-\alpha\mathfrak{t}^{-1}(l_n^-)} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(l_n^+)}\big(G_\alpha^- f\big)\big(l_n\big)\big] \\
&= \sum_{n\geqq 1} E_0\big[e^{-(2\alpha)^{1/2} l_n^-} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(l_n^+)}\big(G_\alpha^- f\big)\big(l_n\big)\big] \\
&= E_0\left[\int_{[0,+\infty)\times(0,+\infty)} \mathfrak{p}(dt\times dl) e^{-(2\alpha)^{1/2}\mathfrak{p}(t-)} e^{-(p_1+\alpha p_3)t}\big(G_\alpha^- f\big)(l)\right] \\
&= \lim_{\varepsilon\downarrow 0} E_0\left[\int_{[\varepsilon,+\infty)\times(0,+\infty)} \mathfrak{p}(dt\times dl) e^{-(2\alpha)^{1/2}\mathfrak{p}(t-\varepsilon)} e^{-(p_1+\alpha p_3)t}\big(G_\alpha^- f\big)(l)\right] \\
&= \int_{[0,+\infty)\times(0,+\infty)} dt\, p_4(dl) \exp\left\{-t\left[p_2(2\alpha)^{1/2} + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2} l}\big)p_4(dl)\right]\right\} \\
&\qquad\qquad \times e^{-(p_1+\alpha p_3)t}\big(G_\alpha^- f\big)(l)
\end{aligned}$$

$$(6) \qquad = \frac{\int_{0+}\big(G_\alpha^- f\big)(l)p_4(dl)}{p_1 + (2\alpha)^{1/2}p_2 + \alpha p_3 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2} l}\big)p_4(dl)}.$$

To compute e_2, use the joint law $\frac{2(b+a)}{(2\pi t^3)^{1/2}} e^{-(b+a)^2/2t} da\, db$ of $\mathfrak{x}^+$ and $\mathfrak{t}^+$:

$$\begin{aligned}
e_2 &= E_0\left[\int_0^{+\infty} e^{-\alpha t} dt \int_0^{+\infty} db \int_0^{+\infty} da\, 2\frac{b+a}{\big(2\pi t^3\big)^{1/2}} e^{-(b+a)^2/2t}\right. \\
&\qquad\qquad\qquad\qquad \left. \times e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(b)} f(a)\right] \\
&= E_0\left[\int_0^{+\infty} e^{-(2\alpha)^{1/2} b} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(b)} db\right] 2\int_{0+} e^{-(2\alpha)^{1/2} a} f(a)\, da \\
&= E_0\left[\int_0^{+\infty} e^{-(2\alpha)^{1/2}\mathfrak{p}} e^{-(p_1+\alpha p_3)t}\mathfrak{p}(dt)\right] 2\int_{0+} e^{-(2\alpha)^{1/2} l} f(l)\, dl \\
&= \frac{p_1 + \alpha p_3}{(2\alpha)^{1/2}} E_0\left[\int_0^{+\infty} e^{-(p_1+\alpha p_3)t}\big(1 - e^{-(2\alpha)^{1/2}\mathfrak{p}}\big)dt\right] 2\int_{0+} e^{-(2\alpha)^{1/2} l} f(l)\, dl
\end{aligned}$$

$$= \frac{(2\alpha)^{1/2}p_2 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)}{p_1 + (2\alpha)^{1/2}p_2 + \alpha p_3 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)} \times \frac{2}{(2\alpha)^{1/2}}\int_{0+} e^{-(2\alpha)^{1/2}l} f(l)\,dl. \tag{7}$$

To compute e_3, use the same manipulations as for e_1 together with the lemma[27]

$$E_0\bigg(\int_0^{\mathfrak{t}^{-1}(l)} e^{-\alpha t} f\big(\mathfrak{x}^+\big)\,dt\bigg) = \big(G_\alpha^+ f\big)(0)\big[1 - e^{-(2\alpha)^{1/2}l}\big], \tag{8}$$

obtaining

$$\begin{aligned} e_3 &= \sum_{n\geqq 1} E_0\big[e^{-(2\alpha)^{1/2}l_n^-} e^{-(p_1+\alpha p_3)\mathfrak{p}^{-1}(l_n^+)}\big(1 - e^{-(2\alpha)^{1/2}l_n}\big)\big]\big(G_\alpha^+ f\big)(0) \\ &= E_0\bigg[\int_{[0,+\infty)\times(0,+\infty)} \mathfrak{p}(dt\times dl) e^{-(2\alpha)^{1/2}\mathfrak{p}(t-)} e^{-(p_1+\alpha p_3)t}\big(1 - e^{-(2\alpha)^{1/2}l}\big)\bigg] \\ &\quad \times \big(G_\alpha^+ f\big)(0) \\ &= \frac{\int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)}{p_1 + (2\alpha)^{1/2}p_2 + \alpha p_3 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)} \\ &\quad \times \frac{2}{(2\alpha)^{1/2}}\int_{0+} e^{-(2\alpha)^{1/2}l} f(l)\,dl. \end{aligned} \tag{9}$$

To compute e_4, use (6) with $f = 1$:

$$\begin{aligned} e_4 &= \frac{p_3 f(0)}{p_1 + \alpha p_3}\bigg[1 - \frac{(2\alpha)^{1/2}p_2 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)}{p_1 + (2\alpha)^{1/2}p_2 + \alpha p_3 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)}\bigg] \\ &= \frac{p_3 f(0)}{p_1 + (2\alpha)^{1/2}p_2 + \alpha p_3 + \int_{0+}\big(1 - e^{-(2\alpha)^{1/2}l}\big)p_4(dl)}. \end{aligned} \tag{10}$$

Combining (5), (6), (7), (9), and (10) verifies (4), and that finishes the proof.

15.16. Bounded Interval: $[-1, +1]$

A Brownian motion on $[-1, +1]$ is defined as in Sect. 15.5 except that

$$\big|\mathfrak{x}^\bullet\big| \leqq 1, \quad t < \mathfrak{m}_\infty^\bullet; \tag{1}$$

the stopped path

$$\begin{gathered} \mathfrak{x}^\bullet(t) : t < \epsilon^\bullet = \lim_{\varepsilon\downarrow 0}\inf\big(t : \big|\mathfrak{x}^\bullet\big| > 1 - \varepsilon\big), \\ -1 < \mathfrak{x}^\bullet(0) = l < +1 \end{gathered} \tag{2a}$$

is now identical in law to the stopped standard Brownian path

$$\mathfrak{x}(t) : t < \epsilon = \min\big(t : |\mathfrak{x}| = 1\big), \quad \mathfrak{x}(0) = l. \tag{2b}$$

Except in the case $P_\cdot^\bullet[|\mathfrak{x}^\bullet(\epsilon^\bullet)| = 1] < 1$ which can be treated as in Sect. 15.6, $C[-1, +1]$ is mapped into itself under the Green operators, $\mathbf{G}_\cdot^\bullet$ can be defined as before, and $D(\mathbf{G}^\bullet)$ can

[27] G_α^+ is the reflecting Brownian Green operator.

be described in terms of six nonnegative numbers $p_{\pm 1}$, $p_{\pm 2}$, $p_{\pm 3}$ and two nonnegative mass distributions $p_{\pm 4}(dl)$ subject to

$$p_{-1}+p_{-2}+p_{-3}+\int_{-1}^{+1}(1+l)p_{-4}(dl)=1, \quad p_{-4}(-1)=0, \tag{3a}$$

$$p_{+1}+p_{+2}+p_{+3}+\int_{-1}^{+1}(1-l)p_{+4}(dl)=1, \quad p_{+4}(+1)=0, \tag{3b}$$

$$p_{-4}(-1,+1]=+\infty \quad \text{in case } p_{-2}=p_{-3}=0, \tag{4a}$$

$$p_{+4}(-1,+1]=+\infty \quad \text{in case } p_{+2}=p_{+3}=0 \tag{4b}$$

as follows. $D(\mathbf{G}^\bullet)$ is the class of functions $u \in C^2[-1,+1]$ subject to

$$\begin{aligned} &p_{-1}u(-1)-p_{-2}u^{+}(-1)+p_{-3}(\mathbf{G}u)(-1) \\ &\qquad = \int_{-1}^{+1}\big[u(l)-u(-1)\big]p_{-4}(dl), \end{aligned} \tag{5a}$$

$$\begin{aligned} &p_{+1}u(+1)+p_{+2}u^{-}(+1)+p_{+3}(\mathbf{G}u)(+1) \\ &\qquad = \int_{-1}^{+1}\big[u(l)-u(+1)\big]p_{+4}(dl); \end{aligned} \tag{5b}$$ [28]

and $\mathbf{G}^\bullet$ is the contraction of $\mathbf{G}=D^2/2$ to $D(\mathbf{G}^\bullet)$;

$$\big(G^\bullet_\alpha f\big)(l)=\big(G^-_\alpha f\big)(l)+e_-(l)\big(G^\bullet_\alpha f\big)(-1)+e_+(l)\big(G^\bullet_\alpha f\big)(+1), \quad |l|\leqq 1, \tag{6}$$

in which

$$\big(G^-_\alpha f\big)(a)=E_a\bigg(\int_0^{\epsilon} e^{-\alpha t}f(\mathfrak{x})\,dt\bigg)=2\int_{-1}^{+1}G(a,b)f(b)\,db, \tag{7a}$$ [29]

$$G(a,b)=G(b,a)=\frac{\sinh(2\alpha)^{1/2}(1+a)\sinh(2\alpha)^{1/2}(1-b)}{(2\alpha)^{1/2}}, \quad a\leqq b, \tag{7b}$$

is the Green operator for the Brownian motion with instant killing at ± 1 and

$$e_-(l)=\frac{\sinh(2\alpha)^{1/2}(1-l)}{\sinh 2(2\alpha)^{1/2}}=E_l\big(e^{-\alpha\mathfrak{m}_{-1}},\ \mathfrak{m}_{-1}<\mathfrak{m}_{+1}\big), \tag{8a}$$

$$e_+(l)=\frac{\sinh(2\alpha)^{1/2}(1+l)}{\sinh 2(2\alpha)^{1/2}}=E_l\big(e^{-\alpha\mathfrak{m}_{+1}},\ \mathfrak{m}_{+1}<\mathfrak{m}_{-1}\big); \tag{8b}$$

and, substituting (6) into (5) and solving for $(G^\bullet_\alpha f)(\pm 1)$, one obtains

$$\begin{aligned} \begin{bmatrix}\big(G^\bullet_\alpha f\big)(-1)\\ \big(G^\bullet_\alpha f\big)(+1)\end{bmatrix} &= \begin{bmatrix} e_{11} & e_{12}\\ e_{21} & e_{22}\end{bmatrix}^{-1} \\ &\cdot \begin{bmatrix} p_{-2}\big(G^-_\alpha f\big)^{+}(-1)+p_{-3}f(-1)+\int_{-1}^{+1}\big(G^-_\alpha f\big)(l)p_{-4}(dl) \\ -p_{+2}\big(G^-_\alpha f\big)^{-}(+1)+p_{+3}f(+1)+\int_{-1}^{+1}\big(G^-_\alpha f\big)(l)p_{+4}(dl)\end{bmatrix}, \end{aligned} \tag{9}$$

[28] $u^{-}(+1)=\lim_{\varepsilon\downarrow 0}\varepsilon^{-1}[u(1)-u(1-\varepsilon)]$.

[29] $P.$, $E.$, $\mathfrak{x}$, $\mathfrak{m}$ are the standard Brownian probabilities, expectations, sample paths, and passage times.

where the exponent -1 indicates the inverse matrix, and

$$e_{11} = p_{-1} - e_-^+(-1)p_{-2} + \alpha p_{-3} + \int_{-1}^{+1} (1-e_-)p_{-4}(dl), \tag{10a}$$

$$e_{12} = -p_{-2}\, e_+^+(-1) - \int_{-1}^{+1} e_+ p_{-4}(dl), \tag{10b}$$

$$e_{21} = p_{+2}\, e_-^-(+1) - \int_{-1}^{+1} e_- p_{+4}(dl), \tag{10c}$$

$$e_{22} = p_{+1} + e_+^-(+1)p_{+2} + \alpha p_{+3} + \int_{-1}^{+1} (1-e_+)p_{+4}(dl), \tag{10d}$$

all of which is due to W. Feller [3, 4]; the proofs can be carried out as in Sect. 15.8.

Coming to the sample paths, let us confine our attention to the case $p_{-4}(-1, +1] = p_{+4}[-1,+1) = +\infty$, leaving the opposite case to the reader.

Given a standard Brownian motion with sample paths $t \to \mathfrak{x}(t)$ and probabilities $P_a(B)$, if f is the map: $R^1 \to [-1,+1]$ defined by folding the line at ± 1, ± 3, ± 5, etc. as in Fig. 15.13, then $\mathfrak{x}^+ = f(\mathfrak{x})$ is the (reflecting) Brownian motion on $[-1,+1]$ associated with (5) in the special case $p_{\pm 1} = p_{\pm 3} = p_{\pm 4} = 0$ ($u^+(-1) = u^-(+1) = 0$); the dot sample path will be made up using $\mathfrak{x}^+$ and its local times

$$\mathfrak{t}^-(t) = \lim_{\varepsilon \downarrow 0} (2\varepsilon)^{-1} \text{ measure } \big(s : \mathfrak{x}^+(s) < -1+\varepsilon,\ s \leqq t\big), \tag{11a}$$

$$\mathfrak{t}^+(t) = \lim_{\varepsilon \downarrow 0} (2\varepsilon)^{-1} \text{ measure } \big(s : \mathfrak{x}^+(s) > 1-\varepsilon,\ s \leqq t\big), \tag{11b}$$

a pair of independent Poisson measures $\mathfrak{p}_\pm(dt \times dl)$ with means $dt\, p_{\pm 4}(dl)$, and the associated differential processes

$$\mathfrak{p}_-(t) = p_{-2}t + \int_{l>-1} (1+l)\mathfrak{p}_-\big([0,\,t] \times dl\big), \tag{12a}$$

$$\mathfrak{p}_+(t) = p_{+2}t + \int_{l<+1} (1-l)\mathfrak{p}_+\big([0,\,t] \times dl\big). \tag{12b}$$

Figure 15.14 depicts the sample paths associated with (5) if $p_{\pm 1} = p_{\pm 3} = 0$: $\mathfrak{x}^\bullet$ and $\mathfrak{x}^+$ agree up to time $\mathfrak{m}_{\pm 1} = \min(t : |\mathfrak{x}^+| = 1)$; if $\mathfrak{m}_{-1} < \mathfrak{m}_{+1}$, as in the picture, then $\mathfrak{x}^\bullet$ changes over into $\mathfrak{p}_-\mathfrak{p}_-^{-1}\mathfrak{t}^- - \mathfrak{t}^- + \mathfrak{x}^+$ until it hits $+1$, at which instant it changes over into $-\mathfrak{p}_+\mathfrak{p}_+^{-1}\mathfrak{t}^+ + \mathfrak{t}^+ + \mathfrak{x}^+$ until it hits -1 for the second time, etc.

If $p_{-1} = p_{+1} = 0$ and $p_{-3} + p_{+3} > 0$, then the desired motion is as in Fig. 15.14 but run with the new clock $\mathfrak{f}^{-1}$ which is the inverse function of

$$\mathfrak{f} = t + p_{-3}\mathfrak{t}^{\bullet -} + p_{+3}\mathfrak{t}^{\bullet +}, \tag{13a}$$

$$\mathfrak{t}^{\bullet \pm} = \mathfrak{p}_\pm^{-1}\mathfrak{t}^\pm, \tag{13b}$$

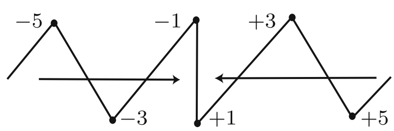

Figure 15.13

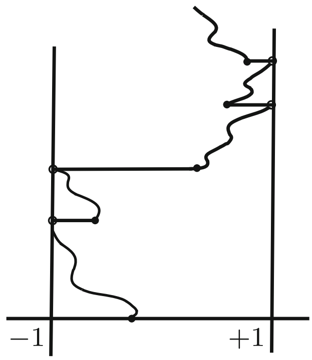

Figure 15.14

while, if $p_{-1}+p_{+1}>0$, then one has just to kill the above motion $\mathfrak{x}^\bullet(\mathfrak{f}^{-1})$ at time $\mathfrak{m}^\bullet_\infty$ with conditional law

(14) $$P_\cdot\big(\mathfrak{m}^\bullet_\infty > t \mid \mathfrak{x}^\bullet\big) = e^{-[p_{-1}\mathfrak{t}^{\bullet -}(\mathfrak{f}^{-1})+p_{+1}\mathfrak{t}^{\bullet +}(\mathfrak{f}^{-1})]};$$

the proofs are left to the industrious reader.

15.17. Two-Sided Barriers

A Brownian motion on R^1 with a *two-sided barrier* at $l=0$ is defined as in Sect. 15.5 except that

(1) $$\mathfrak{x}^\bullet \in R^1, \qquad t < \mathfrak{m}^\bullet_\infty,$$

and the stopped path

(2a) $$\mathfrak{x}^\bullet(t) : t < \epsilon^\bullet = \lim_{\varepsilon\downarrow 0}\inf\big(t : |x^\bullet| < \varepsilon\big), \quad \mathfrak{x}^\bullet(0) = l \in R^1 - 0$$

is identical in law to the stopped standard Brownian motion

(2b) $$\mathfrak{x}(t) : t < \epsilon = \min(t : \mathfrak{x} = 0), \quad \mathfrak{x}(0) = l.$$

Except in the case $P^\bullet_\cdot[\mathfrak{x}^\bullet(\epsilon^\bullet)=0]<1$, which is ignored as before, $C(R^1)$ is mapped into itself under the Green operators, $\mathbf{G}^\bullet$ is the contraction of $\mathbf{G}=D^2/2$ to[30]

(3) $$D\big(\mathbf{G}^\bullet\big) = C^{\bullet 2}\big(R^1\big) \cap \Big(u : p_1 u(0) + p_{-2}u^-(0) - p_{+2}u^+(0) + p_3(\mathbf{G}u)(0\pm) = \int_{|l|>0}\big[u(l)-u(0)\big]p_4(dl)\Big)$$

for some nonnegative numbers p_1, $p_{\pm 2}$, p_3 and some nonnegative mass distribution $p_4(dl)$ subject to

(4a) $$p_1 + p_{-2} + p_{+2} + p_3 + \int \big(|l|\wedge 1\big)p_4(dl) = 1, \quad p_4(0) = 0,$$

(4b) $$p_4\big(R^1\big) = +\infty \quad \text{in case } p_{\pm 2} = p_3 = 0,$$

[30] $C^{\bullet 2}(R^1) = C^2(-\infty,0]\cap C^2[0,+\infty)\cap(u : u''(0-) = u''(0+))$.

and the Green operators are

$$(G^{\bullet}_{\alpha}f)(l) = (G^{-}_{\alpha}f)(l) + e^{-(2\alpha)^{1/2}|l|}(G^{\bullet}_{\alpha}f)(0), \tag{5}$$

where

$$(G^{-}_{\alpha}f)(a) = \int_{ab>0} \frac{e^{-(2\alpha)^{1/2}|b-a|} - e^{-(2\alpha)^{1/2}|b+a|}}{(2\alpha)^{1/2}} f(b)\,db \tag{6}$$

is the Green operator for the Brownian motion with instant killing at $l=0$ and

$$(G^{\bullet}_{\alpha}f)(0) = \frac{-p_{-2}(G^{-}_{\alpha}f)^{-}(0)+p_{+2}(G^{-}_{\alpha}f)^{+}(0)+p_3f(0)+\int_{|l|>0}(G^{-}_{\alpha}f)(l)p_4(dl)}{p_1+(2\alpha)^{1/2}(p_{-2}+p_{+2})+\alpha p_3+\int_{|l|>0}\cdot(1-e^{-(2\alpha)^{1/2}|l|})p_4(dl)}, \tag{7a}$$

$$\pm(G^{-}_{\alpha}f)^{\pm}(0) = 2\int_{\pm l>0} e^{-(2\alpha)^{1/2}|l|}f(l)\,dl. \tag{7b}$$

Coming to the sample paths, P. Lévy [13] proved that if $t \to \mathfrak{x}(t)$ is a standard Brownian path starting at 0 and if $\mathfrak{Z}_1$, $\mathfrak{Z}_2$, etc. are the (open) intervals of the complement of $\mathfrak{Z} = (t : \mathfrak{x} = 0)$, then the *signs* e_1, e_2, etc. of the *excursions* $\mathfrak{x}(t) : t \in \mathfrak{Z}_1$, etc., are independent Bernouilli trials with common law $P_0(e_1 = \pm 1) = \frac{1}{2}$ (standard coin-tossing game), independent of $\mathfrak{Z}$ and of the (unsigned) *scaled excursions*

$$\mathfrak{x}_1(t) = |\mathfrak{Z}_1|^{-1/2}|\mathfrak{x}(t|\mathfrak{Z}_1| + \inf \mathfrak{Z}_1)|, \quad 0 \leqq t \leqq 1, \tag{8}$$
etc.

which are independent, identical in law, and likewise independent of $\mathfrak{Z}$ (see Fig. 15.15).

Given $p_{-2}+p_{+2} > 0$, it is not difficult to see that if e_1, e_2, etc. is now a *skew coin-tossing game* independent of the scaled excursions and of $\mathfrak{Z}$ (i.e., independent of $|\mathfrak{x}|$) with law

$$P_0(e_1 = -1) : P_0(e_1 = +1) = p_{-2} : p_{+2}, \tag{9}$$

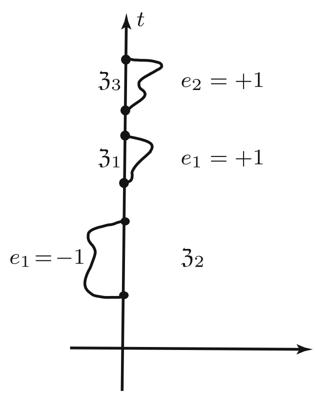

Figure 15.15

then the *skew Brownian motion*

$$\begin{aligned} \mathfrak{x}^\bullet(t) &= e_n\left|\mathfrak{x}(t)\right| && \text{if } t\in\mathfrak{Z}_n,\ n\geqq 1,\\ &= 0 && \text{if } t\in\mathfrak{Z}, \end{aligned} \tag{10}$$

starts afresh at each *constant* time $t\geqq 0$; in addition, its Green operators decompose as in (5), and evaluating $(G^\bullet_\alpha f)(0)$ as[31]

$$\begin{aligned} \left(G^\bullet_\alpha f\right)(0) &= \sum_{n\geqq 1} E_0\left(\int_{\mathfrak{Z}_n} e^{-\alpha t} f\left(e_n|\mathfrak{x}|\right)dt\right)\\ &= \sum_{n\geqq 1}\left(\frac{p_{-2}}{p_{-2}+p_{+2}}E_0\left[\int_{\mathfrak{Z}_n} e^{-\alpha t} f\left(-|\mathfrak{x}|\right)dt\right]\right.\\ &\qquad\left.+\frac{p_{+2}}{p_{-2}+p_{+2}}E_0\left[\int_{\mathfrak{Z}_n} e^{-\alpha t} f\left(+|\mathfrak{x}|\right)dt\right]\right)\\ &= \frac{p_{-2}}{p_{-2}+p_{+2}}E_0\left[\int_0^{+\infty} e^{-\alpha t} f\left(-|\mathfrak{x}|\right)dt\right]\\ &\quad+\frac{p_{+2}}{p_{-2}+p_{+2}}E_0\left[\int_0^{+\infty} e^{-\alpha t} f\left(+|\mathfrak{x}|\right)dt\right]\\ &= \frac{2p_{-2}\int^{0-} e^{-(2\alpha)^{1/2}l} f(l)\,dl + 2p_{+2}\int_{0+} e^{-(2\alpha)^{1/2}l} f(l)\,dl}{(2\alpha)^{1/2}\left(p_{-2}+p_{+2}\right)}\\ &= \frac{-p_{-2}\left(G^-_\alpha f\right)^-(0)+p_{+2}\left(G^-_\alpha f\right)^+(0)}{(2\alpha)^{1/2}\left(p_{-2}+p_{+2}\right)}, \end{aligned} \tag{11}$$

one identifies (10) as the Brownian motion associated with (3) in the special case $p_1=p_3=p_4=0$ $\left(p_{-2}u^-(0)=p_{+2}u^+(0)\right)$.

Coming to the case $p_1=p_{\pm 2}=p_3=0$ $(p_4(R^1-0)=+\infty)$, if $\mathfrak{p}(dt\times dl)$ is a Poisson measure with mean $dt\,p_4(dl)$ independent of the standard Brownian motion $\mathfrak{x}$, if $[l_1^-,l_1^+)$, $[l_2^-,l_2^+)$, etc. are the flat stretches of the inverse function $\mathfrak{p}^{-1}$ of $\mathfrak{p}(t)=\int |l|\mathfrak{p}([0,t]\times dl)$, and if $\mathfrak{t}^+$ is the local time at 0 of the (independent) reflecting Brownian motion $\mathfrak{x}^+=|\mathfrak{x}|$, then the desired motion is

$$\begin{aligned} \mathfrak{x}^\bullet(t) &= \mathfrak{x}(t) && \text{if } t<\mathfrak{m}_0=\min(t:\mathfrak{x}=0),\\ &= \pm\left[\mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+-\mathfrak{t}^++\mathfrak{x}^+\right] && \text{if } t\in\mathfrak{Q}^+,\\ &= 0 && \text{if } \mathfrak{m}_0\leqq t\in\mathfrak{Q}^+, \end{aligned} \tag{12}$$

where $\mathfrak{Q}^+=\bigcup_{n\geqq 1}[\mathfrak{t}^{-1}(l_n^-),\mathfrak{t}^{-1}(l_n^-))$, and the ambiguous sign in the second line is *positive* during the interval $[\mathfrak{t}^{-1}(l_n^-),\mathfrak{t}^{-1}(l_n^+))$ if $l_n=l_n^+-l_n^-$ is a jump of $\mathfrak{p}(dt\times dl\cap(0,+\infty])$ and *negative* otherwise (see Fig. 15.16).

[31] $|\mathfrak{Z}|=0$.

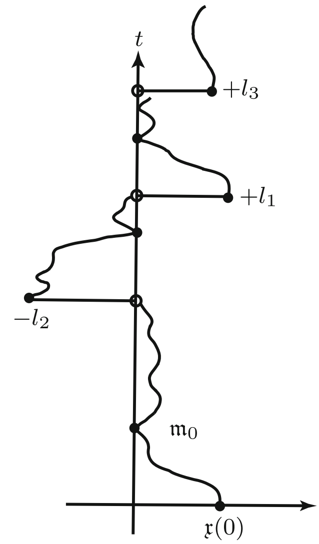

Figure 15.16

Granting that (12) is simple Markov (the proof is left to the reader), it is enough for its identification to evaluate[32]

$$
\begin{aligned}
\big(G^\bullet_\alpha f\big)(0) &= \sum_{n\geqq 1} E_0\bigg(\int_{\mathfrak{t}^{-1}(l_n^-)}^{\mathfrak{t}^{-1}(l_n^+)} e^{-\alpha t} f\big[\pm\big(l_n^+ - \mathfrak{t}^+ = \mathfrak{x}^+\big)\big]\, dt\bigg) \\
&= \sum_{n\geqq 1} E_0\big[e^{-(2\alpha)^{1/2} l_n^-}\big(G_\alpha^- f\big)\big(\pm l_n\big)\big] \\
&= E_0\bigg[\int_0^{+\infty}\int_{R^1-0} \mathfrak{p}(dt\times dl) e^{-(2\alpha)^{1/2}\mathfrak{p}(t-)}\big(G_\alpha^- f\big)(l)\bigg] \\
&= \int_{|l|>0}\big(G_\alpha^- f\big)(l) p_4(dl) \times \bigg[\int_{|l|>0}\big(1-e^{-(2\alpha)^{1/2} l}\big) p_4(dl)\bigg]^{-1}
\end{aligned}
\tag{13}
$$

with the aid of the tricks developed in Sect. 15.15.

Coming to the case $p_1 = p_3 = 0$, it suffices to combine the special cases $p_1 = p_3 = p_4 = 0$ and $p_1 = p_{\pm 2} = p_3 = 0$ as follows.

Given $\mathfrak{p}(dt\times dl)$, $\mathfrak{x}$, and $\mathfrak{t}^+$ as above, if $\mathfrak{x}_2^\bullet$ is the skew Brownian motion based upon $p_{\pm 2}$ and $\mathfrak{x}$, if $\mathfrak{x}_4^\bullet$ is the motion of (12) based upon $\mathfrak{p}^\bullet(t) = \int |l|\mathfrak{p}([0,t]\times dl)$ and $\mathfrak{x}$, if $[l_1^-, l_1^+)$, $[l_2^-, l_2^+)$, etc. are the flat stretches of the inverse function of $\mathfrak{p} = p_2 t + \mathfrak{p}^\bullet$ ($p_2 = p_{-2} + p_{+2}$), and if $\mathfrak{Q}^+ = \bigcup_{n\geqq 1}[\mathfrak{t}^{-1}(l_n^-), t^{-1}(l_n^+))$, then the desired motion is

$$
\begin{aligned}
\mathfrak{x}^\bullet(t) &= \mathfrak{x}(t) && \text{if } t < \mathfrak{m}_0, \\
&= \mathfrak{x}_4^\bullet\big(t^\bullet\big) && \text{if } t \in \mathfrak{Q}^+,\ t^\bullet = \big|\mathfrak{Q}^+ \cap [0,t)\big|, \\
&= \mathfrak{x}_2^\bullet(t) && \text{if } t \in \big[\mathfrak{m}_0, +\infty\big) - \mathfrak{Q}^+.
\end{aligned}
\tag{14}
$$

[32] $|[0,+\infty) - \mathfrak{Q}^+| = 0$ because $\mathfrak{p}(t)$ has no linear part $p_2 t$.

The reader will check that this sample path starts afresh at each *constant* time $t \geqq 0$ and will complete its identification with the aid of

$$
\begin{aligned}
(G_\alpha^\bullet f)(0) &= E_0\left[\int_0^{+\infty} e^{-\alpha t} f(\mathfrak{x}^\bullet)\,dt\right] \\
&= \sum_{n\geqq 1} E_0\left[\int_{\mathfrak{t}^{-1}(l_n^-)}^{\mathfrak{t}^{-1}(l_n^+)} e^{-\alpha t} f[\mathfrak{x}_4^\bullet(t)]\,dt\right] \\
&\quad + E_0\left[\int_0^{+\infty} e^{-\alpha t} f(\mathfrak{x}_2^\bullet)\,dt\right] - \sum_{n\geqq 1} E_0\left[\int_{\mathfrak{t}^{-1}(l_n^-)}^{\mathfrak{t}^{-1}(l_n^+)} e^{-\alpha t} f(\mathfrak{x}_2^\bullet)\,dt\right] \\
&= \sum_{n\geqq 1} E_0\left[e^{-\alpha \mathfrak{t}^{-1}(l_n^-)}(G_\alpha^- f)(\pm l_n)\right] \\
&\quad + E_0\left[\int_0^{+\infty} e^{-\alpha t} f(\mathfrak{x}_2^\bullet)\,dt\right]\left(1 - \sum_{n\geqq 1} E_0\left[e^{-\alpha \mathfrak{t}^{-1}(l_n^-)} - e^{-\alpha \mathfrak{t}^{-1}(l_n^+)}\right]\right) \qquad (15) \\
&= E_0\left[\int_0^{+\infty}\int_{R^1-0} \mathfrak{p}(dt\times dl) e^{-(2\alpha)^{1/2}\mathfrak{p}(t-)}(G_\alpha^- f)(l)\right] \\
&\quad + E_0\left[\int_0^{+\infty} e^{-\alpha t} f(\mathfrak{x}_2^\bullet)\,dt\right] \\
&\qquad \times \left(1 - E_0\left[\int_0^{+\infty}\int_{R^1-0} \mathfrak{p}(dt\times dl) e^{-(2\alpha)^{1/2}\mathfrak{p}(t-)} e^{-(2\alpha)^{1/2}|l|}\right]\right) \\
&= \frac{-p_{-2}(G_\alpha^- f)^-(0) + p_{+2}(G_\alpha^- f)^+(0) + \int (G_\alpha^- f)(l) p_4(dl)}{(2\alpha)^{1/2} p_2 + \int \left(1 - e^{-(2\alpha)^{1/2}|l|}\right) p_4(dl)}.
\end{aligned}
$$

If $p_3 > 0 = p_1$, it is clear that the desired motion is the sample path $\mathfrak{x}^\bullet$ of (14) run with the stochastic clock $\mathfrak{f}^{-1}$ inverse to $\mathfrak{f} = t + p_3 \mathfrak{p}^{-1}\mathfrak{t}^+$ (see Sect. 15.14 for the interpretation of $\mathfrak{p}^{-1}\mathfrak{t}^+$ as a local time), while, if $p_1 > 0$ also, the motion $\mathfrak{x}^\bullet(\mathfrak{f}^{-1})$ has to be annihilated at time $\mathfrak{m}_\infty^\bullet$ with conditional law

$$
P_\cdot\left(\mathfrak{m}_\infty^\bullet > t \mid \mathfrak{x}^\bullet(\mathfrak{f}^-)\right) = e^{-p_1 \mathfrak{p}^{-1}\mathfrak{t}^+ \mathfrak{f}^{-1}(t)}. \qquad (16)
$$

The reader is invited to furnish the proofs.

Brownian motions with the same kind of two-sided barrier can be defined on the unit circle $S^1 = [0, 1)$ as W. Feller [3, 5] pointed out.

Given a standard Brownian motion on R^1, its projection onto[33] $S^1 = R^1/Z^1$ is the so-called *standard circular Brownian motion*; its generator is the contraction of $\mathbf{G} = D^2/2$ to $C^2(S^1)$.

Consider now the general circular Brownian motion with a two-sided barrier at $l = 0$ (i.e., the obvious circular analogue of a Brownian motion with two-sided barrier on R^1), and, as before, single out the case

$$
P_\cdot^\bullet\left[\mathfrak{x}^\bullet(\epsilon^\bullet) = 0\right] = 1, \quad \epsilon^\bullet = \lim_{\varepsilon\downarrow 0} \inf\left(t : |\mathfrak{x}^\bullet| < \varepsilon\right). \qquad (17)
$$

[33] Z^1 is the integers.

$\mathbf{G}^\bullet$ is the contraction of $\mathbf{G} = D_2/2$ to[34]

$$D\big(\mathbf{G}^\bullet\big) = C^{\bullet 2}\big(S^1\big) \cap \Big(u : p_1 u(0) + p_{-2}u^-(0) - p_{+2}u^+(0) + p_3(\mathbf{G}u)(0\pm) = \int \big[u(l) - u(0)\big]p_4(dl)\Big) \tag{18}$$

for some nonnegative numbers p_1, $p_{\pm 2}$, p_3 and some nonnegative mass distribution $p_4(dl)$ subject to

$$p_1 + p_{-2} + p_{+2} + p_3 + \int_0^1 l(1-l)p_4(dl) = 1, \quad p_4(0) = p_4(1) = 0, \tag{19a}$$

$$p_4\big(S^1\big) = +\infty \quad \text{in case } p_{\pm 2} = p_3 = 0, \tag{19b}$$

and an application of (18) to

$$\big(G^\bullet_\alpha f\big)(l) = \big(G^-_\alpha f\big)(l) + \frac{\sinh(2\alpha)^{1/2} l + \sinh(2\alpha)^{1/2}(1-l)}{\sinh(2\alpha)^{1/2}}\big(G^\bullet_\alpha f\big)(0), \quad 0 \leqq l < 1, \tag{20a}$$

$$\big(G^-_\alpha f\big)(a) = 2\int_0^1 G(a,b)f(b)\,db, \quad 0 \leqq a < 1, \tag{20b}$$

$$G(a,b) = G(b,a) = \frac{\sinh(2\alpha)^{1/2} a \sinh(2\alpha)^{1/2}(1-b)}{(2\alpha)^{1/2}\sinh(2\alpha)^{1/2}}, \quad 0 \leqq a \leqq b < 1, \tag{20c}$$

establishes the formula

$$\begin{aligned}\big(G^\bullet_\alpha f\big)(0) = \Big[&2p_{-2}\int_0^1 \frac{\sinh(2\alpha)^{1/2}(1-l)}{\sinh(2\alpha)^{1/2}}f(l)dl \\ &+ 2p_{+2}\int_0^1 \frac{\sinh(2\alpha)^{1/2} l}{\sinh(2\alpha)^{1/2}}f(l)dl + p_3 f(0) + \int_0^1 \big(G^-_\alpha f\big)(l)p_4(dl)\Big]\Big/ \\ &\Big[p_1 + (2\alpha)^{1/2}\frac{\cosh(2\alpha)^{1/2} - 1}{\sinh(2\alpha)^{1/2}}\big(p_{-2} + p_{+2}\big) \\ &+ \alpha p_3 + \int_0^1 \Big(1 - \frac{\sinh(2\alpha)^{1/2} l + \sinh(2\alpha)^{1/2}(1-l)}{\sinh(2\alpha)^{1/2}}\Big)p_4(dl)\Big].\end{aligned} \tag{21}$$

Given a standard circular Brownian motion $\mathfrak{x}$ with local time

$$\mathfrak{t}(t) = \lim_{\varepsilon \downarrow 0}(2\varepsilon)^{-1}\,\text{measure}\,\big(s : \big|\mathfrak{x}(s)\big| < \varepsilon,\ s \leqq t\big) \tag{22}$$

and a (circular) differential process $\mathfrak{p}$ based on $p_{\pm 2}$ and p_4, it is possible to build up the circular Brownian sample paths as in the linear case, but a second method suggests itself: *the method of images.*

Consider for this purpose a Brownian motion on R^1 with two-sided barriers at the integers having as its generator the contraction of $\mathbf{G} = D^2/2$ to the class of functions $u \in C(R^1) \cap C^2(R^1 - Z^1)$ such that

$$(\mathbf{G}u)(n-) = (\mathbf{G}u)(n+), \tag{23a}$$

$$p_1 u(n) + p_{-2}u^-(n) - p_{+2}u^+(n) + p_3(\mathbf{G}u)(n\pm) = \int_0^1 \big[u(l+n) - u(n)\big]p_4(dl) \tag{23b}$$

[34] $C^{\bullet 2}(S^1) = C(S^1) \cap C^2(S^1 - 0) \cap (u : u''(0-) = u''(0+))$.

at each integer $n = 0, \pm 1, \pm 2$, etc. (the reader is invited to build up the sample paths for himself). Because the barriers are periodic, the projection of this motion onto $S^1 = R^1/Z^1$ is (simple) Markov, and its identification as the desired circular Brownian motion is immediate.

15.18. Simple Brownian Motions

Given a *simple Brownian motion* on $[0, +\infty)$, described as in Sect. 15.5 except that *it need not start afresh at nonconstant stopping times*,

$$(1) \qquad \big(G^\bullet_\alpha f\big)(l) = \big(G^-_\alpha f\big)(l) + e^{-(2\alpha)^{1/2} l}\big(G^\bullet_\alpha f\big)(0+), \quad l > 0,$$

as will now be proved with a view to the classification of all such Brownian motions.

Given $\alpha > 0$, a nonnegative Borel function f, and $t_2 \geqq t_1 \geqq 0$,

$$(2) \qquad \begin{aligned} &E^\bullet_{\cdot}\big[e^{-\alpha t_2}\big(G^\bullet_\alpha f\big)\big(\mathfrak{x}^\bullet(t_2)\big) \mid B^\bullet_{t_1}\big] \\ &\quad = e^{-\alpha t_2} E^\bullet_l\big[\big(G^\bullet_\alpha f\big)\big(\mathfrak{x}^\bullet(t)\big)\big], \quad l = \mathfrak{x}^\bullet(t_1),\ t = t_2 - t_1, \\ &\quad = e^{-\alpha t_2}\int_0^{+\infty} e^{-\alpha s} ds\, E^\bullet_l\big(E^\bullet_{\mathfrak{x}^\bullet(t)}\big[f\big(\mathfrak{x}^\bullet(s)\big)\big]\big) \\ &\quad = e^{-\alpha t_2}\int_0^{+\infty} e^{-\alpha s} ds\, E^\bullet_l\big[f\big(\mathfrak{x}^\bullet(t+s)\big)\big] \\ &\quad = e^{-\alpha t_1}\int_t^{+\infty} e^{-\alpha s} ds\, E^\bullet_l\big[f\big(\mathfrak{x}^\bullet(s)\big)\big] \\ &\quad \leqq e^{-\alpha t_1}\big(G^\bullet_\alpha f\big)\big(\mathfrak{x}^\bullet(t_1)\big), \end{aligned}$$

i.e., $e^{-\alpha t}(G^\bullet_\alpha f)(\mathfrak{x}^\bullet)$ is a (nonnegative) *supermartingale*; as such, it possesses one-sided limits as[35] $t = k2^{-n} \downarrow s$ $(s \geqq 0)$, and it follows that if $l > \varepsilon > 0$ and if $\mathfrak{m}^\bullet$ is the crossing time $\inf(t : \mathfrak{x}^\bullet < \varepsilon)$, then

$$(3) \qquad \begin{aligned} \big(G^\bullet_\alpha f\big)(l) &= E^\bullet_l\left[\int_0^{m^\bullet} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big) dt\right] \\ &\quad + \lim_{n\uparrow+\infty}\sum_{k\geqq 0} E^\bullet_l\Big[(k-1)2^{-n} \leqq \mathfrak{m}^\bullet < k2^{-n}, \\ &\qquad\qquad e^{-\alpha k 2^{-n}}\int_0^{m^\bullet_\infty(w^+_{k2^{-n}})} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big(t + k2^{-n}\big)\big)\, dt\Big] \\ &= E^\bullet_l\left[\int_0^{\mathfrak{m}^\bullet} e^{-\alpha t} f\big(\mathfrak{x}^\bullet\big) dt\right] \\ &\quad + \lim_{n\uparrow+\infty}\sum_{k\geqq 0} E^\bullet_l\big[(k-1)2^{-n} \leqq \mathfrak{m}^\bullet < k2^{-n}, \\ &\qquad\qquad e^{-\alpha k2^{-n}}\big(G^\bullet_\alpha f\big)\big(\mathfrak{x}^\bullet\big(k2^{-n}\big)\big)\big] \\ &= E_l\left[\int_0^{\mathfrak{m}} e^{-\alpha t} f(\mathfrak{x})\, dt\right] + E_l\Big[e^{-\alpha\mathfrak{m}} \lim_{k2^{-n}\downarrow\mathfrak{m}} \big(G^\bullet_\alpha f\big)\big(\mathfrak{x}\big(k2^{-n}\big)\big)\Big], \end{aligned}$$

where $\mathfrak{x}$ is a standard Brownian motion, $E_\cdot$ its expectation, and $\mathfrak{m}$ its passage time $\min(t : \mathfrak{x} = \varepsilon)$.

But, in the standard Brownian case, $\lim_{k2^{-n}\downarrow\mathfrak{m}}(G^\bullet_\alpha f)(\mathfrak{x}(k2^{-n}))$ is measurable over $\mathbf{B}_{\mathfrak{m}+}$ and also independent of $\mathbf{B}_{\mathfrak{m}+}$ (i.e, it is measurable over $\mathbf{B}[\mathfrak{x}(t+\mathfrak{m}) : t \geqq 0]$ which is independent

[35] See J. L. Doob [1].

of $\mathbf{B}_{\mathfrak{m}+}$ conditional on the *constant* $\mathfrak{x}(\mathfrak{m}) = \varepsilon$); as such, it is constant, and inserting this information back into (3) and letting $\varepsilon \downarrow 0$ establishes

$$(4) \qquad \big(G^\bullet_\alpha f\big)(l) = \big(G^-_\alpha f\big)(l) + e^{-(2\alpha)^{1/2} l} \times \text{constant},$$

which implies the existence of $(G^\bullet_\alpha f)(0+)$ and leads at once to (1).

Given a bounded function f on $[0, +\infty)$, continuous apart from a possible jump at $l = 0$, define a new function $\widehat{f}$ on $(-1) \cup [0, +\infty)$ as

$$(5) \qquad \begin{aligned} \widehat{f}(l) &= f(0) && \text{if } l = -1, \\ &= f(0+) && \text{if } l = 0, \\ &= f(l) && \text{if } l > 1, \end{aligned}$$

and introduce the new Green operators

$$(6) \qquad \widehat{G}_\alpha \widehat{f} = \big(G^\bullet_\alpha f\big)\widehat{}$$

mapping $C((-1) \cup [0, +\infty))$ into itself.

$\widehat{G}_\alpha$ is the Green operator of a *strict* Markov motion on $(-1) \cup [0, +\infty)$ with sample paths $t \to \widehat{\mathfrak{x}}(t) = \widehat{\mathfrak{x}}(t+) \in (-1) \cup [0, +\infty) \cup \infty$, and $\mathfrak{x}^\bullet$ is identical in law to the projection of $\widehat{\mathfrak{x}}$ under the identification $-1 \to 0$, as the reader can check for himself or deduce from the general embedding of D. Ray [14].

One now computes the domain $D(\widehat{\mathbf{G}})$ of the generator $\widehat{\mathbf{G}}$ of this *covering motion* and finds that it is the class of functions

$$u \in C\big((-1) \cap [0, +\infty)\big) \cup C^2[0, +\infty)$$

subject to

$$(7a) \qquad \begin{gathered} -p_{+2} u^+(0) + p_{+3}(\mathbf{G}u)(0) = \int_{(-1)\cup(0,+\infty)\cup\infty} \big[u(l) - u(0)\big] p_{+4}(dl), \\ p_{+4}(0) = 0 \leqq p_{+2}, p_{+3}, p_{+4}(dl), \\ p_{+2} + p_{+3} + p_{+4}(-1) + \int_{0+} (l \wedge 1) p_{+4}(dl) + p_{+4}(\infty) = 1, \end{gathered}$$

$$(7b) \qquad \begin{gathered} p_{-3}(\widehat{\mathbf{G}}u)(-1) = \int_{[0,+\infty)\cup\infty} \big[u(l) - u(-1)\big] p_{-4}(dl), \\ p_{-4}(-1) = 0 \leqq p_{-3}, p_{-4}(dl), \\ p_{-3} + p_{-4}[0, +\infty) + p_{-4}(\infty) = 1, \end{gathered}$$

where $u(\infty) \equiv 0$.

If $p_{-3} = 0$, the motion starting at -1 begins with a jump $l \in [0, +\infty) \cup \infty$ with law $p_{-4}(dl)$ as in Fig. 15.17, $u(-1) = \int_{[0,+\infty)} u(l) p_{-4}(dl)$, and (7a) goes over into

$$(8a) \qquad p^\bullet_1 u(0) - p^\bullet_2 u^+(0) + p^\bullet_3 (\mathbf{G}u)(0) = \int_{0+} \big[u(l) - u(0)\big] p^\bullet_4(dl),$$

$$(8b) \qquad \begin{gathered} p^\bullet_1 = p_{+4}(\infty) + p_{+4}(-1) p_{-4}(\infty), \\ p^\bullet_2 = p_{+2}, \qquad p^\bullet_3 = p_{+3}, \\ p^\bullet_4(dl) = p_{+4}(dl) + p_{+4}(-1) p_{-4}(dl), \quad l > 0, \end{gathered}$$

i.e., the covering motion *does not land at* -1 which is a superfluous state, and $\widehat{\mathfrak{x}} = \mathfrak{x}^\bullet$ is a *strict Brownian motion on* $[0, +\infty)$ as in Sects. 15.5 to 15.16.

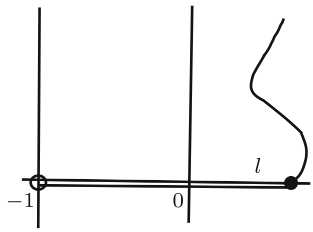

Figure 15.17

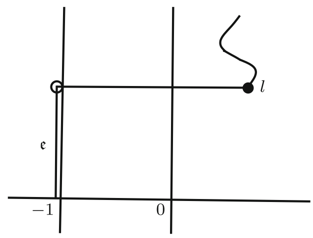

Figure 15.18

If $p_{-3} > 0$, then $\widehat{\mathbf{G}}$ is the contraction of $\mathbf{G} = D^2/2$ to $D(\widehat{\mathbf{G}})$ with the added specification

(9) $$(\widehat{\mathbf{G}}u)(-1) = \int_{[0,+\infty)} \big[u(l) - u(-1)\big] \frac{p_{-4}(dl)}{p_{-3}}, \quad u(\infty) \equiv 0,$$

at -1, and the particle starting at -1 waits there for an exponential holding time ϵ with law $e^{-p_{-4}t/p_{-3}}$ ($p_{-4} = p_{-4}([0,+\infty) \cup \infty)$), and then jumps to $l \in [0,+\infty) \cup \infty$ with law $p_{-4}(dl)/p_{-4}$ as in Fig. 15.18.

If, in addition to $p_{-3} > 0$, one has $p_2 = 0$ and $p_4(0,+\infty) < +\infty$, then the motion starting at 0 is of the same kind, and it is clear that the projection of this motion down to $[0,+\infty)$ $(-1 \to 0)$ cannot even be *simple* Markov unless $p_{-3} = p_{+3}$ and $p_{-4}(dl) = p_{+4}(dl)$ $(l \neq 0)$ up to a common multiplicative constant, in which case the projection is the Brownian motion associated with

(9a) $$p_1 u(0) + p_{+3}(\mathbf{G}u)(0) = \int_{0+} \big[u(l) - u(0)\big] p_{+4}(dl),$$

(9b) $$p_1 = p_{+4}(\infty)$$

studied in Sect. 15.9.

If $p_{-3} > 0$ and either $p_{+2} > 0$ or $p_{+4}(0,+\infty) = +\infty$, the particle starting at -1 waits for an exponential holding time ϵ_1 and then jumps as in Fig. 15.19 to $l_1 \in [0,+\infty) \cup \infty$ and starts afresh; if $0 \leqq l_1 < +\infty$, the particle performs the Brownian motion on $[0,+\infty)$ associated with

(10a) $$p_1 u(0) - p_{+2} u^+(0) + p_{+3}(\mathbf{G}u)(0) = \int_{0+} \big[u(l) - u(0)\big] p_{+4}(dl),$$

(10b) $$p_1 = p_{+4}(-1 \cup \infty)$$

up to the killing time of that motion, at which instant it jumps to $l_2 = \infty$ or -1 with probabilities $p_{+4}(\infty) : p_{+4}(-1)$, and, if $l_2 = -1$, it starts afresh as in Fig. 15.19, while if $l_2 = \infty$, then the motion rests at that place at all later times.

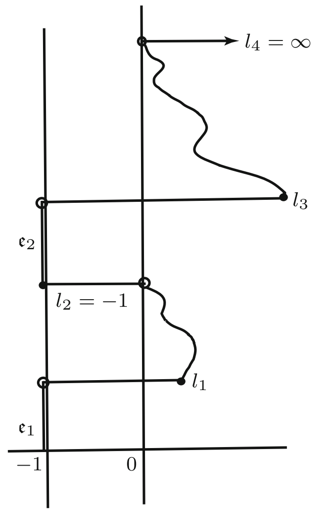

Figure 15.19

Now the projection $\mathfrak{x}^\bullet$ of this motion onto $[0, +\infty)$ $(-1 \to 0)$ is simple Markov if the Brownian motion attached to (10) does not spend positive (Lebesgue) time at $l = 0$; otherwise the knowledge that $\mathfrak{x}^\bullet(s) = 0$ is not sufficient to discriminate between the two possible coverings, and the law of $\mathfrak{x}^\bullet(t) : t \geqq s$ is moot. But if e is the indicator of $l = 0$, and if

(11) $$\mathfrak{x}^\bullet\big(\mathfrak{f}^{-1}\big) \quad \big(t < \mathfrak{m}_\infty^\bullet\big), \qquad \infty \quad \big(t \geqq \mathfrak{m}_\infty^\bullet\big)$$

(12a) $$\mathfrak{f} = t + p_3 \mathfrak{p}^{-1} \mathfrak{t}^+,$$

(12b) $$\mathfrak{x}^\bullet = \mathfrak{p}\mathfrak{p}^{-1}\mathfrak{t}^+ - \mathfrak{t}^+ + \mathfrak{x}^+$$

is the motion attached to (10), then, in the notation of Sect. 15.14,

(13) $$\begin{aligned} &\text{measure}\,\big(s : \mathfrak{x}^\bullet\big(\mathfrak{f}^{-1}\big) = 0,\ s \leqq t\big) \\ &= \int_0^t e\big[\mathfrak{x}^\bullet\big(f^{-1}\big)\big]\,ds = \int_0^{\mathfrak{f}^{-1}(t)} e\big(\mathfrak{x}^\bullet\big)\mathfrak{f}(ds) \\ &= p_{+3} \int_{\mathfrak{z}^+ \cap \mathfrak{Q}^+ \cap [0, \mathfrak{f}^{-1}(t))} \mathfrak{p}^{-1}\mathfrak{t}^+(dt) \\ &= p_{+3}\mathfrak{p}^{-1}\mathfrak{t}^+\big[\mathfrak{Q}^+ \cap \big[0, \mathfrak{f}^{-1}(t)\big)\big], \quad t \leqq \mathfrak{m}_\infty^\bullet, \end{aligned}$$

and this cannot be positive unless $p_{+3} > 0$ and $0 < \mathfrak{t}^+(\mathfrak{Q}^+) = |\mathfrak{Q}|$, i.e., unless $p_{+2} > 0$ also; in short, *the projection is simple Markov unless* $p_{+2}p_{+3} > 0$, and now the classification is complete.

N. Ikeda had conjectured part of our classification (private communication); the case of a two-sided barrier on R^1 is similar except that three covering points lie over 0.

15.19. Feller's Differential Operators

Given a nonnegative mass distribution e on the open half line $(0, +\infty)$ with $0 < e(a, b]$ $(a < b)$, let $D(\mathbf{G})$ be the class of functions $u \in C[0, +\infty)$ such that

(1) $$u^+(b) - u^+(a) = \int_{(a,b]} f\,de, \quad a < b,$$

for some $f \in C[0,+\infty)$, and introduce the differential operator

$$\mathbf{G} : h \longrightarrow (\mathbf{G}u)(a) = \lim_{b \downarrow a} \frac{u^+(b) - u^+(a)}{e(a,b]} = f. \tag{2}$$

W. Feller [5] proved that if $e(0,1] < +\infty$, and if p_1, p_2, p_3, $p_4(dl)$ are nonnegative with $p_4(0) = 0$ and $p_1 + p_2 + p_3 + \int_{0+} (l \wedge 1) p_4(dl) = 1$, then the contraction $\mathbf{G}^\bullet$ of $\mathbf{G}$ to

$$\begin{aligned} D\big(\mathbf{G}^\bullet\big) = D(\mathbf{G}) \cap \bigg(u : p_1 u(0) - p_2 u^+(0) \\ + p_3 (\mathbf{G}u)(0) = \int_{0+} \big[u(l) - u(0)\big] p_4(dl) \bigg) \end{aligned} \tag{3}$$

is the generator of a strict Markov motion (diffusion) on $[0, +\infty)$.

Given a reflecting Brownian motion $\mathfrak{x}^+$ on $[0,+\infty)$, the *local time*

$$\mathfrak{t}^+(t,l) = \big(\text{measure}\big(s : \mathfrak{x}^+(s) \in dl,\ s < t\big)\big)/2\,dl \tag{4}$$

is continuous in the pair $(t,l) \in [0,+\infty)^2$ (see H. Trotter [17]), and the motion associated with $\mathbf{G}^\bullet$ in the special case $p_1 = p_3 = p_4 = 0$ ($u^+(0) = 0$) is identical in law to $\mathfrak{x}^\bullet = \mathfrak{x}^+(\mathfrak{f}^{-1})$ where $\mathfrak{f} = \int_{0+} \mathfrak{t}^+(t,l) e(dl)$ (see V. A. Volkonskiĭ [19] and K. Itô and H. P. McKean, Jr. [9]).

Because $\mathfrak{t}^+(dt,l) = 0$ outside $\mathfrak{Z} = (t : \mathfrak{x}^+ = l)$,

$$\begin{aligned} \int_0^t f\big(\mathfrak{x}^\bullet\big) ds &= \int_0^{\mathfrak{f}^{-1}(t)} f\big(\mathfrak{x}^+\big) \int_{0+} \mathfrak{t}^+(ds,l) e(dl) \\ &= \int_{0+} \bigg(\int_0^{\mathfrak{f}^{-1}(t)} \mathfrak{t}^+(ds,l) \bigg) f(l) e(dl) \\ &= \int_{0+} \mathfrak{t}^+\big[\mathfrak{f}^{-1}(t), l\big] f(l) e(dl); \end{aligned} \tag{5}$$

hence the *local time*

$$\mathfrak{t}^\bullet(t) = \lim_{\varepsilon \downarrow 0} e(0,\varepsilon]^{-1} \text{measure}\big(s : \mathfrak{x}^\bullet(s) < \varepsilon,\ s \leqq t\big) = \mathfrak{t}^+\big(\mathfrak{f}^{-1}, 0\big) \tag{6}$$

exists, and now it is clear that the discussion of the Brownian case can be adapted with little change.

15.20. Birth and Death Processes

Quite a general birth and death process on the nonnegative integers can be changed via a scale substitution into a motion on a discrete series $Q : 0 = l_0 < l_1 < l_2 < \cdots < 1$ having as its generator

$$\mathbf{G}^\bullet u = \big(u^+ - u^-\big)/e, \tag{1}$$

$$u^+\big(l_n\big) = u^-\big(l_{n+1}\big) = \big(l_{n+1} - l_n\big)^{-1} \big[u\big(l_{n+1}\big) - u\big(l_n\big)\big], \tag{2a}$$

$$e = e\big(l_n\big) > 0, \tag{2b}$$

$$e\big(l_0\big) + e\big(l_1\big) + \cdots < +\infty, \tag{2c}$$

subject to

$$u^+(0) = 0, \tag{3a}$$

$$\begin{aligned} p_1 u(1) + p_3 \big(\mathbf{G}^\bullet u\big)(1) = -p_2 u^-(1) + \int_Q \big[u(l) - u(1)\big] p_4(dl), \\ p_1 + p_2 + p_3 + \int_Q (1 - l) p_4(dl) = 1 \end{aligned} \tag{3b}$$

(see W. Feller [6]). In the special case $p_1 = p_3 = p_4 = 0$ the corresponding motion is just the reflecting Brownian motion on $[0, 1]$ run with the inverse function of $\mathfrak{f} = \int_Q \mathfrak{t}^+(t, l)e(dl)$, $\mathfrak{t}^+$ being the reflecting Brownian local time. Once this motion has been obtained, the general path can be built up using local times and differential processes as before.

Bibliography

[1] J. L. Doob. *Stochastic processes.* John Wiley & Sons Inc., New York, 1953.

[2] E. B. Dynkin. Infinitessimal operators of Markov processes (Russian). *Teor. Veroyatnost. i Primenen*, 1:38–60, 1956.

[3] W. Feller. The parabolic differential equations and the associated semi-groups of transformations. *Ann. of Math.*, 55:468–519, 1952.

[4] W. Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77:1–31, 1954.

[5] W. Feller. Generalized second order differential operators and their lateral conditions. *Illinois J. Math.*, 1:459–564, 1957.

[6] W. Feller. The birth and death processes as diffusion processes. *J. Math. Pures Appl.*, 38:301–345, 1959.

[7] G. Hunt. Some theorems concerning Brownian motion. *Trans. AMS*, 81:294–319, 1956.

[8] K. Itô. On stochastic processes. I. (Infinitely divisible laws of probability). *Jap. J. Math.*, 18:261–301, 1942.

[9] K. Itô and H. P. McKean. *Diffusion Processes and Their Sample Paths.* Academic Press, 1965.

[10] M. Kac. On some connections between probability theory and differential and integral equations. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 189–215. University of California Press, 1951.

[11] P. Lévy. Sur les intègrales dont les èlèments sont des variables alèatoires indèpendantes. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (2)*, 3(3–4):337–366, 1934.

[12] P. Lévy. Sur certains processus stochastiques homogènes. *Compositio Math.*, 7:283–339, 1940.

[13] P. Lévy. *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, Paris, 1948.

[14] D. B. Ray. Resolvents, transition functions, and strongly Markovian processes. *Ann. of Math.*, 70:43–72, 1959.

[15] A. V. Skorohod. Stochastic equations for diffusion processes with a boundary. *Teor. Verojatnost. i Primenen*, 6:287–298, 1961.

[16] A. V. Skorohod. Stochastic equations for diffusion processes with boundaries. II. *Teor. Verojatnost. i Primenen*, 7:5–25, 1962.

[17] H. F. Trotter. A property of Brownian motion paths. *Illinois J. Math.*, 2:425–433, 1958.

[18] A. D. Ventcel'. Semigroups of operators that correspond to a generalized differential operator of second order (Russian). *Dokl. Akad. Nauk SSSR*, 111:269–272, 1956.

[19] V. A. Volkonskii. Random time substitution in strong Markov processes. *Teor. Veroyatnost. i Prim.*, 3:332–350, 1958.

[20] N. Wiener. Differential space. *J. Math. Phys.*, 2:131–174, 1923.

[16] The Spectrum of Hill's Equation

[16] The Spectrum of Hill's Equation. *Inv. Math.* **30** (1975), 217–274.

16 The Spectrum of Hill's Equation

H. P. McKean[1] and P. van Moerbeke[2]

16.1. Hill's Equation: Periodic Spectrum

Let [3] $q \in C_1^\infty$ be fixed and let Q be Hill's operator $Q = -d^2/dx^2 + q(x)$. The following information may be found in Magnus and Winkler [26], see also Strutt [32]. Q, acting on the class of functions $f \in L^2(-\infty, \infty)$ with $\|f''\|_2 < \infty$, is self-adjoint with spectrum comprising an infinite number of intervals

$$\left[\lambda_0, \lambda_1\right] \cup \left[\lambda_2, \lambda_3\right] \cup \left[\lambda_4, \lambda_5\right] \cup \left[\lambda_6, \lambda_7\right] \cup \cdots,$$

each of positive width but possibly contiguous:

$$-\infty = \lambda_{-1} < \lambda_0 < \lambda_1 \leqq \lambda_2 < \lambda_3 \leqq \lambda_4 < \lambda_5 \leqq \lambda_6 < \lambda_7 \leqq \cdots \uparrow +\infty.$$

The nature of the solutions of $Qy = \lambda y$ is determined by the placement of λ relative to the spectrum: in the *intervals of instability* $(\lambda_{2i-1}, \lambda_{2i})$ $(i = 0, 1, 2, \ldots)$ no solution is bounded; in the complementary *intervals of stability* $(\lambda_{2i}, \lambda_{2i+1})$ $(i = 0, 1, 2, \ldots)$ every solution is bounded but none is of period 1 or 2; the remaining points $\lambda_0 < \lambda_1 \leqq \lambda_2 < \lambda_3 \leqq \lambda_4 < \cdots$ comprise the *periodic spectrum.* The *principal series* $\lambda_0 < \lambda_3 \leqq \lambda_4 < \lambda_7 \leqq \lambda_8 < \cdots$ is the spectrum of Q acting on the class of functions $f \in L_1^2$ with $\|f''\|_2 < \infty$, while the *complementary series* $\lambda_1 \leqq \lambda_2 < \lambda_5 \leqq \lambda_6 < \cdots$ fills out the spectrum of Q in its action on the class of functions $f \in L_2^2$ with $\|f''\|_2 < \infty$. Thus, λ_0 is always a simple eigenvalue with eigenfunction of period 1, while for $i = 1, 2, 3, \ldots, \lambda_{2i-1}$ or λ_{2i} is a simple or a double eigenvalue according as $\lambda_{2i-1} < \lambda_{2i}$ or $\lambda_{2i-1} = \lambda_{2i}$, with eigenfunction of period 1 or 2 according as $i = 1, 3, 5, \ldots$ or $i = 2, 4, 6, \ldots$. The eigenfunctions f_{2i-1} and f_{2i} for λ_{2i-1} and λ_{2i}, respectively, have precisely i roots in a period $0 \leqq x < 1$. The same classification of spectral points may be described by means of the *discriminant*[4]

$$\Delta(\lambda) = y_1(1, \lambda) + y_2'(1, \lambda),$$

in which $y_1(x, \lambda)[y_2(x, \lambda)]$ is the solution of $Qy = \lambda y$ with $y(0) = 1$, $y'(0) = 0[y(0) = 0, y'(0) = 1]$: in fact, $|\Delta(\lambda)| > 2$ in the intervals of instability, $|\Delta(\lambda)| < 2$ in the intervals of stability, $\Delta(\lambda) = +2$ on the principal series $\lambda_0 < \lambda_3 \leqq \lambda_4 < \lambda_7 \leqq \lambda_8 < \cdots$, and $\Delta(\lambda) = -2$ on the complementary series $\lambda_1 \leqq \lambda_2 < \lambda_5 \leqq \lambda_6 < \cdots$, the multiplicity of the root being the

[1]Courant Institute of Mathematical Sciences, New York University, Mercer Street 251, New York, NY 10012, USA.

[2]Mathematics Department, Université de Louvain, B-1348 Louvain-la-Neuve, Belgium.

[3]C_p^∞ is the class of real infinitely differentiable functions of period p. L_p^2, is defined similarly.

[4]$' = \partial/\partial x$.

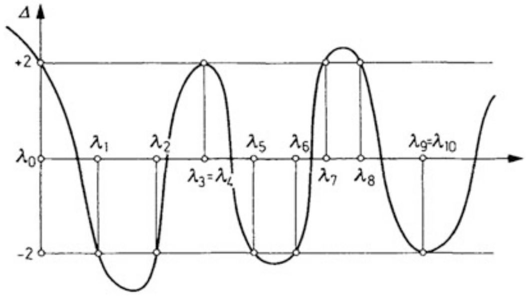

FIGURE 16.1.

same as the spectral multiplicity (= 1 or 2). A typical discriminant is seen in Fig. 16.1. $\Delta(\lambda)$ is an integral function of order 1/2; as such, it is determined by the principal or by the complementary series and by the known estimates

$$y_1(x,\lambda) = \big[1 + o(1)\big] \times \cosh\sqrt{-\lambda}x,$$
$$\sqrt{-\lambda}y_2(x,\lambda) = \big[1 + o(1)\big] \times \sinh\sqrt{-\lambda}x$$

for $\lambda \downarrow -\infty$. The classical estimate $\lambda_{2i-1} \sim \lambda_{2i} \sim i^2\pi^2$ $(i \uparrow \infty)$ is immediate. $\Delta'(\lambda)$ is also an integral function of order 1/2 with simple roots only, comprising *trivial roots* at the double eigenvalues $\lambda_{2i-1} = \lambda_{2i}$ and one *non-trivial root* properly inside each interval of instability $(\lambda_{2i-1}, \lambda_{2i})$ of positive width.

Simple Spectrum. The simple periodic eigenvalues are now required to be finite in number:

$$\lambda_0^o < \lambda_1^o < \lambda_2^o < \lambda_3^o < \lambda_4^o < \cdots < \lambda_{2n-1}^o < \lambda_{2n}^o;$$

apart from $\lambda_0^o = \lambda_0$, they come in pairs $\lambda_{2m-1} < \lambda_{2m}$ $(m = m_1 < \cdots < m_n)$, m_i being the number of roots per period of the eigenfunctions f_{2i-1}^o and f_{2i}^o attached to λ_{2i-1}^o and λ_{2i}^o. The non-trivial roots of $\Delta'(\lambda) = 0$ are now denoted by λ_i' $(i = 1, \ldots, n)$; they interlace the simple spectrum, so:

$$\lambda_1^o < \lambda_1' < \lambda_2^o < \lambda_3^o < \lambda_2' < \lambda_4^o < \cdots < \lambda_{2n-1}^o < \lambda_n' < \lambda_{2n}^o.$$

Hochstadt [17] discussed the remarkable fact that the *simple periodic spectrum* $\lambda_0^o < \cdots < \lambda_{2n}^o$ *determines both the double periodic spectrum and the nontrivial roots* $\lambda_1' < \cdots < \lambda_n'$ of $\Delta'(\lambda) = 0$.

PROOF. The integral function $4 - \Delta^2(\lambda)$ is expressed as a product and its roots $\lambda_0 < \lambda_1 \leqq \lambda_2 < \lambda_3 \leqq \lambda_4 < \cdots$, separated into simple and double spectra[5]:

$$4 - \Delta^2(\lambda) = c_1 \prod_{i=0}^{2n} \left(1 - \frac{\lambda}{\lambda_i^o}\right) \prod_{\substack{\text{double}\\ \text{spectrum}}} \left(1 - \frac{\lambda}{\lambda_{2i}}\right)^2.$$

A similar product is available for $\Delta'(\lambda)$:

$$\Delta'(\lambda) = c_2 \prod_{i=1}^{n} \left(1 - \frac{\lambda}{\lambda_i'}\right) \prod_{\substack{\text{double}\\ \text{spectrum}}} \left(1 - \frac{\lambda}{\lambda_{2i}}\right),$$

[5]To simplify life, it is supposed that $\lambda = 0$ is not a root.

whence

$$\frac{\Delta'(\lambda)}{\sqrt{4-\Delta^2(\lambda)}} = c_3 \frac{\prod_{i=1}^{n}\left(\lambda-\lambda_i'\right)}{\sqrt{\prod_{i=0}^{2n}\left(\lambda-\lambda_i^o\right)}}.$$

Now let $\Delta = 2\cos\psi$. Then the periodic spectrum is determined by $\psi =$ *an integral multiple of* π and from

$$\Delta' = -2\sin\psi \times \psi' = \pm\sqrt{4-\Delta^2}\,\psi',$$

it appears that, with $\psi(\lambda_0) = 0$ and a suitable determination of the signature of the radical,

$$\ell(\lambda) = \sqrt{-\left(\lambda-\lambda_0^o\right)\left(\lambda-\lambda_1^o\right)\cdots\left(\lambda-\lambda_{2n}^o\right)},$$
$$\psi(\lambda) = c_4 \int_{\lambda_0}^{\lambda} \prod_{i=1}^{n}\left(\mu-\lambda_i'\right)\frac{d\mu}{\ell(\mu)}.$$

The constant c_4 is easily evaluated: $\psi(\lambda) \sim 2c_4\sqrt{-\lambda}$ far out, so that $2c_4\sqrt{-\lambda_{2i}} \sim i\pi$ and $\lambda_{2i} \sim i^2\pi^2$ yields $2c_4 = \sqrt{-1}$. The proof is finished by observing that the roots λ_i' $(i = 1, \ldots, n)$ are already determined by the simple spectrum. The point is that $d\psi(\lambda)$ is purely imaginary for $\lambda_{2j-1}^o < \lambda < \lambda_{2j}^o$ so that the increment $\sqrt{-1}\varepsilon = \psi(\lambda_{2j}^o) - \psi(\lambda_{2j-1}^o)$ must vanish in view of

$$\pm 1 = \cos\psi\left(\lambda_{2j}^o\right) = \cos\left(\pm m_j\pi + \sqrt{-1}\varepsilon\right) = \pm\cosh\varepsilon.$$

Thus,

$$\int_{\lambda_{2j-1}^o}^{\lambda_{2j}^o} \prod_{i=1}^{n}\left(\mu-\lambda_i'\right)\frac{d\mu}{\ell(\mu)} = 0 \quad (j = 1, \ldots, n),$$

or what is the same,

$$\sum_{i=1}^{n}\int_{\lambda_{2j-1}^o}^{\lambda_{2j}^o} \mu^{n-i}\frac{d\mu}{\ell(\mu)}(-1)^i\theta_i' = -\int_{\lambda_{2j-1}^o}^{\lambda_{2j}^o} \mu^{n}\frac{d\mu}{\ell(\mu)},$$

in which $\theta_1', \ldots, \theta_n'$ are the elementary symmetric polynomials in $\lambda_1', \ldots, \lambda_n'$. The latter may be determined from the former, and the former can be evaluated since the determinant

$$\det\int_{\lambda_{2j-1}^o}^{\lambda_{2j}^o} \frac{\mu^{n-1}d\mu}{\ell(\mu)}$$
$$= \int_{\lambda_1^o}^{\lambda_2^o}\cdots\int_{\lambda_{2n-1}^o}^{\lambda_{2n}^o} \prod_{i>j}\left(\mu_i-\mu_j\right)\frac{d\mu_1\cdots d\mu_n}{\ell\left(\mu_1\right)\cdots\ell\left(\mu_n\right)}$$

cannot vanish. The result is restated as

Hochstadt's Formula. $\Delta(\lambda) = 2\cos\psi$ *with*

$$\psi(\lambda) = \frac{\sqrt{-1}}{2}\int_{\lambda_0}^{\lambda} \prod_{i=1}^{n}\left(\mu-\lambda_i'\right)\frac{d\mu}{\ell(\mu)};$$

especially, $\psi(\lambda) = \pm m\pi$ *if and only if* $\lambda = \lambda_{2m-1}$ *or* $\lambda = \lambda_{2m}$. □

16.2. Hill's Equation: Auxiliary Spectrum

Besides the periodic spectrum, it is helpful to consider also the spectrum of Q acting on the class of functions $f \in L^2[0,1]$ with $\|f''\|_2 < \infty$ and $f(0) = f(1) = 0$. This spectrum comprises the roots $\mu_1 < \mu_2 < \mu_3 < \cdots$ of $y_2(1,\mu) = 0$, separated into *trivial roots* at the double eigenvalues $\lambda_{2i-1} = \lambda_{2i}$ and *non-trivial roots* $\mu_i^o \in [\lambda_{2i-1}^o, \lambda_{2i}^o]$ $(i = 1, \ldots, n)$; the latter comprise the *auxiliary spectrum:*

$$\lambda_1^o \leqq \mu_1^o \leqq \lambda_2^o < \lambda_3^o \leqq \mu_2^o \leqq \lambda_4^o < \cdots < \lambda_{2n-1}^o \leqq \mu_n^o \leqq \lambda_{2n}^o.$$

The map $q(x) \to q(-x)$ does not change the periodic or the auxiliary spectrum, so q cannot be completely determined by them, but it is *nearly* so, as will be seen in a moment. $\Delta(\lambda) = y_1(1,\lambda) + y_2'(1,\lambda)$ is determined by the periodic spectrum, and if the auxiliary spectrum is also specified, then $y_2(1,\lambda)$ is determined, too. Now let $\lambda = \mu$ be a root of $y_2(1,\lambda) = 0$. Then the Wronskian of $y_1(x,\mu)$ and $y_2(x,\mu)$ at $x = 1$ reduces to $y_1(1,\mu)y_2'(1,\mu) = 1$, so

$$\Delta(\mu) = \left[y_2'(1,\mu)\right]^{-1} + y_2'(1,\mu)$$

may be solved for

$$y_2'(1,\mu) = \frac{\Delta(\mu)}{2} \pm \sqrt{\frac{\Delta^2(\mu)}{4} - 1},$$

up to the ambiguous signature $\pm$, and the norming constant[6]

$$\int_0^1 y_2^2(x,\mu)\,dx = y_2'(1,\mu)y_2^\bullet(1,\mu)$$

is determined thereby, up to the same ambiguity. Borg [1] proved[7] that the roots of $y_2(1,\mu) = 0$ together with the norming constants determine the potential q. Now a *bona fide* ambiguity in the norming constant presents itself only if $|\Delta(\mu)| > 2$, i.e., only for those points of the auxiliary spectrum with $\lambda_{2i-1}^o < \mu_i^o < \lambda_{2i}^o$. Let these be $m \leqq n$ in number. Then there are *at most* 2^m potentials q with simple spectrum $\lambda_0^o < \cdots < \lambda_{2n}^o$ and auxiliary spectrum $\mu_1^o < \cdots < \mu_n^o$, in fact, *there are precisely* 2^m *such*, as will now be confirmed by the method of Gelfand and Levitan [12].

PROOF. It is required to verify that if $\lambda_{2i-1}^o < \mu_i^o < \lambda_{2i}^o$, then there is a bona fide potential q of period 1 having the same periodic and auxiliary spectra as q and the same auxiliary norming constants, except for the i-th in which the signature is reversed. The computation is standard.[8] Let $\mu = \mu_i^o$. The new potential is

$$q(x) - 2\left(\lg\left[1 \pm \frac{\sqrt{\Delta^2(\mu) - 4}}{y_2^\bullet(1,\mu)} \int_0^x y_2^2(x,\mu)\,dx\right]\right)'',$$

the factor $\pm\sqrt{\Delta^2 - 4} \times (y_2^\bullet)^{-1}$ being the change in the reciprocal of the norming constant. The verification that this potential has the same periodic and auxiliary spectra and the proper norming constants is omitted as being more lengthy than instructive: the fact is that $\Delta(\lambda)$ and $y_2(1,\lambda)$ are unchanged while $y_2'(1,\lambda)$ is augmented by

$$\mp \frac{\sqrt{\Delta^2(\mu) - 4} \times y_2(1,\mu)}{y_2^\bullet(1,\mu) \times (\lambda - \mu)},$$

[6] $' = \partial/\partial x$; $^\bullet = \partial/\partial\mu$.

[7] See Levinson [25] for a simpler proof.

[8] Faddeev [7, 100].

which produces the same norming constants for μ_j $(j \neq i)$ and flips the ambiguous signature for μ_i. Observe that the new potential is *really* different if $\Delta^2(\mu) \neq 4$ because the norming constant changed.

The fact that *the auxiliary spectrum can occupy any position* $\mu_1^o < \cdots < \mu_n^o$ *in the cell*

$$C : \left[\lambda_1^o, \lambda_2^o\right] \times \cdots \times \left[\lambda_{2n-1}^o, \lambda_{2n}^o\right]$$

may be proved by a similar computation in which μ_j^o $(j \neq i)$ remain in place and $\mu = \mu_i^o$ is moved to a new location in the interval $\lambda_{2i-1}^o \leqq \mu \leqq \lambda_{2i}^o$. The infinitesimal motion $\mu \to \mu + d\mu$ is regulated by

$$\frac{\partial q(x)}{\partial \mu} = 2\left[\frac{y_2^2(x,\mu)y_1'(1,\mu)}{y_2'(1,\mu) - \left[y_2'(1,\mu)\right]^{-1}} - y_1(x,\mu)y_2(x,\mu)\right]''.$$

The final potential may be computed. The verification that the periodic and auxiliary spectra are unchanged, except for μ_i^o which moves at speed 1, is much the same, only messier.

To sum up from a more geometrical point of view: *the class M of potentials q with fixed simple spectrum $\lambda_0^o < \cdots < \lambda_{2n}^o$ is a 2^n-sheeted covering of the cell C in which the auxiliary spectrum lies. The covering group G is isomorphic to $Z_2 \times \cdots \times Z_2$ (n-fold), the i-th generator being the involution*

$$\sigma_i : q(x) \to q(x) - 2\lg\left[1 \pm \sqrt{\Delta^2(\mu_i) - 4}\ \frac{\int_0^x y_2^2(x,\mu_i^o)\,dx}{y_2^\bullet(1,\mu_i^o)}\right]'',$$

where the ambiguous signature is in accord with the signature in the expression for $y_2'(1,\mu_i^o)$ in terms of $\Delta(\mu_i^o)$. □

Amplification 16.2.1. M can be equipped with either the lifted topology from C or with the inherited topology from C_1^∞; happily, these are the same. The chief point is that $\|q^{(n)}\|_\infty$ is bounded on M for $n = 0, 1, 2, \ldots$, as will appear in Lemma 16.3.8, so that practically any topology you care to place on M is equivalent to the topology inherited from C_1^∞.

Amplification 16.2.2. It is observed that the covering map σ: $q(x) \to q(-x)$ reverses all the signatures in the norming constants $\int_0^1 y_2^2(x,\mu_i^o)\,dx$ $(i = 1,\ldots,n)$ since the eigenfunction $y_2(x,\mu_i^o)$ is changed into $-y_2(1-x,\mu_i^o)/y_2'(1,\mu_i^o)$ and

$$\frac{\int_0^1 y_2^2(1-x,\mu_i^o)\,dx}{\left[y_2'(1,\mu_i^o)\right]^2} = \frac{y_2^\bullet(1,\mu_i^o)}{y_2'(1,\mu_i^o)} = \frac{y_2^\bullet}{2}\left(\Delta \mp \sqrt{\Delta^2 - 4}\right),$$

i.e., $\sigma = \sigma_1, \ldots, \sigma_n$; see §16.5 for more information about σ.

16.3. Hamiltonians

A new theme now enters.

Theorem 16.3.1. *As $t \downarrow 0$, the trace*

$$\vartheta(t) = \sum_{i=0}^\infty e^{-\lambda_i t}$$

of e^{-tQ} acting on functions of period 2 can be expanded into half-integral powers of t, so[9]*:*

$$\vartheta(t) \sim \frac{1}{\sqrt{\pi t}} \sum_{m=0}^\infty \frac{(-t)^m}{(2m-3)\cdots 3\cdot 1} H_{m-1}.$$

[9] $(2m-3)\cdots 3\cdot 1$ is interpreted as unity if $m = 0$ or 1. The rationale for these constants will appear in Theorem 16.3.4.

$H_{-1} \equiv 1$, while for $m \geqq 0$, the so-called Hamiltonian H_m is a nonlinear functional of q of the form

$$H_n = \int_0^1 I_n\big[q(x), q'(x), q''(x), \dots\big]\, dx.$$

I_n is a universal isobaric polynomial of precise degree $n+1$, without constant term, the adjective "isobaric" signifying that q is considered to be of degree 1 and that $'$ =differentiation is considered to raise the degree by $1/2$.

AMPLIFICATION 16.3.2. The result stems from Dikii [4, 5], Gelfand [11], and, more directly, from Buslaev and Faddeev [2]; see also Gardner et al. [10]. The derivation stems from Kac [19]; see also Kac and van Moerbeke [20]. It employs a more detailed expansion embodied in

THEOREM 16.3.3. *The pole of the elementary solution $e(t,x,y)$ of $\partial e/\partial t = -Qe$ on the circle of perimeter 2 can be expanded for $t \downarrow 0$, so:*

$$e(t,x,x) = \frac{1}{\sqrt{4\pi t}} \sum_{m=0}^{\infty} \frac{(-t)^m}{(2m-3)\cdots 3\cdot 1} I_{m-1}$$

with I_m as before, uniformly for $0 \leqq x < 2$.

PROOF OF THEOREM 16.3.3. The key to the proof is the formula of Kac [19]:

$$e(t,x,y) = E_{xy}\left[\exp\left(-\int_0^t \mathfrak{q}\, dt'\right)\right] e^o(t,x,y)$$

with $\mathfrak{q} = q[\mathfrak{x}(t')]$, expressing e in terms of the elementary solution e^o for $q = 0$ and the associated Brownian motion $\mathfrak{x}$, considered modulo 2, "tied", i.e. conditioned so as to begin at $\mathfrak{x}(0) = x$ and end at $\mathfrak{x}(t) = y$. E_{xy} denotes the tied (conditional) mean value. By the Jacobi transformation of the theta function, $e^o(t,x,x) = (4\pi t)^{-1/2}+$ an exponentially small error, so

$$e(t,x,x) \sim E_{xx}\left[\exp\left(-\int_0^t \mathfrak{q}\, dt'\right)\right] \times \frac{1}{\sqrt{4\pi t}},$$

and you may replace $\mathfrak{x}(t')$: $0 \leqq t' \leqq t$ by $x + \sqrt{t}\mathfrak{x}(t'/t)$: $0 \leqq t' \leqq t$ with a new (mod 2) Brownian motion tied at $\mathfrak{x}(0) = 0$ and $\mathfrak{x}(1) = 0$, the two motions being identical in law. Thus,

$$e(t,x,x) \sim \frac{1}{\sqrt{4\pi t}} E_{00}\left[\exp\left(-t\int_0^1 q\big(x + \sqrt{t}\,\mathfrak{x}(t')\big)\, dt'\right)\right].$$

Now replace the exponential by its power series and expand $q(x + \sqrt{t}\,\mathfrak{x})$ into a MacLaurin series about x. This produces a formal sum

$$\begin{aligned}\sqrt{4\pi t}e(t,x,x) \sim \sum_{m=0}^{\infty} (-t)^m \int_0^1 dt'_m \dots \int_0^{t'_3} dt'_2 \int_0^{t'_2} dt'_1 \\ \times \sum_{\ell_1=0}^{\infty} \cdots \sum_{\ell_m=0}^{\infty} \frac{t^{(\ell_1+\cdots+\ell_m)/2}}{\ell_1! \dots \ell_m!} q^{(\ell_1)}(x) \dots q^{(\ell_m)}(x) \\ \times E_{00}\big[\big(\mathfrak{x}(t'_1)\big)^{\ell_1} \dots \big(\mathfrak{x}(t'_m)\big)^{\ell_m}\big],\end{aligned}$$

in which only terms with even $\ell_1 + \cdots + \ell_m$ survive the mean value, $\mathfrak{x} \to -\mathfrak{x}$ being an automorphism of the tied Brownian motion. The expansion is of the proposed form. ☐

PROOF OF THEOREM 16.3.1. Integrate the expansion of the pole over $0 \leqq x < 2$, using the eigenfunction expansion

$$e(t,x,x) = \sum_{i=0}^{\infty} e^{-\lambda_i t} f_i^2(x)$$

and employing the periodicity to replace $\int_0^2 dx$ by $2\int_0^1 dx$. □

Now that the *existence* of the Hamiltonians is established, their evaluation is achieved in a different way.

THEOREM 16.3.4 ([10]). *Define*[11] $X_m q = (\partial H_m/\partial q)'$. *Then*

$$X_m q = \left(qD + Dq - \frac{1}{2}D^3\right)\frac{\partial H_{m-1}}{\partial q} \quad (m = 1, 2, 3, \ldots).$$

LEMMA 16.3.5 ([12]). *The function* $e = e(t,x,x)$ *satisfies* $\left(^\bullet = \frac{\partial}{\partial t}\right)$

$$-2e^{\bullet\prime} = \left(qD + Dq - \frac{1}{2}D^3\right)e.$$

PROOF. From $e^\bullet = e_{11} - qe = e_{22} - qe$, it follows that

$$e^{\bullet\prime} = e_1^\bullet + e_2^\bullet = e_{111} - q'e - qe_1 + e_{222} - q'e - qe_2,$$

so

$$\left(qD + Dq - \frac{1}{2}D^3\right)e = 2qe' + q'e - \frac{1}{2}(e_{111} + 3e_{112} + 3e_{122} + e_{222}).$$

Besides,

$$e_{111} - q'e - qe_1 = e_{221} - qe_1, \qquad e_{222} - q'e - qe_2 = e_{112} - qe_2,$$

and upon substitution the previous expression is reduced to

$$\begin{aligned} 2qe' + q'e - \frac{1}{2}(e_{111} + 3e_{111} - 3q'e + 3e_{222} - 3q'e + e_{222}) \\ = 2qe' + q'e - 2e_{111} + 3q'e - 2e_{222} = -2e^{\bullet\prime}. \end{aligned}$$

□

LEMMA 16.3.6. $\partial\vartheta/\partial q = -te(t,x,x)$.

PROOF. Pick $q^\bullet \in C_1^\infty$. Then from

$$\vartheta = \int_0^2 e(t,x,x)\,dx = \int_0^2 E_{xx}\left[\exp\left(-\int_0^t \mathfrak{q}\,dt'\right)\right]e^o(t,x,x)dx,$$

it is clear that

$$\begin{aligned} \vartheta^\bullet \equiv \int_0^2 \frac{\partial\vartheta}{\partial q}q^\bullet\,dx &= -\int_0^2 E_{xx}\left[\int_0^t \mathfrak{q}^\bullet \exp\left(-\int_0^t \mathfrak{q}\right)\right]e^o(t,x,x)\,dx \\ &= -\int_0^1 dt' \int_0^2 dx\left[\int_0^2 e(t',x,y)q^\bullet(y)e(t-t',y,x)\,dy\right] \\ &= -\int_0^t dt' \int_0^2 \left[\int_0^2 e(t-t',y,x)e(t',x,y)\,dx\right]q^\bullet(y)dy \end{aligned}$$

[10]The fact is due to A. Lenard; see Gardner et al. [10].
[11]$\partial/\partial q$ is the functional gradient; $' = D = d/dx$.
[12]The lemma is due to Menikoff [27].

$$= -\int_0^t dt' \int_0^2 e(t,y,y)q^\bullet(y)dy$$
$$= -t\int_0^2 e(t,y,y)q^\bullet(y)\,dy.$$

The proof is finished. □

PROOF OF THEOREM 16.3.4. The result follows easily from Theorem 16.3.3 and Lemmas 16.3.5 and 16.3.6. A moment's reflection upon the proof of Theorem 16.3.3 will confirm that the expansion of ϑ may be differentiated with regard to q to obtain

$$\frac{\partial\vartheta}{\partial q} = -te(t,x,x) \sim \frac{1}{\sqrt{\pi t}}\sum_{m=0}^{\infty}\frac{(-t)^m}{(2m-3)\cdots 3\cdot 1}\frac{\partial H_{m-1}}{\partial q}.$$

Now differentiate with regard to x and t. A little manipulation produces

$$\frac{1}{\sqrt{\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m t^{m-5/2}(2m-3)}{(2m-3)\cdots 3\cdot 1}\mathsf{X}_{m-1}q = -2e^{\bullet\prime}$$
$$= \left(qD + Dq - \frac{1}{2}D^3\right)e$$
$$= \frac{1}{\sqrt{\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m t^{m-3/2}}{(2m-3)\cdots 3\cdot 1}\left(qD + Dq - \frac{1}{2}D^3\right)\frac{\partial H_{m-1}}{\partial q}.$$

The stated formula is immediate from that. □

The quantities H, $\partial H/\partial q$, and $\mathsf{X}q = (\partial H/\partial q)'$ can now be computed. The results for $m \leqq 5$ are tabulated in table 16.1 below; plainly, the formulas get rapidly out of hand.

COROLLARY 16.3.7. $\int_0^1 \partial H_m/\partial q\,dx = (2m-1)H_{m-1}$ $(m = 0,1,2,\ldots)$.

PROOF. $-t\vartheta = -t\int_0^2 e(t,x,x)\,dx = \int_0^2 \partial\vartheta/\partial q\,dx$; now replace the trace by the development of Theorem 16.3.1. □

A couple of technical lemmas are now prepared for use in §12. Let F_n be the Sobolev space of formal sums

$$f(x) = \sum_{k=-\infty}^{\infty}\hat{f}(k)e^{2\pi i k x}$$

with

$$|f|_n^2 = \sum_{k=-\infty}^{\infty}\left(1+k^2\right)^n\left|\hat{f}(k)\right|^2 < \infty$$

for $n = \ldots,-1,0,1,2,\ldots$.

LEMMA 16.3.8 ([13]). *$|q|_n^2 \leqq 2^{n+1}H_{n+1} + c_n(H_0,\ldots,H_n)$ with a universal function c_n for any $n \neq 1$, while for $n = 1$, the factor $2^{n+1} = 4$ should be replaced by 8. Especially, H_{n+1} is bounded below for fixed $H_0 = h_0,\ldots, H_n = h_n$.*

[13] See Lax [24].

TABLE 16.1.

m	I_m	$\partial H_m/\partial q$	$X_m q$
-1	1	0	0
0	q	1	0
1	$\frac{1}{2}q^2$	q	q'
2	$\frac{1}{2}q^3 + \frac{1}{4}(q')^2$	$\frac{3}{2}q^2 - \frac{1}{2}q''$	$3qq' - \frac{1}{2}q'''$
3	$\frac{5}{8}q^4 + \frac{5}{4}q(q')^2$ $+\frac{1}{8}(q'')^2$	$\frac{5}{2}q^3 - \frac{5}{4}(q')^2 - \frac{5}{2}qq''$ $+\frac{1}{4}q''''$	$\frac{15}{2}q^2q' - 5q'q''$ $-\frac{5}{2}qq''' + \frac{1}{4}q'''''$
4	$\frac{7}{8}q^5 + \frac{35}{8}q^2(q')^2$ $+\frac{7}{8}q(q'')^2$ $+\frac{1}{16}(q''')^2$	$\frac{35}{8}q^4 - \frac{35}{4}q(q')^2$ $-\frac{35}{4}q^2q'' + \frac{21}{8}(q'')^2$ $+\frac{7}{2}q'q''' + \frac{7}{4}qq''''$ $-\frac{1}{8}q''''''$	$\frac{35}{2}q^3q' - \frac{35}{4}(q')^3$ $-35qq'q'' - \frac{35}{4}q^2q'''$ $+\frac{35}{4}q''q''' + \frac{21}{4}q'q''''$ $+\frac{7}{4}qq''''' - \frac{1}{8}q'''''''$
5	$\frac{21}{16}q^6 + \frac{105}{8}q^3(q')^2$ $\frac{63}{16}q^2(q'')^2$ $-\frac{35}{32}(q')^4$ $+\frac{9}{16}q(q''')^2$ $-\frac{5}{8}(q'')^3$ $+\frac{1}{32}(q'''')^2$	$\frac{63}{8}q^5 - \frac{315}{8}q^2(q')^2$ $-\frac{105}{4}q''q^3 + \frac{189}{8}q(q'')^2$ $+\frac{63}{4}(q')^2q''$ $+\frac{63}{2}qq'q''' + \frac{63}{8}q^2q''''$ $-\frac{105}{8}(q')^2q'' - \frac{69}{16}(q''')^2$ $-\frac{57}{8}q''q'''' - \frac{27}{8}q'q'''''$ $-\frac{9}{8}qq''''''$ $+\frac{1}{16}q''''''''$	$\frac{1}{16}q''''''''' - \frac{9}{8}qq'''''''$ $-\frac{9}{2}q'q'''''' + \frac{63}{8}q^2q'''''$ $-\frac{21}{2}q''q''''' + \frac{189}{4}qq'q''''$ $+\frac{315}{4}qq''q''' - \frac{63}{4}q'''q''''$ $-\frac{105}{4}q^3q'''' + \frac{483}{8}(q')^2q'''$ $+\frac{651}{8}q'(q'')^2 - \frac{315}{2}q^2q'q''$ $-\frac{315}{2}q(q')^3$ $+\frac{315}{8}q'q^4$

PROOF. The statement is obvious for $n = 0$, since $|q|_0^2 = 2H_1$, while for $n = 1$, $\|q\|_\infty^2 \leqq 2\|q\|_2^2 + 2\|q'\|_2^2$ implies

$$\begin{aligned} H_2 &= \frac{1}{4}\int_0^1 (q')^2\,dx + \frac{1}{2}\int_0^1 q^3\,dx \\ &\geqq \frac{1}{4}\|q'\|_2^2 - H_1\sqrt{4H_1 + 2\|q'\|_2^2} \\ &\geqq \frac{1}{4}\|q'\|_2^2 - 2H_1^{3/2} - \sqrt{2}H_1\|q'\|_2, \end{aligned}$$

so that

$$\|q'\|_2 \leqq 2\sqrt{2}H_1 + \sqrt{4H_2 + 8H_1^{3/2} + 8H_1^2}$$

and

$$|q|_1^2 \leqq 8H_2 + c_1\big(H_0, H_1\big).$$

The proof for $n \geqq 2$ proceeds by induction. Let $H_0, \dots, H_n$ be fixed so that $|q|_{n-1}$ is controlled, that is to say $\|q^{(i)}\|_\infty$ $(i \leqq n-2)$ and $\|q^{(n-1)}\|_2$ are controlled. Now H_{n+1} is a sum of integrals of monomials in $q, q', \dots, q^{(2n+1)}$ of isobaric degree

$$\begin{aligned} &(i = \text{the number of appearances of } q) \\ &+\frac{1}{2} \times (j = \text{the number of differentiations}) \\ &= n+2, \end{aligned}$$

in which you may take $i \geqq 2$ as such integrals vanish if $i = 1$, and by partial integration, you may arrange to have the largest and next largest number of differentiations applied to a single appearance of q differ by 1 or less. Call them j_1 and $j_2[j_2 \leqq j_1 \leqq j_2 + 1]$. Then $j_1 \leqq n$ since $j_1 > n$ entails too high a degree:

$$i + \frac{1}{2}j \geqq 2 + \frac{1}{2}(n+1) + \frac{1}{2}n = n + \frac{5}{2} > n+2.$$

Now if $j_1 = n$, you must have $j_2 = n$ and $i = 2$ as $i \geqq 3$ leads to degree exceeding $3 + \frac{n}{2} + \frac{1}{2}(n-1) = n + \frac{5}{2}$. You could also have $j_1 = n-1$ or $j_1 \leqq n-2$. Contributions of the second kind are controlled by $\|q^{(i)}\|_\infty$ $(i \leqq n-2)$, while a contribution of the first kind cannot involve $q^{(n-1)}$ to degree $\geqq 3[3 + \frac{3}{2}(n-1) = \frac{3}{2}n + \frac{3}{2} > n+2$, since $n \geqq 2]$ and so is controlled by $\|q^{(i)}\|_\infty$ $(i \leqq n-2)$ and $\|q^{(n-1)}\|_2$. Therefore, $H_{n+1} = \text{constant} \times \|q^{(n)}\|_2^2 +$ terms controlled by $H_0, \dots, H_n$, and the proof may be finished by verifying that the constant multiplier of $\|q^{(n)}\|_2^2$ is 2^{-n-1}. The proof is by induction: ignoring terms with less than a maximal number of differentiations, if $I_n = 2^{-n}[q^{(n-1)}]^2$, then $\partial H_n/\partial q = 2^{-n+1}(-1)^{n-1}q^{(2n-2)}$, and

$$\mathsf{X}_{n+1}q = \left(qD + Dq - \frac{1}{2}D^3\right)\frac{\partial H_n}{\partial q} = (-1)^n 2^{-n} q^{(2n+1)},$$

so that $\partial H_{n+1}/\partial q = (-1)^n 2^{-n} q^{(2n)}$ and $I_{n+1} = 2^{-n-1}[q^{(n)}]^2$. □

LEMMA 16.3.9 ([14]). *Fix $H_0, \dots, H_n$. Then H_{n+1} attains its minimum in the class $C^\infty_{1/m} \subset C^\infty_1$ for any $m = 1, 2, 3, \dots$, and at the minimum $c_1\mathsf{X}_1 + \dots + c_n\mathsf{X}_n + \mathsf{X}_{n+1} : q \to 0$ with suitable coefficients c_i $(i = 1, \dots, n)$.*

[14] See Lax [24].

PROOF FOR $m = 1$. $|q|_n$ stays bounded as H_{n+1} decreases to its infimum in the class C_1^∞ for fixed $H_0, \dots, H_n$, and it is plain from the proof of Lemma 16.3.5 that the infimum is attained by some potential $q \in F_n$. Now the condition of minimization is of the form[15]

$$\sum_{i=0}^{n}(-1)^i D^i \frac{\partial I_{n+1}}{\partial x_i}\left[q, \dots, q^{(n)}\right] + \sum_{j=0}^{n} c_j \sum_{i=0}^{j}(-1)^i D^i \frac{\partial I_j}{\partial x_i}\left[q, \dots, q^{(j-1)}\right] = 0,$$

and apart from the leading term $(-1)^n 2^{-n} q^{(2n)}$, the sum involves $i < n$ differentiations of functions $\partial I_{n+1}/\partial x_i$ of class F_1 or better.[16] Therefore, $q^{(2n)} \in F_{-n+2}$, i.e., $q \in F_{n+2}$. The procedure is now repeated with the result that $q \in F_{n+4}, F_{n+6}, \dots$, i.e., $q \in C_1^\infty$. The condition of minimization can now be written

$$\sum_{j=0}^{n} c_j \frac{\partial H_j}{\partial q} + \frac{\partial H_{n+1}}{\partial q} = 0.$$

The proof is finished by differentiation with regard to x. □

16.4. Vector Fields and Flows

The quantities $\mathsf{X}q = (\partial H/\partial q)'$ of §16.3 are now regarded as vector fields on the class C_1^∞ of infinitely differentiable functions of period 1: if F is any smooth functional on C_1^∞, then

$$\mathsf{X}F(q) = \int_0^1 \frac{\partial F}{\partial q} \mathsf{X}q \, dx.$$

Technical Condition. Let $\mathsf{X} = c_1\mathsf{X}_1 + \cdots + c_n\mathsf{X}_n$. Then $q^\bullet = \mathsf{X}q$ *is assumed to have a nice solution in* C_1^∞; later, this will appear as a simple corollary of classical function theory in the hyperelliptic case; see, especially, Amplification 16.10.11 and §14, below. Notice that for $H = c_1 H_1 + \cdots + c_n H_n$, $q^\bullet = (\partial H/\partial q)'$ has the form of a Hamiltonian flow, $'$ = differentiation being skew-symmetrical.[17]

THEOREM 16.4.1 ([18]). *The periodic spectrum, and so also the Hamiltonian series, is preserved by the Hamiltonian flows $q^\bullet = Xq$.*

PROOF. Under such a flow,

$$\vartheta^\bullet = \int_0^2 \frac{\partial \vartheta}{\partial q} \quad \mathsf{X}q \, dx = -t \int_0^2 e(t, x, x)\mathsf{X}q \, dx.$$

Now the periodic spectrum is surely preserved by the motion of translation $q^\bullet = \mathsf{X}_1 q = q'$, so

$$0 = \int_0^2 e(t, x, x)\mathsf{X}_1 q \, dx.$$

The rest of the proof consists in checking that

$$\frac{d}{dt} \int_0^2 e(t, x, x)\mathsf{X}_m q \, dx = -\frac{1}{2} \int_0^2 e(t, x, x)\mathsf{X}_{m+1} q \, dx$$

[15] x_i stands for $q^{(i)}$. D^i is construed as a map of F_j into F_{j-i}.

[16] I_{n+1} is arranged as in the proof of Lemma 16.3.8.

[17] The Hamiltonian aspect of such flows was first noticed by Gardner et al. [9]; see also Faddeev and Zaharov [33].

[18] The fact is well-known; see, for example, Gardner et al. [10].

for $m = 1, 2, 3, \ldots$. That is easy: by Lemma 16.3.5 and the skew-symmetry of the operator $qD + Dq - \frac{1}{2}D^3$,

$$\begin{aligned}\frac{d}{dt}\int_0^2 e\mathsf{X}_m q\,dx &= \int_0^2 e^\bullet \left(\frac{\partial H_m}{\partial q}\right)' dx \\ &= -\int_0^2 e^{\bullet\prime}\frac{\partial H_m}{\partial q}\,dx \\ &= \frac{1}{2}\int_0^2 \frac{\partial H_m}{\partial q}\left(qD + Dq - \frac{1}{2}D^3\right)e\,dx \\ &= -\frac{1}{2}\int_0^2 \left(qD + Dq - \frac{1}{2}D^3\right)\frac{\partial H_m}{\partial q}\,dx \\ &= -\frac{1}{2}\int_0^2 e\mathsf{X}_{m+1}q\,dx,\end{aligned}$$

as required. □

THEOREM 16.4.2. *The vector fields* X *commute with one another.*

PROOF. This is a standard fact of Hamiltonian mechanics: commuting flows preserve one another's Hamiltonians and vice versa. Fix a smooth functional F of q. Then

$$\mathsf{X}_i F = \int_0^1 \frac{\partial F}{\partial q}\mathsf{X}_i q\,dx = \int_0^1 \frac{\partial F}{\partial q}\left(\frac{\partial H_i}{\partial q}\right)' dx$$

and[19]

$$\begin{aligned}\mathsf{X}_j\mathsf{X}_i F = &\int_0^1\int_0^1 \frac{\partial H_i}{\partial q(x)}\frac{\partial^2 F}{\partial q(x)\partial q(y)}\frac{\partial H_j}{\partial q(y)}\,dx\,dy \\ &+ \int_0^1 \frac{\partial F}{\partial q(x)}\left[\frac{d}{dx}\int_0^1 \frac{\partial^2 H_i}{\partial q(x)\partial q(y)}\left(\frac{\partial H_j}{\partial q(y)}\right)' dy\right] dx.\end{aligned}$$

The first integral is symmetrical in i and j, so what you need is such symmetry in the second integral, also. But the invariance of H_i under $q^\bullet = \mathsf{X}_j q$ entails

$$0 = H_i^\bullet = \int_0^1 \frac{\partial H_i}{\partial q}\mathsf{X}_j q = \int_0^1 \frac{\partial H_i}{\partial q}\left(\frac{\partial H_j}{\partial q}\right)',$$

and you have only to differentiate once more by q and integrate by parts to obtain the necessary identity:

$$0 = \int_0^1 \frac{\partial^2 H_i}{\partial q(x)\partial q(y)}\left(\frac{\partial H_j}{\partial q(y)}\right)' dy - \int_0^1 \left(\frac{\partial H_i}{\partial q(y)}\right)' \frac{\partial^2 H_j}{\partial q(x)\partial q(y)}\,dy.$$

□

The fact that the periodic spectrum is preserved by the flow $q^\bullet = \mathsf{X}q$ implies the existence of a family of unitary operators U in L^2_2 such that $UQU^\dagger = Q^o$, Q^o being Hill's operator for the initial potential. Lax [22] introduced into the subject the formal infinitesimal operator $K = U^\bullet U^\dagger$. The cautionary adjective "formal" refers to the fact that the differentiability of U is unclear. The next theorem presents a formula for K; it will be part of the proof that

[19] $\partial^2 F/\partial q^2$ is the symmetrical kernel defined by

$$F(q + \varepsilon p) = F(q) + \varepsilon\frac{\partial F}{\partial q}\cdot p + \frac{\varepsilon^2}{2}p\cdot\frac{\partial^2 F}{\partial q^2}p + O(\varepsilon^3),$$

in which $\cdot$ stands for the inner product of L^2_1.

a differentiable family U really does exist. The computation is based upon the self-evident identity

$$KQ - QK = (-1) \times \text{multiplication by } \mathsf{X}q,$$

which any such infinitesimal unitary must satisfy.

THEOREM 16.4.3 ([20]). *The infinitesimal unitary operator for* $\mathsf{X} = \mathsf{X}_n$ *is*

$$K_n = \sum_{k=1}^{n} \left(\frac{1}{2} \mathsf{X}_{k-1} q - \frac{\partial H_{k-1}}{\partial q} D \right) (2Q)^{n-k};$$

especially, K_n *is a differential operator of degree* $2n-1$. A short table of these operators will be found in Table 16.1 below.

PROOF. The commutator of $K = K_n$ and Q is as it should be: by Theorem 16.3.4,

$$\begin{aligned} KQ - QK &= \sum_{k=1}^{n} \left[\left(\frac{1}{2} \left(\frac{\partial H_{k-1}}{\partial q} \right)' - \frac{\partial H_{k-1}}{\partial q} D \right) Q \right. \\ &\qquad \left. - Q \left(\frac{1}{2} \left(\frac{\partial H_{k-1}}{\partial q} \right)' - \frac{\partial H_{k-1}}{\partial q} D \right) \right] (2Q)^{n-k} \\ &= \sum_{k=1}^{n} \left[\frac{1}{2} \left(\frac{\partial H_{k-1}}{\partial q} \right)''' - 2 \left(\frac{\partial H_{k-1}}{\partial q} \right)' D^2 - \frac{\partial H_{k-1}}{\partial q} q' \right] (2Q)^{n-k} \\ &= \sum_{k=1}^{n} \left[(\mathsf{X}_{k-1} q) 2Q - \mathsf{X}_k q \right] (2Q)^{n-k} \\ &= (\mathsf{X}_0 q)(2Q)^n - (\mathsf{X}_n q)(2Q)^o \\ &= -\mathsf{X}_n q. \end{aligned}$$

Now solve $f_i^\bullet = K f_i$ for $i = 0, 1, 2, \ldots$, starting from the unit eigenfunctions f_i^o of Q^o. Then

$$\frac{d}{dt} \int f_i f_j = \int f_i K f_j + \int f_j K f_i = 0,$$

so that the f_i form a unit perpendicular family, and

$$\frac{d}{dt} \int f_i Q f_j = \int f_i (-KQ + QK + \mathsf{X}q) f_j = 0,$$

from which it appears, first, that Qf_j belongs to the span of $f_i [\sum_{i=0}^{\infty} |(f_i, Qf_j)|^2 = \|Qf_j\|^2 = \lambda_j^2]$ and, second, that $Qf_j = \lambda_j f_j$ $(j = 0, 1, 2, \ldots)$ with the *same* eigenvalues as for Q^o, $Qf_j - \lambda_j f_j$ being perpendicular to every f_i. A moment's reflection confirms that the spectrum of Q does not change at all, that the f_i span L_2^2, and that the desired unitary family is determined by $U : f_i^o \to f_i$ $(i = 0, 1, 2, \ldots)$. □

AMPLIFICATION 16.4.4. The operators K_n commute in view of Theorem 16.4.2; especially $K = c_1 K_1 + \cdots + c_n K_n$ is the infinitesimal unitary operator for $\mathsf{X} = c_1 \mathsf{X}_1 + \cdots + c_n \mathsf{X}_n$.

AMPLIFICATION 16.4.5. The action of K on the eigenspaces of Q is of special importance. For a simple eigenvalue λ with eigenfunction $f, q^\bullet = \mathsf{X}q$ entails $Qf^\bullet + q^\bullet f = \lambda f^\bullet$ with $f^\bullet = Kf$, i.e., $(Q - \lambda)Kf = \mathsf{X}q \times f$. The special case $\mathsf{X}q = 0$ leads to the simplest result, as in Table 16.2.

[20] Lax [24] computed K_n in another way.

TABLE 16.2.

m	K_m
1	$-D$
2	$2D^3 - 3qD - \frac{3}{2}q'$
3	$-4D^5 + 10qD^3 + 15q'D^2$ $+\frac{25}{2}q''D - \frac{15}{2}q^2D$ $-\frac{15}{2}qq' + \frac{15}{4}q'''$

THEOREM 16.4.6. *If $Xq = 0$, then K annihilates the simple eigenfunctions of Q, while in a double eigenspace with unit perpendicular basis $f_\pm$, it acts as*

$$Kf_- = kf_+$$
$$Kf_+ = -kf_-$$

with a real number k depending upon the eigenvalue.

PROOF. If λ is a simple eigenvalue, then $(Q - \lambda)Kf = 0$ implies that Kf is proportional to f, and the constant of proportionality must vanish, K being skew, i.e., $Kf = 0$. Now let $f_\pm$ be a unit perpendicular basis of a double eigenspace. Then $(Q - \lambda)Kf_- = 0$, as before, so that $Kf_- = c_1 f_- + c_2 f_+$, and from the skewness of K,

$$c_1 = c_1 \int f_-^2 = \int f_- K f_- = 0.$$

Thus, $Kf_- = c_2 f_+$; similarly, $Kf_+ = c_3 f_-$, and

$$c_2 = c_2 \int f_+^2 = \int f_+ K f_- = -\int f_- K f_+ = -c_2 \int f_-^2 = -c_3.$$

The proof is finished. □

The infinitesimal operators K play a central role below; see, especially, §16.6.

16.5. The Manifold of Potentials

Let M be the manifold of potentials $q \in C_1^\infty$ with fixed simple spectrum $\lambda_0^o < \cdots < \lambda_{2n}^o$.

THEOREM 16.5.1 ([21]). *M is an n-dimensional torus identifiable as the quotient of R^n by a lattice L via the map*

$$t \in R^n \to e^{tX} q^o$$

for fixed $q^o \in M$ and $tX = t_1 X_1 + \cdots + t_n X_n$; especially the commutative fields $X_1, \ldots, X_n$ span the tangent space of M at every point.

[21] Lax [24] proved this fact in a somewhat different way.

Step 1. is to recall from §16.2 that M is a 2^n-sheeted covering of the cell $C = [\lambda_1^o, \lambda_2^o] \times \cdots \times [\lambda_{2n-1}^o, \lambda_{2n}^o]$ in which the auxiliary spectrum $\mu_1^o < \cdots < \mu_n^o$ sits, with 2^m covering points over every point of C such that $\lambda_{2i-1}^o < \mu_i^o < \lambda_{2i}^o$ for precisely m values of $1 \leqq i \leqq n$; in particular, *M is compact.*

Step 2. For fixed $q^o \in M$, the map $t \to e^{t\mathsf{X}} q^o$ is $1:1$ in the vicinity of $t = 0$.

PROOF. If not, then $e^{t\mathsf{X}}q^o = q^o + t\mathsf{X}q^o + O(t^2) = q^o$ for arbitrarily small t, i.e., $c\mathsf{X}q^o = 0$ for some non-trivial direction $c \in R^n$, and the corresponding infinitesimal unitary operator $K = c_1K_1 + \cdots + c_nK_n$ annihilates the simple eigenfunctions of Q by Theorem 16.4.6. But the simple eigenfunctions are $2n+1$ in number, while K is a differential operator of degree $2n-1$, and $Kf = 0$ cannot have so many independent solutions. □

Step 3. If q^o is in general position, i.e., if $m = n$, then for small $\delta > 0$,

$$B = \left[e^{t\mathsf{X}}q^o : |t| < \delta\right]$$

covers simply a full neighborhood of q^o, and M has the structure of a smooth n-dimensional manifold in the vicinity; moreover, the commuting fields $\mathsf{X}_1, \ldots, \mathsf{X}_n$ constitute a basis for the tangent space at that point.

Proof. By step 1, the general position of q^o means that the map

$$t \in R^n \to q = e^{t\mathsf{X}}q^o \in M \to \left(\mu_1^o, \ldots, \mu_n^o\right) \in C$$

is locally 1:1. The proof is finished by computing the Jacobian determinant

$$\det \frac{\partial \mu_i^o}{\partial t_j} \neq 0$$

to confirm that $t \in R^n$ is a local coordinate in the vicinity of q^o. The fact is already implicit in §16.2, but the present proof is simpler and useful for the sequel besides. The proof is broken into three simple lemmas.

LEMMA 16.5.2.

$$\frac{\partial \mu_i^o}{\partial t_j} = \mathsf{X}_j \mu_i^o = \sum_{k=1}^{j} \frac{\partial H_{k-1}}{\partial q}(0)\left(2\mu_i^o\right)^{j-k} \mathsf{X}_1 \mu_i^o \quad (i = 1, \ldots, n).$$

PROOF. [22] Let $\mu = \mu_i^o$ and let f be the corresponding eigenfunction with $\int_0^1 f^2\,dx = 1$. Then for any variation $q^\bullet = \mathsf{X}q$, you have $Qf^\bullet + q^\bullet f = \mu^\bullet f + \mu f^\bullet$ and

$$\mu^\bullet = \int f\left(Qf^\bullet + q^\bullet f - \mu f^\bullet\right) = \int f^2 q^\bullet,$$

i.e., $\partial\mu/\partial q = f^2$. Therefore, it suffices to check that

$$\int f^2 \mathsf{X}_j q = 2\mu \int f^2 \mathsf{X}_{j-1} q + \frac{\partial H_{j-1}}{\partial q}(0) \int f^2 \mathsf{X}_1 q.$$

To begin with,

$$\int f^2 \mathsf{X}_1 q = \int f^2 q' = -2\int f f' q = -2\int f'(f'' + f) = -(f')^2\big|_0^1,$$

[22]The computation is similar to the proof of Lemma 16.3.5.

so

$$\begin{aligned}\frac{\partial H_{j-1}}{\partial q}(0)\int f^2\mathsf{X}_1 q &= -(f')^2\frac{\partial H_{j-1}}{\partial q}\bigg|_0^1\\ &= -2\int f'f''\frac{\partial H_{j-1}}{\partial q} - \int (f')^2\left(\frac{\partial H_{j-1}}{\partial q}\right)'\\ &= \int f^2\left(q\frac{\partial H_{j-1}}{\partial q}\right)' - \mu\int f^2\mathsf{X}_{j-1}q - \int (f')^2\mathsf{X}_{j-1}q.\end{aligned}$$

Thus

$$\begin{aligned}&2\mu\int f^2\mathsf{X}_{j-1}q + \frac{\partial H_{j-1}}{\partial q}(0)\int f^2\mathsf{X}_1 q\\ &\quad= \int f(-f''+qf)\mathsf{X}_{j-1}q + \int f^2\left(q\frac{\partial H_{j-1}}{\partial q}\right)' - \int (f')^2\mathsf{X}_{j-1}q\\ &\quad= \int f^2 q\left(\frac{\partial H_{j-1}}{\partial q}\right)' + \int f^2\left(q\frac{\partial H_{j-1}}{\partial q}\right)' - \frac{1}{2}\int (f^2)''\left(\frac{\partial H_{j-1}}{\partial q}\right)'\\ &\quad= \int f^2\left(qD + Dq - \frac{1}{2}D^3\right)\frac{\partial H_{j-1}}{\partial q}\\ &\quad= \int f^2\mathsf{X}_j q,\end{aligned}$$

by Theorem 16.3.4. The proof is finished. □

LEMMA 16.5.3. $\det \mathsf{X}_j\mu_i^o = 2^{n(n-1)/2}\mathsf{X}_1\mu_1^o\cdots\mathsf{X}_1\mu_n^o \times \prod_{i>j}(\mu_i^o - \mu_j^o)$.

PROOF. The formula is self-evident. □

LEMMA 16.5.4. *$\mathsf{X}_1\mu_i^o = 0$ if and only if $\mu_i^o = \lambda_{2i-1}^o$ or λ_{2i}^o.*

PROOF. For $\mu = \mu_i^o$ and $f =$ the associated eigenfunction, $\mathsf{X}_1\mu = -(f')^2(1) + (f')^2(0)$, as in the proof of Lemma 16.5.2. Thus, $\mathsf{X}_1\mu = 0$ only if $f'(0) = \pm f'(1)$, and that happens only if f is of period 1 or 2, as is readily verified by means of the discriminant and the identification of f as a multiple of $y_2(x,\mu)$: namely, you would have $1 = y_2'(0,\mu) = \pm y_2'(1,\mu)$ and $y_1(1,\mu) = [y_2'(1,\mu)]^{-1} = \pm 1$, also, so that $\Delta(\mu) = \pm 2$. □

AMPLIFICATION 16.5.5. The content of Lemma 16.5.4 may be clarified as follows: f_{2i-1}^o of f_{2i}^o has i roots per period $0 \leqq x < 1$. Therefore by translation through one full period, $\mu_i^o(x)$ must hit each of λ_{2i-1}^o and λ_{2i}^o precisely i times; see Amplification 16.6.4 for a picture and more details.

Step 4. is to extend step 3 to points $q^o \in M$ in special position $(m < n)$. The idea is that $e^{t\mathsf{X}}$ will move q^o into general position for some choice of $t \in R^n$ and that $e^{t\mathsf{X}}$ and $e^{-t\mathsf{X}}$ are smooth and inverse to each other, so that $e^{t\mathsf{X}}$ is a diffeomorphism of M. The proof is easy: if $q = e^{t\mathsf{X}}q^o$ is in special position for every $t \in R^n$, then some point from the auxiliary spectrum of q sits at one end of its ambient interval of instability for every t from so small a ball $|t| < \delta$ that step 2 applies. Then you would have a 1:1 continuous map of that n-dimensional ball into a lower-dimensional part of 2^n copies of C. That is impossible.

Step 5. M is now seen to be a compact n-dimensional manifold without boundary on which the commuting vector fields $\mathsf{X}_1,\ldots,\mathsf{X}_n$ act nontrivially at every point and so span the tangent space. The connected piece of M in which a fixed point q^o sits may now be identified as a torus R^n/L, L being the lattice of points $t \in R^n$ such that $e^{t\mathsf{X}}q^o = q^o$. M is therefore a sum of $\ell = 1,2,3,\ldots$ such tori and it remains to prove only that M is connected, i.e., $\ell = 1$. This

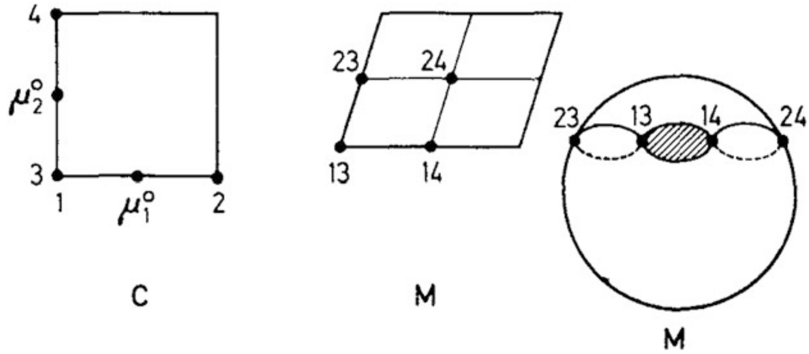

FIGURE 16.2.

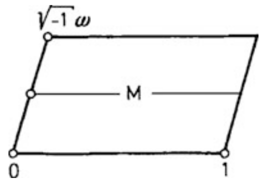

FIGURE 16.3.

is geometrically clear from the fact that if $\mu_i = \lambda_{2i-1}^o$ $(i = 1, \dots, n)$, then M contains just *one* covering point, and from the computations of §16.2 which make it plain that any point of M may be joined to this special point by a smooth curve. A sketch of M for $n = 2$ is seen in Figs. 16.2 and 16.3.

The next item is an immediate corollary of Theorem 16.5.1.

COROLLARY 16.5.6. *The vector field X_{n+1} also acts upon M, so there exists a relation $\mathsf{X} = c_1\mathsf{X}_1 + \cdots + c_n\mathsf{X}_n + \mathsf{X}_{n+1} = 0$ at each point of M. The coefficients are unique because $\mathsf{X}_1, \dots, \mathsf{X}_n$ are independent; moreover, they do not depend upon q because $\exp(t_1\mathsf{X}_1 + \cdots + t_n\mathsf{X}_n)$ is transitive on M and commutes with X. $\mathsf{X}q = 0$ represents a nonlinear ordinary differential equation for q of degree $2n+1$.*[23]

COROLLARY 16.5.7. *The operator Q has $2n + 1 < \infty$ simple eigenvalues if and only if $\mathsf{X}q = 0$ for some choice of $c_1, \dots, c_n$ and n is minimal in this regard.*

PROOF. $\mathsf{X}q = 0$ implies that K annihilates the simple eigenfunctions of Q, so their number cannot exceed $\deg K = 2n + 1$. The rest will be plain. □

EXAMPLE 16.5.8. For $n = 0$, this proves the result of Borg [1]: *Q has just one simple eigenvalue if and only if q is constant, namely $q = \lambda_0$.* Then M comprises just this single point.

EXAMPLE 16.5.9. The result for $n = 1$ was discovered by Hochstadt [16] in a different way. Then $0 = c_1\mathsf{X}_1 q + \mathsf{X}_2 q = c_1 q' + 3qq' - \frac{1}{2}q'''$, and the only solutions of class C_1^∞ are translates of $q(x) = -c_1/3 + 2p(x + \sqrt{-1}\omega'/2)$, in which p is the Weierstrassian elliptic function with primitive periods $\omega =$ the reciprocal of an integer $m = 1, 2, 3, \dots$ and $\sqrt{-1}\omega'$; see Example 16.6.3 in which it is proved that ω' is real and that the invariants of p are

[23] Flaschka [8] proved $\mathsf{X}q = 0$ in a non-geometrical way; see also Goldberg [14]. The case $n = 1$ was discovered by Hochstadt [16].

$e_1 = -\lambda_0^o$, $e_2 = -\lambda_1^o$, $e_3 = -\lambda_2^o$. The general shape of q for $n = 2, 3, \dots$ is still obscure, though a large body of numerical information has been obtained by J. M. Hyman of CIMS; for example, he finds that for $n = 2$, the number of peaks and valleys is usually 4 and, on occasion 5.

Involution. A useful choice of origin in M is to make $\mu_i^o = \lambda_{2i-1}^o$ $(i = 1, \dots, n)$. The involution $\sigma : q(x) \to q(-x)$ is a covering map of M over C inducing a map of R^n modulo periods; the matter will be discussed from this point of view.

PROPOSITION 16.5.10. *By proper choice of the origin of M, σ fixes half-periods and nothing else.*

PROOF. $\sigma X_i = -X_i \sigma$ $(i = 1, 2, \dots, n)$, so

$$e^{t\mathsf{X}} = \sigma e^{-t\mathsf{X}} \sigma.$$

Let q^o be the origin of M as described above. Then $q = e^{-t\mathsf{X}} q^o \in M$ is even if and only if

$$e^{-t\mathsf{X}} q^o = q = \sigma q = \sigma e^{-t\mathsf{X}} \sigma q^o = e^{t\mathsf{X}} q^o,$$

which is to say $e^{2t\mathsf{X}} = I$, i.e., $2t$ is a period. The proof is finished. □

PROPOSITION 16.5.11. *The half-periods of M are precisely the places at which M is 2^n-fold ramified over C, i.e., the places at which $\mu_i^o = \lambda_{2i-1}^o$ or λ_{2i}^o $(i = 1, \dots, n)$. They are 2^n in number.*

PROOF. If $t \in R^n$ is a half-period, then $q(x) = q(-x)$ and each eigenfunction of Q with $f(0) = f(1) = 0$ is even or odd about $x = 1/2$ and so of period 1 or 2, alternately. This proves $\mu_i^o = \lambda_{2i-1}^o$ or λ_{2i}^o $(i = 1, \dots, n)$. The converse is just as easy: if the eigenfunctions are as above, then $Q = \sum_{i=1}^{\infty} \mu_i f_i \otimes f_i$ commutes with $\sigma : f \to f(-\bullet)$, and that happens only if $q(x) = q(-x)$. □

Usage. The origin q^o is now fixed as proposed above $[\mu_i^o = \lambda_{2i-1}^o$ $(i = 1, \dots, n)]$, and q is regarded as a function of $t \in R^n : q(t) = e^{t\mathsf{X}} q^o$ *evaluated at* $x = 0$. The actual potential $q(x)$ attached to a point of $M = R^n/L$ is then obtained as $q(t_1 + x, t_2, \dots, t_n)$, X_1 being the infinitesimal operator of translation.

16.6. Eigenfunctions

The purpose of the present article is to use the infinitesimal unitary operator $K = c_1 K_1 + \dots + c_n K_n + K_{n+1}$ associated with $\mathsf{X} = c_1 \mathsf{X}_1 + \dots + c_n \mathsf{X}_n + \mathsf{X}_{n+1}$, as in §16.4, to study the periodic eigenfunctions of Q under the condition of Corollary 16.5.6: $\mathsf{X} q = 0$. The results will be used to compute the coefficients $c_1, \dots, c_n$ in terms of $\lambda_0^o < \dots < \lambda_{2n}^o$ in §16.7 and to determine the period lattice L in §16.8. K is a differential operator of degree $2n + 1$ annihilating the simple eigenfunctions of Q and acting as

$$Kf_- = 2^n k f_+$$
$$Kf_+ = -2^n k f_-$$

on a double eigenspace with unit perpendicular basis $f_\pm$.[24] The simple eigenfunctions are $2n+1$ in number, so they comprise the whole null space of K. Now for *any* eigenvalue λ with

[24] See Theorem 16.4.6; the notation is a little changed from §16.4 by the insertion of the factor 2^n.

eigenfunction f,[25]

$$Kf = \sum_{i=1}^{n+1} c_i \sum_{j=1}^{i} \left(\frac{1}{2}X_{j-1}q - \frac{\partial H_{j-1}}{\partial q}D\right)(2\lambda)^{i-j} f$$
$$= 2^n \left(\frac{1}{2}m' - mD\right) f$$

with

$$m(x) = m(x,\lambda) = 2^{-n} \sum_{i=1}^{n+1} c_i \sum_{j=1}^{i} \frac{\partial H_{j-1}}{\partial q}(2\lambda)^{i-j}.$$

The properties of the function m will now be explained in a series of simple propositions.

PROPOSITION 16.6.1. *$m(x,\lambda) = (\lambda-\mu_1^o)\dots(\lambda-\mu_n^o)$, the auxiliary spectrum $\mu_1^o < \cdots < \mu_n^o$ being computed for the original potential q translated by the amount $0 \leqq x < 1$ so that each μ_i^o $(i = 1,\dots,n)$ is regarded as a function of $0 \leqq x < 1$.*[26]

PROOF. $m(x,\lambda)$ is a polynomial in λ of degree n with top coefficient $2^{-n}c_{n+1} \times \partial H_0/\partial q \times 2^n = 1$, and by the formula of Lemma 16.5.2, translated in the amount x, $0 = 2^{-n}\mathsf{X}\mu_i^o = m(x,\mu_i^o)\mathsf{X}_1\mu_i^o$ for $i = 1,\dots,n$. This identifies the roots of $m(x,\lambda)$. The proof is finished. □

PROPOSITION 16.6.2. *If λ is a simple eigenvalue of Q, then the associated eigenfunction f is a constant multiple of $\sqrt{m(x,\lambda)}$ with an appropriate determination of the radical.*

PROOF. For such λ, $m(x,\lambda)$ is of one signature in $0 \leqq x < 1$ since $\lambda_{2i-1}^o \leqq \mu_i^o \leqq \lambda_{2i}^o$ $(i = 1,\dots,n)$. Now $Kf = 0$ implies $\frac{1}{2}m'f = mf'$, or what is the same, $m'f^2 = m(f^2)'$. Therefore, m and f^2 are proportional, and it must only be proved that m does not vanish identically in $0 \leqq x < 1$. But if it did, you would have

$$0 = m' = \sum_{i=1}^{n+1} c_i \sum_{j=1}^{i} \mathsf{X}_{j-1}q(2\lambda)^{i-j} = \left(c_1'\mathsf{X}_1 + \cdots + c_n'\mathsf{X}_n\right)q,$$

contradicting the independence of the vector fields X_i $(i \leqq n)$ on M. The signature of the radical may now be determined so that f is proportional to $\sqrt{m}$. □

EXAMPLE 16.6.3. Recall from Example 16.5.9 that $q = -c_1/3 + 2p(x + \sqrt{-1}\omega'/2)$ for $n = 1$. The simple eigenfunctions f_i^o $(i = 0, 1, 2)$ are easily computed from

$$m(x,\lambda) = \frac{1}{2}\left(c_1 + 2\lambda + q\right) = p + \lambda + \frac{c_1}{3};$$

f_i^o is proportional to $\sqrt{m(x,\lambda_i^o)}$ and vanishes once in a period for $i = 1, 2$, so $\lambda_1^o + c_1/3$ and $\lambda_2^o + c_1/3$ can only be the Weierstrassian invariants $-e_2$ and $-e_3$; moreover, $-c_1 = \sigma_1 = \lambda_0^o + \lambda_1^o + \lambda_2^o$, as will be proved in §16.7, so $\lambda_0^o + c_1/3 = -e_1[e_1 + e_2 + e_3 = 0]$, and the simple eigenfunctions are proportional to the 3 Jacobian elliptic functions $\sqrt{p - e_i}$ $(i = 1, 2, 3)$. Because e_1, e_2, e_3 are real, the complex primitive period is pure imaginary; in fact, the primitive periods are

$$\omega = \int_{\lambda_1^o}^{\lambda_2^o} \frac{d\mu}{\ell(\mu)} = \frac{1}{m_1},$$
$$\sqrt{-1}\omega' = \int_{\lambda_2^o}^{\infty} \frac{d\mu}{\ell(\mu)} = \int_{-\infty}^{e_3} \frac{dp}{\sqrt{\left(p - e_1\right)\left(p - e_2\right)\left(p - e_3\right)}}.$$

[25] See Theorem 16.4.3; $c_{n+1} = 1$.

[26] This device was introduced into the subject by Hochstadt [16].

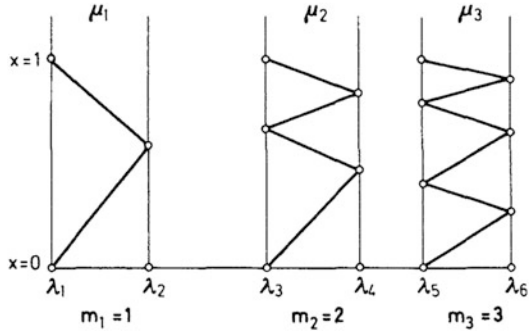

FIGURE 16.4.

AMPLIFICATION 16.6.4. From Propositions 16.6.1 and 16.6.2, you obtain an overview of the motion of the auxiliary spectrum $\mu_1^o < \cdots < \mu_n^o$ under translation by $0 \leqq x < 1$. Let f_i^o $(i = 0, \dots, 2n)$ be the simple eigenfunctions of Q. Then for $i = 1, 2, \dots, n$, f_{2i-1}^o is proportional to $\mu_i^o - \lambda_{2i-1}^o$, while f_{2i}^o is proportional to $\lambda_{2i}^o - \mu_i^o$, and both of these eigenfunctions have precisely m_i roots per period. Thus, μ_i^o meets λ_{2i-1}^o and λ_{2i}^o each m_i times as x runs from 0 to 1 and the values of x at which such meetings take place are interlaced. Figure 16.4 depicts the case $n = 3$, $m_1 = 1$, $m_2 = 2$, $m_3 = 3$, $q =$ the origin of M; naturally, the picture is merely schematic.

PROPOSITION 16.6.5. $\frac{1}{4}(m')^2 - \frac{1}{2}mm'' + m^2(q-\lambda) = \ell^2(\lambda) = -(\lambda - \lambda_2^o)\dots(\lambda - \lambda_{2n}^o)$.

PROOF. The left-hand side is a polynomial in λ of degree $2n+1$ with top coefficient -1, so it suffices to identify the roots as $\lambda_0^o, \dots, \lambda_{2n}^o$. This is done by observing that for $\lambda = \lambda_i^o$ $(i = 0, \dots, 2n)$, $m(x, \lambda)$ is proportional to the square of the corresponding eigenfunction f, whereupon the left-hand side is seen to vanish:

$$\frac{1}{4}(2ff')^2 - \frac{1}{2}f^2\big(2(f')^2 + 2ff''\big) + f^4(q-\lambda) = -f^3\big(f'' + (\lambda - q)f\big) = 0.$$

□

PROPOSITION 16.6.6.

$$y_1(x,\lambda) = \sqrt{\frac{m(x,\lambda)}{m(0,\lambda)}} \times \cos\left[\sqrt{-1}\ell(\lambda)\int_0^x m^{-1}(y,\lambda)dy\right] - \frac{1}{2}m'(0,\lambda)\frac{\sin\left[\sqrt{-1}\ell(\lambda)\int_0^x m^{-1}(y,\lambda)dy\right]}{\sqrt{-1}\ell(\lambda)},$$

$$y_2(x,\lambda) = \sqrt{m(0,\lambda)m(x,\lambda)}\frac{\sin\left[\sqrt{-1}\ell(\lambda)\int_0^x m^{-1}(y,\lambda)dy\right]}{\sqrt{-1}\ell(\lambda)}$$

with a proper determination of the radicals, provided λ is outside the intervals of instability $[\lambda_{2i-1}^o, \lambda_{2i}^o]$ $(i = 1, \dots, n)$.

PROOF. The proviso on λ keeps m from vanishing in $0 \leqq x < 1$. $Qy = \lambda y$ for $y = y_1$ or y_2 is easily verified with the help of Proposition 16.6.5, and you check $y_1(0) = 1$, and $y_1'(0) = 0$, $y_2(0) = 0$, and $y_2'(0) = 1$ by inspection. □

PROPOSITION 16.6.7. $\Delta(\lambda) = y_1(1,\lambda) + y_2'(1,\lambda) = 2\cos[\sqrt{-1}\ell(\lambda)\int_0^1 m^{-1}(x,\lambda)dx]$ *off the intervals of instability* $[\lambda_{2i-1}^o, \lambda_{2i}^o]$ $(i = 1,\dots,n)$*; especially,* λ *is a double eigenvalue* $\lambda_{2i-1} = \lambda_{2i}$ *if and only if*

$$\sqrt{-1}\ell(\lambda)\int_0^1 \frac{dx}{(\lambda-\mu_1^o)\dots(\lambda-\mu_n^o)} = \pm i\pi.$$

PROOF. The formula for $\Delta(\lambda)$ is immediate from Proposition 16.6.6, and the rest follows from the fact that λ is a double eigenvalue $\lambda_{2i-1} = \lambda_{2i}$ if and only if $y_2(x,\lambda) = 0$ has precisely i roots in a period $0 \leqq x < 1$, i.e., if and only if $\sqrt{-1}\ell(\lambda)\int_0^1 m^{-1}dx = \pm i\pi$. □

AMPLIFICATION 16.6.8. A comparison of Proposition 16.6.7 with Hochstadt's formula of §16.1 yields the remarkable identity

$$\frac{1}{2}\int_{\lambda_0}^{\lambda}\prod_{i=1}^{n}(\mu-\lambda_i')\frac{d\mu}{\ell(\mu)} = \ell(\lambda)\int_0^1 \frac{dx}{(\lambda-\mu_1^o)\cdots(\lambda-\mu_n^o)};$$

see Corollary 16.10.9 for additional information.

PROPOSITION 16.6.9. *If* λ *is a double eigenvalue of* Q*, then the functions*

$$\begin{matrix} f_-(x) \\ f_+(x) \end{matrix} = \sqrt{m(x,\lambda)}\begin{matrix}\sin\\ \cos\end{matrix}\left[\sqrt{-1}\ell(\lambda)\int_0^x m^{-1}(y,\lambda)\,dy\right]$$

constitute a (non-unit) perpendicular basis for the eigenspace.

PROOF. f_- may be chosen as the function y_2 of Proposition 16.6.6. Then f_+ is proportional to

$$Kf_- = 2^n\left(\frac{1}{2}m' - mD\right)f_- = \text{constant} \times \sqrt{m} \times \text{the cosine}.$$

□

AMPLIFICATION 16.6.10. The multipliers needed to make the double eigenfunctions $f_\pm$ of Proposition 16.6.9 have unit length are the *same* for both: $2^n k||f_+||^2 = \int f_+ K f_- = 2^n k||f_-||^2$, so the common multiplier is the reciprocal of $\int_0^2 m(x,\lambda)\,dx = 2\int_0^1 m(x,\lambda)dx$; the same multiplier serves for the simple eigenfunctions of Proposition 16.6.2. The evaluation[27]

$$\int_0^1 m(x,\lambda)\,dx = (\lambda-\lambda_1')\cdots(\lambda-\lambda_n')$$

is carried out in Corollary 16.10.7. The upshot is that the simple and double eigenfunctions of Q can be expressed as

$$\sqrt{2}f(x) = \sqrt{\frac{(\lambda-\mu_1^o)\dots(\lambda-\mu_n^o)}{(\lambda-\lambda_1')\dots(\lambda-\lambda_n')}}$$

and

$$\begin{aligned}\begin{matrix} f_-(x) \\ f_+(x) \end{matrix} &= \sqrt{\frac{(\lambda-\mu_1^o)\cdots(\lambda-\mu_n^o)}{(\lambda-\lambda_1')\cdots(\lambda-\lambda_n')}} \\ &\quad\times \begin{matrix}\sin\\ \cos\end{matrix}\left[\sqrt{-1}\ell(\lambda)\int_0^x \frac{dx}{(\lambda-\mu_1^o)\cdots(\lambda-\mu_n^o)}\right],\end{aligned}$$

respectively.

[27] $\lambda_1' < \cdots < \lambda_n'$ are the nontrivial roots of $\Delta'(\lambda) = 0$ introduced in §16.1.

PROPOSITION 16.6.11. *In a double eigenspace, $k^2 = -\ell^2(\lambda)$; especially,*

$$K^2 = 2^{2n}\ell^2(Q) = -2^{2n}\big(Q-\lambda_0^o\big)\cdots\big(Q-\lambda_{2n}^o\big).$$

PROOF. The evaluation of k^2 falls out from Proposition 16.6.5: in a double eigenspace,

$$\begin{aligned} -k^2 &= \left(\frac{1}{2}m' - mD\right)^2 = \frac{1}{4}(m')^2 - \frac{1}{2}mm'' - m'mD + mm'D + m^2D^2 \\ &= \frac{1}{4}(m')^2 - \frac{1}{2}mm'' + m^2(q-\lambda) = \ell^2(\lambda). \end{aligned}$$

□

PROPOSITION 16.6.12. *The following function classes have the same span:*

1°. *the $n+1$ elementary symmetric polynomials in $\mu_1^o, \ldots, \mu_n^o$;*
2°. *any $n+1$ of the squares of the simple eigenfunctions $f_0^o, \ldots, f_{2n}^o$;*
3°. *the gradients $\frac{\partial H_0}{\partial q}, \ldots, \frac{\partial H_n}{\partial q}$.*

The dimension of the common span is precisely $n+1$.

PROOF. The polynomial $m(x,\lambda)$ of degree $n+1$ may be reconstituted from its values at any $n+1$ of the points $\lambda_0^o, \ldots, \lambda_{2n}^o$. The identification of span 1° and span 2° is now plain from Propositions 16.6.1 and 16.6.2, while the identification with span 3° is an easy consequence of the formula

$$2^n m(x,\lambda) = \sum_{i=1}^{n+1} c_i \sum_{j=1}^{i} \frac{\partial H_{j-1}}{\partial q}(2\lambda)^{i-j}.$$

The only point at issue is the value of the dimension, but if that were $\leqq n$, there would exist a dependence among $\partial H_0/\partial q, \ldots, \partial H_n/\partial q$ entailing a dependence among $\mathsf{X}_1, \ldots, \mathsf{X}_n$ on M. That is impossible. □

AMPLIFICATION 16.6.13. A number of interesting formulas may be derived from the same ideas; for example, $m(x,\lambda)$ can be interpolated from its values at λ_i^o $(i = 0, \ldots, n)$, with the help of Amplification 16.6.8:

$$m(x,\lambda) = \sum_{i=0}^{n} \prod_{\substack{j\neq i \\ 0\leqq j\leqq n}} \frac{\lambda-\lambda_j^o}{\lambda_i^o-\lambda_j^o} 2\prod_{k=1}^{n}\big(\lambda_i^o-\lambda_k'\big)\big[f_i^o(x)\big]^2.$$

Now match coefficients of λ^n and λ^{n-1} to produce

$$1 = \sum_{i=0}^{n} \varepsilon_i'\big(f_i^o\big)^2 \text{ and } -\big(\mu_1^o + \cdots + \mu_n^o\big) = \sum_{i=0}^{n}\big[\lambda_i^o - \big(\lambda_0^o + \cdots + \lambda_n^o\big)\big]\varepsilon_i'\big(f_i^o\big)^2$$

with suitable coefficients ε_i' $(i = 0, \ldots, n)$. The second sum is identified as $\frac{1}{2}(q + c_n)$ by the formula for m in terms of $\partial H_0/\partial q, \ldots, \partial H_n/\partial q$. These two formulas may be combined with the evaluation $-c_n = \sigma_1 = \lambda_0^o + \cdots + \lambda_{2n}^o$ of §16.7 to prove[28]

$$\begin{aligned} q = \sigma_1 - 2\big(\mu_1^o + \cdots + \mu_n^o\big) &= -\big(\lambda_0^o + \cdots + \lambda_n^o\big) \\ &+ \big(\lambda_{n+1}^o + \cdots + \lambda_{2n}^o\big) + 2\sum_{i=0}^{n}\lambda_i^o\varepsilon_i'\big(f_i^o\big)^2. \end{aligned}$$

[28]Flaschka also proved a formula of this kind [private communication from J. Moser].

There are similar formulas involving $(f_i^o)^4$ $(i = 0, \dots, 2n)$: $m^2(x, \lambda)$ is interpolated from its values at λ_i^o $(i = 0, \dots, 2n)$ and the coefficients of λ^{2n} and λ^{2n-1} are matched to obtain

$$1 = \sum_{i=0}^{2n} \varepsilon_i'' \big(f_i^o\big)^4 \text{ and } -2\big(\mu_1^o + \cdots + \mu_n^o\big) = \sum_{i=0}^{2n} \big(x_i^o - \sigma_1\big)\varepsilon_i'' \big(f_i^o\big)^4$$

with new coefficients ε_i'' $(i = 0, \dots, 2n)$. These are combined to obtain

$$q = \sum_{i=0}^{2n} \lambda_i^o \varepsilon_i'' \big(f_i^o\big)^4.$$

The general formulas

$$\sum_{j=0}^{\infty} \frac{\partial H_j}{\partial q} \varepsilon^j = \frac{\sqrt{\prod_{i=0}^n \big(1 - 2\varepsilon\lambda_i^o\big)}}{\sqrt{\prod_{i=0}^n \big(1 - 2\varepsilon\lambda_{n+1}^o\big)}} \sum_{i=1}^{n} \frac{\varepsilon_i'}{1 - 2\varepsilon\lambda_i^o} \big(f_i^o\big)^2$$

$$\left(\sum_{j=0}^{\infty} \frac{\partial H_j}{\partial q} \varepsilon^j\right)^2 = \sum_{i=0}^{2n} \frac{\varepsilon_i''}{1 - 2\varepsilon\lambda_i^o} \big(f_i^o\big)^4$$

may be deduced from Corollary 16.12.7.

Amplification 16.6.14. Lax [22] remarked that if $H = c_1 H_1 + \cdots + c_n H_n + H_{n+1}$, then[29] $\partial^2 H / \partial q^2 D$ annihilates the *squares* of the simple eigenfunctions f_i^o $(i = 0, \dots, 2n)$. The proof is simple: $\partial H / \partial q = -c_0$ is constant on the torus M, so

$$\frac{\partial^2 H}{\partial q^2} \mathsf{X}_i q = \frac{\partial^2 H}{\partial q^2} D \frac{\partial H_i}{\partial q} = 0 \quad (i = 0, \dots, n)$$

and $(f_i^o)^2$ $(i = 0, \dots, 2n)$ belongs to the span of $\partial H_i / \partial q$ $(i = 0, \dots, n)$. This fact suggests the existence of some relation between $\partial^2 H / \partial q^2 D$ and the infinitesimal unitary operator $K : f_i^o \to 0$ $(i = 0, \dots, 2n)$, but this unknown.

16.7. Symmetric Polynomials

A number of important symmetric polynomials now make their first appearance.

Theorem 16.7.1. *The coefficients $c_1, \dots, c_n$ of the differential equation $\mathsf{X}q = 0$ are (universal) symmetric polynomials in the simple periodic eigenvalues $\lambda_0^o, \dots, \lambda_{2n}^o$; see below for detailed formulas. An additional $n + 1$ symmetric polynomials appear as integrals of $\mathsf{X}q = 0$.*

Proof. The proof depends upon the identity of Proposition 16.6.5:

$$\begin{aligned}
\ell^2(\lambda) &= \frac{1}{4}(m')^2 - \frac{1}{2} m m'' + m^2(q - \lambda) \\
&= 2^{-2n} \sum_{k=1}^{n+1} \sum_{\ell=1}^{n+1} \sum_{i=1}^{k} \sum_{j=1}^{\ell} c_k c_\ell \\
&\quad \times \left[\frac{1}{4} \mathsf{X}_{i-1} q \mathsf{X}_{j-1} q - \frac{1}{4} \big(\mathsf{X}_{i-1} - q\big)' \frac{\partial H_{j-1}}{\partial q} \right. \\
&\qquad \left. - \frac{1}{4} \frac{\partial H_{i-1}}{\partial q} \big(\mathsf{X}_{j-1} q\big)' + (q - \lambda) \frac{\partial H_{i-1}}{\partial q} \frac{\partial H_{j-1}}{\partial q} \right] \times (2\lambda)^{k+\ell-i-j}.
\end{aligned}$$

To proceed further some lemmas are needed; they stem from Lax [22].

[29] $\partial^2 H / \partial q^2$ is the symmetric operator $f \to \int_0^1 [\partial^2 H / \partial q(x) \partial q(y)] f(y) dy$.

TABLE 16.3.

J_{ij}	$j=1$	2	3	4
$i=1$	q	$\frac{1}{2}q^2$	$\frac{1}{3}q^3-\frac{1}{4}(q')^2$	$\frac{5}{8}q^4-\frac{5}{4}q(q')^2$ $-\frac{1}{8}(q'')^2+\frac{1}{4}q'q'''$
2	$\frac{3}{2}q^2-\frac{1}{2}q''$	$q^3+\frac{1}{4}(q')^2$ $-\frac{1}{2}qq''$	$\frac{9}{8}q^4-\frac{3}{4}q^2q''$ $+\frac{1}{8}(q'')^2$	
3	$\frac{5}{2}q^3-\frac{5}{4}(q')^2$ $-\frac{5}{2}qq''$ $+\frac{1}{4}q''''$	$\frac{15}{8}q^4-\frac{5}{2}q^2q''$ $+\frac{1}{8}(q'')^2$ $-\frac{1}{4}q'q''''$ $+\frac{1}{4}qq''''$	$\frac{9}{4}q^5-\frac{15}{4}q^3q'''$ $+\frac{3}{8}q^2q''''$ $-\frac{3}{4}qq'q'''+q(q'')^2$ $+\frac{3}{4}(q')^2q''$ $-\frac{1}{8}q''q''''$ $+\frac{1}{16}(q''')^2$	

LEMMA 16.7.2. *$\mathsf{X}_i q\,\partial H_{j-1}/\partial q$ is the derivative with regard to x of a universal polynomial J_{ij} in q, q', q'', etc., without constant term. J_{ij} is uniquely determined thereby.* A short table of these quantities appears in table 16.3.

PROOF. $\int \mathsf{X}_i q\,\partial H_{j-1}/\partial q = 0$, so the first part is plain. Now let $q^{(m)}$ be the highest derivative of q involved in the difference I of two determinations of J_{ij}. Then with the temporary notation $q^{(i)} = x_i$ $(i=0,\dots,m+1)$,

$$DI = x_1\frac{\partial I}{\partial x_0} + x_2\frac{\partial I}{\partial x_1} + \dots + x_{m+1}\frac{\partial I}{\partial x_m} = 0,$$

so that $\partial I/\partial x_m = \dots = \partial I/\partial x_0 = 0$, and $I \equiv 0$ by the absence of a constant term. The proof is finished. □

LEMMA 16.7.3.

$$\frac{1}{4}\mathsf{X}_{i-1} - q\mathsf{X}_{j-1}q - \frac{1}{4}\big(\mathsf{X}_{i-1}q\big)'\frac{\partial H_{j-1}}{\partial q} - \frac{1}{4}\frac{\partial H_{i-1}}{\partial q}\big(\mathsf{X}_{j-1}q\big)' + q\frac{\partial H_{i-1}}{\partial q}\frac{\partial H_{j-1}}{\partial q}$$
$$= \frac{1}{2}\big(J_{ij}+J_{ji}\big) \quad \textit{for } i,j \geqq 1.$$

PROOF. Both sides are universal polynomials in q, q', q'', etc. without constant term, so it suffices to check the derivative:

$$
\begin{aligned}
&\frac{1}{4}(\mathsf{X}_{i-1}q)'\mathsf{X}_{j-1}q+\frac{1}{4}\mathsf{X}_{i-1}q(\mathsf{X}_{j-1}q)' \\
&\qquad -\frac{1}{4}\left(\frac{\partial H_{i-1}}{\partial q}\right)'''\frac{\partial H_{j-1}}{\partial q}-\frac{1}{4}(\mathsf{X}_{i-1}q)'\mathsf{X}_{j-1}q \\
&\qquad -\frac{1}{4}\frac{\partial H_{i-1}}{\partial q}\left(\frac{\partial H_{j-1}}{\partial q}\right)'''-\frac{1}{4}\mathsf{X}_{i-1}q(\mathsf{X}_{j-1}q)' \\
&\qquad +q'\frac{\partial H_{i-1}}{\partial q}\frac{\partial H_{j-1}}{\partial q}+q\left(\frac{\partial H_{i-1}}{\partial q}\right)'\frac{\partial H_{j-1}}{\partial q}+q\frac{H_{i-1}}{\partial q}\left(\frac{\partial H_{j-1}}{\partial q}\right)' \\
&\quad =\frac{1}{2}\left(qD+Dq-\frac{1}{2}D^3\right)\frac{\partial H_{i-1}}{\partial q}\frac{\partial H_{j-1}}{\partial q} \\
&\qquad +\frac{1}{2}\frac{\partial H_{i-1}}{\partial q}\left(qD+Dq-\frac{1}{2}D^3\right)\frac{\partial H_{j-1}}{\partial q} \\
&\quad =\frac{1}{2}\mathsf{X}_iq\frac{\partial H_{j-1}}{\partial q}+\frac{1}{2}\frac{\partial H_{i-1}}{\partial q}\mathsf{X}_jq=\frac{1}{2}J'_{ij}+\frac{1}{2}J'_{ji}.
\end{aligned}
$$

□

LEMMA 16.7.4. $J_{i-1,j}+J_{j-1,i}=\frac{\partial H_{i-1}}{\partial q}\frac{\partial H_{j-1}}{\partial q}$ *for* $i,j\geqq 1$, *provided* J_{01} *is redefined to be* $1/2$.

PROOF. The proviso makes the formula hold for $i=j=1$. Otherwise, either i or j exceeds 2, and both sides are universal polynomials in q, q', q'', etc. without constant term, and the derivatives match. Now multiply

$$0=\mathsf{X}q=c_1\mathsf{X}_1q+\cdots+c_n\mathsf{X}_nq+\mathsf{X}_{n+1}q$$

by $\partial H_{j-1}/\partial q$ and integrate to obtain

$$c_1J_{1j}+\cdots+c_nJ_{nj}+J_{n+1j}=-c_{1-j}\quad (j=1,\ldots,n+1),$$

in which $c_0,\ldots,c_{-n}$ are constants of integration depending possibly upon the individual point $q\in M$. It will be proved that $c_0,\ldots,c_{-n}$, and likewise $c_1,\ldots,c_n$, are universal (non-elementary) symmetric polynomials in $\lambda_0^o,\ldots,\lambda_{2n}^o$; especially, they are the *same* at every point of M. To do that, it is necessary to rewrite the formula for $-\ell^2(\lambda)$ displayed at the beginning of the proof in terms of the polynomials J_{ij} and the elementary symmetric polynomials σ_i ($i\leqq 2n+1$) in $\lambda_0^o,\ldots,\lambda_{2n}^o$, so:

$$
\begin{aligned}
2^m\ell^2(\lambda)&=-2^{2n}\big(\lambda^{2n+1}-\sigma_1\lambda^{2n}+\cdots-\sigma_{2n+1}\big)\\
&=\sum_{\substack{1\leqq i\leqq k\leqq n+1\\1\leqq j\leqq \ell\leqq n+1}}c_kc_\ell\left[\frac{1}{2}\big(J_{ij}+J_{ji}\big)-\lambda\big(J_{i-1j}+J_{j-1i}\big)\right](2\lambda)^{k+\ell-i-j}\\
&=\sum_{\substack{1\leqq i\leqq k\leqq n+1\\1\leqq j\leqq \ell\leqq n+1}}c_kc_\ell\big[J_{ij}-2\lambda J_{i-1j}\big](2\lambda)^{k+\ell-i-j}
\end{aligned}
$$

$$= \sum_{1 \leqq k, \ell \leqq n+1} c_k c_\ell \sum_{j=1}^{\ell} J_{kj} (2\lambda)^{\ell - j} - \sum_{1 \leqq k, \ell \leqq n+1} c_k c_\ell J_{01} (2\lambda)^{k+\ell-1}$$

$$= \sum_{1 \leqq j \leqq \ell \leqq n+1} c_{1-j} c_\ell (2\lambda)^{\ell-j} - \frac{1}{2} \sum_{1 \leqq k, \ell \leqq n+1} c_k c_\ell (2\lambda)^{k+\ell-1},$$

by the self-evident telescoping in line 3 and the fact that $J_{oj} = 0$ for $j > 1$. Now match like powers of λ to obtain

$$2^m (-1)^m \sigma_m = 2 \sum_{\substack{1 \leqq j \leqq \ell \leqq n+1 \\ \ell - j = 2n+1-m}} c_{1-j} c_\ell + \sum_{\substack{1 \leqq k, \ell \leqq n+1 \\ k+\ell = 2n+2-m}} c_k c_\ell$$

$$= \sum_{\substack{k+\ell = 2n+2-m \\ k, \ell \leqq n+1}} c_k c_\ell$$

for $m = 0, 1, \ldots, 2n+1$, and solve for $c_n, \ldots, c_{-n}$ in terms of $\sigma_1, \ldots, \sigma_{2n}$, as in the table 16.4 below. The proof is finished. □

AMPLIFICATION 16.7.5. A neater way of expressing $c_{n+1} = 1$, $c_n, \ldots, c_{-n}$ is to introduce the polynomial

$$c(\varepsilon) = \sum_{i=0}^{2n+1} c_{n+1-i} \varepsilon^i,$$

and to remark that

$$c^2(\varepsilon) = \prod_{i=0}^{2n} \left(1 - 2\varepsilon \lambda_i^o\right) \quad \text{modulo}\, \varepsilon^{2n+2},$$

so that

$$c(\varepsilon) = \sqrt{\prod_{i=0}^{2n} \left(1 - 2\varepsilon \lambda_i^o\right)} \quad \text{modulo}\, \varepsilon^{2n+2}.$$

The present formula for $c(\varepsilon)$, combined with Proposition 16.6.1, leads at once to

$$\sum_{j=0}^{\infty} \frac{\partial H_j}{\partial q} \varepsilon^j = \frac{\prod_{i=1}^{n} \left(1 - 2\varepsilon \mu_i^o\right)}{\sqrt{\prod_{i=0}^{2n} \left(1 - 2\varepsilon \lambda_i^o\right)}} \quad \text{modulo}\, \varepsilon^{n+1};$$

in Corollary 16.12.7, it will be seen that the proviso, modulo ε^{n+1}, is unnecessary.

COROLLARY 16.7.6. *$c_{n+1} = 1$, $c_n, \ldots, c_{-n}$ is an algebraic basis for the class of symmetric polynomials in $\lambda_0^o, \ldots, \lambda_{2n}^o$; in particular, these quantities determine the simple spectrum.*

PROOF. See Table 16.4. □

COROLLARY 16.7.7. *Let q be a smooth solution of $\mathsf{X}q = 0$ of period 1, not satisfying any relation $c_1' \mathsf{X}_1 q + \cdots + c_n' \mathsf{X}_n q = 0$, and let it admit the $n+1$ integrals*

$$\sum_{i=1}^{n+1} c_i J_{ij} = -c_{1-j} \quad (j = 1, \ldots, n+1).$$

Then q appears as a point of M, i.e., M represents a complete list of such potentials.

PROOF. See Corollaries 16.5.7 and 16.7.6 above. □

TABLE 16.4.

i	c_{n+1-i}
0	1
1	$-\sigma_1$
2	$2\sigma_2 - \frac{1}{2}\sigma_1^2$
3	$-4\sigma_3 + 2\sigma_1\sigma_2 - \frac{1}{2}\sigma_1^3$
4	$8\sigma_4 - 4\sigma_1\sigma_3 - \frac{5}{2}\sigma_1^2\sigma_2 - 2\sigma_2^2 - \frac{5}{8}\sigma_1^4$

16.8. More About Symmetric Polynomials

The symmetric polynomials $c_{n+1}, c_n, \ldots, c_{-n}$ of Theorem 16.7.1 can be interpreted in another way. The differential equation $c_1\mathsf{X}_1 q + \cdots + c_n\mathsf{X}_n q + \mathsf{X}_{n+1} q = 0$ is integrated once with regard to x, bringing in a constant of integration c_0:

$$c_0 + c_1\frac{\partial H_1}{\partial q} + \cdots + c_n\frac{\partial H_n}{\partial q} + \frac{\partial H_{n+1}}{\partial q} = 0;$$

the latter is a function of the simple spectrum, only, as appears from Corollary 16.3.7:

$$c_0 + c_1 H_0 + \cdots + c_n(2n-1)H_{n-1} + (2n+1)H_n = 0.$$

Next, apply $qD + Dq - \frac{1}{2}D^3$ to the identity $c_0 + c_1\partial H_1/\partial q + \cdots = 0$, integrate with regard to x to obtain

$$c_{-1} + c_0\frac{\partial H_1}{\partial q} + \cdots + c_n\frac{\partial H_{n+1}}{\partial q} + \frac{\partial H_{n+2}}{\partial q} = 0,$$

and use Corollary 16.3.7 once more to check that the new constant of integration c_{-1} is likewise a function of the simple spectrum:

$$c_{-1} + c_0 H_0 + \cdots + c_n(2n+1)H_n + (2n+3)H_{n+1} = 0.$$

Now you may do this as many times as you like. The result is

$$\sum_{-\infty}^{n+1} c_i \frac{\partial H_{i+j}}{\partial q} = 0 \quad (j = 0, 1, 2, \ldots),$$

with the understanding that $H_n = 0$ for $n < 0$, or what is the same in the language of formal power series,

$$\begin{aligned} 0 &= \sum_{j=0}^{\infty}\sum_{i=-\infty}^{n+1} c_i\varepsilon^{n+1-i}\frac{\partial H_{i+j}}{\partial q}\varepsilon^{i+j} \\ &= \sum_{i=-\infty}^{n+1} c_i\varepsilon^{n+1-i}\sum_{j=i}^{\infty}\frac{\partial H_j}{\partial q}\varepsilon^j, \end{aligned}$$

and with a little manipulation, this may be brought to the form

$$\begin{aligned}\sum_{-\infty}^{n+1} c_i\varepsilon^{n+1-i}\sum_{j=0}^{\infty}\frac{\partial H_j}{\partial q}\varepsilon^j &= \varepsilon^n \times \sum_{i=1}^{n+1}\sum_{j=1}^{i} c_i\frac{\partial H_{j-1}}{\partial q}\varepsilon^{1+i-j}\\ &= (2\varepsilon)^n m\left(x,\frac{1}{2\varepsilon}\right)\\ &= \prod_{i=1}^{n}\left(1-2\varepsilon\mu_i^o\right).\end{aligned}$$

But by Corollary 16.12.7, this is the same as to say

$$\sum_{i=0}^{\infty} c_{n+1-i}\varepsilon^i = \sqrt{\prod_{i=0}^{2n}\left(1-2\varepsilon\lambda_i^o\right)},$$

proving that the present $c_0,\ldots,c_{-n}$ are the same as before; see also Amplification 16.7.5.

16.9. An Algebraic Variety

The relations of §16.7

$$\sum_{i=1}^{n+1} c_i J_{ij} + c_{1-j} = 0 \quad (j=1,\ldots,n+1)$$

are polynomials in $x_0 = q$, $x_1 = q', \ldots, x_{2n} = q^{(2n)}$, and so define an algebraic variety V of dimension $2n+1-n-1 = n$ in R^{2n+1}. Moreover, the n-dimensional torus M sits inside V, as is plain from the observation that the map

$$q \in M \to \left[q(0), q'(0), \ldots, q^{(2n)}(0)\right] \in V$$

is 1:1 since the image specifies the solution of $\mathsf{X}q = 0$ uniquely.

EXAMPLE 16.9.1 ([30]). For $n=1$, the variety is defined in R^3 by the relations

1°. $0 = c_0 + c_1x_0 + \frac{3}{2}x_0^2 - \frac{1}{2}x_2$,

2°. $0 = c_{-1} + c_1\frac{1}{2}x_0^2 + x_0^3 - \frac{1}{2}x_0x_2 + \frac{1}{4}x_1^2$,

from which you may eliminate x_2, so:

$$c_1\frac{1}{2}x_0^2 + \frac{1}{2}x_0^3 + c_0x_0 - c_{-1} = \frac{1}{4}x_1^2.$$

This defines a complex torus in C^2 of which the real part is a circle. Therefore, $V = M$.

EXAMPLE 16.9.2 ([31]). For $n=2$, the situation is already quite complicated. The variety is defined in R^5 by the relations

1°. $0 = c_0 + c_1x_0 + c_2(\frac{3}{2}x_0^2 - \frac{1}{2}x_2) + \frac{5}{2}x_0^3 - \frac{5}{4}x_1^2 - \frac{5}{2}x_0x_2 + \frac{1}{4}x_4$

2°. $0 = c_- + c_1\frac{1}{2}x_0^2 + c_2(x_0^3 - \frac{1}{2}x_0x_2 + \frac{1}{4}x_1^2) + \frac{15}{8}x_0^4 - \frac{5}{2}x_0x_2 + \frac{1}{8}x_2^2$
$-\frac{1}{4}x_1x_3 + \frac{1}{4}x_0x_4$

3°. $0 = c_{-2} + c_1(\frac{1}{3}x_0^3 - \frac{1}{4}x_1^2) + c_2(\frac{9}{8}x_0^4 - \frac{3}{4}x_0^2x_2 + \frac{1}{8}x_2^2)$
$+\frac{9}{4}x_0^5 - \frac{15}{4}x_0^3x_3 + \frac{3}{8}x_0^2x_4 - \frac{3}{4}x_0x_1x_3$
$+x_0x_2^2 + \frac{3}{4}x_1^2x_2 - \frac{1}{8}x_2x_4 + \frac{1}{16}x_3^2$.

CONJECTURE 16.9.3. $M = V$; compare Lax [24].

[30] See Table 16.3 above.
[31] See Table 16.3 above.

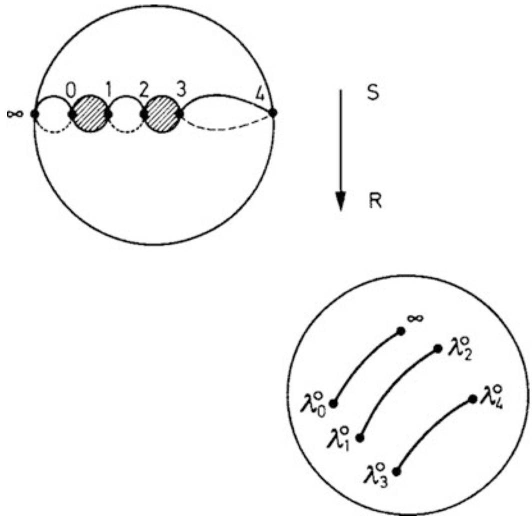

FIGURE 16.5.

16.10. Periods of M

To compute the periods of the real torus $M = R^n/L$, it is necessary to make more explicit the relation between the auxiliary spectrum $(\mu_1^o, \dots, \mu_n^o) \in C$ and the global coordinate $(t_1, \dots, t_n) \in R^n$. Let $\mu_i^o = \mu_i$ $(i = 1, \dots, n)$ for ease of notation. Recall from Propositions 16.6.1 and 16.6.5 that $m(x, \lambda) = (\lambda - \mu_1) \dots (\lambda - \mu_n)$ satisfies the identity

$$\ell^2(\lambda) = -\big(\lambda - \lambda_0^o\big) \dots \big(\lambda - \lambda_{2n}^o\big) = \frac{1}{4}(m')^2 - \frac{1}{2}mm'' + m^2(q - \lambda)$$

and let $\lambda = \mu_i$ for some $i = 1, \dots, n$. Then $m(x, \mu_i) = 0$, so that $m'(x, \mu_i) = 2\sqrt{\ell^2(\mu_i)} = 2\ell(\mu_i)$ with a suitable determination of the radical, and by differentiating $m(x, \mu_i)$ at $x = 0$, you get the supplementary formula[32]

$$\mathsf{X}_1\mu_i(0) = -\frac{m'\big(0, \mu_i\big)}{m^\bullet\big(0, \mu_i\big)} = \frac{2\ell\big(\mu_i\big)}{m^\bullet\big(0, \mu_i\big)} \quad (i = 1, \dots, n).$$

Warning. The signature of the radical $\ell(\mu_i^o)$ is ambiguous. The correct attitude is to construe the formula on the Riemann surface S of the irrationality $\ell(\lambda) = \sqrt{-(\lambda - \lambda_0^o) \dots (\lambda - \lambda_{2n}^o)}$ with change of sheet each time μ_i hits λ_{2i-1}^o or λ_{2i}^o so as to keep $\lambda_{2i-1}^o \leqq \mu_i \leqq \lambda_{2i}^o$. S is a sphere with n handles forming a 2-sheeted covering of the Riemann sphere R of λ, simply ramified over $\infty, \lambda_0^o, \dots, \lambda_{2n}^o$. Figure 16.5 depicts the case $n = 2$; see §16.13.

Now fix $\omega \in R^n$, let $\partial q/\partial t = \omega \mathsf{X} q$ with $\omega X = \omega_1 X_1 + \dots + \omega_n X_n$, and compute $d\mu_i = \omega \mathsf{X} \mu_i \, dt$ $(i = 1 \dots, n)$ with the aid of Lemma 16.5.2:

$$\frac{d\mu_i}{dt} = \sum_{j=1}^n \omega_j \mathsf{X}_j \mu_i = \sum_{j=1}^n \omega_j \sum_{k=1}^j \frac{\partial H_{k-1}(0)}{\partial q} \big(2\mu_i\big)^{j-k} \times \frac{2\ell\big(\mu_i\big)}{m^\bullet\big(0, \mu_i\big)},$$

or what is the same for fixed $1 \le k \le n$,

$$\sum_{i=1}^n \mu_i^{\ell-1} \frac{d\mu_i}{\ell\big(\mu_i\big)} = \sum_{j=1}^n \Lambda_{\ell j} \omega_j \, dt$$

[32] $' = \partial/\partial x$; $^\bullet = \partial/\partial \mu$.

with

$$\Lambda_{\ell j} = \sum_{k=1}^{j} \frac{\partial H_{k-1}}{\partial q}(0) 2^{j-k+1} \sum_{i=1}^{n} \frac{\mu_i^{j-k+\ell-1}}{m^\bullet(0,\mu_i)}$$
$$= \sum_{k=1}^{j} \frac{\partial H_{k-1}}{\partial q}(0) 2^{j-k+1} \frac{1}{2\pi\sqrt{-1}} \oint \frac{\lambda^{j-k+\ell-1}}{m(0,\lambda)}\, d\lambda,$$

the integral being taken about a big circle.

PROPOSITION 16.10.1. *Λ can be expressed by means of the first n elementary symmetric polynomials $\sigma_1, \ldots, \sigma_n$ in the simple eigenvalues of Q; see Table 16.5 below for explicit formulas.*

PROOF. The adjusted entry

$$2^{\ell-2}\Lambda_{\ell j} = \sum_{k=1}^{j} \frac{\partial H_{k-1}}{\partial q}(0) \frac{1}{2\pi\sqrt{-1}} \oint \frac{(2\lambda)^{\ell+j-k-1}\, d\lambda}{m(0,\lambda)} \equiv 2^{n-1} c^*_{\ell+j-n-1}$$

depends upon $\ell + j$ alone as the sum suggests because $\oint \lambda^{\ell+j-k-1} m^{-1}\, d\lambda$ vanishes for $k > j$ as that makes the power $\leqq \deg m - 2$; moreover, $\Lambda_{\ell j} = 0$ for $\ell + j \leqq n$ for the same reason, so $c_i^* = 0$ for $i < 0$. The summation over k is now extended up to $k = n+1$, and the definition of $c^*_{\ell+j-n-1}$ so obtained is adopted for *all* $\ell + j$. The nontrivial quantities $c_0^*, \ldots, c_{n-1}^*$ may now be computed in terms of $\sigma_1, \ldots, \sigma_n$ from Table 16.4, using the identity

$$2^{n-1} \sum_{j=1}^{n+1} c_j c^*_{\ell+j-n-1} = \frac{1}{2\pi\sqrt{-1}} \oint (2\lambda)^{\ell-1} \frac{2^n m(0,\lambda)}{m(0,\lambda)}\, d\lambda$$
$$= 0 \quad (\ell = 1, \ldots, n),$$

which follows from the next to last display of §16.8, the evaluation

$$2^{n-1} c_0^* = \frac{1}{2\pi\sqrt{-1}} \oint \frac{(2\lambda)^{n-1}}{m(0,\lambda)}\, d\lambda = 2^{n-1},$$

and the recursion

$$-c_\ell^* = c_{n-\ell+1} c_0^* + c_{n-\ell+2} c_1^* + \cdots + c_n c^*_{\ell-1},$$

with the results seen in Table 16.5. The proof is finished. ☐

TABLE 16.5.

ℓ	c_ℓ^*
0	1
1	σ_1
2	$\frac{3}{2}\sigma_1^2 - 2\sigma_2$
3	$\frac{5}{2}\sigma_1^3 - 6\sigma_1\sigma_2 + 4\sigma_3$

AMPLIFICATION 16.10.2. Λ may now be displayed as:

$$\begin{bmatrix} 0 & \dots & 0 & 0 & 2^n c_0^* \\ 0 & \dots & 0 & 2^{n-1}c_0^* & 2^{n-1}c_1^* \\ 0 & \dots & 2^{n-2}c_0^* & 2^{n-2}c_1^* & 2^{n-2}c_2^* \\ \vdots & & & & \\ 2c_0^* & \dots & 2c_{n-3}^* & 2c_{n-2}^* & 2c_{n-1}^* \end{bmatrix};$$

in particular, $\det \Lambda = 2^{n(n+1)/2} \neq 0$.

AMPLIFICATION 16.10.3. A more elegant description of $c_0^*, \dots, c_{n-1}^*$ is by means of the polynomial

$$c^*(\varepsilon) = \sum_{i=0}^{n-1} c_i^* \varepsilon^i,$$

to wit, $c^*(\varepsilon) = c^{-1}(\varepsilon)$ modulo ε^n, with $c(\varepsilon)$ as in Amplification 16.7.5. The formula[33] $\Lambda_{ij}^{-1} = c_{i+j}2^{j-1-n}$ $(1 \leqq i, j \leqq n)$ is another way of saying the same thing.

The computation of the periods is easy now. Fix $i = 1, \dots, n$ and pick a primitive period ω so that, under the flow $\partial q/\partial t = \omega \mathsf{X} q$ starting from the origin, μ_j $(j \neq i)$ returns to λ_{2j-1}^o at time $t = 1/2$ without change of sheet, while μ_i starts at λ_{2i-1}^o and ends at λ_{2i}^o at time $t = 1/2$, also without change of sheet.[34] Then

$$\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} \frac{\mu^{\ell-1}\, d\mu}{\ell(\mu)} = \sum_{j=1}^n \Lambda_{\ell j}\omega_j \times \frac{1}{2}.$$

This proves

THEOREM 16.10.4. *A complete set of primitive periods of M is given by the formula*

$$2\Lambda^{-1} \int_{\lambda_{2i-1}^o}^{\lambda_{2\ell}^o} \left(1, \mu, \dots, \mu^{n-1}\right)^\dagger \frac{d\mu}{\ell(\mu)}$$

as i runs from 1 to n.

The differential equations relating $t \in R^n$ and $(\mu_1, \dots, \mu_n) \in C$ are restated as

THEOREM 16.10.5. *For the flow $\partial q/\partial t = \mathsf{X}_j q$,*

$$\sum_{i=1}^n \left(1, \mu_i, \dots, \mu_i^{n-1}\right)^\dagger \frac{d\mu_i}{\ell(\mu_i)} = (\text{the } j\text{th column of } \Lambda) \times dt.$$

COROLLARY 16.10.6. *The volume element of M inherited from R^n may be expressed as*

$$d^n t = 2^{-n(n+1)/2} \prod_{i>j} (\mu_i - \mu_j) \left|\ell(\mu_1) \dots \ell(\mu_n)\right|^{-1} d^n \mu.$$

The volume of M is then

$$2^{-n(n-1)/2} \det \int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} \frac{\mu^{j-1} d\mu}{\ell(\mu)};$$

an extra factor of 2^n comes in front because M is a 2^n-sheeted covering of C.

COROLLARY 16.10.7 ([35]). $\int_0^1 m(x, \lambda)\, dx = \prod_{i=1}^n (\lambda - \lambda_i')$.

[33] $c_{i+j} = 0$ if $i + j > n + 1$.

[34] The desired effect can be achieved by moving along a suitable curve in M lifted up from C, and the latter can bc deformed into rectilinear motion in some direction ω.

[35] $\lambda_1' < \dots < \lambda_n'$ are the nontrivial roots of $\Delta'(\lambda) = 0$ introduced in §16.1.

PROOF. The left-hand side is constant on M because

$$2^n \int_0^1 m\,dx = \int_0^1 \sum_{i=1}^{n+1} \sum_{j=1}^{i} c_i \frac{\partial H_{j-1}}{\partial q} (2\lambda)^{i-j}\,dx = \sum_{i=1}^{n+1} \sum_{j=1}^{i} c_i (2j-3) H_{j-2} (2\lambda)^{i-j},$$

by Corollary 16.3.7, and the final expression is a function of the simple spectrum only. Therefore, the period average $\int_0^1 m\,dx$ can be replaced by the average of $(\lambda-\mu_1)\ldots(\lambda-\mu_n)$ over M with regard to the volume element $d^n t$, and the latter can be computed with the aid of Corollary 16.10.6, the upshot being that you have only to verify the vanishing of

$$\int_C \prod_{i=1}^{n} (\lambda-\mu_i) \frac{\prod_{i>j}(\mu_i-\mu_j)}{\ell(\mu_1)\ldots\ell(\mu_n)}\,d^n\mu = \det \int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} (\lambda-\mu)\mu^{j-1} \frac{d\mu}{\ell(\mu)}$$

for $\lambda = \lambda_1', \ldots, \lambda_n'$. The latter are determined so as to make

$$\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} \prod_{\ell=1}^{n} (\mu-\lambda_\ell') \frac{d\mu}{\ell(\mu)} = 0 \quad (i=1,\ldots,n).$$

Now the monomials $(\lambda_j'-\mu)\mu^{k-1}$ $(k=1,\ldots,n)$ can be regarded as a basis for polynomials of degree $\leqq n$ vanishing at $\mu=\lambda_j'$. Thus, for fixed $1 \leqq j \leqq n$,

$$\prod_{\ell=1}^{n} (\mu-\lambda_\ell') = -(\lambda_j'-\mu)\mu^{n-1} + \sum_{k=1}^{n-1} \varepsilon_k (\lambda_j'-\mu)\mu^{k-1}$$

with coefficients ε_k $(k=1,\ldots,n-1)$ depending upon j, and

$$\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} (\lambda_j'-\mu)\mu^{n-1} \frac{d\mu}{\ell(\mu)} = \sum_{k=1}^{n-1} \varepsilon_k \int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} (\lambda_j'-\mu)\mu^{k-1} \frac{d\mu}{\ell(\mu)},$$

which is to say that for $\lambda=\lambda_j'$ the bottom row of the determinant belongs to the span of the preceding rows. The determinant vanishes for this reason. □

AMPLIFICATION 16.10.8. The evaluation

$$2^n(\lambda-\lambda_1')\ldots(\lambda-\lambda_n') = \sum_{i=1}^{n+1}\sum_{j=1}^{i} c_i(2j-3)H_{j-2}(2\lambda)^{i-j}$$

is equivalent to

$$\sum_{j=0}^{\infty}(2j-1)H_{j-1}\varepsilon^j = \frac{\prod_{i=1}^{n}(1-2\varepsilon\lambda_i')}{\sqrt{\prod_{i=0}^{2n}(1-2\varepsilon\lambda_i^o)}}, \quad \text{modulo } \varepsilon^{n+1},$$

in agreement with the more detailed identity of Corollary 16.12.8.

COROLLARY 16.10.9. *The double spectrum is determined from the simple spectrum by the rule*

$$\sqrt{-1}\ell(\lambda-\mu_1)\frac{\int_C\left[(\lambda-\mu_1)\ldots(\lambda-\mu_n)\right]^{-1} d\text{ volume}}{\text{volume of } C}$$

$$= \sqrt{-1}\ell(\lambda)\frac{\det\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o}(\lambda-\mu)^{-1}\mu^{j-1}\frac{d\mu}{\ell(\mu)}}{\det\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o}\mu^{j-1}\frac{d\mu}{\ell(\mu)}} = \pm m\pi \quad (m \neq m_1,\ldots,m_n).$$

PROOF. The proof is the same as that of Corollary 16.10.7: you have only to rewrite Proposition 16.6.7. The point is that $\sqrt{-1}\ell(\lambda)\int_0^1 m^{-1}(x,\lambda)\,dx = \cos^{-1}[1/2\Delta(\lambda)]$ is constant on M so that the averaging of $m^{-1}(x,\lambda)$ can be done on the whole of M and not just over the period $0 \leqq x < 1$. □

EXAMPLE 16.10.10. For $n = 2$ and $m_1 = 1$, the ratio in Corollary 16.10.9 reduces to

$$\sqrt{-1}\ell(\lambda)\int_{\lambda_1^o}^{\lambda_2^o}(\lambda-\mu)^{-1}\frac{d\mu}{\ell(\mu)}$$

which is to be compared to the expression

$$\frac{\sqrt{-1}}{2}\int_{\lambda_0^o}^{\lambda}(\mu-\lambda_1')\frac{d\mu}{\ell(\mu)}$$

figuring in Hochstadt's formula from §6.1. The fact that they are the same may be proved directly by the addition theorem for the Weierstrassian elliptic function p. Take $\sigma_1 = 0$ so that $-\lambda_0^o = e_1$, $-\lambda_1^o = e_2$, $-\lambda_2^o = e_3$ are the Weierstrassian invariants, as per Example 16.6.3, and substitute $\lambda = -p(x)$. Then, with the notation $\omega_2 = 1/2 + \sqrt{-1}\omega'/2$ and $\omega_3 = \sqrt{-1}\omega'/2$, you have to verify that

$$\ell(\lambda)\int_{\lambda_1^o}^{\lambda_2^o}\frac{1}{\lambda-\mu}\frac{d\mu}{\ell(\mu)} = p'(x)\int_{\omega_2}^{\omega_3}\frac{dy}{p(y)-p(x)}$$

is the same as

$$\frac{1}{2}\int_{\lambda_0^o}^{\lambda}(\mu-\lambda_1')\frac{d\mu}{\ell(\mu)} = \int_{1/2}^{x}p(y)dy + \lambda_1'\left(x-\frac{1}{2}\right).$$

But this follows (with tears) from a variant of the addition formula, namely

$$\frac{1}{2}\frac{p'(x)-p'(y)}{p(x)-p(y)} + \int_{1/2}^{x+y}p - \frac{1}{2}\int_{1/2}^{x}p - \frac{1}{2}\int_{1/2}^{y}p + \int_{\omega_2}^{\omega_3}p = 0,$$

upon integrating from $y = \omega_2$ to $y = \omega_3$ and recognizing

$$\lambda_1' = \int_{\lambda_1^o}^{\lambda_2^o}\frac{\mu d\mu}{\ell(\mu)} \text{ as } 2\int_{\omega_2}^{\omega_3}p.$$

Presumably, the general identity

$$\frac{1}{2}\int_{\lambda_0}^{\lambda}\prod_{i=1}^{n}(\mu-\lambda_i')\frac{d\mu}{\ell(\mu)} = \ell(\lambda)\frac{\int_C\left[(\lambda-\mu_1)\dots(\lambda-\mu_n)\right]^{-1}d\text{ volume}}{\text{volume of } C}$$

is a consequence of the addition theorem of the hyperelliptic abelian integral.

AMPLIFICATION 16.10.11. The formula of Theorem 16.10.4 reduces the solution of $q^\bullet = \omega X q$ to a simple problem in ordinary differential equations; see §6.14 for explicit formulas for q as a function of $t \in R^n$; the latter may be regarded as substantiating the technical condition of §6.4.

THEOREM 16.10.12. *The series $\lambda_0^o < \cdots < \lambda_{2n}^o$ is the simple spectrum of a potential of class C_1^∞ if and only if*

$$\sum_{i=1}^{n}m_i\int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o}\mu^{j-1}\frac{d\mu}{\ell(\mu)} = \begin{matrix} 0 & (j<n) \\ \pm 1 & (j=n)\end{matrix}$$

for some integral $0 < m_1 < \cdots < m_n$.

AMPLIFICATION 16.10.13. $m_1 < m_2 < \cdots < m_n$ are unique or nonexistent because

$$\det \int_{\lambda^o_{2i-1}}^{\lambda^o_{2i}} \mu^{j-1} \frac{d\mu}{\ell(\mu)} = 2^{n(n-1)/2} \times \text{volume of } M \neq 0.$$

Note that the potential is of primitive period $1/m$ if and only if m is the greatest common divisor of m_i $(i = 1, \ldots, n)$; compare Borg [1] who proved that q is of period $1/2$ if and only if $\lambda_{2i-1} = \lambda_{2i}$ for odd $i = 1, 3, 5, \ldots$.

PROOF OF THE NECESSITY. Translate the function q^o sitting at the origin of M through a full period $0 \leqq x < 1$. Then both $\mu_i = \lambda^o_{2i-1}$ and $\mu_i = \lambda^o_{2i}$ have m_i roots per period, with change of sheet and accompanying change of signature in the radical $\ell(\mu_i)$ at each root; see Amplification 16.6.4 and the accompanying Fig. 16.4. Now integrate the differential equations of Theorem 16.10.5 over $0 \leqq x < 1$ with $t_2 = \cdots = t_n = 0$ fixed to obtain

$$\sum_{i=1}^{n} 2m_i \int_{\lambda^o_{2i-1}}^{\lambda^o_{2i}} \mu^{j-1} \frac{d\mu}{\ell(\mu)} = \Lambda_{j1} \quad (j = 1, \ldots, n)$$

and recall that $\Lambda_{j1} = 0$ for $j < n$ while $\Lambda_{n1} = 2$. □

PROOF OF THE SUFFICIENCY. Let $\lambda^o_0, \ldots, \lambda^o_{2n}$ meet the stated condition and form the torus $M = R^n/L$ with the periods of Theorem 16.10.4. Recall Amplification 16.6.13, fix $\mu_i(0) = \lambda_{2i-1}$ $(i = 1, \ldots, n)$, and solve

$$\sum_{j=1}^{n} \mu_j^{i-1} \frac{d\mu_j}{\ell(\mu_j)} = \begin{matrix} 0 & (i < n) \\ 2\,dx & (i = n) \end{matrix}$$

for $\mu_i(x)$ $(i = 1, \ldots, n)$ up to $x = 1$. Then the function

$$q(x) = \lambda^o_0 + \cdots + \lambda^o_{2n} - 2\big[\mu_1(x) + \cdots + \mu_n(x)\big]$$

is smooth and of period 1 by the condition. To see this, you have only to solve for[36]

$$\mu_i' = \frac{2\ell(\mu_i)}{m^\bullet(\mu_i)} \quad (i = 1, \ldots, n)$$

and to study the ensuing periodic motion in the cell C. It remains to prove that the Hill's operator $Q = -d^2/dx^2 + q(x)$ has $\lambda^o_0, \ldots, \lambda^o_{2n}$ as its simple spectrum, and nothing more. The chief step is to prove the identity

$$\ell^2(\lambda) = \frac{1}{4}(m')^2 - \frac{1}{2}mm'' + m^2(q - \lambda).$$

The periodic spectrum is then read off as follows: the identity implies that $f^o_i = \sqrt{m(x, \lambda^o_i)}$ is a simple (non-unit) eigenfunction of Q with eigenvalue λ^o_i for $i = 0, \ldots, 2n$. Besides, the functions y_1 and y_2 are given by the formulas of Proposition 16.6.6 off the intervals $[\lambda^o_{2i-1}, \lambda^o_{2i}]$, and now the formula of Proposition 16.6.7 for the discriminant shows that the rest of the periodic spectrum of Q is double. The proposed identity is equivalent to

$$\frac{1}{2}\left(\frac{m'}{m}\right)' + \frac{1}{4}\left(\frac{m'}{m}\right)^2 = q - \lambda - \frac{\ell^2}{m^2}.$$

Now[37]

$$\frac{1}{2}\frac{m'}{m} = -\frac{1}{2}\sum_{i=1}^{n} \frac{\mu_i'}{\lambda - \mu_i} = -\sum_{i-1}^{n} \frac{1}{\lambda - \mu_i}\frac{\ell(\mu_i)}{m^\bullet(\mu_i)},$$

[36] $m(\lambda) = (\lambda - \mu_1) \ldots (\lambda - \mu_n)$ as in §16.6; $^\bullet = \partial/\partial\mu$.
[37] $' = \partial/\partial x$; $^\bullet = \partial/\partial\mu$.

and

$$
\begin{aligned}
\frac{1}{2}\left(\frac{m'}{m}\right)' = &-\sum_{i=1}^{n} \frac{1}{(\lambda-\mu_i)^2}\frac{2\ell^2(\mu_i)}{m^{\bullet 2}(\mu_i)} \\
&+\sum_{i=1}^{n}\frac{\ell(\mu_i)}{\lambda-\mu_i}\sum_{j\neq i}\prod_{k\neq i,j}\frac{1}{\mu_i-\mu_k}\left[\frac{2\ell(\mu_i)}{m^\bullet(\mu_i)}-\frac{2\ell(\mu_j)}{m^\bullet(\mu_j)}\right]\frac{1}{(\mu_i-\mu_j)^2} \\
&-\sum_{i=1}^{n}\frac{1}{\lambda-\mu_i}\frac{1}{m^\bullet(\mu_i)}\times\frac{1}{2}\frac{1}{\ell(\mu_i)}\times\ell^{\bullet 2}(\mu_i)\times\frac{2\ell(\mu_i)}{m^\bullet(\mu_i)} \\
= &-2\sum_{i=1}^{n}\frac{1}{(\lambda-\mu_i)^2}\frac{\ell^2(\mu_i)}{m^{\bullet 2}(\mu_i)} \\
&+2\sum_{j\neq i}\frac{1}{\lambda-\mu_i}\frac{\ell^2(\mu_i)}{m^{\bullet 2}(\mu_i)}\frac{1}{\mu_i-\mu_j} \\
&-2\sum_{j\neq i}\frac{1}{\lambda-\mu_i}\frac{\ell(\mu_j)}{m^\bullet(\mu_i)}\frac{\ell(\mu_j)}{m^\bullet(\mu_j)}\frac{1}{\mu_i-\mu_j} \\
&-\sum_{i=1}^{n}\frac{1}{\lambda-\mu_i}\frac{\ell^2(\mu_i)}{m^{\bullet 2}(\mu_i)} \\
= &-2\sum_{i=1}^{n}\frac{1}{(\lambda-\mu_i)^2}\frac{\ell^2(\mu_i)}{m^{\bullet 2}(\mu_i)}+\sum_{i=1}^{n}\frac{1}{\lambda-\mu_i}\ell^2(\mu_i)\frac{m^{\bullet\bullet}(\mu_i)}{m^{\bullet 3}(\mu_i)} \\
&-\sum_{j\neq i}\frac{1}{\lambda-\mu_i}\frac{1}{\lambda-\mu_j}\frac{\ell(\mu_i)}{m^\bullet(\mu_i)}\frac{\ell(\mu_i)}{m^\bullet(\mu_j)}-\sum_{i=1}^{n}\frac{1}{\lambda-\mu_i}\frac{\ell^{\bullet 2}(\mu_i)}{m^{\bullet 2}(\mu_i)},
\end{aligned}
$$

so

$$
\begin{aligned}
&\frac{1}{2}\left(\frac{m'}{m}\right)'+\frac{1}{4}\left(\frac{m'}{m}\right)^2 \\
&\quad=-\sum_{i=1}^{n}\frac{1}{(\lambda-\mu_i)^2}\frac{\ell^2(\mu_i)}{m^{\bullet 2}(\mu_i)}+\sum_{i=1}^{n}\frac{1}{\lambda-\mu_i}\left[\frac{\ell^2(\mu_i)m^{\bullet\bullet}(\mu_i)}{m^{\bullet 3}(\mu_i)}-\frac{\ell^{\bullet 2}(\mu_i)}{m^{\bullet 2}(\mu_i)}\right],
\end{aligned}
$$

and this is to be compared to $q-\lambda-\ell^2/m^2$. Both are rational functions of λ vanishing at ∞, and it is an elementary exercise to check that they have the same principal parts at their common poles $\lambda=\mu_i$ $(i=1,\ldots,n)$. Therefore, they are the same. The proof is finished. □

COROLLARY 16.10.14 ([38]). *The potential q is of period $1/m$ if and only if m divides m_i $(i=1,\ldots,n)$, i.e., if and only if $\lambda_{2i-1}=\lambda_{2i}$ for i incongruent to 0 modulo m.*

16.11. The Korteweg-de Vries Flow

An amusing consequence of Theorem 16.10.4 is that *the Korteweg-de Vries flow $q^\bullet = X_2 q = 3qq' - \frac{1}{2}q'''$ is periodic on M with period T if and only if*

$$\sum_{i=1}^{n} 2m_i' \int_{\lambda_{2i-1}^o}^{\lambda_{2i}^o} \mu^{j-1}\frac{d\mu}{\ell(\mu)} = T\times\Lambda_{j2} \quad (j=1,\ldots,n)$$

[38] Borg [1] proved this for $n=\infty$ and $m=2$.

for some integral $m'_1, \dots, m'_n$; the latter are unique if they exist. From the formula $\Lambda_{ij} = 2^{n+1-i} c^*_{i+j-n-1}$ of §16.10, it appears that the condition is equivalent to

$$\begin{aligned} \sum_{i=1}^{n} m'_i \int_{\lambda^o_{2i-1}}^{\lambda^o_{2i}} \mu^{j-1} \frac{d\mu}{\ell(\mu)} &= 0 \qquad \text{if } 1 \leqq j \leqq n-2 \\ &= 2T \qquad \text{if } j = n-1 \\ &= \sigma_1 T \quad\ \text{if } j = n. \end{aligned}$$

This is automatic for $n = 1$; in fact, $c_1 \mathsf{X}_1 q + \mathsf{X}_2 q = 0$ so that $\partial q/\partial t = \mathsf{X}_2 q$ is equivalent to translation at speed $-c_1$. For $n = 2$, the matter is less simple but still manageable.The period relations of Theorem 16.10.12 are

$$\begin{aligned} m_1 \int_{\lambda^o_1}^{\lambda^o_2} \frac{d\mu}{\ell(\mu)} + m_2 \int_{\lambda^o_3}^{\lambda^o_4} \frac{d\mu}{\ell(\mu)} &= 0, \\ m_1 \int_{\lambda^o_1}^{\lambda^o_2} \mu \frac{d\mu}{\ell(\mu)} + m_2 \int_{\lambda^o_3}^{\lambda^o_4} \mu \frac{d\mu}{\ell(\mu)} &= 1 \end{aligned}$$

and for the periodicity under the X_2 flow, it is required that

$$\begin{aligned} m'_1 \int_{\lambda^o_1}^{\lambda^o_2} \frac{d\mu}{\ell(\mu)} + m'_2 \int_{\lambda^o_3}^{\lambda^o_4} \frac{d\mu}{\ell(\mu)} &= 2T \\ m'_1 \int_{\lambda^o_1}^{\lambda^o_2} \mu \frac{d\mu}{\ell(\mu)} + m'_2 \int_{\lambda^o_3}^{\lambda^o_4} \mu \frac{d\mu}{\ell(\mu)} &= \sigma_1 T. \end{aligned}$$

These relations are inverted by

$$(1^\circ) \qquad \int_{\lambda^o_1}^{\lambda^o_2} \frac{d\mu}{\ell(\mu)} = m_2 \times 8 \text{ volume } C$$

$$(2^\circ) \qquad \int_{\lambda^o_3}^{\lambda^o_4} \frac{d\mu}{\ell(\mu)} = -m_1 \times 8 \text{ volume } C$$

$$(3^\circ) \qquad 2 \int_{\lambda^o_3}^{\lambda^o_4} \mu \frac{d\mu}{\ell(\mu)} - \sigma_1 \int_{\lambda^o_3}^{\lambda^o_4} \frac{d\mu}{\ell(\mu)} = \frac{m'_1}{T} \times 8 \text{ volume } C$$

$$(4^\circ) \qquad 2 \int_{\lambda^o_1}^{\lambda^o_2} \mu \frac{d\mu}{\ell(\mu)} - \sigma_1 \int_{\lambda^o_1}^{\lambda^o_2} \frac{d\mu}{\ell(\mu)} = -\frac{m'_2}{T} \times 8 \text{ volume } C$$

in which

$$8 \text{ volume } C = \int_{\lambda^o_1}^{\lambda^o_2} \frac{d\mu}{\ell(\mu)} \int_{\lambda^o_3}^{\lambda^o_4} \mu \frac{d\mu}{\ell(\mu)} - \int_{\lambda^o_1}^{\lambda^o_2} \mu \frac{d\mu}{\ell(\mu)} \int_{\lambda^o_3}^{\lambda^o_4} \frac{d\mu}{\ell(\mu)}.$$

Corollary 16.11.1. *T can only be an integral multiple of 4 vol C.*

Proof. Multiply (3°) by m_2, (4°) by m_1, and add to obtain

$$2 = \left(m_2 m'_1 - m_1 m'_2\right) \times T^{-1} \times \text{volume } C.$$

□

Corollary 16.11.2. *Write*

$$\begin{aligned} a_1 &= \int_{\lambda^o_1}^{\lambda^o_2} \ell^{-1}\, d\mu, & a_2 &= \int_{\lambda^o_3}^{\lambda^o_4} \ell^{-1}\, d\mu, \\ b_1 &= \int_{\lambda^o_1}^{\lambda^o_2} \mu \ell^{-1}\, d\mu, & b_2 &= \int_{\lambda^o_3}^{\lambda^o_4} \mu \ell^{-1}\, d\mu. \end{aligned}$$

Then the X_2 *flow is periodic on* M *if and only if*

$$\frac{2b_2 - \left(\lambda_0^o + \cdots + \lambda_4^o\right)a_2}{2b_1 - \left(\lambda_0^o + \cdots + \lambda_4^o\right)a_1}$$

is a rational number.

PROOF. The periodicity of the X_2 flow is equivalent to $3^\circ/4^\circ = -m_1'/m_2' =$ a rational number. The rest of the proof is plain. □

For $n \geqq 3$, the matter becomes quite complicated and should be investigated further as to the possible periods, etc.

16.12. Trace Formulas

The trace formulas referred to express power sums in $\lambda_0^o < \cdots < \lambda_{2n}^o$ and $\mu_1 < \cdots < \mu_n$ as polynomials in $q(0), q'(0), q''(0), \ldots$. They stem from Gelfand [11] and Dikii [4, 5]; see also Gelfand and Levitan [13], Buslaev and Faddeev [2], and Faddeev and Zaharov [33]. Let $\mu_1 < \mu_2 < \cdots$ denote the *full* series of roots of $y_2(1,\mu) = 0$. The first few trace formulas are

$$\lambda_0 + \sum_{i=1}^{\infty} \left(\lambda_{2i-1} + \lambda_{2i} - 2\mu_i\right) = q(0)$$

$$\lambda_0^2 + \sum_{i=1}^{\infty} \left(\lambda_{2i-1}^2 + \lambda_{2i}^2 - 2\mu_i^2\right) = q^2(0) - \frac{1}{2}q''(0)$$

$$\lambda_0^3 + \sum_{i=1}^{\infty} \left(\lambda_{2i-1}^3 + \lambda_{2i}^3 - 2\mu_i^3\right) = q^3(0) - \frac{15}{16}q'(0)^2 - \frac{3}{2}q(0)q''(0) + \frac{3}{16}q''''(0).$$

AMPLIFICATION 16.12.1. The sums are finite if $n < \infty$, as will be assumed to simplify the technical aspect of the derivation, though the formulas have a general validity. Notice that $\sum(\lambda_{2i-1}^p + \lambda_{2i}^p - 2\lambda_i^p) < \infty$ for every $p = 1, 2, 3, \ldots$ if and only if $\lambda_{2i} - \lambda_{2i-1}$ is rapidly decreasing as $i \uparrow \infty$. Hochstadt [16] proved that this condition is met if $q \in C_1^\infty$. Hochstadt [16] found and exploited a less explicit version of the first trace formula; see also Flaschka [8].

The trace formulas are closely allied to the "zeta" function

$$Z(s) = \lambda_0^{-s} + \sum_{i=1}^{\infty} \left(\lambda_{2i-1}^{-s} + \lambda_{2i}^{-s} - 2\mu_i^{-s}\right);$$

the latter is an integral function with values

$$Z(-p) = \lambda_0^p + \sum_{i=1}^{\infty} \left(\lambda_{2i-1}^p + \lambda_{2i}^p - 2\mu_i^p\right) \quad (p = 0, 1, 2, 3, \ldots).$$

The proof parallels the discussion of the classical zeta function; see Dikii [4, 5] and Flaschka [8]. The trace formula for $p = 1$ was already proved in Amplification 16.6.13 on the basis of Proposition 16.6.1; the formulas for $p = 2, \ldots, n$ can be derived in the same way.

AMPLIFICATION 16.12.2. The trace formulas lead to a whole series of amusing little facts. The point is that $q(0)$, $q^2(0) - 2q''(0)$, etc. reach their maxima [minima] on M when $\mu_i = \lambda_{2i}[\lambda_{2i-1}]$ $(i = 1, 2, \ldots)$; for example, the absolute maximum of $q(0)$ is $\lambda_0 + \sum_{i=1}^{\infty}(\lambda_{2i} - \lambda_{2i-1})$.

THEOREM 16.12.3. *The power sums* $\lambda_0^p + \sum_{i=1}^{\infty}(\lambda_{2i-1}^p + \lambda_{2i}^p - 2\mu_i^p) \equiv L_p$ *may be read off from the formal power series identity*

$$\sum_{p=1}^{\infty} \frac{(2\varepsilon)^p}{p} L_p = 2\log\left[\sum_{j=0}^{\infty} \frac{\partial H_j(0)}{\partial q}\varepsilon^j\right].$$

PROOF. Let $e^\infty(t,x,y)$ be the elementary solution of $\partial e/\partial t = -Qe$ in $0 \leqq x \leqq 1$ with $e=0$ at $x=0$ and $x=1$. Then

$$\vartheta^\infty(t) \equiv \sum_{i=1}^\infty e^{-\mu_i t} = \int_0^1 e^\infty(t,x,x)\,dx,$$

and $\partial\vartheta^\infty/\partial q = -te^\infty(t,x,x)$ as in §16.3, so that for any Hamiltonian flow $q^\bullet = \mathsf{X}q$,

$$\sum_{i=1}^\infty e^{-\mu_n t}\mu_n^\bullet = \int_0^1 e^\infty(t,x,x)\mathsf{X}q\,dx.$$

Now as $t \downarrow 0$, it is easy to see from the method of images that

$$e^\infty(t,x,x) = \frac{1}{\sqrt{4\pi t}} - \frac{e^{-x^2/t}}{\sqrt{4\pi t}} - \frac{e^{-(1-x)^2/t}}{\sqrt{4\pi t}} + o(1)$$

uniformly in $0 \leqq x \leqq 1$. This proves □

LEMMA 16.12.4.

$$\sum_{i=1}^\infty \mu_i^\bullet = -\lim_{t\downarrow 0}\left[\int_0^1 \frac{e^{-x^2/t}}{\sqrt{4\pi t}}\mathsf{X}q\,dx + \int_0^1 \frac{e^{-(1-x)^2/t}}{\sqrt{4\pi t}}\mathsf{X}q\,dx\right]$$
$$= -\frac{1}{2}\mathsf{X}q(0).$$

LEMMA 16.12.5. *For any $p = 1,2,3,\dots$, $L_p \equiv \lambda_0^p + \sum_{i=1}^\infty(\lambda_{2i-1}^p + \lambda_{2i}^p - 2\mu_i^p)$ is a universal polynomial in $q(0), q'(0), q''(0), \dots$, without constant term.*

PROOF. The proof is based upon the identity (see §16.3 for the definition of ϑ)

$$e^{-\lambda_0 t} + \sum_{i=1}^\infty \left(e^{-\lambda_{2i-1}t} + e^{-\lambda_{2i}t} - 2e^{-\mu_i t}\right) = \vartheta - 2\vartheta^\infty$$
$$= 2\int_0^1 e(t,x,x)\,dx - 2\int_0^1 e^\infty(t,x,x)\,dx,$$

e being the elementary solution of $\partial e/\partial t = -Qe$ for period 2, as in §16.3. A more elaborate use of the method of images shows that if $e^-[e^+]$ is the elementary solution of $\partial e/\partial t = -Qe$ on the whole line with q extended from $x > 0[x < 1]$ into $x \leqq 0[x \geqq 1]$ so as to be symmetrical about $x = 0[x = 1]$, then

$$e^\infty(t,x,x) = e(t,x,x) - e^-(t,x,x) - e^+(t,x,2-x) + o(1),$$

with an error $o(1)$ which is exponentially small for $t \downarrow 0$, uniformly in $0 \leqq x \leqq 1$, so that

$$\vartheta - 2\vartheta^\infty \sim 2\int_0^1 e^-(t,x,-x)\,dx + 2\int_0^1 e^+(t,x,2-x)\,dx,$$

with a similar error. The right-hand side may now be developed in the style of Theorem 16.3.3 using

$$q^-(x) = q(0) + q'(0)|x| + \frac{1}{2}q''(0)x^2 + \frac{1}{6}q'''(0)|x|^3 + \cdots$$

in estimating e^- and

$$q^+(x) = q(0) - q'(0)|x-1| + \frac{1}{2}q''(0)(x-1)^2 - \frac{1}{6}q'''(0)|x-1|^3 + \cdots$$

in estimating e^+. This leads to an expansion of $\vartheta - 2\vartheta^\infty$ in whole integral powers of t, valid for $t \downarrow 0$, with coefficients $(-1)^p L_p/p!$ of the stated kind. The point is that any monomial in $q(0), q'(0), q''(0), \dots$ attached to a half-integral power of t in the expansion of $\int_0^1 e^-(t,x,x)\,dx$, involves an odd number of differentiations and is balanced by a like monomial of opposite signature in the expansion of $\int_0^1 e^+(t,x,2-x)\,dx$, so that only *whole* integral powers of t survive. The rest of the proof will be plain. □

PROOF OF THEOREM 16.12.3. By Lemma 16.5.2 and the present Lemma 16.12.4,

$$-\frac{1}{2}\mathsf{X}_j q = \sum_{i=1}^{\infty} \mathsf{X}_j \mu_i = \sum_{k=1}^{j} \frac{\partial H_{k-1}}{\partial q} \sum_{i=1}^{\infty} (2\mu_i)^{i-k} \mathsf{X}_1 \mu_i,$$

first for $x=0$ and then, by translation, for any $0 \leqq x \leqq 1$, thinking of μ_i $(i \geqq 1)$ as functions of $0 \leqq x \leqq 1$ in the manner of Proposition 16.6.1. The same attitude is adopted towards

$$L_p = \lambda_0^p + \sum_{i=1}^{\infty} \left(\lambda_{2i-1}^p + \lambda_{2i}^p - 2\mu_i^p\right),$$

which is now differentiated with regard to x to obtain

$$L_p' = -2\sum_{i=1}^{\infty} p\mu_i^{p-1} \mathsf{X}_1 \mu_i.$$

The latter is put back into the formula for $-\frac{1}{2}\mathsf{X}_j q$ to confirm

$$\mathsf{X}_j q = \sum_{k=1}^{j} \frac{\partial H_{k-1}}{\partial q} \frac{2^{j-k}}{j-k+1} L_{j-k+1}',$$

from which is derived the formal power series

$$\begin{aligned} \left(\sum_{j=0}^{\infty} \frac{\partial H_j}{\partial q} \varepsilon^j\right)' &= \sum_{j=0}^{\infty} \mathsf{X}_j q \varepsilon^j \\ &= \frac{1}{2} \sum_{j=0}^{\infty} \frac{\partial H_j}{\partial q} \varepsilon^j \sum_{p=1}^{\infty} \frac{(2\varepsilon)^p}{p} L_p'. \end{aligned}$$

The formula of Theorem 16.12.3 follows by integration. Here, you could have an additive constant of integration, but it is not so: the coefficient of ε^p on either side is a universal polynomial in $q, q', q'', \ldots$ without constant term and these match as soon as their derivatives do so. The proof is finished. □

AMPLIFICATION 16.12.6. The recipe of Theorem 16.12.3 produces

$$\begin{aligned} L_1 &= \frac{\partial H_1}{\partial q}, \\ L_2 &= \frac{\partial H_2}{\partial q} - \frac{1}{2}\left(\frac{\partial H_1}{\partial q}\right)^2, \\ L_3 &= \frac{3}{4}\frac{\partial H_3}{\partial q} - \frac{3}{4}\frac{\partial H_2}{\partial q}\frac{\partial H_1}{\partial q} + \frac{1}{4}\left(\frac{\partial H_1}{\partial q}\right)^3 \end{aligned}$$

recapitulating the first three trace formulas. The next one is

$$\begin{aligned} L_4 =& \frac{1}{2}\frac{\partial H_4}{\partial q} - \frac{1}{4}\left(\frac{\partial H_2}{\partial q}\right)^2 - \frac{1}{2}\frac{\partial H_1}{\partial q}\frac{\partial H_3}{\partial q} \\ &+ \frac{1}{2}\left(\frac{\partial H_1}{\partial q}\right)^2 \frac{\partial H_2}{\partial q} - \frac{1}{8}\left(\frac{\partial H_1}{\partial q}\right)^4. \end{aligned}$$

COROLLARY 16.12.7. *The formula of Theorem 16.12.3 can be expressed as*

$$\sum_{j=0}^{\infty} \frac{\partial H_j}{\partial q} \varepsilon^j = \frac{\prod_{i=1}^{n} (1 - 2\varepsilon \mu_i^o)}{\sqrt{\prod_{i=0}^{2n} (1 - 2\varepsilon \lambda_i^o)}}.$$

COROLLARY 16.12.8.

$$\sum_{j=0}^{\infty} (2j-1) H_{j-1} \varepsilon^j = \frac{\prod_{i=1}^{n} (1 - 2\varepsilon \lambda_i')}{\sqrt{\prod_{i=0}^{2n} (1 - 2\varepsilon \lambda_i^o)}}.$$

PROOF. The proof of Corollary 16.12.7 consists in recognizing the left-hand side of the formula of Theorem 16.12.3 as

$$\sum_{p=1}^{\infty} \frac{(2\varepsilon)^p}{p} \left[(\lambda_0^o)^p + \sum_{i=1}^{n} \left((\lambda_{2i-1}^o)^p + (\lambda_{2i}^o)^p - 2(\mu_i^o)^p \right) \right]$$
$$= -2 \log \frac{\prod_{i=1}^{n} (1 - 2\varepsilon \mu_i^o)}{\sqrt{\prod_{i=0}^{2n} (1 - 2\varepsilon \lambda_i')}}.$$

Then Corollary 16.12.8 follows from Corollaries 16.10.7 and 16.3.7 by integration over a period $0 \leqq x < 1$. □

AMPLIFICATION 16.12.9. The formula of Corollary 16.12.8 makes explicit how the simple spectrum determines the Hamiltonian series, and vice versa by determination of branch points.

COROLLARY 16.12.10. *The development of Theorem 16.3.1 can be put into the alternative shape*

$$\vartheta = \sum_{i=0}^{\infty} e^{-\lambda_i t}$$
$$\sim \frac{\sqrt{-1}}{\pi} \left(\int_{\lambda_0^o}^{\lambda_1^o} + \int_{\lambda_2^o}^{\lambda_3^o} + \cdots + \int_{\lambda_{2n-2}^o}^{\lambda_{2n-1}^o} + \int_{\lambda_{2n}^o}^{\infty} \right) e^{-\lambda t} \prod_{i=1}^{n} (\lambda - \lambda_1') \frac{d\lambda}{\ell(\lambda)}.$$

PROOF. For $\lambda > 0$,

$$\frac{(-1)^m}{\sqrt{\pi}(2m-3)\cdots 3 \cdot 1} \int_0^{\infty} e^{-\lambda t} t^{m-1/2}\, dt = \frac{(-1)^m (2m-1)}{(2\lambda)^m \sqrt{\lambda}},$$

so

$$\int_0^{\infty} e^{-\lambda t} \frac{1}{\sqrt{\pi t}} \sum_{m=0}^{\infty} \frac{(-t)^m}{(2m-3)\cdots 3 \cdot 1} H_{m-1}\, dt$$
$$= \sum_{m=0}^{\infty} \frac{(-1)^n (2m-1) H_{m-1}}{(2\lambda)^m \sqrt{\lambda}} = \frac{\sqrt{-1} \prod_{i=1}^{n} (\lambda + \lambda_i')}{\sqrt{\prod_{i=0}^{2n} (\lambda + \lambda_i^o)}},$$

by Corollary 16.12.8 above. This Laplace transform is now inverted by means of the formula

$$\frac{1}{\sqrt{\pi t}} \sum_{m=0}^{\infty} \frac{(-t)^m}{(2m-3)\cdots 3 \cdot 1} H_{m-1} = \frac{1}{2\pi\sqrt{-1}} \int e^{\lambda t} \frac{\prod_{i=1}^{n} (\lambda + \lambda_i')}{\sqrt{\prod_{i=0}^{2n} (\lambda - \lambda_i^o)}}\, d\lambda,$$

the integral being extended over a vertical line in the complex plane to the right of $-\lambda_0^o$. Now deform that line into the curve of Fig. 16.6, cutting the plane along the line segments $[-\infty, -\lambda_{2n}^o]$ and $[-\lambda_{2i+1}^o, -\lambda_{2i}^o]$ $(i = 0, \ldots, n-1)$. A check on the signature of the radical shows that the intervening intervals $[-\lambda_{2i}^o, -\lambda_{2i-1}^o]$ $(i = 1, \ldots, n)$ contribute nothing, while the cuts produce the desired formula. □

FIGURE 16.6.

16.13. Gelfand Trace Formulas

Additional trace formulas in the style of Gelfand [11], Gelfand and Levitan [13], and Dikii [4, 5] may be obtained from the development

$$\sum_{i=1}^{\infty} e^{-\mu_i t} \sim \frac{1}{\sqrt{4\pi t}} \sum_{m=0}^{\infty} \frac{(-t)^m}{(2m-3)\cdots 3\cdot 1} H_{m-1} - \frac{1}{2}\sum_{p=0}^{\infty} \frac{(-t)^p}{p!} L_p,$$

implicit in the proof of Theorem 16.12.3. The trick is to replace[39]

$$\frac{1}{\sqrt{4\pi t}}$$

by

$$\frac{1}{2} + \sum_{l=1}^{\infty} e^{-\ell^2\pi^2 t}$$

for $t \downarrow 0$, as you may do up to an exponentially small error by the Jacobi identity for the theta function. Then you bring $\sum e^{-\ell^2\pi^2 t} \times \sum (-t)^m H_{m-1}/(2m-3)\cdots 3\cdot\ 1$ to the left and equate like powers of t, with the result that

$$2\sum_{\ell=1}^{\infty}\left[\mu_\ell^p - \sum_{i+j=p} \frac{p!\left(\ell^2\pi^2\right)^i H_{j-i}}{i!(2j-3)\cdots 3\cdot 1}\right] = \frac{p! H_{p-1}}{(2p-3)\cdots 3\cdot 1} - L_p.$$

A comparison with the trace formulas of §16.12 now yields similar information about the periodic spectrum:

$$\lambda_0^p + \sum_{\ell=1}^{\infty}\left[\lambda_{2\ell-1}^p + \lambda_{2\ell}^p - 2\sum_{i+j=p} \frac{p!\left(\ell^2\pi^2\right)^i H_{j-1}}{i!(2j-3)\cdots 3\cdot 1}\right] = \frac{p! H_{p-1}}{(2p-3)\cdots 3\cdot 1}.$$

The first two formulas are especially attractive:

$$\lambda_0 + \sum_{\ell=1}^{\infty}\left[\lambda_{2\ell-1} + \lambda_{2\ell} - 2\int_0^1 \left(q+\ell^2\pi^2\right) dx\right] = H_0,$$

$$\lambda_0^2 + \sum_{\ell=1}^{\infty}\left[\lambda_{2\ell-1}^2 + \lambda_{2\ell}^2 - 2\int_0^1 \left(q+\ell^2\pi^2\right)^2 dx\right] = 2H_1;$$

the third is more complicated:

$$\lambda_0^3 + \sum_{\ell=1}^{\infty}\left[\lambda_{2\ell-1}^3 + \lambda_{2\ell}^3 - 2\int_0^1 \left(q+\ell^2\pi^2\right)^3 dx - \int_0^1 (q')^2\, dx\right] = 2H_2.$$

An immediate consequence is a development of $\lambda_{2\ell-1}$ or $\lambda_{2\ell}$ as $\ell \uparrow \infty$ in the form[40]

$$(\ell\pi)^2 + k_0 + k_1(\ell\pi)^{-2} + k_2(\ell\pi)^{-4} + k_3(\ell\pi)^{-6} + \cdots$$

[39]This replacement embodies the Hamiltonian expansion for $q \equiv 0$.
[40]Compare Dikii [5].

with

$$k_0 = H_0, \qquad k_1 = H_1 - \frac{1}{2}H_0^2, \qquad 3k_2 = 2H_2 - 6H_0H_1 + 2H_0^3, \dots$$

by substitution of the expansion into

$$\lambda_{2\ell}^p = \sum_{i+j=p} \frac{p!\left(\ell^2\pi^2\right)^i H_{j-1}}{i!(2j-3)\cdots 3\cdot 1} + o(1) \quad (\ell \uparrow \infty);$$

naturally, the same development is valid for μ_ℓ $(\ell \uparrow \infty)$.

16.14. The Jacobian Variety

Take new coordinates $\mathfrak{x} = (x_1, \dots, x_n)$ on M defined by

$$\sum_{j=1}^n \mu_j^{i-1} \frac{d\mu_j}{\ell(\mu_j)} = dx_i \quad (i = 1, 2, \dots, n),$$

with $\mu_i^o = \mu_i$ $(i = 1, \dots, n)$ for ease of writing. This is nothing but $\mathfrak{x} = \Lambda t$; compare Theorem 16.10.5. Theorem 16.10.4 now states that a complete set of primitive periods ω of M relative to the new coordinate is given by

$$\omega = \oint d\mathfrak{x} = \oint \left(1, \mu, \dots, \mu^{n-1}\right)^\dagger \frac{d\mu}{\ell(\mu)}$$

in which the integral is taken about a loop enclosing $[\lambda_{2j-1}^o, \lambda_{2j}^o]$ for each $j = 1, \dots, n$ in turn. Recall from §16.8 the Riemann surface S of the irrationality $\ell(\mu) = \sqrt{-(\mu - \lambda_0^o)\dots(\mu - \lambda_{2n}^o)}$ with points $\mathfrak{p} = (\mu, \ell(\mu))$. This is a sphere with n handles constituting a 2-sheeted covering of the Riemann sphere R where μ sits, ramified over the $2n+2$ places ∞, $\lambda_0^o, \dots, \lambda_{2n}^o$. The periods of S comprise the n *real* periods ω of M plus n additional *pure imaginary* periods

$$\sqrt{-1}\omega' = \oint d\mathfrak{x} = \oint \left(1, \mu, \dots, \mu^{n-1}\right)^\dagger \frac{d\mu}{\ell(\mu)}$$

in which the integral is taken about a loop enclosing $[\lambda_{2j-2}^o, \lambda_{2j-1}^o]$ for each $j = 1, \dots, n$ in turn, and the map[41]

$$\left(\mathfrak{p}_1, \dots, \mathfrak{p}_n\right) \to \mathfrak{x} = \sum_{j=1}^n \int_{\mathfrak{o}_j}^{\mathfrak{p}_j} \left(1, \mu, \dots, \mu^{n-1}\right)^\dagger \frac{d\mu}{\ell(\mu)},$$

considered modulo the lattice $\mathfrak{L} = L + \sqrt{-1}L'$ of periods ω and $\sqrt{-1}\omega'$, is an application of the n-fold symmetric product S^n of the Riemann surface onto the associated complex "Jacobi variety" $\mathfrak{M} = C^n/\mathfrak{L}$ of which $M = R^n/L$ is the real part so to say.[42] The map $(\mathfrak{p}_1, \dots, \mathfrak{p}_n) \to \mathfrak{M}$ is 1:1 unless $\mathfrak{p}_i = (\mu, \ell(\mu))$ and $\mathfrak{p}_j = (\mu, -\ell(\mu))$ for some $i \neq j$ and $\ell(\mu) \neq 0$. The complex variety plays no further role until §16.15.

[41] $\mathfrak{o}_j$ $(j = 1, \dots, n)$ are fixed points of S, so chosen as to make $\mu(\mathfrak{o}_j) = \lambda_{2j-1}^o$, i.e., so that $(\mathfrak{o}_1, \dots, \mathfrak{o}_n)$ corresponds to the origin of M.

[42] See Siegel [31] for this and other information from classical function theory employed below. The set up of the map $S^n \to M$ is not quite the conventional one, and it is necessary to check that it is insensitive to permutations of $\mathfrak{p}_1, \dots, \mathfrak{p}_n$, for example, you must have $\int_{\mathfrak{o}_1}^{\mathfrak{p}_1} + \int_{\mathfrak{o}_2}^{\mathfrak{p}_2} = \int_{\mathfrak{o}_2}^{\mathfrak{p}_1} + \int_{\mathfrak{o}_1}^{\mathfrak{p}_2}$ modulo periods, and that is so because $\int_{\mathfrak{o}_1}^{\mathfrak{p}_1} - \int_{\mathfrak{o}_2}^{\mathfrak{p}_1} + \int_{\mathfrak{o}_2}^{\mathfrak{p}_2} - \int_{\mathfrak{o}_1}^{\mathfrak{p}_2} = 2\int_{\mathfrak{o}_1}^{\mathfrak{o}_2}$ and $\int_{\mathfrak{o}_1}^{\mathfrak{o}_2}$ is a half-period, by choice of origin.

Now take q with simple spectrum $\lambda_0^o, \ldots, \lambda_{2n}^o$ and auxiliary spectrum $\mu, \ldots, \mu_n$. Over each μ lie two points $\mathfrak{p} = (\mu, \pm\ell(\mu))$, or only one (ramified) point if $\ell(\mu) = 0$, and in the case of ambiguity when $\ell(\mu) \neq 0$, the signature of the radical is specified by $\mathsf{X}_1\mu = 2\ell(\mu)/m^\bullet(0,\mu)$; in short, q specifies a particular "divisor" $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$. This recipe can be inverted: X_1 is infinitesimal translation, so you have only to solve $\mathsf{X}_1\mu_i = 2\ell(\mu_i)/m^\bullet(0,\mu_i)$ for $i = 1, \ldots, n$ and to invoke the trace formula $q(0) = \sigma_1 - 2(\mu_1 + \ldots + \mu_n)$ of §16.12, updated by translation in the amount x, to produce the whole of $q(x)$ for $0 \leq x < 1$.

You may ask: What is the point? The object is to integrate the complicated flows $\partial q/\partial t = \mathsf{X}q$, but the corresponding motion of divisors is complicated, too, as per Lemma 16.5.2. But look: The map from the divisor to $\mathfrak{x}$ is also 1:1 since $\mu_i \neq \mu_j$ for $i \neq j$, and Theorem 16.10.5 says that $\mathfrak{x}$ responds to the flow $\partial q/\partial t = \mathsf{X}_j q$ as in $\mathfrak{x}(t) = \mathfrak{x}(0) + t \times$ (the jth column of Λ); in short, the composite map $q \to$ divisor $\to \mathfrak{x}$ converts the general flow $\partial q/\partial t = \mathsf{X}q$ into explicit straight-line motion at constant speed. The only question is now: How to go from $\mathfrak{x} \in M$ back to q? What is needed is Riemann's theta function.

The first trace formula of §16.12 now comes into its own: The formula states that

$$q(0) = \sigma_1 - 2\big(\mu_1 + \cdots + \mu_n\big),$$

exhibiting $q(0)$ as a symmetric function of $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$, i.e., *as a function on* S^n, and it is a fact of classical function theory that any such "Abelian" function may be viewed as a locally rational function on M and may be expressed by means of quotients of theta functions,[43] but more of that below. For the moment, let us cite only the startling fact that $q(0)$, *regarded as an n-fold periodic function on the real torus $M = R^n/L$, extends to a $2n$-fold periodic function on its complexification* $\mathfrak{M} = C^n/\mathfrak{L}$; this is the proper generalization of the discovery of Hochstadt [16] that q is a Weierstrassian elliptic function in the special case $n = 1$; see Example 16.5.9.

To explain the connection with theta functions, it is necessary to introduce some notation. Let P be the $n \times n$ matrix whose columns are the real periods $\omega = \oint d\mathfrak{x}$ taken about the intervals of instability $[\lambda_{2i-1}^o, \lambda_{2i}^o]$ $(i = 1, \ldots, n)$ and let $\sqrt{-1}P'$ be the $n \times n$ matrix whose columns are the imaginary periods $\sqrt{-1}\omega' = \oint d$ taken about the intervals of stability $[\lambda_{2i-2}^o, \lambda_{2i-1}^o]$ $(i = 1, \ldots, n)$. Then $\mathfrak{L} = L + \sqrt{-1}L'$ is the integral span of the columns of P and P', and by the classical period relations,[44] $Q = P'P^\dagger$ is symmetric $[Q^\dagger = Q]$, positive $[kQk > 0$ if $k \neq 0]$, and maps the dual[45] $L^\dagger$ of L onto L'. The theta function is now declared to be

$$\vartheta(\mathfrak{x}) = \sum e^{2\pi i k\cdot\mathfrak{x} - \pi k Q k^\dagger},$$

the sum being extended over the points k of the dual lattice $L^\dagger$, and it is easy to check that

$$\vartheta(\mathfrak{x} + \omega) = \vartheta(\mathfrak{x});$$

$$\vartheta(\mathfrak{x} + \omega') = e^{-2\pi i k\cdot\mathfrak{x} + \pi k Q k^\dagger}\vartheta(\mathfrak{x})$$

for $\omega \in L$, $\omega' \in \sqrt{-1}L'$, and $k = Q^{-1}\omega' \in L^\dagger$.

Now it is a fact of classical function theory[46] that if $f(\mathfrak{p})$ is a function of rational character on S with roots $\mathfrak{a}_1, \ldots, \mathfrak{a}_m$ and poles $\mathfrak{b}_1, \ldots, \mathfrak{b}_m$, equal in number of necessity, and if paths of

---

[43] See Siegel [31, 4.11], Farkas and Rauch [29], and for the older literature Göpel [15], Rosenhain [30] and Krazer and Wirtinger [21].

[44] Siegel [31, 2.3].

[45] $k \in L^\dagger$ if and only $k \cdot \omega$ is a whole number for every $\omega \in L$.

[46] Siegel [31, 4.11].

integration from $\mathfrak{a}_i$ to $\mathfrak{b}_i$ $(i = 1, \ldots, m)$ are chosen so as to make

$$\sum_{i=1}^{m} \int_{\mathfrak{a}_i}^{\mathfrak{b}_i} d\mathfrak{x} = 0,$$

this possibility being present also of necessity, then the product $f(\mathfrak{p}_1)\ldots f(\mathfrak{p}_n)$ can be expressed as a constant multiple of

$$\prod_{i=1}^{m} \frac{\vartheta(\mathfrak{x} - \mathrm{a}_i)}{\vartheta(\mathfrak{x} - \mathrm{b}_i)}$$

in which

$$\mathfrak{x} = \sum_{j=1}^{m} \int_{\mathfrak{o}_j}^{\mathfrak{p}_j} d\mathfrak{x}$$

as usual, and

$$\mathrm{a}_i = \int_{\mathfrak{o}}^{\mathfrak{a}_i} d\mathfrak{x}, \quad \mathrm{b}_i = \int_{\mathfrak{o}}^{\mathfrak{b}_i} d\mathfrak{x} \quad (i = 1, \ldots, m),$$

$\mathfrak{o}$ being the pole of $\mu(\mathfrak{p})$. This result is now applied to the function $f(\mathfrak{p}) = 1 - \mu(\mathfrak{p})/\lambda$ for $\lambda \neq 0$: on S, it has simple roots at $\mathfrak{p}_\pm = (\lambda, \pm\ell(\lambda))$, or else a double root if $\ell(\lambda) = 0$, and a double pole at $\mathfrak{o}$, so $f(\mathfrak{p}_1)\ldots f(\mathfrak{p}_n) = \lambda^{-n} m(0, \lambda)$ can be expressed as

$$\frac{\vartheta(\mathfrak{x} - \mathfrak{x}(\mathfrak{p}_+))\vartheta(\mathfrak{x} - \mathfrak{x}(\mathfrak{p}_-))}{\vartheta^2(\mathfrak{x})},$$

in which $\mathfrak{x}(\mathfrak{p}) = \int_0^{\mathfrak{p}} d\mathfrak{x}$ and no constant multiplier is present since $\mathfrak{p}_\pm$ tends to $\mathfrak{o}$ as λ tends to ∞. But also

$$\frac{1}{2\pi\sqrt{-1}} \oint \lambda^{-n} m d\lambda = -(\mu_1 + \ldots + \mu_u),$$

the integral being taken about a big circle, so

$$q(0) = \sigma_1 - 2 \times \frac{1}{2\pi\sqrt{-1}} \oint \frac{\vartheta(\mathfrak{x} - \mathfrak{x}(\mathfrak{p}_+))\vartheta(\mathfrak{x} - \mathfrak{x}(\mathfrak{p}_-))}{\vartheta^2(\mathfrak{x})} d\lambda$$

into which the straight-line motion of x described above may be substituted to express the general flow $\partial q/\partial t = X_q$; in particular, the **Technical Condition** of §16.4 is now justified.

Example 16.14.1. The formula is classical if $n = 1$[47]: Then

$$q(x) = \sigma_1/3 + 2p\big(x + \sqrt{-1}\omega'/2\big),$$

p being the Weierstrassian elliptic function with primitive periods $1/m_1$ and $\sqrt{-1}\omega'$, and the formula is equivalent to any one of the classical identities[48]

$$p = e_1 + \left[\frac{\vartheta_1'(0)\vartheta_2}{\vartheta_1\vartheta_2(0)}\right]^2 = e_2 + \left[\frac{\vartheta_1'(0)\vartheta_3}{\vartheta_1\vartheta_3(0)}\right]^2 = e_3 + \left[\frac{\vartheta_1'(0)\vartheta_4}{\vartheta_1\vartheta_4(0)}\right]^2.$$

[47] See Amplification 16.5.5.
[48] ϑ_i $(i = 1, 2, 3, 4)$ are the customary functions of Jacobi.

Note added in proof, November 2008. The present §16.14 replaces the original unsatisfactory §16.14 of McKean-van Moerbeke; it employs only the methods developed there. Its-Matveev [Hill's operator with a finite number of lacunae. *Funkt. Anal. Pril.* **9** (1975) 69–70] had found a much better way to express $q(0)$ in terms of $\mathfrak{x}$ unknown to us then, based upon a formula of H. F. Baker [*Abel's Theorem and the Allied Theory, including the Theory of the Theta Functions*, Cambridge U. Press, Cambridge, 1897], namely

$$q(0) = -2\mathsf{X}_1^2[\log\vartheta(\mathfrak{x})] + \sum_0^{2n}\lambda_i^o - 2\sum_1^n \lambda_i',$$

into which the straight-line motion of $\mathfrak{x}$ may be substituted as above. That's how it should be done.

16.15. Elliptic Functions

Think of q as a point of the real torus M sitting inside the complex torus $\mathfrak{M}$ and look at the map $\sigma^{1/2}$: $q(x) \to -q(\sqrt{-1}x)$; the latter is a root of the involution $\sigma : q(x) \to q(-x)$, as the notation suggests.

Lemma 16.15.1. *$\mathsf{X}_m\sigma^{1/2}q = (\sqrt{-1})^{2m+3}\sigma^{1/2}\mathsf{X}_m q$ $(m = 1, 2, 3, \ldots)$.*

Proof. $\mathsf{X}_m q$ is an isobaric polynomial in q, q', q'' etc., of degree $m + 1/2$, counting q as of degree 1 and differentiation as augmenting the degree by $1/2$, as in §16.3, and for any monomial $\mathsf{X}q$ of $\mathsf{X}_m q$ involving i factors q and j differentiations, you have

$$\begin{aligned}\mathsf{X}\sigma^{1/2}q(x) &= (-1)^i\big(\sqrt{-1}\big)^j\mathsf{X}q\big(\sqrt{-1}x\big)\\ &= -\big(\sqrt{-1}\big)^{2i+j}\sigma^{1/2}\mathsf{X}q(x)\end{aligned}$$

with $2i + j = 2m + 1$. □

Lemma 16.15.2. *$q^* = \sigma^{1/2}q$ is a solution of $c_1^*\mathsf{X}_1 + \cdots + c_n^*\mathsf{X}_n + \mathsf{X}_{n+1} : q \to 0$, in which $c_1^*, \ldots, c_n^*$ are the symmetric polynomials of §16.7 formed with $\lambda_0^* = -\lambda_{2n}^o, \ldots, \lambda_{2n}^* = -\lambda_0^o$ in place of $\lambda_0^o, \ldots, \lambda_{2n}^o$, and n is minimal in this regard, i.e., q^* cannot satisfy any equation $c_1^{**}\mathsf{X}_1 + \cdots + \mathsf{X}_{m+1} : q \to 0$ with $m < n$.*

Proof. Use Lemma 16.15.1. □

Lemma 16.15.3. *$q^*(x)$ is real in the vicinity of $x = 0$ if and only if it corresponds to a half-period $\omega/2$ of M.*

Proof. $q(x)$ is real analytic in $0 \leqq x < 1$, so $q^*(x)$ is real in the vicinity of $x = 0$ if and only if $q(x) = q(\sigma x)$ first locally and then in a whole period $0 \leqq x < 1$. Thus, q is a fixed point of σ, and the latter are known to be the half-periods of M; see Proposition 16.5.10. □

Now fix the half-period $\omega/2$ and[49] think of $q^*(x) = -\sigma_1 + 2\mu_1(\sqrt{-1}x) + \cdots + 2\mu_n(\sqrt{-1}x)$ as a function on the line $\sqrt{-1}x + \omega/2$ on the complex torus $\mathfrak{M}$.

Lemma 16.15.4. *$q^*(x)$ stays real and smooth for $-\infty < x < \infty$ only for those 2^{n-1} half-periods with $\mu_n = \lambda_{2n-1}^o$.*

[49]As before, $\mu_i = \mu_i^o$ $(i = 1, \ldots, n)$ for ease of writing.

PROOF. Under the motion $\dot{q} = \sqrt{-1}\mathsf{X}_1 q$, the power sums of $\mu_1, \ldots, \mu_n$ remain real in the vicinity of $x = 0$ by Lemma 16.15.3 and the trace formulas of §16.12. Therefore $\mu_1, \ldots, \mu_n$ remain real, and from

$$\mu_i^\bullet = \sqrt{-1}\mathsf{X}_1\mu_i = \frac{2\sqrt{-1}\ell(\mu_i)}{\prod_{j\neq i}(\mu_i - \mu_j)} \quad (i = 1, \ldots, n),$$

it appears that the irrationalities $\ell(\mu_i)$ $(i = 1, \ldots, n)$ are pure imaginary, which is to say that each μ_i, starting from λ^o_{2i-1} or λ^o_{2i}, moves within the adjacent interval of stability $[\lambda^o_{2i-2}, \lambda^o_{2i-1}]$ or[50] $[\lambda^o_{2i}, \lambda^o_{2i+1}]$. Now there are several possibilities. If $\mu_n(0) = \lambda^o_{2n}$, then μ_n moves steadily to the right with x and blows up at a finite parameter $x = x_\infty$ because $\mu_n^\bullet \sim 2\mu_n^{3/2}$ far out, $\mu_1, \ldots, \mu_{n-1}$ being bounded. Therefore, to keep q^* smooth, you must take $\mu_n = \lambda^o_{2n-1}$. At the opposite extreme, if $\mu_i(0) = \lambda^o_{2i-1}$ $(i = 1, \ldots, n)$, then μ_i moves inside $[\lambda^o_{2i-2}, \lambda^o_{2i-1}]$, turning around in a smooth manner whenever it comes to λ^o_{2i-2} or returns to λ^o_{2i-1}. The remaining possibility is that $\mu_i = \lambda^o_{2i}$ and $\mu_{i+1} = \lambda^o_{2i+1}$ for one or more values of $i < n$. Let $i = 1$ for example. Then μ_1 moves to the right and μ_2 to the left, colliding at a finite parameter $x = x_{12}$ with $\mu_1^\bullet = +\infty$ and $\mu_2^\bullet = -\infty$ at the collision. This looks bad, but actually $\mu_1^\bullet + \mu_2^\bullet$ remains well behaved through the collision, so that $\mu_1 + \mu_2$ is smooth, μ_1 traveling up to λ^o_3 and back to λ^o_2 and μ_2 down to λ^o_2 and back to λ^o_3 with smooth turns at these places. Thus, each half-period with $\mu_n = \lambda^o_{2n-1}$ produces a smooth q^*, and only these will serve. □

The proof of the next theorem is now self-evident.

THEOREM 16.15.5. *In order that $\lambda^*_0 = -\lambda^o_{2n}, \ldots, \lambda^*_{2n} = -\lambda^o_0$ be the simple spectrum of a potential $q^* \in C^\infty$ of period T, it is necessary and sufficient that*

$$\sum_{j=0}^{n-1} m^*_{n-j} \int_{\lambda^o_{2j}}^{\lambda^o_{2j+1}} \mu^{i-1}\frac{d\mu}{\ell(\mu)} = \begin{matrix} 0 & (i < n) \\ \sqrt{-1}T & (i = n) \end{matrix}$$

for some necessarily unique integers $0 < m^*_1 < \cdots < m^*_n$; in such a case, $q^*(x)$ can be taken as $-q(\sqrt{-1}x + \omega/2)$ with any of the permissible half-periods for which $\mu_n = \lambda^o_{2n-1}$, and the pure imaginary translates of the latter comprise a complete list of such q^*.

AMPLIFICATION 16.15.6. Let $q(x)$ have a complex period T as a function of x. Then

$$\sum_{j=1}^{n} m'_j \int_{\lambda_{2j-1}}^{\lambda^o_{2j}} \mu^{n-i}\frac{d\mu}{\ell(\mu)} + m''_j \int_{\lambda^o_{2j-2}}^{\lambda^o_{2j-2}} \mu^{n-i}\frac{d\mu}{\ell(\mu)} = \begin{matrix} 0 & (i < n) \\ T & (i = n) \end{matrix}$$

with suitable integers m'_j, m''_j $(j = 1, \ldots, n)$, and it is clear that the real and imaginary parts of T are periods, separately. Therefore, such potentials are elliptic functions with a pure imaginary primitive period, the same for every point of M; these are the potentials of Theorem 16.15.5.

EXAMPLE 16.15.7. The condition is automatically satisfied for $m = 1$, the sole permissible half-period being $1/2\ m_1 + (\sqrt{-1}/2)\omega'$. It may be remarked that in the problem of the simple pendulum, the map $cn(t) \to -cn(\sqrt{-1}t)$ is well known,[51] reflecting a reversal of the gravitational field; see Copson [3, p.417].

[50] $\lambda^o_{2n\cdot 1} = \infty$.
[51] *cn* is the customary Jacobi function.

The $2n$ conditions placed upon $\lambda_0^o < \cdots < \lambda_{2n}^o$ by double periodicity in x indicate that, for fixed $m_1^*, \ldots, m_n^*$, and T, the spectrum admits only one degree of freedom, *to wit*, the freedom of translation $\lambda_i^o \to \lambda_i^o + k$ $(i = 0, \ldots, 2n)$; in particular, if you also fix $\lambda_0 = 0$, say, it seems likely that the simple spectrum is unique for each $n \geqq 0$. The remark of Lax [22], that the Hill's operator[52] $Q = -d^2/dx^2 + (2kK)^2 \cdot n(n+1) \operatorname{sin}^2(2Kx, k)$ has $2n+1$ simple eigenvalues, should enable one to determine these spectra effectively.

16.16. Geometry of Simple Spectra

The purpose of this article is to explain the geometry of the space Λ_n of simple spectra $\lambda_0^o < \cdots < \lambda_{2n}^o$. It turns out that such a spectrum has only $n+1$ degrees of freedom and that $H_0, \ldots, H_n$ is a local coordinate on Λ_n. From this viewpoint, the problem is to find nice moduli for the tori M; the geometrical problem of how C_1^∞ is fibered by these tori (or by their ∞-dimensional counterparts for $n = \infty$) is still obscure.

THEOREM 16.16.1. *Λ_n is an open $(n+1)$-dimensional manifold in R^{2n+1}, and $(H_0, \ldots, H_n)$ is a local coordinate upon it.*

PROOF. Let f be a simple eigenfunction of Q of unit length with eigenvalue $\lambda = \lambda_i^o$ $(i = 0, \ldots, 2n)$ and let $q^\bullet$ be an infinitely small variation of $q \in C_1^\infty$. Then $Qf^\bullet + q^\bullet f = \lambda f^\bullet + \lambda^\bullet f$ and $(f, f^\bullet) = 0$, so

$$\int q^\bullet f^2 = (f, Qf^\bullet) + \int q^\bullet f^2 = \lambda^\bullet \int f^2 = \lambda^\bullet,$$

i.e., $\partial\lambda/\partial q = f^2$ as noted before. The fact that $\dim \Lambda_n = n+1$ now follows from Proposition 16.6.12: exactly $n+1$ of the functions $\partial\lambda_i^o/\partial q = (f_i^o)^2$ $(i = 0, \ldots, 2n)$ are independent. As to the local coordinate, Proposition 16.6.2 implies that

$$2^{-n} \sum_{1 \leqq j \leqq i \leqq n+1} c_i \frac{\partial H_{j-1}}{\partial q} \left(2\lambda_k^o\right)^{i-j} = 2 \prod_{\ell=1}^{n} \left(\lambda_k^o - \lambda_\ell'\right) \frac{\partial \lambda_k^o}{\partial q},$$

or what is the same,

$$2^{-n} \sum_{1 \leqq j \leqq i \leqq n+1} c_i \left(2\lambda_k^o\right)^{i-j} dH_{j-1} = 2 \prod_{\ell=1}^{n} \left(\lambda_k^o - \lambda_\ell'\right) d\lambda_k^o,$$

and for any choice of $n+1$ values of $j, k = 0, \ldots, 2n$,

$$\det_{jk} \sum_{i \geqq j} c_i \left(2\lambda_k^o\right)^{i-j} = 2^{n(n+1)/2} \prod_{k>j} \left(\lambda_k^o - \lambda_j^o\right) \neq 0,$$

so that $(\lambda_0^o, \ldots, \lambda_{2n}^o) \to (H_0, \ldots, H_{n+1})$ is a local diffeomorphism. □

EXAMPLE 16.16.2. The tangent space of M may be computed directly from the period relations of Theorem 16.10.12. For example, if $n = 1$, then there is only the one period relation

$$\int_{\lambda_1^o}^{\lambda_2^o} \frac{d\mu}{\ell(\mu)} = \int_0^1 \frac{1}{\sqrt{\left(\lambda_1^o - \lambda_0^o\right) + \left(\lambda_2^o - \lambda_1^o\right)x}} \frac{dx}{\sqrt{x(1-x)}} = \frac{1}{m_1},$$

and the tangent space to $\Lambda_1 \subset R^3$ at the point $(\lambda_0^o, \lambda_1^o, \lambda_2^o)$ is determined by

$$0 = \int_0^1 \frac{-d\lambda_0^o + (1-x)d\lambda_1^o + x d\lambda_2^o}{\sqrt{\left(\lambda_1^o - \lambda_0^o\right) + \left(\lambda_2^o - \lambda_1^o\right)x}} \frac{dx}{\sqrt{x(1-x)}}.$$

[52] K is the complete elliptic integral $\int_0^{k/2} (l - k^2 \sin^2\theta)^{-1/2}\, d\theta$.

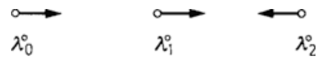

FIGURE 16.7.

For $n = 0$, the potential is constant and the map of $\lambda_0^o \to H_0 = \lambda_0^o$ is onto the whole line which is now designated by D_0. For $n = 1$, the situation is already more complicated. To begin with $H_1 = \frac{1}{2}\int_0^1 q^2 \geqq \frac{1}{2}(\int_0^1 q)^2 = \frac{1}{2}H_0^2$, and the equality is excluded, so the map of $(\lambda_0^o, \lambda_1^o, \lambda_2^o) \to (H_0, H_1)$ cannot cover more than the figure

$$D_1 = \left(h_1 > 2h_0^2\right) \subset R^2.$$

THEOREM 16.16.3. *Λ_1 is an unramified ∞-sheeted covering of D_1, the sheets being distinguished by the value*

$$\int_{\lambda_1^o}^{\lambda_2^o} \frac{d\mu}{\sqrt{\left(\mu - \lambda_0^o\right)\left(\mu - \lambda_1^o\right)\left(\lambda_2^o - \mu\right)}} = \frac{1}{m}$$

of the primitive real period of q; especially, each sheet maps 1:1 onto D_1.

PROOF. Fix $m = 1, 2, 3, \ldots$ and note that the period relation determines λ_0^o as a smooth function of λ_1^o and λ_2^o. Let $H_0 = h_0$ be fixed and move H_1. Then the differential equations of the proof of Theorem 16.16.1 reduce to

$$dH_1 = 4\left(\lambda_i^o - \lambda_1'\right)d\lambda_1^o \quad (i = 0, 1, 2),$$

and for $H_1^\bullet = -1$, it appears from Fig. 16.7 that $[\lambda_1^o, \lambda_2^o]$ collapses $\lambda_0^o\lambda_1^o\lambda_2^o$ into a single point at a finite parameter T. At that moment, $H_0 = \lambda_0^o$, the period relation reduces to

$$\frac{\pi}{\sqrt{\lambda_1^o - \lambda_0^o}} = \frac{1}{m},$$

or what is the same,

$$\lambda_1^o = \lambda_0^o + m^2\pi^2,$$

and H_1 is at its minimum $\frac{1}{2}h_0^2$. Now reverse the flow by taking $H_1^\bullet = +1$. Then the motion of Fig. 16.7 is reversed and continues forever while $H_1 \uparrow \infty$. This proves that the m-th sheet of Λ_1 maps *onto* D_1 as soon as you notice that *any* value of H_0 can be achieved by adding a constant to q. To see that the map is 1:1, it suffices to prove that there is exactly one trajectory of $4(\lambda_i^o - \lambda_1')d\lambda_i^o = dt$ $(i = 0, 1, 2)$ with

$$\frac{1}{m} = \int_{\lambda_1^o}^{\lambda_2^o} \frac{d\mu}{\sqrt{\left(\mu - \lambda_0^o\right)\left(\mu - \lambda_1^o\right)\left(\lambda_2^o - \mu\right)}},$$

$$\lambda_1' = m\int_{\lambda_1^o}^{\lambda_2^o} \frac{\mu d\mu}{\sqrt{\left(\mu - \lambda_0^o\right)\left(\mu - \lambda_1^o\right)\left(\lambda_2^o - \mu\right)}}$$

issuing from each point $(\lambda_0^o, \lambda_1^o, \lambda_2^o)$ with $\lambda_1^o = \lambda_2^o = \lambda_0^o + m^2\pi^2$. Now the problem is two-dimensional, λ_0^o being eliminated by the period relation; it is also invariant under translation $\lambda_i^o \to \lambda_i^o + k$ $(i = 0, 1, 2)$, and trajectories cannot cross while $\lambda_1^o < \lambda_2^o$. The statement is obvious from that, but it will be preferable to make a different proof. To do that, modify the time scale in the region $\lambda_1^o < \lambda_2^o$ by the factor $4\sqrt{(\lambda_2^o - \lambda_1')(\lambda_1' - \lambda_1^o)}$. The differential equations become

$$d\lambda_1^o = -\sqrt{\frac{\lambda_2^o\lambda_1'}{\lambda_1'\lambda_1^o}}\,dt,$$

$$d\lambda_2^o = +\sqrt{\frac{\lambda_1'\lambda_1^o}{\lambda_2^o\lambda_1^o}}\,dt,$$

and it is claimed that this system is smooth up to and including $\lambda_1^o = \lambda_2^o$; in fact, if $\lambda_1^o < \lambda_2^o$, then

$$\lambda_1' = m\int_0^1 \frac{\lambda_1^o + \left(\lambda_2^o - \lambda_1^o\right)x}{\sqrt{\lambda_1^o - \lambda_0^o + \left(\lambda_2^o - \lambda_1^o\right)x}}\frac{dx}{\sqrt{x(1-x)}},$$

so that

$$\frac{\lambda_2^o - \lambda_1'}{\lambda_1' - \lambda_1^o} = \frac{\int_0^1 \frac{1-x}{\sqrt{\lambda_1^o-\lambda_0^o+\left(\lambda_2^o-\lambda_1^o\right)x}}\frac{dx}{\sqrt{x(1-x)}}}{\int_0^1 \frac{x}{\sqrt{\lambda_1^o-\lambda_0^o+\left(\lambda_2^o-\lambda_1^o\right)x}}\frac{dx}{\sqrt{x(1-x)}}}$$

which makes everything plain. The proof is finished. □

Amplification 16.16.4. The sheets of Λ_1 are depicted in Fig. 16.8: for a general point $\lambda_0^o < \lambda_1^o < \lambda_2^o$ thereof, $H_1 - 1/2H_0^2$ is the time it takes to drive the initial spectrum $\lambda_0^o =$

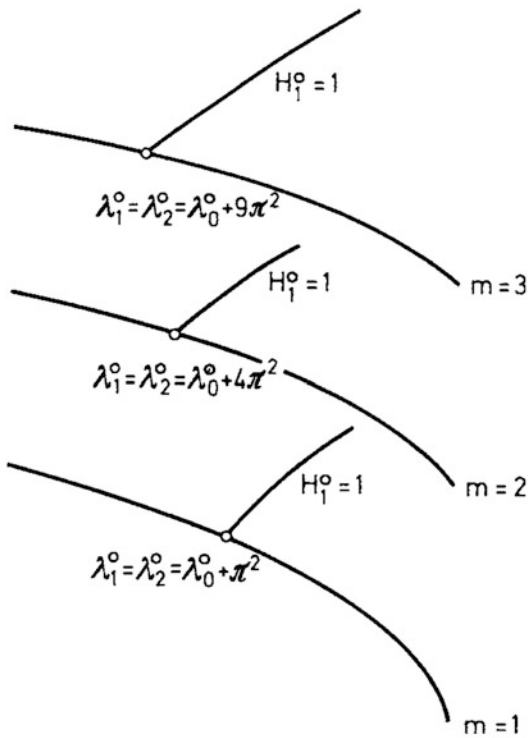

Figure 16.8.

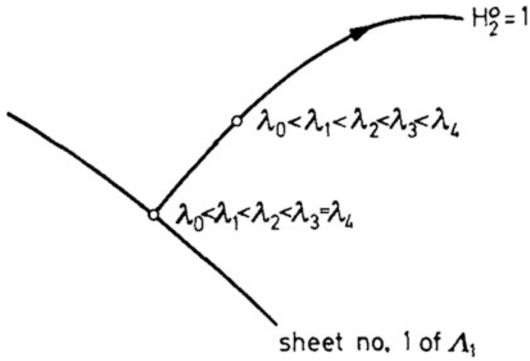

FIGURE 16.9.

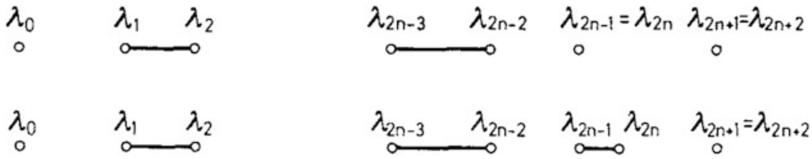

FIGURE 16.10.

$H_0, \lambda_1^o = \lambda_2^o = H_0 + m^2\pi^2$ to that position by means of $4(\lambda_i^o - \lambda_1')d\lambda_i^o = dt$ $(i = 0, 1, 2)$. An incidental bonus is the realization that Λ_0 can be viewed as the common boundary of all the sheets of Λ_1.

THEOREM 16.16.5. *Fix $(\lambda_0^o, \ldots, \lambda_{2n}^o) \in \Lambda_n$. Then the associated double eigenvalues can be smoothly opened up, one at a time, into bona fide pairs of simple eigenvalues, keeping $H_i = h_i$ $(i \leqq n)$ fixed. Moreover, the opening can be done in only one way, i.e., the opened spectra form a smooth curve in Λ_{n+1} issuing from the original spectrum in Λ_n, and H_{n+1} is an adequate coordinate upon it; see Fig. 16.9 for the case $n = 1$, $m_1 = 1$, $m_2 = 2$.*

PROOF. The idea is already contained in the proof of Theorem 16.16.3. The opening depicted in Fig. 16.10 is typical. To achieve it, you use the differential equations

$$2^{n+1} \prod_{k=1}^{n} (\lambda_i - \lambda_n')\lambda_i^\bullet = H_{n+1}^\bullet = +1 \quad (i = 0, \ldots, 2n)$$

under which motion $H_i = h_i$ $(i \leqq n)$ stay fixed. The claim is that this system has a unique solution in Λ_n issuing from the first spectrum of Fig. 16.10. To begin with, by the substitution $\mu_n = \lambda_{2n-1} + (\lambda_{2n} - \lambda_{2n-1})x$,

$$\begin{aligned}\prod_{k=1}^{n} (\lambda - \lambda_k') &= \frac{\int \prod_{k=1}^{n} (\lambda - \mu_k)\, d \text{ volume}}{\text{volume of } C} \\ &= \int_{\lambda_1}^{\lambda_2} \cdots \int_{\lambda_{2n-3}}^{\lambda_{2n-2}} \frac{\prod_{k=1}^{n-1} (\lambda - \mu_k) \prod_{i<j<n} (\mu_j - \mu_i)\, d^{n-1}\mu}{\ell(\mu_1) \ldots \ell(\mu_{n-1})} \\ &\quad \times \int_0^1 \frac{[\lambda - \lambda_{2n-1} + (\lambda_{2n} - \lambda_{2n-1})x] \prod_{i<n} (\mu_n - \mu_i)}{\sqrt{(\mu_n - \lambda_0) \ldots (\mu_n - \lambda_{2n-2})}} \frac{dx}{\sqrt{x(1-x)}}\end{aligned}$$

$$\times \left(\int_{\lambda_1}^{\lambda_2} \cdots \int_{\lambda_{2n-3}}^{\lambda_{2n-2}} \frac{\prod_{i<j<n} (\mu_j - \mu_i)\, d^{n-1}\mu}{\ell(\mu_1) \dots \ell(\mu_{n-1})} \right.$$

$$\left. \times \int_0^1 \frac{\prod_{i<n} (\mu_n - \mu_i)}{\sqrt{(\mu_n - \lambda_0) \cdots (\mu_n - \lambda_{2n-2})}} \frac{dx}{\sqrt{x(1-x)}} \right)^{-1}$$

Therefore, in the vicinity of $\lambda_0 < \cdots < \lambda_{2n-2} < \lambda_{2n-1} = \lambda_{2n}$ and for $\lambda < \lambda_{2n-1}$ or λ_{2n}, $\prod_{k=1}^{n}(\lambda - \lambda'_k)$ can be expressed as the product of $\lambda_{2n} - \lambda_{2n-1}$ and a smooth nonvanishing function of $\lambda_0, \dots, \lambda_{2n}$, and after modification of the time scale by the factor $\lambda_{2n} - \lambda_{2n-1}$, the differential equations take the form

$$\lambda_i^{\bullet} = f_i(\lambda_0, \dots, \lambda_{2n}) \quad (i = 0, \dots, 2n)$$

with smooth functions f_i such that, for $\lambda_{2n-1} = \lambda_{2n}$, $f_i = 0$ $(i \leqq 2n-2)$, but $f_{2n-1} < 0$ and $f_{2n} > 0$. This gets the solution off the ground Λ_{n-1} and moving out into Λ_n with H_{n+1} measuring the elapsed time. □

Recall Lemma 16.3.9 and for fixed h_i $(i = 0, 1, \dots, n)$, let $h_k^o(1/m)$ denote the minimum of H_k in the subclass of $C^\infty_{1/m}$ with $H_i = h_i$ $(i < k)$, assuming this class to be non-void. Obviously, a Hamiltonian series $H_i = h_i$ $(i \leqq n)$ arising from Λ_n must fall inside the region $D_n \subset R^{n+1}$ where $h_1 > h_1^o(1), \dots, h_n > h_n^o(1)$. For $n = 2$, more can be said, but first a couple of lemmas.

Lemma 16.16.6 ([53]). *$h_2^o(1) < h_2^o(1/2) < h_2^o(1/3) < \cdots \uparrow \infty$.*

Proof. To study $h_2^o(1/m)$, let $q = p(mx)$ with $p \in C_1^\infty$. Then

$$H_0 = \int_0^1 q\, dx = \int_0^1 p\, dx,$$

$$H_1 = \frac{1}{2}\int_0^1 q^2\, dx = \frac{1}{2}\int_0^1 p^2\, dx,$$

$$H_2 = \int_0^1 \left[\frac{1}{2}q^3 + \frac{1}{4}(q')^2\right] dx = \int_0^1 \left[\frac{1}{2}p^3 + \frac{m^2}{4}(p')^2\right] dx,$$

and you have to minimize the latter for fixed $H_0 = h_0$ and $H_1 = h_1 > \frac{1}{2}h_0^2 = h_1^o(1)$. Clearly, $h_2^o(1/m)$ is steadily increasing with m. Now $\|p\|_\infty^2 \leqq 4h_1 + 2\|p'\|_2^2$, so that

$$H_2 \geqq \frac{m^2}{4}\|p'\|_2^2 - \frac{1}{2}h_1\sqrt{4h_1} + 2\|p'\|_2^2$$

and from $2h_1 - h_0^2 = c > 0$, you have

$$c = \int_0^1 \left| \int_0^1 p(x+y)\, dy - p(x) \right|^2 dx \leqq \|p'\|_2^2,$$

whence

$$h_2^o(1/m) \geqq am^2 - b$$

for $m \uparrow \infty$. □

Lemma 16.16.7. *Let $H_0 = h_0$ and $H_1 = h_1 > h_1^o(1)$ be fixed and let $\lambda_0^o < \lambda_1^o < \lambda_2^o$ be the corresponding simple spectrum on the m-th sheet of Λ_1. Then $H_2 = h_2^o(1/m)$, and conversely: the Hamiltonian series $H_0, H_1, H_2, \dots$ begins with $H_0 = h_0$, $H_1 = h_1$ and $H_2 = h_2^o(1/m)$ only if $n = 1$, $m_1 = m$, and you have the simple spectrum $\lambda_0^o < \lambda_1^o < \lambda_2^o$.*

[53] For $n = 1$, this kind of thing does not happen: $h_1^o(1/m)$ is independent of $m = 1, 2, 3, \dots$.

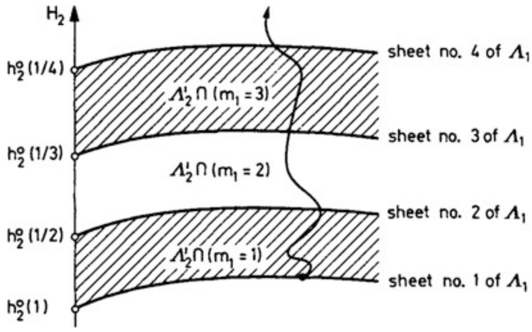

FIGURE 16.11.

PROOF. The stated simple spectrum comes from a potential with primitive period $1/m$, and if H_2 exceeded $h_2^o(1/m)$, you could minimize it by means of Lemma 16.3.9 to obtain a nonconstant potential $[h_1 > \frac{1}{2}h_0^2]$ with the *same* primitive period. But the simple spectrum of the latter lies on the same sheet of Λ_1 and is determined by h_0 and h_1 by Theorem 16.16.3; see also Amplification 16.16.4. The rest of the proof runs along the same lines. □

The moral of Lemmas 16.16.6 and 16.16.7 is that Λ maps into

$$D_2' = D_2 \cap \big(H_2 \neq h_2^o(1/m) : m = 2, 3, 4, \dots\big).$$

Let Λ_2' denote the part of Λ_2 in which $m_2 = m_1 + 1$.

THEOREM 16.16.8. *Λ_2' is diffeomorphic to D_2'; the geometrical details are summarized in Fig. 16.11.*

PROOF. Λ_2 is already known to be locally diffeomorphic to D_2'. It suffices to show that a smooth curve in D_2' connecting points (h_0, h_1, h_2) and (h_0, h_1, h_2') with $2h_1 > h_o^2$ and $h_2^o(1) < h_2 < h_2^o(1/2) < h_2' < h_2^o(1/3)$, say, comes from a smooth curve in Λ_2' connecting a simple spectrum $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$ with $m_1 = 1$ and $m_2 = 2$ to a spectrum of type $\lambda_0 < \lambda_1 = \lambda_2 < \lambda_3 < \lambda_4 < \lambda_5 < \lambda_6$ with $m_1 = 2$ and $m_2 = 3$ which crosses the sheet no. 2 of Λ_1 corresponding to $H_2 = h_2^o(1/2)$. The proof is made by driving H_2 up from $h_2^o(1)$, keeping H_0 and H_1 fixed. This gives rise to an unlimited curve as in Fig. 16.11, commencing on sheet 1 of Λ_1, ultimately crossing every sheet of Λ_1. To put the matter a little differently, a simple spectrum $\lambda_0^o < \lambda_1^o < \lambda_2^o < \lambda_3^o < \lambda_4^o$ with $m_1 = m$ and $m_2 = m + 1$ can be driven by the differential equations

$$8 \prod_{k=1}^{2} \big(\lambda_i^o - \lambda_k'\big)\big(\lambda_i^o\big)^\bullet = H_2^\bullet = \pm 1 \quad (i = 0, \dots, 4)$$

to a point $\lambda_0^o < \lambda_1^o < \lambda_2^o[< \lambda_3^o = \lambda_4^o]$ on the m-th sheet of Λ_1 by choice of $H_2^\bullet = -1$, and to a point $\lambda_0^o[< \lambda_1^o = \lambda_2^o] < \lambda_3^o < \lambda_4^o$ on the $(m+1)$st sheet of Λ_1 by choice of $H_2^\bullet = +1$. The special case $H_2^\bullet = -1$, $m = 1$ is typical and will suffice; see Fig. 16.12. The differential equations determine a motion of $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4$ during which H_0 and H_1 stay fixed and the period relations

$$\int_{\lambda_1}^{\lambda_2} \frac{d\mu}{\ell(\mu)} + 2\int_{\lambda_3}^{\lambda_4} \frac{d\mu}{\ell(\mu)} = 0,$$
$$\int_{\lambda_1}^{\lambda_2} \frac{d\mu}{\ell(\mu)} + 2\int_{\lambda_3}^{\lambda_4} \frac{d\mu}{\ell(\mu)} = 1$$

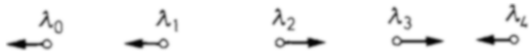

FIGURE 16.12.

are maintained, leaving one degree of freedom, as it should. Now the only way to *prevent* the outcome $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 = \lambda_4$ is to have *either* (a) λ_0 escape to $-\infty$, *or* (b) λ_1 collide with λ_0, *or else* (c) λ_2 collide with λ_3, at or simultaneously with the collision of λ_3 and λ_4. Consider the motion up to time $T =$ the smallest of the escape time of λ_0 and the collision times of λ_0 and λ_1, λ_2 and λ_3, or λ_3 and λ_4, and verify

$$\frac{\lambda_3^\bullet - \lambda_2^\bullet}{\lambda_3 - \lambda_2} \geqq \frac{1}{8} \frac{1}{(\lambda_2 - \lambda_2')(\lambda_2 - \lambda_1')(\lambda_3 - \lambda_1')} = \frac{-\lambda_2^\bullet}{\lambda_3 - \lambda_1'}.$$

Now let $\lambda_0 \downarrow -\infty$, i.e., let (a) hold. Then $\lambda_1' \sim \lambda_0$ because of

$$\lambda_0 + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - 2\lambda_1' - 2\lambda_2' = H_0 = \lambda_0 + \lambda_1 - 2\lambda_1' + O(1)$$

and

$$\lambda_1^\bullet - \lambda_0^\bullet = \frac{1}{8} \frac{1}{(\lambda_0 - \lambda_1')(\lambda_0 - \lambda_2')} - \frac{1}{8} \frac{1}{(\lambda_1 - \lambda_1')(\lambda_1 - \lambda_2')} < 0,$$

so that

$$\log(\lambda_3 - \lambda_2) \geqq -\int_0 \frac{\lambda_2^\bullet}{\lambda_3 - \lambda_1'}\, dt + O(1)$$

is bounded from below and λ_2 cannot collide with λ_3. This permits (a) to be eliminated: taking $\lambda_1 < 0 < \lambda_2$ for simplicity, the second period relation

$$\int_{\lambda_1}^{\lambda_2} \frac{\mu d\mu}{\ell(\mu)} + 2\int_{\lambda_3}^{\lambda_4} \frac{\mu d\mu}{\ell(\mu)} = 1$$

cannot be maintained in the face of

$$\left| \int_{\lambda_3}^{\lambda_4} \frac{\mu d\mu}{\ell(\mu)} \right| \leqq \frac{\lambda_4}{\sqrt{(\lambda_3 - \lambda_0)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)}} \int_{\lambda_3}^{\lambda_4} \frac{d\mu}{\sqrt{(\mu - \lambda_3)(\mu_4 - \lambda)}}$$
$$= O(|\lambda_0|^{-1}),$$

and[54]

$$\left| \int_{\lambda_1}^{\lambda_2} \frac{\mu d\mu}{\ell(\mu)} \right| \geqq -\int_{\lambda_1}^{o} \frac{\mu d\mu}{\sqrt{(\mu - \lambda_0)(\mu - \lambda_1)(\lambda_2 - \mu)(\lambda_3 - \mu)(\lambda_4 - \mu)}} - O(|\lambda_0|^{-1})$$
$$\geqq \frac{1}{\sqrt{\lambda_1 - \lambda_0}} \times |\lambda_0|^{-1/2}.$$

The upshot is that $\lambda_0 = O(1)$, and from

$$\lambda_0 + \cdots + \lambda_4 - 2\lambda_1' - 2\lambda_2' = H_0,$$
$$\lambda_0^2 + \cdots + \lambda_4^2 - 2(\lambda_1')^2 - 2(\lambda_2')^2 = H_1,$$

[54] $\lambda_1^\bullet - \lambda_0^\bullet < 0$ implies $\lambda_1 \sim \lambda_0$.

you see that the limiting values $\lambda_1'(T-)$ and $\lambda_2'(T-)$ exist. Now (b) $\lambda_1 - \lambda_0 = o(1)$ cannot take place without (c) $\lambda_3 - \lambda_2 = o(1)$ as that would unbalance the first period relation:

$$\pm\infty \sim \int_{\lambda_1}^{\lambda_2} \frac{d\mu}{\ell(\mu)} = -2\int_{\lambda_3}^{\lambda_4} \frac{d\mu}{\ell(\mu)} = O(1).$$

But in the face of (c),

$$\int_0^T \frac{\lambda_2^\bullet}{\lambda_3 - \lambda_1'}\, dt \geqq \log \frac{\lambda_3(0) - \lambda_2(0)}{\lambda_3(T-) - \lambda_2(T-)} = \infty$$

implies $\lambda_3 - \lambda_1' = o(1)$, too, with the result that

$$\frac{\lambda_1^\bullet - \lambda_0^\bullet}{\lambda_1\lambda_0} = \frac{1}{8}\frac{\lambda_0 + \lambda_1 - \lambda_1' - \lambda_2'}{(\lambda_0 - \lambda_1')(\lambda_0 - \lambda_2')(\lambda_1 - \lambda_1')(\lambda_1 - \lambda_2')} = O(1).$$

This eliminates (b). Now in the presence of (c), the first period relation implies $\lambda_4 - \lambda_3 = o(1)$: if not, the ratio of

$$\pm\int_{\lambda_1}^{\lambda_2} \frac{d\mu}{\ell(\mu)} \sim \frac{1}{\sqrt{(\lambda_2 - \lambda_0)(\lambda_2 - \lambda_1)(\lambda_4 - \lambda_2)}} \log\frac{\lambda_2 - \lambda_1}{\lambda_3 - \lambda_2},$$

and

$$\pm 2\int_{\lambda_3}^{\lambda_4} \frac{d\mu}{\ell(\mu)} \sim \frac{2}{\sqrt{(\lambda_3 - \lambda_0)(\lambda_3 - \lambda_1)(\lambda_4 - \lambda_3)}} \log\frac{\lambda_4 - \lambda_3}{\lambda_3 - \lambda_2},$$

would tend to $1/2$ and not to 1, violating the first period relation. This leaves you in the case $-\infty < \lambda_0 < \lambda_1 < \lambda_1' = \lambda_2 = \lambda_3 = \lambda_2' = \lambda_4$ at time $t = T-$, and that is ruled out by observing that

$$2H_1 - H_0^2 = (\lambda_0^2 + \lambda_1^2 - \lambda_2^2) - (\lambda_0 + \lambda_1 - \lambda_2)^2 = -2(\lambda_2 - \lambda_0)(\lambda_2 - \lambda_1) < 0,$$

violates $H_0^2 \leqq 2H_1$. The proper period relation for sheet 1 of Λ_1:

$$\int_{\lambda_1}^{\lambda_2} \frac{d\mu}{\sqrt{(\mu - \lambda_0)(\mu - \lambda_1)(\lambda_2 - \mu)}} = 1$$

is automatically met by the collapsed spectrum $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 = \lambda_4$ since the old period relations collapse to

$$\int_{\lambda_1}^{\lambda_2} \frac{1}{\lambda_3 - \mu}\frac{d\mu}{\sqrt{(\mu - \lambda_0)(\mu - \lambda_1)(\lambda_2 - \mu)}} = \frac{2\pi}{\sqrt{(\lambda_3 - \lambda_0)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)}} = 0$$

and

$$\int_{\lambda_1}^{\lambda_2} \frac{1}{\lambda_3 - \mu}\frac{\mu d\mu}{\sqrt{(\mu - \lambda_0)(\mu - \lambda_1)(\lambda_2 - \mu)}} - \frac{2\pi\lambda_3}{\sqrt{(\lambda_3 - \lambda_0)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)}} = \pm 1.$$

The proof is finished by use of Lemma 16.16.7 to fill in the geometrical details of Fig. 16.12. □

Amplification 16.16.9. It will be plain from the above that with more effort along similar lines, the disposition of the sheets of Λ_2 over D_2' could be completely clarified; in particular, it seems likely that the sheets can be labelled by means of the numbers m_1, m_2 figuring in the period relations, or, equivalently, by means of suitable critical values of H_3, though the matter may be more complicated than that. For $n = 1$, the period relation on the m-th sheet of Λ_1:

$$m \int_{\lambda_1}^{\lambda_2} \frac{d\mu}{\ell(\mu)} = 1$$

states that q is of period $1/m$, and you have $H_2 = h_2^o(1/m)$. Similarly, for $n = 2$, the value of H_3 is determined locally by $H_i = h_i$ $(i \leqq 2)$, and it would be a source of satisfaction to identify that value as the minimum of H_3 over some simple subclass of $C_1^\infty \cap (H_0 = h_0) \cap (H_1 = h_1) \cap (H_2 = h_2)$ determined by m_1, m_2. Clearly, all the same difficulties arise for $n \geqq 3$.

Amplification 16.16.10. The full Hamiltonian series $H_0 H_1 H_2 \ldots$ is always a global coordinate on Λ_n; see Corollary 16.12.8.

Bibliography

[1] G. Borg. Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe. Bestimmung der Differentialgleichung durch die Eigenwerte. *Acta Math.*, 78:1–96, 1946.

[2] V. S. Buslaev and L. D. Faddeev. Formulas for traces for a singular Sturm-Liouville differential operator. *Dokl. Akad. Nauk SSSR*, 132:13–16, 1960.

[3] E. T. Copson. *An Introduction to the Theory of Functions of a Complex Variable.* Oxford University Press, Oxford, 1960.

[4] L. A. Dikiĭ. The zeta function of an ordinary differential equation on a finite interval. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 19:187–200, 1955.

[5] L. A. Dikiĭ. Trace formulas for Sturm-Liouville differential operators. *Amer. Math. Soc. Trans. (2)*, 18:81–115, 1961.

[6] B. Dubrovin and S. Novikov. Periodic and conditionally periodic analogues of multi-soliton solutions of the Kortweg-de Vries equation. *Dokl. Akad. Nauk SSSR*, 6:2131–2144, 1974.

[7] L. D. Faddeev. The inverse problem in the quantum theory of scattering. *Uspehi Mat. Nauk.*, 14:57–119, 1959. Translation in *J. Mathematical Phys.* 4:72–104, 1963.

[8] H. Flaschka. On the inverse problem for Hill's operator. *Arch. Rational Mech. Anal.*, 54:293–309, 1975.

[9] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Method for solving the Korteweg-de Vries equation. *Phys. Rev. Letters*, 19:1095–1097, 1967.

[10] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura. Korteweg-de Vries equation and generalizations. VI: methods for exact solution. *Comm. Pure Appl. Math.*, 27:97–133, 1974.

[11] I. M. Gelfand. On identities for eigenvalues of a differential operator of second order. *Uspehi Mat. Nauk (N.S.)*, 11:191–198, 1956.

[12] I. M. Gelfand and B. M. Levitan. On the determination of a differential equation from its spectral function. *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 15:309–360, 1951. Translation: American Mathematical Translation, 1 (1955), 253–304.

[13] I. M. Gelfand and B. M. Levitan. On a simple identity for the characteristic values of a differential operator of second order. *Dokl. Akad. Nauk SSSR*, 88:593–596, 1953.

[14] W. Goldberg. Necessary and sufficient conditions for determining a Hill's equation from its spectrum. *Bull. Amer. Math. Soc.*, 81:423–424, 1975.

[15] A. Göpel. Entwurf einer Theorie der Abel'schen Transcendenten erster Ordnung. *Ostwald's Klassiker der exacten Wissenschaften*, 67:W. Engelmann, Leipzig, 1895.

[16] H. Hochstadt. Function-theoretic properties of the discriminant of Hill's equation. *Math. Zeit.*, 82:237–242, 1963.

[17] H. Hochstadt. On the determination of Hill's equation from its spectrum. *Arch. Rat. Mech. Anal.*, 19:353–362, 1965.

[18] A. R. Its and V. B. Matveev. Hill operators with a finite number of lacunae. *Funkcional. Anal. i Priložen.*, 9(1):69–70, 1975.

[19] M. Kac. On distributions of certain Wiener functionals. *Trans. Amer. Math. Soc.*, 65:1–13, 1949.

[20] M. Kac and P. van Moerbeke. On some isospectral second order differential operators. *Proc. Nat. Acad. Sci. U.S.A.*, 71:2350–2351, 1974.

[21] A. Krazer and W Wirtinger. *Abelsche Funktionen und allgemeine Thetafunktionen*, volume II,2 of *Encyklopädie der Mathematischen Wissenschaften.* B. T. Teubner, Leipzig, 1921.

[22] P. Lax. Integrals of non-linear equations and solitary waves. *Comm. Pure Appl. Math.*, 21:467–490, 1968.

[23] P. Lax. Periodic solutions of the KdV equations. In *Nonlinear Wave Motion (Proc. AMS-SIAM Summer Sem., Clarkson Coll. Tech., Postdam, N.Y. 1972)*, volume 15 of *Lectures in Appl. Math,*, pages 85–96. Amer. Math. Soc., Rhode Island, 1975.

[24] P. Lax. Periodic solutions of the Korteweg-de Vries equation. *Comm. Pure Appl. Math.*, 28:141–188, 1975.

[25] N. Levinson. The inverse Sturm-Liouville problem. *Mat. Tidsskr. B*, 1949:25–30, 1949.

[26] W. Magnus and W. Winkler. *Hill's equation.* Interscience Wiley, New York, 1st edition, 1966.

[27] A. Menikoff. The existence of unbounded solutions of the Korteweg-de Vries equation. *Comm. Pure Appl. Math.*, 25:407–432, 1972.

[28] S. P. Novikov. The periodic problem for the Korteweg-de Vries equation. *Funct. Anal. Appl.*, 8:236–246, 1974.

[29] H. E. Rauch and H. M. Farkas. *Theta functions with applications to Riemann Surfaces.* The Williams & Wilkins Co., Baltimore, 1974.

[30] G. Rosenhain. Abhandlung über die Functionen zweier Variabler mit vier Perioden, welche die Inversen sind der ultra-elliptischen Integrale erster Klasse. *Ostwald's Klassiker der exacten Wissenschaften*, 65:W. Engelmann, Leipzig, 1895.

[31] C. L. Siegel. *Topics in Complex Function Theory II. Automorphic Functions and Abelian Integrals.* Wiley-Interscience, 1969.

[32] M. J. O. Strutt. *Lamesche-, Mathieusche- und verwandte Funktionen in Physik und Technik.* Springer-Veralg, 1932.

[33] V. E. Zakharov and L. D. Faddeev. The Korteweg-de Vries equation is a fully integrable Hamiltonian system. *Funkcional Anal. i Priložen*, 5:18–27, 1974.