Surprising contraction properties of Wiener-Hopf operators, 1

BY DAVID WILLIAMS AND DAVID S MARLES

Dept. of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, U.K.

dw@maths.bath.ac.uk, dsm@maths.bath.ac.uk

A much studied type of Wiener-Hopf theory may be put in a clean, rigorous form because of a surprising contraction property of certain 'half-winding' operators. The key idea for the proof is presented in §2 here; and it gives more precise information (and is more interesting) than the contraction property itself.

1. Introduction

We have made this paper fully rigorous (except for the Appendix, §4) and as self-contained as possible; and because (except for that Appendix) all discussion of the probability is deferred to a later occasion, we hope that the paper will be accessible to a wide audience. (Probabilist fans of stable laws, *pace*!) The one reader we most hope will like this has never needed any concessions, and that's Henry McKean; we hope that he will regard this paper and its sequels as a small 'Thanks' for many years of inspiration.

It is now possible to simplify very greatly, and to generalize, the type of Wiener-Hopf *theory* surveyed in Williams (1991) (= [W]), though deriving closed-form answers remains as challenging as ever. A major reason for the mentioned simplification is that certain infinite-dimensional Wiener-Hopf operators have desirable contraction properties not shared by their 'seemingly obvious' finite-dimensional analogues.

What we mean by a Wiener-Hopf factorization of an operator, and why certain contraction properties matter, will be clarified further in §1a, but here is part of the story. The typical operator which we 'factorize' in this paper has the form $\mathrm{sgn}(x)^{-1}\mathcal{A}$, where $\mathcal{A}$ is a generalized second-order elliptic operator on functions on $\mathbb{R}$. (In the so-called 'canonical case', $\mathcal{A} = \frac{1}{2}\Delta$ on $(0,\infty)$ and $\mathcal{A} = \frac{1}{2}c^{-2}\Delta$ on $(-\infty,0)$ where c is a positive constant.) All 'eigenvalues' of our operator $\mathrm{sgn}(x)^{-1}\mathcal{A}$ are real. Roughly speaking, the 'half-winding' operator Π^{-+} which is our chief interest tells how functions in the span of all 'eigenfunctions corresponding to negative eigenvalues' extend from $(0,\infty)$ to the remainder of $\mathbb{R}$. Thus, if f_θ is an 'eigenfunction' of $\mathrm{sgn}(x)^{-1}\mathcal{A}$ corresponding to a negative 'eigenvalue' $-\frac{1}{2}\theta^2$, then $\Pi^{-+}f_\theta^+ = f_\theta^-$, where f_θ^+ is the restriction of f_θ to $(0,\infty)$ and f_θ^- that of f_θ to $(-\infty,0]$. The reason that we have put 'quotes' around 'eigenvalue', 'eigenfunction', etc, is that f_θ will generally not be in the Hilbert space with which we wish to work.

The history of the results presented here is as follows. Ivan Graham and John Toland (both at Bath) suggested the relevance of Mellin transforms to the much-studied canonical case, because the half-winding operator is then a *multiplier* in Mellin-transform space. (Of course, the first paper on half-winding operators, McKean (1963), is a *tour de force* of transform theory.) Somewhat to our surprise, it turned out for the canonical case that the natural L^2 norm of the half-winding operator is never greater than $2^{-\frac{1}{2}}$. We then studied the discrete approximation via atomic measures on $(-\infty,0]$ which was used in London, McKean, Rogers & Williams (1982a) (= [LMRW1]) to derive results for the general problem when $\mathcal{A} = \frac{1}{2}\Delta$ on $(0,\infty)$ studied there. The requisite new tools for studying this 'atomic' case were developed in Kennedy & Williams (1990) (= [KW]). To our initial astonishment, we found a mysterious reason why the L^2 norm of the half-winding operator for the 'atomic' situation again never exceeds $2^{-\frac{1}{2}}$. This reason forms the main part, §2 of this paper.

The $2^{-\frac{1}{2}}$ is associated with those cases in which, on $(0,\infty)$, $\mathcal{A}$ acts as a constant multiple of the Laplace operator. In §1d, we examine a situation where this property of $\mathcal{A}$ no longer holds; and the result there both illuminates some aspects of the subject and deepens the mystery of others.

(*a*) *The 'general' problem in [LMRW1]*

As promised, we give a completely analytic statement. Let m be a measure on (Borel subsets of) $\mathbb{R}$ with restrictions m^+ to $(0,\infty)$ and m^-

to $(-\infty, 0]$ where

$$\begin{aligned} m^+ &= \text{Leb} := \text{Lebesgue measure on } (0,\infty), \\ m^-[a,0] &< \infty \quad \text{ for every } a \in (-\infty, 0]. \end{aligned} \tag{1.1}$$

To achieve a perfect tie-up with Krein's theory of strings (for which see, for example, Dym & McKean (1976)), a less restrictive, less easily stated, condition on m^- was imposed in [LMRW1].

Let $\mathcal{A}$ be the operator

$$\mathcal{A} := \tfrac{1}{2}\frac{\mathrm{d}}{\mathrm{d}m}\frac{\mathrm{d}}{\mathrm{d}x}. \tag{1.2}$$

We understand that for $f \in \mathcal{D}(\mathcal{A})$, the one-sided derivatives

$$(D^+f)(x) := \lim_{y\downarrow x}\frac{f(y)-f(x)}{y-x}, \text{ and } (D^-f)(x) := \lim_{y\uparrow x}\frac{f(x)-f(y)}{x-y} \tag{1.3}$$

exist, and for $-\infty < b < c < \infty$,

$$(D^+f)(c) - (D^-f)(b) = 2\int_{[b,c]}(\mathcal{A}f)(x)m(\mathrm{d}x), \tag{1.4}$$

so that the function $\mathcal{A}f$ is only defined almost everywhere (m), and $\mathcal{A}f$ is locally in $L^1(m)$.

Remarks. We think of $\mathcal{A}$ as *locally defined* in that we do not impose 'Banach' or 'Hilbert' requirements on its domain. To try to help set the 'global' scene, however, we make some remarks in which, to avoid technical nonsenses, we shall assume that the support of m is the whole of $\mathbb{R}$.

Ultimately, though not in this paper, we wish to think of the closure of the essentially self-adjoint operator $\mathcal{A}_2$, the restriction of $\mathcal{A}$ to the dense domain

$$\mathcal{D}(\mathcal{A}_2) = \{f \in \mathcal{D}(\mathcal{A}) \cap L^2(\mathbb{R}, m) : \mathcal{A}f \in L^2(\mathbb{R}, m)\}$$

in the Hilbert space $L^2(\mathbb{R}, m)$, as the generator of a strongly continuous one-parameter semigroup $\{P_t\}$ of self-adjoint contraction operators on $L^2(\mathbb{R}, m)$. Alternatively, if $C_0(\mathbb{R})$ denotes the Banach space of continuous functions on $\mathbb{R}$ which tend to 0 at $\pm\infty$, with the usual supremum norm, then the restriction $\mathcal{A}_0$ of $\mathcal{A}$ to the dense domain

$$\mathcal{D}(\mathcal{A}_0) = \{f \in \mathcal{D}(\mathcal{A}) \cap C_0(\mathbb{R}) : \mathcal{A}f \text{ may be chosen in } C_0(\mathbb{R})\}$$

generates a strongly continuous one-parameter semigroup $\{T_t\}$ of positive operators on $C_0(\mathbb{R})$, each of norm 1. $\square$

For $\theta \in \mathbb{R} \setminus \{0\}$, there exists a unique *bounded continuous* function f_θ on $\mathbb{R}$ with

$$(\mathcal{A} f_\theta)(r) = -\tfrac{1}{2}\theta^2 \operatorname{sgn}(r) f_\theta(r), \quad (D^+ f_\theta)(0) = 1, \tag{1.5}$$

where we define (the detail matters when $m(0) := m(\{0\}) > 0$)

$$\operatorname{sgn}(r) := \begin{cases} -1 & \text{if } r \leq 0, \\ +1 & \text{if } r > 0, \end{cases} \tag{1.6}$$

so that, in particular, $\operatorname{sgn}(0) := -1$. Note that from the form of m^+ in (1.1), we must have

$$f_\theta(y) = f_\theta(0) \cos\theta y + \theta^{-1} \sin\theta y \qquad (y > 0). \tag{1.7}$$

Remarks. The only difficulty is in establishing the existence and uniqueness of f_θ on $(-\infty, 0)$. In that region, we have

$$(\tfrac{1}{2}\theta^2 - \mathcal{A}) f_\theta = 0,$$

and so it is not surprising that we can express f_θ in terms of the action of the resolvent of either $\{P_t\}$ or $\{T_t\}$ on a function with support in $(0, \infty)$. For example, if h is a positive element of $C_0(\mathbb{R})$ with support in $(0, \infty)$ and $u := R(\frac{1}{2}\theta^2, \mathcal{A}_0)h = (\frac{1}{2}\theta^2 - \mathcal{A}_0)^{-1}h$, then $f_\theta \equiv$ constant $\times\, u$ on $(-\infty, 0]$. This is perhaps the best way of understanding the 'Sturm-Liouville' result on existence and uniqueness of f_θ; and it shows that $f_\theta > 0$ on $(-\infty, 0]$. $\square$

For a function f on $\mathbb{R}$, let f^+ [respectively, f^-] denote the restriction of f to $(0, \infty)$ [respectively, $(-\infty, 0]$]. It is known from [LMRW1] that there exists a unique positive continuous function

$$\pi : (-\infty, 0) \times (0, \infty) \to [0, \infty) \tag{1.8}$$

such that, for $\theta \in \mathbb{R} \setminus \{0\}$,

$$\Pi^{-+} f_\theta^+ = f_\theta^-, \tag{1.9}$$

this meaning that

$$\int_{(0,\infty)} \pi(x, y) f_\theta(y) m^+(\mathrm{d}y) = f_\theta(x), \qquad \text{for } x \in (-\infty, 0). \tag{1.10}$$

If $m(0) > 0$, then we can and do extend π continuously to $(-\infty, 0] \times (0, \infty)$

so that (1.10) holds for $x = 0$. As explained in [W], for example, the trigonometric form (1.7) of f_θ^+ makes the problem of solving (1.10) into a Riemann-Hilbert problem. It is automatic that, for $x \in (-\infty, 0)$ (or $x \in (-\infty, 0]$) if $m(0) > 0$),

$$\Pi^{-+}\mathbf{1} \leq \mathbf{1} \text{ in that } \int_{(0,\infty)} \pi(x,y)m^+(\mathrm{d}y) \leq 1, \qquad \text{for } x \in (-\infty, 0). \tag{1.11}$$

We wish to consider Π^{-+} (defined in the now-obvious fashion from $\pi(\cdot,\cdot)$) as a map

$$\Pi^{-+} : L^2\big((0,\infty), m^+\big) \to L^2\big((-\infty, 0], m^-\big), \tag{1.12}$$

and especially to study its norm $\|\Pi^{-+}\|_2$ (which at this stage we have not shown to be finite).

Remarks on Wiener-Hopf factorization of $\mathrm{sgn}(\cdot)^{-1}\mathcal{A}$**.** Let $E^+ = (0, \infty)$, $E^- = (-\infty, 0]$ and $E = \mathbb{R}$. We have the canonical identification

$$\tilde{L}^2 = \tilde{L}^2_+ \oplus \tilde{L}^2_-, \quad \text{where } \tilde{L}^2 = L^2(E, m), \ \tilde{L}^2_\pm = L^2(E^\pm, m^\pm). \tag{1.13}$$

The tildes help emphasize that the inner products on $\tilde{L}^2$ and $\tilde{L}^2_-$ are not the standard ones relative to Lebesgue measure. Let $I^\pm$ be the identity operator on $\tilde{L}^2_\pm$.

The operator Π^{-+} arises in the factorization (corresponding to the decomposition (1.13))

$$J^{-1}\mathrm{sgn}(\cdot)^{-1}\mathcal{A}J = \begin{pmatrix} \mathcal{G}_+ & 0 \\ 0 & -\mathcal{G}_- \end{pmatrix}, \quad \text{where } J = \begin{pmatrix} I^+ & \Pi^{+-} \\ \Pi^{-+} & I^- \end{pmatrix} \tag{1.14}$$

In this paper, $\mathcal{A}$ is essentially self-adjoint on $\tilde{L}^2$, and the heuristic duality principle (proved later for all of the examples discussed in this paper) says that Π^{+-} is the Hibert-space adjoint of Π^{-+}. The operator $\mathcal{G}_+$ is a negative-semi-definite self-adjoint operator on functions on E^+ relative to the inner product

$$\langle f, g\rangle_+ = \langle f, (I^+ - \Pi^{+-}\Pi^{-+})g\rangle_{m^+},$$

the $\langle\cdot,\cdot\rangle_{m^+}$ on the right-hand side being the standard inner product for $\tilde{L}^2_+$. Analogously, $\mathcal{G}_-$ is a negative-semi-definite self-adjoint operator on functions on E^- relative to the inner product

$$\langle f, g\rangle_- = \langle f, (I^- - \Pi^{-+}\Pi^{+-})g\rangle_{m^-},$$

the $\langle\cdot,\cdot\rangle_{m^-}$ on the right-hand side now being the standard inner product for $\tilde{L}^2_-$. It is immediately clear that we need the operators $(I^+ - \Pi^{+-}\Pi^{-+})$ and $(I^- - \Pi^{-+}\Pi^{+-})$ to be invertible; and that is one key use of the fact that we have $\|\Pi^{-+}\|_2 < 1$ in the situations considered in this paper. The operators $\mathcal{G}_+$ and $\mathcal{G}_-$ generate contraction semigroups of self-adjoint operators (relative to the $\langle\cdot,\cdot\rangle_\pm$ inner products) which have probabilistic significance.

(*b*) *The canonical case*

This is the case when in §1a we have, for some $c > 0$,

$$m^- = c^2 \times \text{Lebesgue measure on } (-\infty, 0]. \tag{1.15}$$

We see that we require $(D^- f_\theta)(0) = (D^+ f_\theta)(0) = 1$, and hence

$$f_\theta(x) = (c\theta)^{-1}\mathrm{e}^{c\theta x} \quad (x \le 0); \quad f_\theta(0) = (c\theta)^{-1}.$$

Then, as may be verified by contour integration, the solution of equation (1.10) is given by N. Baker's formula (see [W] for one direct method of arriving at this)

$$\pi(x,y) = \frac{2c^\alpha |x|^\alpha y^{1-\alpha} \sin\frac{1}{2}\pi\alpha}{\pi(c^2x^2 + y^2)}, \tag{1.16}$$

where

$$0 < \alpha < 1 \text{ and } \tan\tfrac{1}{2}\pi\alpha = c^{-1}. \tag{1.17}$$

THEOREM 1.1. *In this case, we have* $\Pi^{-+}\mathbf{1} = \mathbf{1}$ *and*

$$\|\Pi^{-+}\|_2 = \frac{(2c)^{\frac{1}{2}}}{1+c} \le 2^{-\frac{1}{2}},$$

with $\|\Pi^{-+}\|_2 = 2^{-\frac{1}{2}}$ *if and only if* $c = 1$.

(*c*) *The case when* m^- *consists of finitely many atoms*

Suppose now that we have the situation of §1a where on $(-\infty, 0]$, m^- consists just of $n+1$ $(n \ge 0)$ atoms of strictly positive masses

$$m(0), m(b_1), m(b_2), \ldots, m(b_n) \tag{1.18}$$

at positions

$$b_0 = 0, b_1, b_2, \ldots, b_n, \text{ where } b_n < b_{n-1} < \cdots < b_1 < b_0 = 0. \tag{1.19}$$

THEOREM 1.2. *For this atomic case,*

$$\|\Pi^{-+}\|_2 = 2^{-\frac{1}{2}}.$$

If we allowed $m_0 = 0$ *(when* $n \geq 1$*), then we would have* $\|\Pi^{-+}\|_2 \leq 2^{-\frac{1}{2}}$*.*

The proof of Theorem 1.2 hinges on a strange algebraic identity (first discovered via *numerical* computer investigation prompted by vague probabilistic intuition) which is a lot more interesting than the theorem itself. The proof also relies heavily on a duality property proved for the case of symmetrizable Markov chains in the sequel [LMRW2] to [LMRW1], the general unproved case of which has become part of the folklore. We *prove* the duality for the 'atomic' situation. (Incidentally, the duality is *obvious* for the cases covered by Theorems 1.1 and 1.3 by symmetry considerations.)

(d) A case with 'power-law' measures

This case does not fit into the [LMRW1] regime. We take $\beta > -1$ and $c > 0$ and

$$m^+(\mathrm{d}y) = y^\beta \mathrm{d}y \quad (y > 0); \qquad m^-(\mathrm{d}x) = c^{2+\beta}|x|^\beta \mathrm{d}x \quad (x \leq 0). \tag{1.20}$$

Then, as was first shown in Rogers & Williams (1984), the appropriate half-winding operator has

$$\pi(x, y) = K \frac{|x|^\alpha y^{1-\alpha}}{(c|x|)^{2+\beta} + y^{2+\beta}}, \tag{1.21}$$

for $x \in (-\infty, 0)$ and $y \in (0, \infty)$, where α is the unique root in $(0, 1)$ of the equation

$$\sin(1-\alpha)\delta = c \sin \alpha\delta, \text{ where } \delta := \pi/(2+\beta), \tag{1.22}$$

and

$$K = \pi^{-1}(2+\beta)c^\alpha \sin \alpha\delta. \tag{1.23}$$

The case when $\beta = 0$ is our earlier 'canonical case'. The case when $\beta = 1$ and $c = 1$, an important case in practice, was solved by McKean (1963). We may generalize Theorem 1.1 as follows.

THEOREM 1.3. *We have* $\Pi^{-+}\mathbf{1} = \mathbf{1}$*. The map*

$$\Pi^{-+} : L^2\big((0, \infty), m^+\big) \to L^2\big((-\infty, 0], m^-\big),$$

has norm

$$\|\Pi^{-+}\|_2 = 2\frac{c^{\frac{1}{2}}}{1+c}\sin\frac{\pi}{4+2\beta}.$$

Note that for fixed β, this norm is maximized when $c = 1$. If c is fixed at 1, then $\|\Pi^{-+}\|_2$ decreases from 1 to 0 as β increases from -1 to ∞. There is good sense in all this.

Proof. We may as well prove this theorem here, thereby allowing §2 to concentrate on the more interesting 'atomic' case. We leave the reader to check that $\Pi^{-+}\mathbf{1} = \mathbf{1}$ by using an integral occurring later in the proof.

The map S with

$$(Sf)(y) := y^{\frac{1}{2}\beta} f(y) \quad (y > 0)$$

is an isometric isomorphism from $L^2((0,\infty), m^+)$ to $L^2(0,\infty) := L^2((0,\infty), \mathrm{Leb})$. The map T with

$$(Th)(z) := c^{1+\frac{1}{2}\beta} z^{\frac{1}{2}\beta} h(-z) \quad (z > 0)$$

is an isometric isomorphism from $L^2((-\infty,0), m^-)$ to $L^2(0,\infty)$. We have

$$\Pi^{-+} = T^{-1}BS, \quad B = T\Pi^{-+}S^{-1}$$

where

$$B : L^2(0,\infty) \to L^2(0,\infty)$$

is given by

$$\begin{aligned}
(Bg)(z) &= c^{1+\frac{1}{2}\beta}\int_0^\infty z^{\frac{1}{2}\beta}\pi(-z,y)y^{-\frac{1}{2}\beta}g(y)y^\beta \,\mathrm{d}y \quad (z>0)\\
&= Kc^{1+\frac{1}{2}\beta}\int_0^\infty \frac{z^{\alpha+\frac{1}{2}\beta}y^{1-\alpha+\frac{1}{2}\beta}}{(cz)^{2+\beta}+y^{2+\beta}}\, g(y)\,\mathrm{d}y\\
&= \int_0^\infty b(r)g(rz)\,\mathrm{d}r
\end{aligned}$$

where

$$b(r) := Kc^{1+\frac{1}{2}\beta}\frac{r^{1-\alpha+\frac{1}{2}\beta}}{c^{2+\beta}+r^{2+\beta}} \geq 0.$$

Of course, $\|\Pi^{-+}\|_2 = \|B\|_2$, and hence, by a theorem of Hardy, Littlewood and Pólya (1934, pp. 229–234) (see Remarks below),

$$\|\Pi^{-+}\|_2 = \|B\|_2 = \int_0^\infty b(r)r^{-\frac{1}{2}}\,\mathrm{d}r = Kc^{1+\frac{1}{2}\beta}\int_0^\infty \frac{r^{\frac{1}{2}-\alpha+\frac{1}{2}\beta}}{c^{2+\beta}+r^{2+\beta}}\,\mathrm{d}r.$$

Putting $r = cx^{1/(2+\beta)}$, we see that, with

$$\gamma = \frac{\frac{3}{2} - \alpha + \frac{1}{2}\beta}{2+\beta}$$

we have, with $\delta := \pi/(2+\beta)$ as before,

$$\begin{aligned}\|\Pi^{-+}\|_2 &= \frac{Kc^{\frac{1}{2}-\alpha}}{2+\beta}\int_0^\infty \frac{x^{\gamma-1}}{1+x}\,\mathrm{d}x = \frac{Kc^{\frac{1}{2}-\alpha}}{2+\beta} \times \frac{\pi}{\sin\pi\gamma} \\ &= \frac{c^{\frac{1}{2}}\sin\alpha\delta}{\sin(\frac{3}{2}-\alpha+\frac{1}{2}\beta)\delta} = \frac{c^{\frac{1}{2}}\sin\alpha\delta}{\cos(\frac{1}{2}-\alpha)\delta}.\end{aligned}$$

However, from (1.22),

$$(c+1)\sin\alpha\delta = \sin(1-\alpha)\delta + \sin\alpha\delta = 2\sin\tfrac{1}{2}\delta\cos(\tfrac{1}{2}-\alpha)\delta,$$

and the theorem now follows. ■

Of course, Theorem 1.1 is now also proved.

Remarks. The Hardy-Littlewood-Pólya theorem says that if $b(\cdot) > 0$ on $(0,\infty)$ and

$$(Bg)(z) := \int_0^\infty b(r)g(rz)\,\mathrm{d}r,$$

then considering $B : L^2(0,\infty) \to L^2(0,\infty)$, we have

$$\|B\|_2 = \int_0^\infty b(r)r^{-\frac{1}{2}}\,\mathrm{d}r, \quad Bh = \|B\|_2 h \text{ where } h(u) := u^{-\frac{1}{2}}, \quad h \notin L^2(0,\infty).$$

Of course, the substitutions $r = \mathrm{e}^x$, $\tilde{h} = \mathrm{e}^{\frac{1}{2}x}h(\mathrm{e}^x)$, $h \mapsto \tilde{h}$ being an isomorphism of $L^2(0,\infty)$ to $L^2(-\infty,\infty)$, lead to

$$(\tilde{B}\tilde{g})(a) := (\widetilde{Bg})(a) = \int_{-\infty}^\infty \tilde{b}(x)\tilde{g}(x+a)\,\mathrm{d}x$$

and so the H-L-P result is the familiar result for convolution operators that

$$\|\tilde{B}\|_2 = \|\tilde{b}\|_1,$$

provided that $\tilde{b}(\cdot) \geq 0$.

2. Proof of Theorem 1.2

In §2a we recall facts from Kennedy & Williams (1990) (= [KW]). By using the methods of [KW], *we enrich the structure*: for example, the Γ_+ operator does not feature in the [LMRW] story. However, it is the case that $\mathcal{G}_- = -\frac{1}{2}\Gamma_-^2$. The relevance of §2a to Theorem 1.2 will soon become apparent.

(*a*) *Wiener-Hopf factorization of a quadratic*

Let S be a finite set. Let m be measure on S with each m_i strictly positive. The measure m will later be the m^- of the 'atomic' situation. For an $S \times S$ matrix G, let G^* denote the adjoint of G with respect to $L^2(S, m)$, so that, if M denotes the diagonal $S \times S$ matrix

$$M := \operatorname{diag}(m_i : i \in S), \tag{2.1}$$

then, for column vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^S$, and with $\cdot^T$ denoting transpose,

$$< \mathbf{u}, G^*\mathbf{v} >:= \mathbf{u}^T MG^*\mathbf{v} =< G\mathbf{u}, \mathbf{v} >= \mathbf{u}^T G^T M\mathbf{v}, \text{ so that } G^* = M^{-1}G^T M. \tag{2.2}$$

Let $\mathcal{Q}$ denote the set of real $S \times S$ matrices with non-negative off-diagonal elements and with non-positive row sums; and let $\mathcal{Q}_0$ consist of those elements of $\mathcal{Q}$ for which all row sums are exactly 0. Let $Q \in \mathcal{Q}_0$, and assume that Q is *m-symmetrizable* in that

$$Q = Q^*; \quad \text{equivalently, } MQ \text{ is symmetric.} \tag{2.3}$$

Let V be a real diagonal $S \times S$ matrix.

In [KW] it is shown that there exists a unique pair (Γ_-, Γ_+) of elements of $\mathcal{Q}$ such that

$$t^2 I + tV + Q = (tI - \Gamma_-^*)(tI + \Gamma_+) = (tI + \Gamma_+^*)(tI - \Gamma_-), \tag{2.4}$$

I denoting the identity $S \times S$ matrix. As consequences, we have

$$\Gamma_+ - \Gamma_-^* = V, \tag{2.5}$$

and

$$\Gamma_-^*\Gamma_+ = -Q = \Gamma_+^*\Gamma_-. \tag{2.6}$$

It is immediate to verify that

$$\Gamma_+^2 - V\Gamma_+ + Q = 0, \qquad \Gamma_-^2 + V\Gamma_- + Q = 0. \tag{2.7}$$

We shall be interested only in the case when Q is *irreducible* (in that for any disjoint partition $S = S_1 \cup S_2$ of S into non-empty sets, we can find $i \in S_1$ and $j \in S_2$ with $Q(i,j) > 0$) and V has non-negative diagonal entries which are not all zero. We now assume that we are working with that case. Then $\Gamma_+ \in \mathcal{Q}_0$, and $\Gamma_- \in \mathcal{Q} \setminus \mathcal{Q}_0$. We can find a basis $(\mathbf{e}_j : j \in S)$ of $\mathbb{R}^S$ and strictly negative numbers which we write $(-\lambda_j : j \in S)$ such that

$$(\lambda_j^2 I - \lambda_j V + Q)\mathbf{e}_j = \mathbf{0}; \quad \Gamma_- \mathbf{e}_j = -\lambda_j \mathbf{e}_j. \tag{2.8}$$

(b) An algebraic identity

We continue with the notation of the last section. The key step in the paper will be to apply the following lemma to the case when $W = \Pi^{-+}(\Pi^{-+})^*$.

Lemma 2.1. *Suppose that the eigenvalues $(-\lambda_j : j \in S)$ of Γ_- are distinct. Suppose also that an $S \times S$ matrix W satisfies*

$$\langle W\mathbf{e}_k, \mathbf{e}_j \rangle = \frac{\langle V\mathbf{e}_j, \mathbf{e}_k \rangle}{\lambda_j + \lambda_k} \qquad (j, k \in S). \tag{2.9}$$

Then

$$(I - 2W)\Gamma_- = \Gamma_+. \tag{2.10}$$

Proof. From (2.8) and (2.6), we have

$$\lambda_j^2 \langle \mathbf{e}_j, \mathbf{e}_k \rangle - \lambda_j \langle V\mathbf{e}_j, \mathbf{e}_k \rangle = -\langle Q\mathbf{e}_j, \mathbf{e}_k \rangle = \langle \Gamma_- \mathbf{e}_j, \Gamma_+ \mathbf{e}_k \rangle, \tag{2.11}$$

and since $\lambda_j \neq 0$ and $\Gamma_- \mathbf{e}_j = -\lambda_j \mathbf{e}_j$, we have

$$\lambda_j \langle \mathbf{e}_j, \mathbf{e}_k \rangle - \langle V\mathbf{e}_j, \mathbf{e}_k \rangle = -\langle \mathbf{e}_j, \Gamma_+ \mathbf{e}_k \rangle, \tag{2.12}$$

Next, from (2.11) and the fact that $Q = Q^*$ and $V = V^*$, we have

$$\lambda_j^2 \langle \mathbf{e}_j, \mathbf{e}_k \rangle - \lambda_j \langle V\mathbf{e}_j, \mathbf{e}_k \rangle = \lambda_k^2 \langle \mathbf{e}_j, \mathbf{e}_k \rangle - \lambda_k \langle V\mathbf{e}_j, \mathbf{e}_k \rangle,$$

so that

$$\langle \mathbf{e}_j, \mathbf{e}_k \rangle = \frac{\langle V\mathbf{e}_j, \mathbf{e}_k \rangle}{\lambda_j + \lambda_k} \quad (j \neq k). \tag{2.13}$$

By combining (2.12) with (2.13), we have

$$\langle \Gamma_+ \mathbf{e}_k, \mathbf{e}_j \rangle = \frac{\lambda_k}{\lambda_j + \lambda_k} \langle V\mathbf{e}_j, \mathbf{e}_k \rangle = \lambda_k \langle \mathbf{e}_k, \mathbf{e}_j \rangle \quad (j \neq k). \tag{2.14}$$

From (2.12) with $k = j$, we have

$$\langle \Gamma_+ \mathbf{e}_j, \mathbf{e}_j \rangle = -\lambda_j \langle \mathbf{e}_j, \mathbf{e}_j \rangle + \langle V\mathbf{e}_j, \mathbf{e}_j \rangle. \tag{2.15}$$

Next, from (2.9),

$$\langle (I - 2W)\Gamma_- \mathbf{e}_k, \mathbf{e}_j \rangle = -\lambda_k \langle (I - 2W)\mathbf{e}_k, \mathbf{e}_j \rangle = -\lambda_k \langle \mathbf{e}_k, \mathbf{e}_j \rangle + \frac{2\lambda_k}{\lambda_j + \lambda_k} \langle V\mathbf{e}_j, \mathbf{e}_k \rangle.$$

Hence, using (2.13) and (2.14),

$$\langle (I - 2W)\Gamma_- \mathbf{e}_k, \mathbf{e}_j \rangle = \lambda_k \langle \mathbf{e}_k, \mathbf{e}_j \rangle = \langle \Gamma_+ \mathbf{e}_k, \mathbf{e}_j \rangle \quad (j \neq k),$$

and, from (2.9), and (2.15),

$$\langle (I - 2W)\Gamma_- \mathbf{e}_j, \mathbf{e}_j \rangle = -\lambda_j \langle \mathbf{e}_j, \mathbf{e}_j \rangle + \langle V\mathbf{e}_j, \mathbf{e}_j \rangle = \langle \Gamma_+ \mathbf{e}_j, \mathbf{e}_j \rangle.$$

Hence, the lemma is proved. ■

(c) *Beginning the link with Theorem 1.2*

Let S now denote the set $\{b_0, b_1, b_2, \ldots, b_n\}$ ($b_0 = 0$) from §1c, and let m^- be the measure there. Let

$$c_k := (b_k - b_{k+1})^{-1} \quad (0 \leq k \leq n - 1),$$

and let C be the *symmetric* $S \times S$ matrix

$$C := \begin{pmatrix} -c_0 & c_0 & 0 & . & 0 & 0 & 0 \\ c_0 & -(c_0 + c_1) & c_1 & . & 0 & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & c_{n-2} & -(c_{n-2} + c_{n-1}) & c_{n-1} \\ 0 & 0 & 0 & . & 0 & c_{n-1} & -c_{n-1} \end{pmatrix} \tag{2.16}$$

Let U be the $S \times S$ matrix with

$$U_{00} := 1, \quad U_{ij} := 0 \text{ for all other pairs } (i, j). \tag{2.17}$$

Let $M := \operatorname{diag}\{m(b_i)\}$, and

$$V := M^{-1}U, \quad Q := M^{-1}C. \tag{2.18}$$

We factorize $(t^2 I + tV + Q)$ as in §2a, with m^- now playing the rôle of m there. Let f_θ be as in (1.5). Consider f_θ on $(-\infty, 0]$. Clearly f_θ is a piecewise-linearization of its values on S, and f_θ is constant on $(-\infty, b_n]$. (Recall that

f_θ must be bounded.) We suppose that the eigenvectors $\mathbf{e}_j$ of Γ_- are all normalized (see Note below) so that

$$\mathbf{e}_j(0) = 1, \tag{2.19}$$

whence, since $V(0,0) = m(0)^{-1}$, we have

$$\langle V\mathbf{e}_j, \mathbf{e}_k \rangle = 1 = \langle \mathbf{e}_j, V\mathbf{1} \rangle, \tag{2.20}$$

$\mathbf{1}$ here denoting the vector $(1, 1, \ldots, 1)^T$ on S.

Note. Suppose that $\mathbf{e}_j$ is an eigenvector of Γ_- such that $\mathbf{e}_j(0) = 0$. Then $V\mathbf{e}_j = \mathbf{0}$, and

$$\mathbf{0} = (\lambda_j^2 I - \lambda_j V + Q)\mathbf{e}_j = (\lambda_j^2 I + Q)\mathbf{e}_j.$$

But the tri-diagonal nature of Q would now imply that $\mathbf{e}_j = \mathbf{0}$. Hence the normalization (2.19) is possible. □

Equation (1.5) on S now reads

$$(\theta^2 I - Q)\mathbf{f}_\theta = V\mathbf{1} \qquad (\theta \neq 0), \tag{2.21}$$

$\mathbf{f}_\theta$ denoting the restriction of f_θ to S as a column vector on S.

(d) The magic key to duality

Since $Q = Q^*$ and $\Gamma_-^2 + V\Gamma_- + Q = 0$, we have from (2.21) and (2.20), for $\theta \neq 0$, and with $\langle \cdot, \cdot \rangle$ as the m^- scalar product,

$$\begin{aligned} 1 = \langle \mathbf{e}_j, V\mathbf{1} \rangle &= \langle \mathbf{e}_j, (\theta^2 I - Q)\mathbf{f}_\theta \rangle \\ &= \langle (\theta^2 I - Q)\mathbf{e}_j, \mathbf{f}_\theta \rangle = \langle (\theta^2 I + \Gamma_-^2 + V\Gamma_-)\mathbf{e}_j, \mathbf{f}_\theta \rangle \\ &= (\theta^2 + \lambda_j^2)\langle \mathbf{e}_j, \mathbf{f}_\theta \rangle - \lambda_j \langle V\mathbf{e}_j, \mathbf{f}_\theta \rangle = (\theta^2 + \lambda_j^2)\langle \mathbf{e}_j, \mathbf{f}_\theta \rangle - \lambda_j f_\theta(0), \end{aligned}$$

whence

$$\begin{aligned} \langle \mathbf{e}_j, \mathbf{f}_\theta \rangle &= \frac{\lambda_j}{\theta^2 + \lambda_j^2} f_\theta(0) + \frac{1}{\theta^2 + \lambda_j^2} \\ &= \int_0^\infty \mathrm{e}^{-\lambda_j y} \left\{ f_\theta(0) \cos\theta y + \theta^{-1} \sin\theta y \right\} \, \mathrm{d}y \\ &= \int_0^\infty \mathrm{e}^{-\lambda_j y} f_\theta(y) \, \mathrm{d}y. \end{aligned}$$

But

$$\mathbf{f}_\theta(\cdot) = \int_0^\infty \pi(\cdot, y) f_\theta(y) \, \mathrm{d}y,$$

and we have the following theorem.

THEOREM 2.1. *The map $\pi(\cdot, y)$ on S is characterized by*

$$\langle \mathbf{e}_j, \pi(\cdot, y)\rangle = \mathrm{e}^{-\lambda_j y},$$

so that if ε_j is the function in $(0, \infty)$ with

$$\varepsilon_j(y) = \mathrm{e}^{-\lambda_j y},$$

then with $(\Pi^{-+})^* : L^2(S, m) \to L^2(0, \infty)$ the Hilbert-space adjoint of Π^{-+}, we have

$$(\Pi^{-+})^* \mathbf{e}_j = \varepsilon_j.$$

This analytic expression of the 'duality principle' really states that the *other* half-winding operator Π^{+-} is just $(\Pi^{-+})^*$. For we have

$$\operatorname{sgn}(\cdot)^{-1} \mathcal{A} \varepsilon_j = +\tfrac{1}{2}\lambda_j^2 \varepsilon_j \quad \text{on } (0, \infty) \cup S,$$

if we regard $\mathbf{e}_j$ as the extension to S of ε_j.

Remark. The algebra which we have developed leads to new insights into the algebraic study of various rational functions in [LMRW1]. This will be taken up elsewhere.

(*e*) *Completion of proof of Theorem 1.2*

Suppose that $-\lambda$ is an eigenvalue of Γ_- and that $\Gamma_- \mathbf{e} = -\lambda \mathbf{e}$. Then, by equation (2.4),

$$(\lambda^2 I + Q)\mathbf{e} = \lambda V \mathbf{e}.$$

The tri-diagonal nature of Q now implies that the eigenspace of Γ_- corresponding to eigenvalue $-\lambda$ is 1-dimensional, whence all eigenvalues of Γ_- are distinct.

Let W denote the full-winding operator

$$W := \Pi^{-+}(\Pi^{-+})^*,$$

so that $W = W^*$ as an $S \times S$ matrix. We need only show that

$$\|W\|_2 = \tfrac{1}{2}.$$

Now, we have, using Theorem 2.1 and equation (2.20),

$$\begin{aligned}\langle \mathbf{e}_k, W\mathbf{e}_j\rangle &= \langle \mathbf{e}_k, \Pi^{-+}(\Pi^{-+})^*\mathbf{e}_j\rangle_{m^-} = \langle (\Pi^{-+})^*\mathbf{e}_k, (\Pi^{-+})^*\mathbf{e}_j\rangle_{m^+}\\ &= \langle \varepsilon_k, \varepsilon_j\rangle_{m^+} = \int_0^\infty \mathrm{e}^{-\lambda_j y}\mathrm{e}^{-\lambda_k y}\,\mathrm{d}y = \frac{1}{\lambda_j+\lambda_k} = \frac{\langle V\mathbf{e}_j, \mathbf{e}_k\rangle}{\lambda_j+\lambda_k}.\end{aligned}$$

Since we have proved that Γ_- has distinct eigenvalues, we can now apply Lemma 2.1 to deduce that

$$(I-2W)\Gamma_- = \Gamma_+.$$

Now $\Gamma_+ \in \mathcal{Q}_0$ and so $\Gamma_+\mathbf{1} = 0$, so that the Perron-Frobenius eigenvalue of Γ_+ is 0. Thus there exists a positive row vector $\mathbf{p}$ on S with $\mathbf{p}\Gamma_+ = \mathbf{0}$. Since Γ_- is invertible, we have $\mathbf{p}(I-2W) = \mathbf{0}$, whence

$$\mathbf{p}W = \tfrac{1}{2}\mathbf{p}.$$

But W is a positive matrix, and hence its Perron-Frobenius eigenvalue is $\frac{1}{2}$. The proof of Theorem 1.2 is complete. ∎

Our method extends to the situation where m^+ is Lebesgue measure on $\mathbb{R}\setminus[a,b]$, m^- is atomic on $[a,b]$ with $m(a)>0$ and $m(b)>0$, and where sgn is replaced by the function which is 1 on $\mathbb{R}\setminus[a,b]$ and -1 on $[a,b]$. That is why we stated Lemma 2.1 with a slightly general boundary operator V.

3. A brief return to the canonical case

It is planned to resolve in sequels some of the many mysteries brought to light by this paper. Here, we indicate briefly the relation between §2 and the canonical case. So, return to the assumptions of Theorem 1.1. A moment's puzzlement and a little thought show what must happen ... . See the Appendix (§4) for details.

The σ-finite measure $\mathcal{P}$ (unique up to scalar multiples) with $\mathcal{P}\Gamma_+ = 0$ has density p relative to m^- satisfying

$$p(x) \equiv |x|^{-\alpha} \quad \text{on } (-\infty, 0],$$

where, as at (1.17),

$$0<\alpha<1 \text{ and } \tan\tfrac{1}{2}\pi\alpha = c^{-1}.$$

We have if $W := \Pi^{-+}(\Pi^{-+})^*$,

$$pW(x) = \int_z \int_y m^-(\mathrm{d}z)p(z)\pi(z,y)m^+(\mathrm{d}y)\pi(x,y) = \Lambda^2\, p(x), \tag{3.1}$$

where $\Lambda = 2^{-\frac{1}{2}}$. Now, $p > 0$ almost everywhere (m^-). By Schur's test (see Toland & Williams (1998) and references therein), this positivity of p, the positivity of π and (3.1) already imply that $\|\Pi^{-+}\|_2 \le \Lambda = 2^{-\frac{1}{2}}$. We know from Theorem 1.1 that if $c \neq 1$, then we can find a better 'test function' than p, namely $x^{-\frac{1}{2}}$, leading to the best-possible norm estimate.

But the hope is to use the Schur test (and some harder analysis!) to prove that $\|\Pi^{-+}\|_2 \le 2^{-\frac{1}{2}}$ for the general [LMRW1] situation, and, more importantly, clarify the rôle of p. There are many reasons why the analysis will be tricky. Note, for example, that the integral on the left-hand side of the equation (4.2) (in the Appendix) which determines p, is only conditionally convergent. To end, we quote a theorem from the Toland-Williams paper.

THEOREM 3.1. (Toland & Williams (1998)) *Suppose that*

$$m = \text{Leb on } (0,\infty); \quad m^- = c^2 \times \text{Leb on } [-1,0], \quad m^- = 0 \text{ on } (-\infty,-1).$$

Then

$$\|\Pi^{-+}\|_2 = \begin{cases} \frac{(2c)^{\frac{1}{2}}}{1+c} & \textit{if } c \le 1, \\ 2^{-\frac{1}{2}} & \textit{if } c \ge 1. \end{cases}$$

4. Appendix

We explain (briefly and heuristically) how to calculate Γ_+ and p for the canonical case by standard probabilistic methods found for example in Revuz & Yor (1991) or in Rogers & Williams (1987).

One has to be extremely careful about normalizations, especially in more difficult cases than this one. The probabilistically correct equation for an eigenvector $\mathbf{g}_\theta$ of Γ_+ in the 'atomic' situation is

$$(\tfrac{1}{2}\theta^2 + \tfrac{1}{2}\theta V + \tfrac{1}{2}Q)\mathbf{g}_\theta = 0, \qquad \Gamma_+\mathbf{g}_\theta = -\theta\mathbf{g}_\theta, \tag{4.1}$$

$\theta \ge 0$ chosen to make the left-most matrix singular of course. The correct Markov chain on S for the [KW] work has Q-matrix $\frac{1}{2}Q$ because of the $\frac{1}{2}$

in $\mathcal{A} = \frac{1}{2}\dfrac{\mathrm{d}}{\mathrm{d}m}\dfrac{\mathrm{d}}{\mathrm{d}x}$. The $\frac{1}{2}V$ focuses on a 'local time' at 0 which is half of the occupation density relative to m^-.

In the canonical case, the Markov chain on S is replaced by a scaled reflecting Brownian motion on $(-\infty, 0]$. We may regard X as given by an Itô equation

$$\mathrm{d}X = c^{-1}\mathrm{d}\beta - \mathrm{d}L^X,$$

L^X, a normalization of the local time of X at 0, being a continuous adapted increasing process which grows only when X is at 0. Alternatively, we can think of X as $c^{-1}R$, where R is a standard reflecting Brownian motion on $(-\infty, 0]$. It is well known that

$$R + L^R \text{ is a martingale}$$

where

$$\begin{aligned} L_t^R &= \lim_{\varepsilon \downarrow 0} (2\varepsilon)^{-1} \mathrm{meas}\{s : 0 \le s \le t;\ R_s \in [-\varepsilon, 0]\} \\ &= \lim_{\varepsilon \downarrow 0} (2\varepsilon)^{-1} \mathrm{meas}\{s : 0 \le s \le t;\ X_s \in [-c^{-1}\varepsilon, 0]\}, \end{aligned}$$

whence, putting $\varepsilon = c\delta$,

$$L_t^X = c^{-1}L_t^R = \lim_{\delta \downarrow 0} (2c^2\delta)^{-1} \mathrm{meas}\{s : 0 \le s \le t;\ X_s \in [-\delta, 0]\},$$

so that L^X *is* half the occupation-density local time at 0 for X relative to m^-.

Let B be a Brownian motion independent of X. Then, according to the [KW] philosophy, Γ_+ is the infinitesimal generator of the time-transform of X via the fluctuating additive functional $\varphi := L^X + B$. To be precise, Γ_+ is the generator of $\{X(\tau_t)\}$ where $\tau_t := \inf\{u : \varphi_u > t\}$. For $\theta > 0$, we seek a bounded g_θ such that

$$\exp(\theta\varphi_t) g_\theta(X_t) \quad \text{ is a martingale.}$$

Itô's formula shows that we require

$$\theta g_\theta(0) - g_\theta'(0) = 0, \qquad \tfrac{1}{2}\theta^2 g_\theta + \tfrac{1}{2}c^{-2}g_\theta'' = 0 \text{ on } (-\infty, 0),$$

the first ('elastic boundary condition') to annihilate the $\mathrm{d}L^X$ term. (Of course, the pair of equations just obtained forms the analogue of the first equation at (4.1).) Hence, we have (up to constant multiples)

$$g_\theta(x) = c\ \cos\theta cx + \sin\theta cx \quad (x \leq 0).$$

By the stopping-time principle,

$$g_\theta \in \mathcal{D}(\Gamma_+) \quad \text{and} \quad \Gamma_+ g_\theta = -\theta g_\theta.$$

The condition $\mathcal{P}\Gamma_+ = 0$ leads to $\mathcal{P}\Gamma_+ g_\theta = 0 = -\theta \mathcal{P} g_\theta \ \ (\theta > 0)$, so that

$$\mathcal{P} g_\theta = \int_{-\infty}^{0} m^-(\mathrm{d}x) p(x) g_\theta(x) = 0 \qquad (\theta > 0), \tag{4.2}$$

and contour integration verifies that $p(x) = |x|^{-\alpha}$ is a solution. That (3.1) holds with $\Lambda = 2^{-\frac{1}{2}}$ is easily verified via the formulae used to prove Theorem 1.3.

In the case of Theorem 3.1 studied by Toland and Williams, we have $p(x) = |\sin\frac{1}{2}\pi x|^{-\alpha}$, where, once more, $\tan\frac{1}{2}\pi\alpha = c^{-1}$. The Toland-Williams paper contains a discussion of the relation to the work here.

Francis Burstall, Ivan Graham, John Toland and Chris Rogers richly deserve our thanks. DSM is supported by an EPSRC Research Studentship.

References

Dym, H. & McKean, H.P. 1976 *Gaussian processes, function theory, and the inverse spectral problem.* NewYork: Academic Press.

Hardy, G.H., Littlewood, J.E. & Pólya, G. 1934 *Inequalities.* 2nd edn, Cambridge: Cambridge University Press.

Kennedy, J. & Williams, D. 1990 Probabilistic factorization of a quadratic matrix polynomial. *Math. Proc. Camb. Phil. Soc.* **107**, pp. 591–600.

London, R.R., McKean, H.P., Rogers, L.C.G. & Williams, D. 1982a A martingale approach to some Wiener-Hopf problems. I. *Séminaire de probabilités XVI* (ed. J. Azéma & M. Yor). *Lecture notes in mathematics*, **920**, pp. 41–67, Berlin: Springer.

London, R.R., McKean, H.P., Rogers, L.C.G. & Williams, D. 1982b A martingale approach to some Wiener-Hopf problems. II. *Séminaire de probabilités XVI* (ed. J. Azéma & M. Yor). *Lecture notes in mathematics*, **920**, pp. 68–90, Berlin: Springer.

McKean, H.P. 1963 A winding engine for a resonator driven by white noise. *J. Math. Kyoto Univ.* **2**, pp. 227–235.

Revuz, D. & Yor, M. 1991 *Continuous Martingales and Brownian Motion.* Berlin: Springer-Verlag.

Rogers, L.C.G. & Williams, D. 1984 A differential equation in Wiener-Hopf theory. *Stochastic analysis and applications* (ed. A. Truman & D. Williams). *Lecture notes in mathematics*, **1095**, pp. 187–199, Berlin: Springer.

Rogers, L.C.G. & Williams, D. 1979 *Diffusions, Markov Processes, and Martingales. Vol 1. Foundations.* 2nd edn, Chichester: Wiley.

Rogers, L.C.G. & Williams, D. 1987 *Diffusions, Markov Processes, and Martingales. Vol 2. Itô Calculus.* Chichester: Wiley.

Williams, D. 1991 Some aspects of Wiener-Hopf factorization. *Phil. Trans. R. Soc. Lond. A*, **335**, pp. 593–608.

Williams, D. & Toland, J.F. 1998 On the Schur test for L_2-boundedness of positive integral operators with a Wiener-Hopf example. (submitted).