Rejecta Mathematica

Volume 2, Number 1 — June 2011

ISSN 1948-8351

math.rejecta.org

Rejecta Mathematica is an open access, online journal that publishes papers that have been rejected from peer-reviewed journals in the mathematical sciences. Each paper is accompanied by an open letter from its authors discussing the original review process and stating the case for its value to the research community.

Articles

Classification via incoherent subspaces 1
Karin Schnass and Pierre Vandergheynst

Explanation of low Hurst exponent for Riemann zeta zeros 19
Oruganti Shanker

On the distribution of Carmichael numbers 24
Aran Nayebi

Extended real number system in measure theory 44
Satish Shirali

The problematic nature of Gödel's theorem 49
Hermann Bauer and Christoph Bauer

Scattering, determinants, hyperfunctions in relation to $\frac{\Gamma(1-s)}{\Gamma(s)}$ 59
Jean-François Burnol

ARTICLE

Karin Schnass and Pierre Vandergheynst
Classification via incoherent subspaces
***Rejecta Mathematica*, Vol. 2, No. 1, pp. 1–18, June 2011**
2011 Rejecta Publications
math.rejecta.org

An open letter concerning

Classification via incoherent subspaces

Karin Schnass

It all started with a talk by Efi, a fellow PhD-student (now Dr. Effrosyni Kokiopoulou), on classification and dimensionality reduction which gave me an idea. Classification was foreign territory to me so I discussed it with her and in a moment when my other projects where going nowhere, I sat down and thought about my idea in detail. As a result I developed a mathematical model of how signals in different classes could be represented and regarding how you could then try to classify them. And I was happy for a while. Unfortunately the characterization I had come up with was at the same time too complicated and too vague to be actually calculable. So when I wanted to test the idea on some real data I had to simplify to a problem I could solve. The implementation with all the optimizations in the intermediate steps turned out to be a nightmare but after many long months of suffering the algorithm spit out something that seemed to work pretty well, and I wrote up my results and included them in my thesis.

As a next step we submitted the paper to *IEEE Trans. on Pattern Analysis and Machine Intelligence*. The reviews were late and ambiguous. Reviewer 1 thought that the paper was garbage, because it did not beat state of the art and was mathematically unsound. In his opinion putting N d-dimensional vectors as columns into a matrix did not result in a $d \times N$ matrix. Reviewer 2 was not too thrilled but gave us the benefit of doubt. Reviewer 3 seemed to understand and appreciate the idea and gave a lot of useful comments. We were asked to make a major rewrite. We clarified the model on the coefficients, included test results on another database, emphasized that computational complexity should be a criterion when talking about state of the art, etc., and resubmitted. We got rejected — the general opinion seemed to be that it was too mathy and did not fit into the narrow view of 2 out of 3 reviewers.

So we thought we would try to submit to a journal with a broader scope: *IEEE Trans. on Signal Processing*. The reviews were again late and ambiguous. Reviewer 1 and 2 again thought that it was too mathy, though the idea novel and the paper technically sound. Reviewer 3 again understood and appreciated the idea and made a lot of useful comments. The general verdict was to reject with encouragement to resubmit. Since we do not think that another rewrite will improve the paper, we made the changes requested by Reviewer 3 and now hope that the paper will make a nice contribution to *Rejecta Mathematica*. After all, it was mainly rejected because it contained "too much math".

Finally, I have learned one thing: never try to introduce new ideas into an old field that is not your own.

Classification via incoherent subspaces

Karin Schnass*and Pierre Vandergheynst†‡

Abstract

This article presents a new classification framework that can extract individual features per class. The scheme is based on a model of incoherent subspaces, each one associated to one class, and a model on how the elements in a class are represented in this subspace. After the theoretical analysis an alternate projection algorithm to find such a collection is developed. The classification performance and speed of the proposed method is tested on the AR and YaleB databases and compared to that of Fisher's LDA and a recent approach based on on ℓ_1 minimisation. Finally connections of the presented scheme to already existing work are discussed and possible ways of extensions are pointed out.

1 Introduction

A general approach in classification is to select features of the signal at hand and to get a decision by comparing them to the equivalent features of already labelled signals with a simple classifier like nearest neighbour, e.g. [3], or nearest subspace, cp [9]. This of course raises the question which features to take. For face recognition, which is the example we will use here, some classic and simple, because linear, features are Eigen, [19], Fisher, [6], or Laplace features, [7]. However, as these classifiers are very simple and the features not adjusted to them, their performance is somehow disappointing, and researchers turned to the development of more complicated nonlinear features and kernel methods, [10, 15].

Here we start from the point of view that the potential of linear methods and simple classifiers is not exhausted. In order to achieve better results, we propose to give up the uniformity of features over classes and mix the feature selection with the classifier. To motivate the idea of class specific features let us have a look at classical nearest neighbour (NN) and nearest subspace (NS) classification using linearly selected features and give it a new interpretation.

Assume we have N already labelled training signals $y \in \mathbb{R}^d$ belonging to c classes, where each class i contains n_i elements, i.e. $\sum_i n_i = N$. We denote the j-th signal in class i as y_i^j, $i = 1 \dots c$, $j = 1 \dots n_i$. For each class i we collect all its training signals as columns in the $d \times n_i$ class matrix Y_i, i.e. $Y_i = (y_i^1 \dots y_i^{n_i})$, and these class matrices in turn are combined into a big $d \times N$ data matrix $Y = (Y_1 \dots Y_c) = (y_1^1 \dots y_1^{n_1} \dots y_c^1 \dots y_c^{n_c})$. Given a new signal y_{new} the goal is to decide which class it belongs to with the help of the already labelled training signals.

The classical first step is to select relevant features f_{new} from y_{new} via a linear transform A, where

*Karin Schnass is with the Johann Radon Institute for Computational and Applied Mathematics (RICAM), Altenbergerstrasse 69, 4040 Linz, Austria, E-mail: karin.schnass@oeaw.ac.at

†Pierre Vandergheynst is with the Signal Processing Laboratory 2, Swiss Federal Institute of Technology (EPFL), Station 7, 1015 Lausanne, Switzerland, E-mail: pierre.vandergheynst@epfl.ch

‡This work was partly supported by NSF grant 200021-117884/1

A is a $d \times d$ matrix of rank $r \leq d$.

$$\textbf{Feature Selection:} \;\; f_{new} = Ay_{new}. \tag{1}$$

The exact shape of the transform is determined by the training signals and their labels. For instance for Fisher's LDA A is chosen as the orthogonal projection that maximises the ratio of between-class scatter to that of within-class scatter, [6].
In the second step these features are compared to the features $f_i^j := Ay_i^j$ of the training signals y_i^j. In case of the nearest neighbour classifier this means that the new signal will get the label of the training signal which has features that maximally correlate with the features of the new signal, ie.

$$\textbf{NN Labelling:} \;\; i_{new} = \operatorname*{argmax}_i \max_j |\langle f_i^j, f_{new}\rangle|. \tag{2}$$

If by analogy we define the feature matrix for a class i as $F_i = AY_i$ we can rewrite the expression, for which we are seeking the maximal argument, and combine the feature selection with the labelling step,

$$\begin{aligned}\max_j |\langle f_i^j, f_{new}\rangle| &= \max_j |\langle Ay_i^j, Ay_{new}\rangle| \\ &= \max_j |\langle A^\star Ay_i^j, y_{new}\rangle| \\ &= \|(A^\star AY_i)^\star \, y_{new}\|_\infty,\end{aligned} \tag{3}$$

where the matrix $M^\star$ denotes the transpose of M, the p-norm of a vector is defined by $\|v\|_p := (\sum_k v(k)^p)^{1/p}$ for $1 \leq p < \infty$ and $\|v\|_\infty := \max_k |v(k)|$ and the qp-norm of a matrix by $\|M\|_{q,p} = \max_{\|v\|_q=1} \|Mv\|_p$. Thus another way of looking at the classification procedure is to say that for every class we have a set of sensing signals $A^\star Ay_i^j$ and the new signal belongs to the class which has the sensing signal closest to it. From this point of view we also see that the scheme will work stably only if two conditions are fullfilled. Every new signal is well represented by one vector in its class, i.e. a lot of its energy is captured by the projection on one vector, and no two sensing vectors from different classes are the same or close to each other, i.e.

$$\max_{i\neq k,j,l} |\langle A^\star Ay_i^j, A^\star Ay_k^l\rangle| = \|(A^\star AY_i)^\star (A^\star AY_k)\|_{1,\infty} \leq \mu. \tag{4}$$

Let us do the same analysis for the nearest subspace classifier. Again the features of the new signal are compared to those of the training signals. For each class the features of the training signals in it span a subspace and the new signal will get the label of the class for which the orthogonal projection of the features of the new signal on the corresponding subspace has the highest energy. Let Q_i be an orthonormal system[1] spanning the subspace for class i, then i.e.

$$\textbf{NS Labelling:} \;\; i_{new} = \operatorname*{argmax}_i \|Q_i^\star f_{new}\|_2. \tag{5}$$

Again we can combine the feature selection with the labelling by manipulating the expression, we want to maximise,

$$\|Q_i^\star f_{new}\|_2 = \|(A^\star Q_i)^\star y_{new}\|_2.$$

If we compare to NN classification we see that again for every class we get a set of sensing signals, the columns of the matrix $A^\star Q_i$, and that the new signal belongs to the class for which the sensing

[1] Q_i can for instance be found via a (reduced) qr-decomposition of AY_i

signals can take out the most energy (or for which the biorthogonal system to $(AQ_i)^\star$ provides the best representation). Again this leads to two conditions for the classification to work, which are however more complex. First every new signal should be comparatively well represented in the biorthogonal system $(Q_i^\star A)^\dagger$ determined by its class and second no signal which is in the span of the sensing signals of one class should be well representable in the biorthogonal system of another class.

$$\max_{i\neq k}\max_{\|x\|=1}\|(A^\star Q_i)^\dagger A^\star Q_j x\|_2 = \|(A^\star Q_i)^\dagger A^\star Q_j x\|_{2,2} \leq \mu. \tag{6}$$

Summarising our findings for both nearest neighbour and nearest subspace classification we see that in both cases for every class we have a set of sensing signals or a subspace defined by the feature selection transform. There is a model how signals from this class are represented by this subspace, which implicitly determines which norm is used for the classification, $\|\cdot\|_\infty$ for NN, $\|\cdot\|_2$ for NS, and at the same requires that the interaction of subspaces measured by a corresponding matrix norm is small, i.e. that they are incoherent.

The classification scheme presented in this paper is based on the following idea. We give up the restriction that the subspaces associated to each class are generated canonically as a function of the feature selection transform and the training samples, i.e. $A^\star AY_i$ in the case of nearest neighbour classification, but generate them individually. This idea can also be motivated using the example of face recognition, to which we will apply our scheme later. Uniform feature extraction would mean realising that in general the most relevant parts of a face are the regions of the eyes, nose and mouth. Thus in order to classify a person we would focus on the eyes, nose and mouth regions while ignoring the hairstyle and comparing them to the eyes, nose and mouth regions of all the candidates. While this makes sense in general it will fail as soon as the set of candidates contains identical twins which can only be distinguished by the birth mark one has on his cheek. So while for most people the cheek is not a very distinguishing feature for the twins it is and it would be better to remember for them the cheek instead of for instance the nose. Even without the extreme example of the identical twins individual features are natural considering that the people we meet every day all have eyes, mouths and noses but not all of them have distinguishing eyes, mouths and noses. Instead they may have distinguishing birthmarks, scars, chins, etc. and a representation using these features will characterise them well but nobody else.

In the next section we will introduce the mathematical framework on which we base our classification scheme. It consists of a model of subspaces associated to each class and a model of how the elements in this class are represented in this subspace, which together lead to a natural choice of the norm we have to use for the classification and an incoherence requirement on the subspaces. In Section 3 we will develop a comparatively simple algorithm to learn these subspaces from the training signal, which we will use to classify faces in Section 4. In the last section we summarise our findings, point out connections to related approaches and outline possibilities for future work.

2 Class Model

The most general model for the subspaces we can think of is to ascribe to every class i a set of s_i vectors f_i^j, $j = 1\dots s_i$, which are collected as columns in the matrix $F_i = (f_i^1 \dots f_i^{s_i})$. These correspond to the features that characterise elements of this class, so every element y_i in class i can be written as a combination of these class specific features with coefficients x_i and some residual r_i, orthogonal to the feature span,

$$y_i = F_i x_i + r_i, \qquad r_i^k \perp sp(F_i). \tag{7}$$

The condition that features of a class well characterise the elements in it translates into a property of the coefficients x_i, i.e. when measuring their strength in some norm it is higher than the strength of the coefficients we would obtain trying to represent the element by features of the wrong class. Since without further restrictions on the set of features per class it is not straightforward to calculate the coefficients of the best representation of signal in a class, we will sacrifice generality for simplicity and for the rest of the analysis assume that for every class we have the same number of features s and that they form an orthonormal system, i.e. $F_i^\star F_i = I_s$. We will point out how to deal with the more general situation in the last section. Given that the features of each class form an orthonormal system we can easily calculate the coefficients of the best representation of a general signal y in class i as $F_i^\star y$. The question is now how should we measure these coefficients in order to correctly classify our images, i.e. which norm $\|\cdot\|$ should we choose such that for all y_i in class i we have

$$\frac{\|F_j^\star y_i\|}{\|F_i^\star y_i\|} < 1, \, \forall j \neq i. \tag{8}$$

To answer this question we will introduce three models on the coefficients, each leading to a certain p-norm as optimal measure.

2.1 1-Sparse Coefficients

Assume that all signals we want to classify can be well represented by one element of one class, i.e.

$$y_i = F_i x_i + r_i \;\text{ with }\; \|x_i\|_0 = 1, \tag{9}$$

where $\|\cdot\|_0$ counts the number of non-zeros entries. An example for this situation would be trying to sort pictures of monkeys, snails, cucumbers and broccoli into animal and vegetable pictures. Even though monkeys and snails are both animals their shapes are very different, meaning that we can think of them as orthogonal, and the same goes for the shapes of cucumbers and broccoli in the other class. Let x be the absolute value of the only non-zero component of the coefficients x_i. We immediately see that whatever p-norm we apply using the correct class the response is always equal to x, $\|F_i^\star y_i\|_p = \|x_i\|_p = x$. Therefore to find out which p-norm is best we will use a trick that involves estimating the ratio we need to be smaller than 1 for successful classification with the triangular equation and a matrix norm bound. So for general $1 \leq p, q \leq \infty$ we get,

$$\frac{\|F_j^\star y_i\|_p}{\|F_i^\star y_i\|_p} \leq \|F_j^\star F_i\|_{q,p} \frac{\|x_i\|_q}{\|x_i\|_p} + \frac{\|F_j^\star r_i\|_p}{\|x_i\|_p}. \tag{10}$$

In the special case where the coefficients are 1-sparse and thus $\|x\|_q = \|x\|_p, \forall p, q$, this means that

$$\frac{\|F_j^\star y_i\|_p}{\|F_i^\star y_i\|_p} \leq \|F_j^\star F_i\|_{q,p} + \frac{\|F_j^\star r_i\|_p}{x}.$$

The smallest qp norm of a matrix is obtained when $p = \infty$ and $q = \infty$. Then it corresponds to the maximal absolute entry of the matrix $F_j^\star F_i$, i.e. the maximal absolute correlation between two features from different classes. Since in that case also the response from the residual $\|F_j^\star r_i\|_p$ is minimal we get the best bound choosing the ∞-norm for the classification. Summarising our findings we see that in case of a sparse model on the coefficients, the ∞ norm is optimal and that the incoherence requirement we get for classification to work stably is that no two features from two different classes are too similar, but it does not matter if a feature is moderately close to all features in a different class or even representable by them. Thinking to the example of the animal vs. vegetable pictures this means that even though you can approximate the shape of a snail combining the shape of the cucumber and the broccoli, classification using the ∞ norm will work well because no animal shape alone closely resembles a vegetable shape and vice versa.

2.2 Flat Coefficients

Let us now assume the completely opposite distribution of the coefficients, i.e. to represent one element in a class we need to combine all features of that class with equal magnitudes, i.e.

$$y_i = F_i x_i + r_i \text{ with } |x_i(k)| = x, \text{ for } k = 1 \dots s. \tag{11}$$

An example would be trying to label pictures of national flags and with the corresponding countries. For simplicity assume that the only flags in question are those of the Netherlands, Germany, Estonia, Lithuania and Gabon, which all consist of three horizontal stripes in various colours, i.e. red, white and blue for the Netherlands, black, red and yellow for Germany, blue, black and white for Estonia, yellow, green and red for Lithuania and green, yellow and blue for Gabon, cp. Figure 1. Good features in this example are the colours of the stripes. Each national flag has its three distinctive colours which appear in an equal amount but are not exclusive to this flag.

Netherlands Germany Estonia Lithuania Gabon

Figure 1: National Flags

In this case we get the maximal response from the correct class when choosing $p = 1$, i.e. $\|x\|_1 = sx$. Also from Inequality (10) we see that using $q = \infty$ and $p = 1$ gives a very beneficial bound.

$$\frac{\|F_j^\star y_i\|_1}{\|F_i^\star y_i\|_1} \leq \frac{\|F_j^\star F_i\|_{\infty,1}}{s} + \frac{\|F_j^\star r_i\|_1}{sx}.$$

Remembering that $\|F_j^\star F_i\|_{\infty,1}$ is smaller than the absolute sum of all the correlations between features in one class and features in another class we get a less sharp version of the above bound,

$$\frac{\|F_j^\star y_i\|_1}{\|F_i^\star y_i\|_1} \leq \frac{\sum_{k,l} |\langle f_i^k, f_j^l \rangle|}{s} + \frac{\|F_j^\star r_i\|_1}{sx}.$$

which shows that for the case of flat coefficients we have a quite different coherence constraint. Even if a few features in a class are very close to features in another class or actually the same this is not a problem as long as the majority of features from two different classes are not very correlated.

In the example of the flags this means that even though two different national flags might share up to two colours, as long as we take into account that all three colours have to appear to the same degree, we can still identify the country from a picture of the flag.

2.3 Unstructured Coefficients

The last case we are going to discuss is probably the most common and concerns coefficients which follow neither of the two extreme distributions discussed above or where the exact distribution is unknown. An example is the task of face recognition, i.e. identifying a person from a picture. Obvious features in this case are noses, eyes and mouths. In any picture most of these features will be visible but their strength will largely depend on the facial expression and lighting conditions. To choose a good p-norm for the classification in this case, we again bound the norm ratio we need to be small.

$$\frac{\|F_j^\star y_i\|_p}{\|F_i^\star y_i\|_p} \leq \frac{\|F_j^\star F_i x_i\|_p}{\|x_i\|_p} + \frac{\|F_j^\star r_i\|_p}{\|x_i\|_p}. \tag{12}$$

Since we do not have information about the shape of the coefficients, the first term on the right hand side can be as big as $\|F_j^\star F_i\|_{p,p} = \max_{\|x\|_p=1} \|F_j^\star F_i\|_p$. Taking into account the orthogonality of the features in the matrices F_i, we see that for $p = 2$ this term can only be equal to one if two classes overlap, meaning that there is a signal whose features in its own class can be represented by features in a different class. For $p = 1$, $p = \infty$, however, the corresponding term is equal to the maximum absolute column/row sum of the $F_j^\star F_i$ and it can be easily seen that this can be larger than one, even if for no signal the features in its own class can be fully represented by features in a different class. Similar results hold for all other $p \neq 2$, thus making $p = 2$ the best choice in this case. Observe also that $p = 2$ corresponds to measuring the energy captured by the features of a class. Thus if the features are well chosen also the second term in Inequality (12) can be expected to be small.
Finally we see that choosing $p = 2$ puts the following incoherence constraint on the feature spaces. No signal that can be constructed from features in one class should be well representable by features in another class. This constraint is the strongest we have encountered so far, which is only natural since we do not have an assumption on coefficient distribution. Coming back to our example it also corresponds quite naturally to what one would expect from face recognition, ie. that in all pictures enough distinctive features are visible and no matter the lighting condition or facial expression two people can always be uniquely identified from their features.

Of course there is ample opportunity to develop more class models, assuming different distributions on the coefficients and using more exotic norms. Also one could use different assumptions on the features, i.e. non-orthogonal. However, in this paper we will focus on finding a practical way to calculate sensing or feature matrices for classification based on the three main models.

3 Finding Feature/Sensing Matrices

From the analysis in the last section we can derive two types of conditions that the collection of features or subspaces F_i needs to satisfy. The first type describes how features from different classes should interact, i.e. the interplay measured in the appropriate matrix norm should be small, and the second type how the features should interact with the training data, i.e. the ratio of the response without to within class should be small. The problem with both kinds of conditions is they are not linear and difficult to handle. For instance calculating the $(2, 2)$-norm is equivalent to finding the largest singular value and calculating the $(\infty, 1)$-norm is even NP-hard. We will therefore start with a very simple approach that will lead to a reasonably fast algorithm, and in the last section point out how to extend it to include more complicated constraints. Instead of requiring explicitly that the interplay between features from different classes is small, hereby avoiding to investigate what small means quantitatively, we use the intuition that this should come as free side effect from regulating the interaction with the training data, and simply ask that F is a collection of orthonormal systems F_i each of rank s. What we would actually like to do about the interaction of the features with the training data is to minimise the ratio between the response of the training data without to within class. However, a constraint involving the ratio is not linear and very hard to handle. We will therefore split it into two constraints that guarantee that the ratio is small if they are fulfilled. The first constraint is that the response within class is equal to a constant β_p which we choose to be the maximally achievable value given the rank of the orthonormal systems and p. The second constraint is that the response without class is smaller than a constant μ_p, whose dependence on s, p, d is more complicated and will be discussed later. Define the two sets $\mathcal{F}_s$ and $\mathcal{F}_\mu$ as

$$\begin{aligned} \mathcal{F}_s &:= \{F = (F_1, \dots, F_c) : \; F_i^\star F_i = I_s\} \\ \mathcal{F}_\mu &:= \{F : \; \|F_i^\star y_i^k\|_p = \beta_p, \\ &\qquad\qquad \|F_j^\star y_i^k\|_p \leq \mu_p, \, \forall k, i, j \neq i\}, \end{aligned} \tag{13}$$

then our problems could be summarised as finding a matrix in the intersection of the two sets, i.e. $F \in \mathcal{F}_s \cap \mathcal{F}_\mu$. However, since this intersection might be empty, we should rather look for a pair of matrices, each belonging to one set, with minimal distance to each other measured in some matrix norm, eg. the Frobenius norm, denoted by $\|\cdot\|_{\mathbf{2}}$[2],

$$\min \|F_s - F_\mu\|_{\mathbf{2}} \text{ s.t. } F_s \in \mathcal{F}_s, \, F_\mu \in \mathcal{F}_\mu. \tag{14}$$

One line of attack is to use an alternate projection method, i.e. we fix a maximal number of iterations, an initialisation for F_s^0 and then in each iterative step do:

- find a matrix $F_\mu^k \in \text{argmin}_{F \in \mathcal{F}_\mu} \|F_s^{k-1} - F\|_{\mathbf{2}}$
- check if $\|F_s^{k-1} - F_\mu^k\|_{\mathbf{2}}$ is smaller than the distance of any previous pair and if yes store F_s^{k-1}
- find a matrix $F_s^k \in \text{argmin}_{F \in \mathcal{F}_s} \|F_\mu^k - F\|_{\mathbf{2}}$
- check if $\|F_s^k - F_\mu^k\|_{\mathbf{2}}$ is smaller than the distance of any previous pair and if yes store F_s^k

If both sets are convex, the outlined algorithm is known as Projection onto Convex Sets (POCS) and guaranteed to converge. Non convexity of possibly both sets, as is the case here, results in much more complex behaviour. Instead of converging, the algorithm just creates a sequence (F_μ^k, F_s^k) with at least one accumulation point. We will not discuss all the possible difficulties here but refer to [18], where all details, proofs and background information can be found and wherein the authors conclude that alternate projection is a valid strategy for solving the posed problem.

To keep the flow of the paper, we will not discuss the two minimisation problems that need to be alternatively solved here. The interested reader can find them, including the exact parameter settings in the simulations of the next section, in the appendix. Instead we will discuss how to set the parameters β_p, μ_p and possible choices for the initialisation F_s^0.

As mentioned above we choose β_p to be the maximally achievable value. An orthonormal system of s feature vectors can maximally take out all the energy of a signal,

$$\|F_i^\star y_i\|_2 \leq \|y_i\|_2. \tag{15}$$

As the signals are assumed to have unit norm, this energy is at most one and we set $\beta_2 = 1$. The maximal 1-norm of the vector $F_i^\star y_i$ of length s with energy 1 is $\sqrt{s}$. This is attained when all features of one class take out the same energy, i.e. the absolute values of the entries in $F_i^\star y_i$ are all equal to $1/\sqrt{s}$. This leads to $\beta_1 = \sqrt{s}$. The infinity norm $F_i^\star y_i$ corresponds to the maximal inner product between one of the feature vectors and the signal. As both the feature vector and

[2] We use this notation instead of the more common variant $\|\cdot\|_F$ to avoid confusion.

the signals are normalised, this can be at most one and so we set $\beta_\infty = 1$.
From the discussion in the last section we see that the parameter μ reflects the incoherence we require between features from different classes. If we have $d \geq c \cdot s$, it is theoretically possible to have c subspaces of dimension s which are mutually orthogonal to each other, and μ could be zero. As soon as the above inequality is reversed, because for instance the actual dimension of the span of all features, i.e. $rank(F)$, is smaller than d, not all subspaces corresponding to the different classes can be orthogonal but will have to overlap. How the size of this overlap, i.e. coherence, should be measured, is determined by the choice of p-norm for classification. For instance for $p = 2$ the coherence was measured by $\|F_j^\star F_i\|_{2,2}$ and from theory about Grassmannian manifolds, see [18], we know that the maximal coherence between two of c subspaces of dimension s embedded in the space $\mathbb{R}^d$ can be lower bounded by

$$\max_{i \neq j} \|F_j^\star F_i\|_{2,2}^2 \geq \frac{s \cdot c - d}{d(c-1)}. \tag{16}$$

The problem with setting μ as above is that we are not controlling the interaction between the sets of features directly but only indirectly over the training data. There the worst case might not be assumed and so μ as above would be too large. Also for the cases $p = 1, \infty$ we do not have a similar bound. Therefore instead of trying to analyse theoretically how to set μ, where we have to deal with too many unknowns, we use the above bound as an indication of order of magnitude and, when testing our scheme on real data, vary the parameter μ. Lastly for the initialisation for each class we choose the orthogonal system that maximises the energy taken from this class opposed to the energy taken from the other classes, i.e.

$$F_{s,i}^0 = \operatorname*{argmin}_{F_i^\star F_i = I_s} \|F_i^\star Y_i\|_\mathbf{2}^2 - \sum_{j \neq i} \|F_i^\star Y_j\|_\mathbf{2}^2. \tag{17}$$

This problem can be easily solved, by considering the rewritten version of the function to minimise,

$$\min_{F_i^\star F_i = I_s} \operatorname{trace}\big(F_i^\star (Y_i Y_i^\star - \sum_{j \neq i} Y_j Y_j^\star) F_i\big). \tag{18}$$

If $UDU^\star$ is an eigenvalue decomposition of the symmetric (Hermitian) matrix $Y_i Y_i^\star - \sum_{j \neq i} Y_j Y_j^\star$, then the minimum is attained for $F_{s,i}^0$ consisting of the s eigenvectors corresponding to the s largest eigenvalues.

4 Testing

To test the proposed scheme we use two face databases, the AR-database, [13] and the extended Yale B database, [1]. First we will test the validity of all three approaches on the AR-database, even though it is intuitively clear that the most appropriate model for faces corresponds to $p = 2$. Using the experience from the AR-database we will then run similar tests on the extended Yale B database using only the most appropriate model $p = 2$.

4.1 AR-Database

For the test we used a subset of images from the AR-database. For each of the 126 people there are 26 frontal images of size 165×120 taken in two separate sessions. The images include changes in illumination, facial expression and disguises. For the experiment we selected 50 male and 50 female subjects and for each of them took the 14 images with just variations in illumination and facial

$s\backslash\frac{\mu}{\sqrt{s}}$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
2	60	56	56	57	60	58	60	61	66	64	69
3	52	**46**	48	**46**	51	51	53	58	62	61	61
4	62	52	54	55	55	56	56	54	55	57	61
5	64	59	56	56	55	58	61	63	66	68	68
6	61	54	57	54	56	59	62	58	61	71	71
7	57	55	57	55	59	57	58	62	61	68	69

Table 1: Number of misclassified images on the AR-database for $p = 1$ and varying values s and μ.

$s\backslash\mu$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
1	57	58	59	58	60	59	59	58	58	58	62
2	51	49	51	51	51	55	57	57	59	58	56
3	47	42	45	50	53	53	54	61	62	61	64
4	46	42	41	41	47	48	51	62	63	61	63
5	48	43	**40**	44	50	51	52	55	55	59	61
6	49	45	42	45	49	48	51	54	54	57	58
7	45	43	43	43	45	45	48	53	51	54	52

Table 2: Number of misclassified images on the AR-database for $p = 2$ and varying values s and μ.

expression, neutral, light from the right and left, front light, angry, happy, sleepy. The all together 700 images from the first session were used as training data and the 699 images[3] from the second session for testing. Every image was converted to grayscale and then stored as a 19800 dimensional column vector. The images from the first session were stored in the 19800×700 matrix Y^1 and those from the second in the 19800×699 matrix Y^2. All images (columns) in Y^1 were re-scaled to have unit norm. In order to speed up the calculations, we first applied a unitary transform, which does not change the geometry of the problem, but reduces the size of the matrices, i.e. we did a reduced QR-factorisation decomposing Y^1 into the 19800×700 matrix Q with orthogonal columns and the 700×700 upper triangular matrix R and set $\tilde{Y}^1 = Q^\star Y^1 = R$ and $\tilde{Y}^2 = Q^\star Y^2$.
We tested the proposed scheme for all three choices of p and varying values of μ_p scaling from 0 to 10% of β_p and number of features per class varying from 1 to 7. The choice of the maximal outside-class contribution $\mu_{\max} = 0.1\beta_p$ was inspired by the bound in (16). If we take as effective signal dimension $d = 700$ and assume that the space should not only accommodate the 100 different people in our training set but all people, i.e. we let c go to infinity, the bound approaches $\sqrt{s/d}$ which is 0.1 if $s = 7$ and 0.0378 if $s = 1$. The maximal number of features per class is 7, since we only have 7 test images and so it does not make sense to look for spaces of higher dimension containing all test images. Note also that for $s = 1$ the three schemes are the same, so the results are only displayed once. For each set of parameters we calculated the corresponding feature matrix using the algorithm described in the last section on the images from the first session. We then classified the images from the second session using the appropriate p-norm. The results are shown in Tables 1, 2 and 3.

As we can see we get the best performance for $p = 2$, followed by $p = 1$ and $p = \infty$. This comes as no surprise when considering the structure of our data. Intuitively the important features of a face are eyes, nose and mouth. Since the people in the pictures have different facial expression, usually not all of these features will be active explaining why $p = 1$ is not the most appropriate

[3] 700 minus corrupted image w-027-14.bmp

$s\backslash\mu$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
2	55	62	59	54	56	**52**	54	61	63	64	62
3	55	63	58	56	60	58	59	63	65	69	69
4	55	64	60	57	59	58	58	61	67	70	67
5	55	60	59	55	58	57	57	60	66	71	69
6	55	61	59	54	57	56	56	65	67	72	69
7	55	61	59	55	56	54	55	66	66	71	70

Table 3: Number of misclassified images on the AR-database for $p = \infty$ and varying values s and μ.

model. On the other hand we can expect to have more than one feature active at the same time even if not to the same extent. Using $p = \infty$ we lose the information given by these secondary active features while with $p = 2$ we still incorporate it into the final decision.
We can also see that 0.1% of μ as maximally allowed outside class 'energy' seemed to have been a good choice as we can always see a small decrease and large increase of the error going from 0 to 0.1, with the best range for $p = 1$ and $p = 2$ between 0.01 and 0.03 and for $p = \infty$ between 0.02 and 0.06. For $p = 1$ we get better performance for the lower dimensions, which seems reasonable because there the equal energy distribution over the features is easier achieved. For $p = 2$ on the other hand the better performance is achieved with higher dimensions, which are able to capture more important side details. Finally for $p = 2$ the results seem equal for all dimensions. A possible explanation is given by the initialisation, which ensures that for all dimensions the first, most promising direction is included.
Still in all three cases in the most promising ranges the proposed scheme outperforms a standard method like Fisher's LDA, [6]. The best result by LDA is obtained when using the original (not-normalised) images and the highest possible number of discriminant axes $c - 1 = 99$. In this case nearest neighbour classification, corresponding to $p = \infty$ but with non orthogonal features, fails to identify 59 images, and nearest subspace classification, corresponding to $p = 2$ fails to identify 71 images. When concentrating on the results for $p = 2$, which is the most sensible choice given the structure of the data, $p = 2$, we also see that the scheme performs well in comparison to a recent, successful method based on ℓ_1 minimisation, [20]. The best result reported there is a success rate of 94.99%, meaning 35 misclassified images, which is 5 images better than our best case of 40 errors.
Encouraged by the promising results we now turn to testing our scheme on the extended Yale B database.

4.2 Extended Yale B Database

From the extended Yale B database we used the 2414 frontal face images, about 64 images taken under varying illumination conditions for each of the 38 people. For the test we randomly split the set of images per person into an equal number of training and test images, using one more training than test image in case of an odd number of images per class. We then ran our classification scheme with the number of features per class varying from 2 to 5 and thanks to the experience gained from the AR-database with the values of μ running only from 0 to 0.05. For the computation of the feature matrices we used the same simplifications as described for the AR-database. For comparison we ran Fisher's LDA with 37 and 30 discriminative axes in combination with the nearest neighbour classifier. This procedure was repeated 19 times and the mean of all 20 runs was computed.

The results of our method can be found in Table 4. While Fisher's LDA on average missclassi-

$s\backslash\mu$	0	0.01	0.02	0.03	0.04	0.05
2	19.80 ± 5.74	20.30 ± 5.80	22.25 ± 7.20	23.85 ± 6.81	25.25 ± 6.66	26.25 ± 6.61
3	14.15 ± 4.37	**13.60 ± 4.22**	13.85 ± 3.73	15.85 ± 5.25	16.40 ± 4.78	17.55 ± 6.00
4	15.75 ± 3.82	14.05 ± 3.49	13.95 ± 3.55	15.35 ± 3.95	16.45 ± 4.10	16.95 ± 4.30
5	15.70 ± 4.78	15.00 ± 4.91	14.45 ± 4.30	15.30 ± 3.34	17.60 ± 4.65	17.65 ± 4.55

Table 4: Mean ± standard deviation of misclassified images on the Extended Yale B database for $p = 2$ and varying values s and μ.

fied 23.30 ± 6.42 images (success rate of 98.07 ± 0.53%) using 37 discriminant axes and 231.55 ± 23.48 images (success rate 80.78 ± 1.95%) using 30 discriminant axes, our method in the best case only misclassified 13.60 ± 4.22 images (success rate 98.87 ± 0.35%). In general it outperformed Fisher's LDA for a wide range of values for μ and s.
Comparison to the ℓ_1-minimisation scheme in [20] is harder, as it seems that there only a single run was used. However, their best success rate of 98.26%, achieved at the same time as Fisher's LDA with 30 discriminant axes achieved 87.57% (the maximal rate for Fisher's LDA we encountered in 20 runs was 84.73%), is still below our best average rate of 98.87%.

To illustrate the results in Figure 2 and to confirm the motivation in the introduction for using different features for different classes, we show what happens to the training images of two different subjects when projected on the features of their own class and the other subject's class, using the settings that gave the best performance. As expected the projections on features of their own class nicely filter out common traits like eyes, mouths and noses, but on top of that the features of the first subject capture the very distinctive birth mark on his right cheek. The projections on the wrong class on the other hand are not only much weaker (note the difference in scale) but also less clear. Two overlapping sets of features seems to appear at the same time, the ones that belong to the subject in the image and the ones that the projection is trying to filter out.

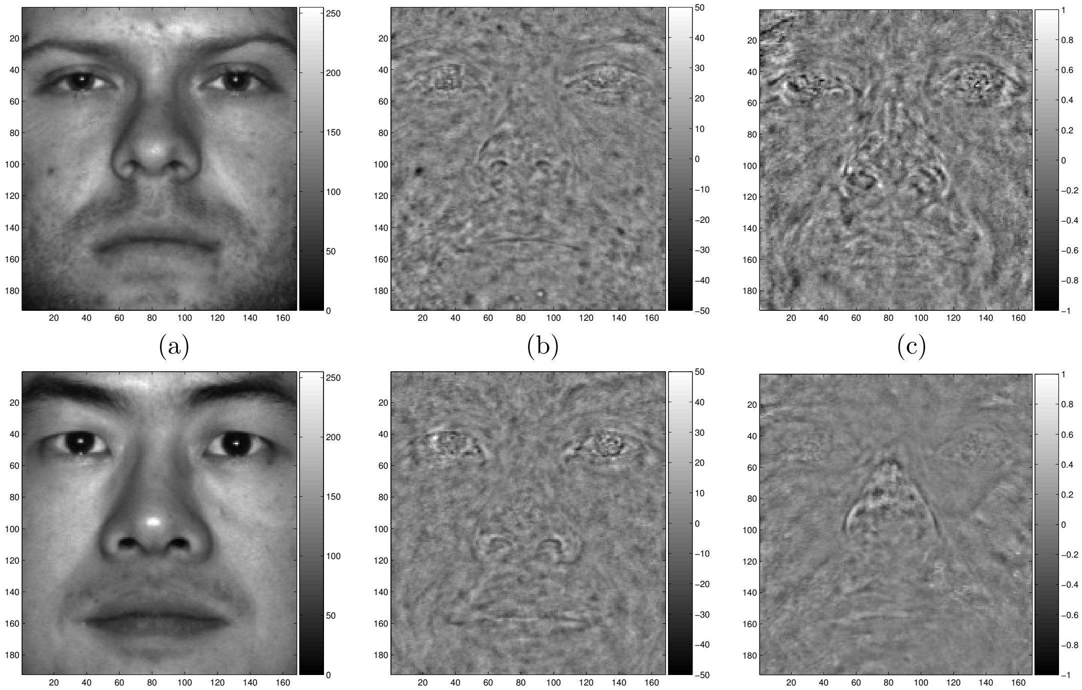
Summarising the results, we can say that our method outperforms a classic scheme like Fisher's LDA. In comparison to the ℓ_1-minimisation scheme in [20] its best performance is slightly worse on the AR-database but seem to be better on the YaleB-database. However it has one big advantage over the ℓ_1-minimisation scheme, which is its low computational complexity. Not taking the calculation of the feature matrices into account, as this is part of the pre-processing, basically all that has to be done to classify a new data vector is to multiply it with the feature matrix and calculate some statistics on the resulting vector. The ℓ_1 minimisation method on the other hand requires on top of extracting the features the solution of a convex optimisation problem

$$\min \|z\|_1 \text{ s.t. } \|f_{new} - Fz\|_2 \leq \varepsilon, \tag{19}$$

where F in this case is the $d_f \times N$ matrix containing the features of all the training data. For comparison in [20] the authors state that the classification of one image takes a few seconds on a typical 3 GHz Pc. At the same time for classifying 1205 images of size 192×168, using our method with 4 feature dimensions per class, MATLAB takes less than half a minute on a Dual 1.8Ghz PowerPC G5, which is less than 25ms per image.

5 Discussion

We have presented a classification scheme based on a model of incoherent subspaces, each one associated to one class, and a model on how the elements in a class are represented in this subspace.



(a) (b) (c)

(d) (e) (f)

Figure 2: Images of two subjects, original (a) & (d), projected onto the span of features from their own class (b) & (e), projected onto the span of features of the wrong class (c) & (f) - note different scales!!

From a more practical viewpoint we have developed an algorithm to calculate these subspaces, i.e. the feature matrices, and shown that the scheme gives promising results on the AR database and in the best case even outperforms a state of the art method like the ℓ_1-minimisation scheme in [20] on the YaleB-database. However, of the drawbacks of the scheme so far is that we did not specify how to choose the parameter μ (the parameter β in the end was always set to 1). To apply it to a concrete problem it would therefore be necessary to split your training data again into a set for training and one for tuning the parameter μ.

The idea that each class should have its own representative system, learned from the training data can already be found in [17]. There frames or dictionaries for texture classification are learned, such that each provides a sparse representation for its texture class. The new texture then gets the label of the texture frame providing the sparsest representation. In [12], the same basic idea is used but the learning is guided by the principle that the dictionaries should also be discriminant, while in [16] both learning principles are combined, i.e. the dictionaries should be discriminant and approximative. This third scheme can be considered as a more general and more complicated version of our approach. Alternatively our approach can be considered to be a hybrid of Nearest Subspace respectively Nearest Neighbour and the discriminative and approximative frame scheme, in so far as it is linear but has individual features for every class.

The idea to use a collection of subspaces for data analysis can also be found in [11], where the subspaces are used to model homogenous subsets of high-dimensional data which together can capture the heterogenous structures.

For the future there remain some interesting directions to explore. Firstly the possibilities of the subspace classification approach do not seem exhausted using the proposed algorithm. Ironically this fact revealed itself through a bug in the minimisation procedure, resulting in matrix pairs with distances larger than the optimal ones, and sensing matrices giving better classification results, i.e. in the best case an error of only 35 misclassified images. The main difference of these fake optimal matrices to the sensing matrices corresponding to the actual minima, seemed to be that, while capturing approximately the same 'energy' within class, they were more accurate in respecting the without class energy bound, i.e. less overshooting of the maximally allowed value μ. This overshooting for the real minima is a result of imposing not only $\|F_i y_j^k\|_2 \leq \mu$ but also $\|F_i y_i^k\|_2 = \beta$, which forces the optimal feature matrix to balance the error incurred by not attaining β within class and the error incurred by being larger than μ without class. A promising idea to avoid the overshooting would be to change the problem formulation and ask to maximise the 'energy' within class subject to keeping the 'energy' without class small, i.e. in the case $p = 2$ solve,

$$\max \sum_i \|F_i^\star Y_i\|_{\mathbf{2}}^2$$
$$\text{s.t. } F_i^\star F_i = I_s \text{ and}$$
$$\|F_i x_j^k\|_2 \leq \mu, \forall k, j \neq i.$$

Lastly our approach allows to impose additional constraints on F, like incoherence of the subspaces between each other, e.g. $\|F_i^\star F_j\|_{2,2} \leq \nu$ for $p = 2$, different sizes of the subspaces but with appropriate weighting, or low rank of the whole feature matrix to reduce the cost of calculating $F^\star y_{new}$. Also one could think of replacing the orthogonality constraint with an incoherence constraint of the form $\langle f_i^j, f_i^k \rangle < \nu$ for $j \neq k$, which could be beneficial when using the ∞ or 1-norm or more exotic norms, but would led to a quite large increase of complexity of the scheme. Finally there is the possibility to reduce computational cost if d and N are very large, especially in the training step. A promising and computationally efficient strategy would be to first take random samples of the training data, which reduce their dimension but very likely preserve the geometrical structure, as described in [2] and used in [20]. Alternatively to reduce the dimension of F one can apply our scheme on classical features, like Eigen or Laplace features, instead of directly on the raw training data.

A Solution Sketches for the Minimisation Problems

In order to use the alternate projection method for calculating the feature matrices we need to find the projection of a matrix $\hat{F}$ onto $\mathcal{F}_s$ and onto $\mathcal{F}_\mu$ in the three cases $p = 1, 2, \infty$. We will start with the easier of the two problems

$$\text{find: } F_s \in \operatorname*{argmin}_{F \in \mathcal{F}_s} \|F - \hat{F}\|_{\mathbf{2}}. \tag{20}$$

Since the minimisation problem is invariant under squaring of the objective function and thus equivalent to

$$\min_{F \in \mathcal{F}_s} \|F - \hat{F}\|_{\mathbf{2}}^2 = \min_{F \in \mathcal{F}_s} \sum_{i=1}^{c} \|F_i - \hat{F}_i\|_{\mathbf{2}}^2, \tag{21}$$

it splits into c independent problems

$$\min_{F_i^\star F_i = I_s} \|F_i - \hat{F}_i\|_{\mathbf{2}}^2. \tag{22}$$

The solution of these problems is straightforward. If $\hat{F}_i$ has the reduced singular value decomposition $\hat{F}_i = U_i S_i V_i$ then the orthonormal system F_i of same rank closest to it is $F_i = U_i V_i$, see e.g. [8].
The second minimisation problem

$$\text{find: } F_\mu \in \operatorname*{argmin}_{F \in \mathcal{F}_\mu} \|F - \hat{F}\|_{\mathbf{2}}.$$

is more complicated to solve. Assume that the number of training signals is larger than the dimension of the signals and span the whole space, so that the $d \times N$ matrix Y has rank $d \leq N$. If not we embed the training signals into a lower dimensional space corresponding to the rank of Y via a reduced QR-decomposition of Y and set $\tilde{Y} = Q^\star Y = R$ before starting the alternating projection procedure. Afterwards we set $F = Q^\star \tilde{F}$, where $\tilde{F}$ is the feature matrix calculated from the lower dimensional embedded data. Since Y has rank d we have $YY^\dagger = I_d$ and can reformulate the problem to solve as

$$\min_{F \in \mathcal{F}_\mu} \|F - \hat{F}\|_{\mathbf{2}} = \min_{F \in \mathcal{F}_\mu} \|(F^\star Y - \hat{F}^\star Y)Y^\dagger\|_{\mathbf{2}}. \tag{23}$$

The advantage of this formulation is that it is in terms of $F^\star Y$, which is also used to describe $\mathcal{F}_\mu$. To further exploit this property we define the set $\mathcal{G}_\mu$, which is of the form $F^\star Y$ with $F \in \mathcal{F}_\mu$. To characterise the set $\mathcal{G}_\mu$ we assume the following notation. Let G_{ij} refer to the $s \times n_j$ submatrix that corresponds to $F_i^\star Y_j$ inside $F^\star Y$ and denote the k-th column of G_{ij} by $G_{ij}(:,k)$. We can then define

$$\begin{aligned} \mathcal{G}_\mu := \{G : \|G_{ii}(:,k)\|_p &= \beta_p, \\ \|G_{ij}(:,k)\|_p &\leq \mu_p, \forall k, i, j \neq i\}. \end{aligned} \tag{24}$$

Set $\hat{G} = \hat{F}^\star Y$ then the problem in (23) is equivalent to

$$\min_{G \in \mathcal{G}_\mu} \|(G - \hat{G})Y^\dagger\|_{\mathbf{2}}. \tag{25}$$

To attack this problem we will use resolvents or proximity operators which are a generalisation of projection operators. Given a Hilbertspace $\mathcal{H}$ and a function f from $\mathcal{H}$ to $]-\infty, +\infty]$ that is lower semicontinuous, convex and not identical to $+\infty$, i.e. belonging to $\Gamma_0(\mathcal{H})$ the proximity operator prox_f is defined by

$$\text{prox}_f(x) = \operatorname*{argmin}_{\mathcal{H}} f(y) + \frac{1}{2}\|x - y\|^2_{\mathcal{H}}.$$

Proximity operators were first studied by Moreau in [14], who developed a theory of proximal calculus, and recently have been used to solve optimisation problems in signal processing, [4]. Here we will use the forward backward splitting approach as described in [5]. Assume that we can write the function to minimise as the sum of two functions f_1, f_2 in $\Gamma_0(\mathcal{H})$, i.e.

$$\min_{x \in \mathcal{H}} f_1(x) + f_2(x). \tag{26}$$

If f_2 is differentiable with a β-Lipschitz continuous gradient for $\beta > 0$ then the sequence generated by fixing $x_0 \in \mathcal{H}$ and iterating

$$x^{n+1} = \text{prox}_{\gamma^n f_1}(x^n - \gamma^n \nabla f_2(x^n)) \tag{27}$$

converges weakly to a minimum of (26) if $\gamma < 2/\beta$.
To apply the concept to our problem we take as Hilbert space the set of all $c \cdot s \times N$ matrices G equipped with the Frobenius norm and define the indicator function $\mathbb{I}_{\mathcal{G}_\mu}$ of the set $\mathcal{G}_\mu$ by

$$\mathbb{I}_{\mathcal{G}_\mu}(G) := \begin{cases} 1 & \text{if } G \in \mathcal{G}_\mu \\ +\infty & \text{else} \end{cases} .$$

The we can replace problem (25) by

$$\min_G \mathbb{I}_{\mathcal{G}_\mu}(G) + \|(G - \hat{G})Y^\dagger\|_{\mathbf{2}}^2. \tag{28}$$

The slight imperfection of this approach is that the set $\mathcal{G}_\mu$ is not convex, therefore $\mathbb{I}_{\mathcal{G}_\mu}(G)$ is not convex and the sequence generated applying (27) is not guaranteed to converge. Finding only a local minimum is however not such a big problem, since the procedure is only part of a bigger iterative scheme, as long as in each step we get some improvement.
What remains to be done is to calculate the proximity operators for $\gamma f_1 = \gamma \mathbb{I}_{\mathcal{G}_\mu} = \mathbb{I}_{\mathcal{G}_\mu}$, the gradient of $f_2(G) = \|(G - \hat{G})Y^\dagger\|_{\mathbf{2}}^2$ and decide about the initialisation G_0 and the step sizes γ^n. A straightforward calculation shows that $\nabla f_2(G) = 2(G - \hat{G})Y^\dagger (Y^\dagger)^\star$. Since $\mathbb{I}_{\mathcal{G}_\mu}$ is an indicator function the proximity operator is simply the orthogonal projection onto $\mathcal{G}_\mu$, i.e.

$$\operatorname*{argmin}_G \mathbb{I}_{\mathcal{G}_\mu}(G) + \frac{1}{2}\|G^n - G\|_{\mathbf{2}}^2 = \operatorname*{argmin}_{G \in \mathcal{G}_\mu} \|G^n - G\|_{\mathbf{2}}^2$$

Because of the structure of $\mathcal{G}_\mu$, see (24), the problem above splits into the smaller problems

$$\min_{\|G_{ii}(:,k)\|_p = \beta_p} \|G_{ii}^n(:,k) - G_{ii}(:,k)\|_2^2,\ \forall i$$
$$\text{and} \qquad \min_{\|G_{ij}(:,k)\|_p \le \mu_p} \|G_{ij}^n(:,k) - G_{ij}(:,k)\|_2^2,\ \forall i \neq j.$$

In other words for $p = 1,\ 2,\ \infty$ we need to solve problems of the form

$$\min_{\|g\|_p = \beta_p} \|g - h\|_2^2 \qquad \text{and} \qquad \min_{\|g\|_p \le \mu_p} \|g - h\|_2^2. \tag{29}$$

The solutions are collected in the following Theorem.

Theorem 1 *Denote by g_{β_p} the minimal argument of the first problem and by g_{μ_p} the minimal argument of the second problem in* (29).
$p = 1:$ *Set $\sigma(i) = \operatorname{sign}(h(i))$ if $h(i) \neq 0$ and $\sigma(i) = 1$ else, and denote by m the length of the h, then*

$$g_{\beta_1}(i) = h(i) + \sigma(i)\lambda, \ \ \text{where } \lambda = \frac{\beta - \|h\|_1}{m}.$$

If $\|h\|_1 \le \mu$ set $g_{\mu_1} = h$. Otherwise set $g^0 = h$ and iteratively shrink

$$g_{\beta_1}^k(i) = \sigma(i)\max(|g^{k-1}(i)| - \lambda^k, 0),$$
$$\text{where } \lambda^k = \frac{\|g^{k-1}\|_1 - \mu}{\sharp\{g_i^{k-1} \neq 0\}}.$$

until g^k *with* $\|g^k\|_1 = \mu$ *is found and set* $g_{\mu_1} = g^k$.
$p = 2$:

$$g_{\beta_2} = \beta_2 \cdot \frac{h}{\|h\|_2},$$

$$g_{\mu_2} = \begin{cases} h & \text{if } \|h\|_2 \leq \mu_2 \\ \mu_2 \cdot \frac{h}{\|h\|_2} & \text{else} \end{cases} .$$

$p = \infty$: *Let* $i_{\max}$ *be the index of (one of) the largest absolute component of* h *then*

$$g_{\beta_\infty}(i) = \begin{cases} 1 & \text{if } i = i_{\max} \\ h(i) & \text{else} \end{cases} ,$$

$$g_{\mu_\infty}(i) = \begin{cases} h(i) & \text{if } |h(i)| \leq \mu_\infty \\ \mu_\infty & \text{else} \end{cases} .$$

Lastly as initialisation G^0 we choose the projection of $\hat{G}$ onto $\mathcal{G}_\mu$, i.e. $G^0 = \text{prox}_{f_1}(\hat{G})$. Finding the correct step-sizes is usually a matter or trial and error. For the application considered here we used $\gamma^n = \|G^n\|_\mathbf{2}/(20\|\nabla f_2(G^n)\|_\mathbf{2})$, which worked better for small μ, and $\gamma^n = 1/\|\nabla f_2(G^n)\|_\mathbf{2}$, which worked better for large μ. The iteration was stopped when the relative improvement in each step was below 10^{-4}. The number of iterations for the alternative projections was set to 10.

Thanks: We would like to thank John Wright for helping us getting the cropped version of the AR-database faces.

References

[1] Georghiades A., P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[2] D. Achlioptas. Database-friendly random projections. In *Proc. 20th Annual ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems*, pages 274–281, 2001.

[3] R. Brunelli and T. Poggio. Face recognition: Features vs. templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1053, October 1993.

[4] P.L. Combettes and Pesquet J.-C. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, December 2007.

[5] P.L. Combettes and V.R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[6] R.A. Fisher. The use of multiple measures in taxonomic problems. *Ann. Eugenics*, 7:179–188, 1936.

[7] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. In *Proc. IEEE CVPR*, 2005.

[8] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[9] K. Lee, J. Ho, and D.J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[10] C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006.

[11] Y. Ma, A. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3):413–458, August 2008.

[12] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. IMA Preprint Series 2212, University of Minnesota, 2008.

[13] A.M. Martinez and Benavente R. The AR face database. Technical Report 24, CVC, 1998.

[14] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Ser. A. Math.*, 255:2897–2899, 1962.

[15] P. Phillips, W. Scruggs, A. O'Tools, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale results. NISTIR 7408, NIST, 2007.

[16] F. Rodriguez and G. Sapiro. Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries. IMA Preprint Series 2213, University of Minnesota, 2008.

[17] K. Skretting and J.H. Husoy. Texture classification using sparse frame-based representations. *EURASIP Journal on Applied Signal Processing*, 2006:11, 2006.

[18] J. Tropp, I. Dhillon, R Heath Jr, and T. Strohmer. Designing structured tight frames via an alternating projection method. *IEEE Transactions on Information Theory*, 51(1):188–209, January 2005.

[19] M. Turk and A. Pentland. Eigenfaces for recognition. In *Proc. IEEE CVPR*, 1991.

[20] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 2009.

ARTICLE

Oruganti Shanker
Explanation of low Hurst exponent for Riemann zeta zeros
***Rejecta Mathematica*, Vol. 2, No. 1, pp. 19–23, June 2011**
2011 Rejecta Publications
math.rejecta.org

An open letter concerning

Explanation of low Hurst exponent for Riemann zeta zeros

O. Shanker

In 2006 a striking result was published (Generalised Zeta Functions and Self-Similarity of Zero Distributions, J. Phys. A 39(2006), 13983-13997) about the statistics of the zeros of the Riemann zeta function. The paper (by the current author) applied rescaled range analysis, and found that the zeros exhibited an unusably low Hurst Exponent. While that paper discussed some possible explanations, no clear reason for the low Hurst Exponent emerged. This was particularly interesting because the most obvious explanation, that the differences of the zeros have a very large anti-correlation, would be very unusual indeed.

Recently the author came up with empirical evidence for another less radical explanation, and submitted the explanation for publication, and it was rejected. The main reason for the rejection appears to be the limited content. While the content is empirical and limited to a single point, the author nevertheless feels that the result should be published, not as an important new result, but because it provides a plausible explanation for the original findings. The referee also commented that the author did not mention that the distribution of the fluctuating part of the zero counting function has been known to be Gaussian distributed, and did not explain what the Hurst exponent was. The author agrees that the paper may benefit by having more detail and references. However, the original article had detailed references and discussion about the literature concerning the distribution of the zeros of the Riemann zeta function, so the reader may rely on the original paper for details and references. The author is not aware of any errors in the paper, and has not made any changes in the paper.

The referee also mentioned that there may be many distributions other than the Gaussian that would produce the same results. The author agrees: the key point of the paper is that the rescaled range analysis does not necessarily imply a long-range anti-correlation in the differences of the zeros, and to give an indication of other possible, less radical explanations.

Explanation of low Hurst exponent for Riemann zeta zeros

O. Shanker *

Abstract

We discuss a possible explanation for the low Hurst exponent extracted from a rescaled range analysis of the large height Riemann zeta function zeros.

Physicists have studied the zeros of the Riemann zeta function because of its relation to the spectra of random matrix theories (RMT) [1, 2, 3, 4] and the spectra of classically chaotic quantum systems [5, 6, 7, 8]. A rescaled range analysis of the large height Riemann zeta zeros leads to a very low Hurst exponent (~ 0.1) over several orders of magnitude variation (from 10^7 to 10^{22}) in the heights of the zeros [9]. So far there has been no explanation for this behaviour. One possible explanation is that the zeros represent a long range anti-correlation. However, one has to be careful in coming to that conclusion, since such a large anti-correlation is rather special [10]. In this work we argue that the low Hurst exponent is not due to an anti-correlation, but can be explained by superposing a random Gaussian correction term to Riemann's approximation for the variation of the number of zeros with height.

We first briefly set up the notation. The Riemann Zeta function is defined for $\mathrm{Re}(s) > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_{p} \left(1 - p^{-s}\right)^{-1}. \tag{1}$$

$\zeta(s)$ has a continuation to the complex plane and satisfies a functional equation

$$\xi(s) := \pi^{-s/2}\,\Gamma(s/2)\,\zeta(s) = \xi(1-s); \tag{2}$$

$\xi(s)$ is entire except for simple poles at $s = 0$ and 1. We write the zeroes of $\xi(s)$ as $1/2 + i\gamma$. The Riemann Hypothesis [11, 12, 13, 14] asserts that γ is real for the non-trivial zeroes. We order the γs in increasing order, with

$$\ldots\ldots\gamma_{-1} < 0 < \gamma_1 \leq \gamma_2 \ldots. \tag{3}$$

Then $\gamma_j = -\gamma_{-j}$ for $j = 1, 2, \ldots$, and γ_1, γ_2, $\ldots$ are roughly 14.1347, 21.0220, ... The Hurst exponent is extracted by applying rescaled range analysis to the distribution of the spacings [9] $\delta_j = \gamma_{j+1} - \gamma_j$.

From Riemann's time it is known that the mean number of zeros with height less than γ (the smoothed Riemann zeta staircase) is approximately [14, 6]

$$< \mathcal{N}_{\mathcal{R}}(\gamma) >= (\gamma/2\pi)(ln(\gamma/2\pi) - 1) + \frac{7}{8}. \tag{4}$$

Thus, the mean spacing of the zeros at height γ is $2\pi(\ln(\gamma/2\pi))^{-1}$. Eqn. 4 can be inverted to give an approximation $\gamma_{a,i}$ for the i^{th} zero of the Riemann zeta function, if we set $< \mathcal{N}_{\mathcal{R}} >= i$. Let us write the i^{th} zero γ_i as

$$\gamma_i = \gamma_{a,i} + X_i, \tag{5}$$

*O. Shanker is with Hewlett-Packard (oshanker@gmail.com)

Table 1: Hurst Exponent for Riemann zeroes and for the values from Eqn. 5 with two different seed values for the random number generator. The sample size is 10000 and the standard deviation for X_i in Eqn. 5 is 0.25

Range of zeroes	Riemann zeros	Eqn. 5 seed A	Eqn. 5 seed B
$10^4 < j \leq 2 * 10^4$	0.51	0.51	0.51
$10^5 < j \leq 10^5 + 10^4$	0.21	0.21	0.20
$10^6 < j \leq 10^6 + 10^4$	0.10	0.13	0.11
$10^{12} < j \leq 10^{12} + 10^4$	0.12	0.12	0.11
$10^{22} < j \leq 10^{22} + 10^4$	0.12	0.12	0.11

where X_i is a correction term. We show that the observed behaviour of the rescaled range analysis is reproduced even if there is no anti-correlation between the Riemann zeta zeros, all that is needed is to assume that the correction term X_i is distributed normally. With this assumption, Table 1 shows the Hurst exponent extracted from the Riemann zeta zeros, and for comparision the Hurst exponent for the sequence given by Eqn. 5 with X_i distributed normally. We ran the analysis using two different seed values for the random number generator. The standard deviation of the normal distribution was taken to be 0.25 . We see from the table that Eqn. 5 reproduces the observed behaviour of the Riemann zeros fairly well. It is also not too sensitive to the assumed seed value used to generate the Gaussian term X_i in Eqn. 5. Finally, it reproduces the observed low Hurst exponent for the large height zeros, since for these zeros the first term in Eqn. 5 becomes essentially linear, and the Hurst exponent is then determined by the normal term, independent of the height. Thus, it appears that the observed low value for the large height Riemann zeta zeros is not due to an anti-correlation, but instead it is due to the fact that for these zeros the dependence on height is essentially linear, with a normal correction term superposed on the linear variation.

Figure 1 shows the rescaled range analysis for the zeros in the range $100000 \ldots 110000$. The horizontal axis is the log of the bin size used for the rescaled range analysis, and the vertical axis is the log of the mean rescaled range for the given values of bin size. The low slope at the low values of the bin size is due to the X_i term in Eqn. 5, and the increase in slope at higher bin sizes is due to the $\gamma_{a,i}$ term in Eqn. 5. Eqn. 5 gives a fairly good representation of the rescaled range behaviour of the actual Riemann zeta zeros.

In conclusion, we have presented evidence that the remarkable behaviour of the Riemann zeta zeros under rescaled range analysis can be explained by Riemann's approximation for the variation of the number of zeros with height coupled with a random Gaussian correction term.

References

[1] E. Wigner, "Random Matrices in Physics," *Siam Review*, **9**, 1-23, (1967).

[2] M. Gaudin, M. Mehta, "On the Density of Eigenvalues of a Random Matrix," *Nucl. Phys.*, **18**, 420-427, (1960).

[3] M. Gaudin, "Sur la loi Limite de L'espacement de Valuers Propres D'une Matrics Aleatiore," *Nucl. Phys.*, **25**, 447-458, (1961).

[4] F. Dyson, "Statistical Theory of Energy Levels III," *J. Math. Phys.*, **3**, 166-175, (1962).

[5] M. V. Berry, "Semiclassical theory of spectral rigidity," *Proc. R. Soc.*, **A 400** , 229-251, (1985).

[6] M. V. Berry, "Riemann's zeta function: a model for quantum chaos?," *Quantum chaos and statistical nuclear physics (Springer Lecture Notes in Physics)*, **263** , 1-17, (1986).

[7] M. V. Berry, "Quantum Chaology," *Proc. R. Soc.* , **A 413** , 183-198, (1987).

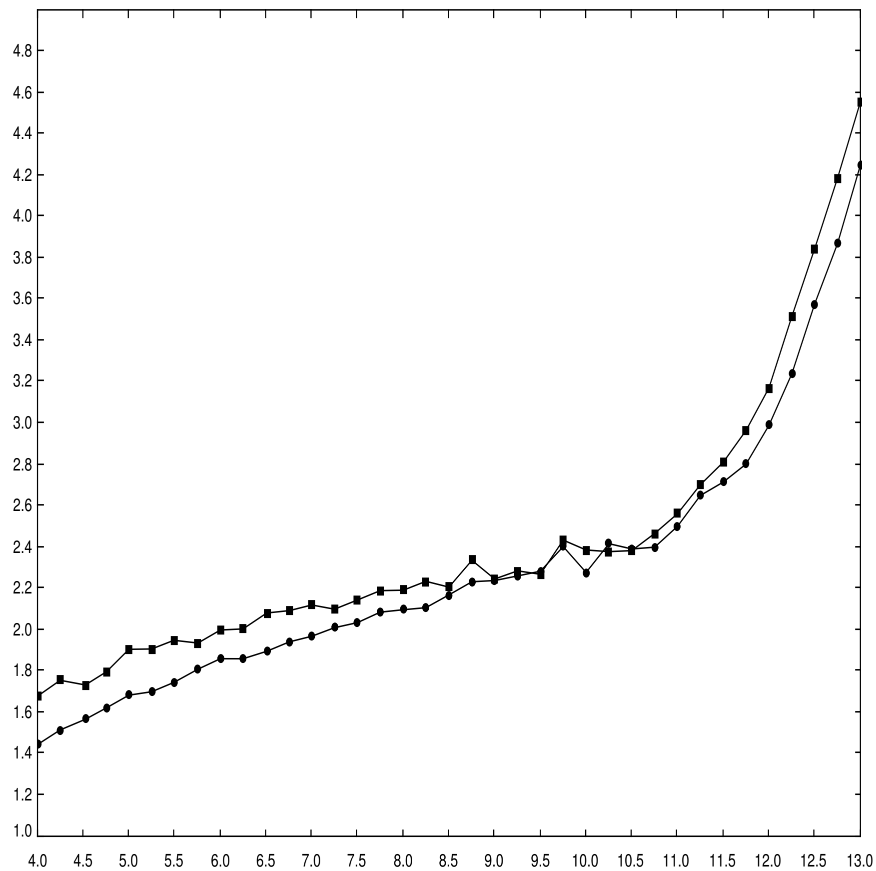


Figure 1: Rescaled range analysis for the zeros in the range 100000 . . . 110000. The y axis is the log of the rescaled range and the x axis is the log of the bin size. Squares represent the values for the Riemann zeros and circles represent the values for the sequence in Eqn. 5

.

[8] M. V. Berry, 'Number variance of the Riemann zeros',” *NonLinearity* , **1** ,399-407 , (1988).

[9] O. Shanker, Generalised Zeta Functions and Self-Similarity of Zero Distributions, *J. Phys. A* **39**(2006), 13983-13997.

[10] Govind Rangarajan and Mingzhou Ding, “An Integrated Approach to the Assessment of Long Range Correlation in Time Series Data”, *Phys. Rev.*, **E61**, 4991-5001, (2000).

[11] B. Riemann, “Über die Anzahl der Primzahlen uter
Einer Gegebenen Gröbe,” *Montasb. der Berliner Akad.*, (1858), 671-680.

[12] B. Riemann, “Gesammelte Werke”, Teubner, Leipzig, (1892).

[13] E. Titchmarsh, “The Theory of the Riemann Zeta Function,” Oxford University Press, Second Edition, (1986).

[14] H. M. Edwards, “Riemann's Zeta Function,” Academic Press, (1974).

ARTICLE

Aran Nayebi
On the distribution of Carmichael numbers
***Rejecta Mathematica*, Vol. 2, No. 1, pp. 24–43, June 2011**
2011 Rejecta Publications
math.rejecta.org

An open letter concerning

On the distribution of Carmichael numbers

Aran Nayebi

In my paper, entitled "On the distribution of Carmichael numbers", I investigate the distribution of Carmichael numbers. The importance of Carmichael numbers is that they test the limits of Fermat's primality test, which ultimately led mathematicians to formulate more effective primality tests in the twentieth century. There have been two important conjectures regarding the distribution of these numbers up to sufficiently large bounds, one made by Paul Erdős in 1956 and a subsequent sharpening of this conjecture by Carl Pomerance in 1981. However, neither of these conjectures are well-supported by the Carmichael number counts famously performed by Richard Pinch up to 10^{21}. The inaccuracies of these two aforementioned conjectures are understandable, since not too much is known about Carmichael numbers. In fact, after a century of investigation regarding these numbers, it was only a decade ago that the infinitude of Carmichael numbers was proven! In this paper, I present two conjectures (which sharpen Erdős' and Pomerance's conjectures) regarding the distribution of Carmichael numbers that fit proven bounds, are roughly supported by Pinch's data (as well as data from other papers and resources), that closely model the true distribution of Carmichael numbers, and are supported by many theorems and conjectures put forth by renowned mathematicians such as Alford, Erdős, Galway, Granville, Harman, Pomerance, Wagstaff, Selfridge, and Szymiczek. The reader may wonder why *two* conjectures are presented. The reason is that due to the lack of information regarding Carmichael numbers and their distribution, both conjectures are viable to their own merit.

Unfortunately, although I feel that the results in this paper are important and would satisfy the interests of the mathematical community, the paper was rejected by three journals.

The first journal the paper was submitted to was *Mathematics of Computation*. The referee stated that "the paper deals with interesting topics and might be generally appropriate for Math. Comp. However, the paper is written very poorly and it needs a lot of work before it can be properly considered." Thus, I humbly took the advice of the referee, and I spent the better part of two months revising the paper rigorously with a colleague of mine. I made the paper more readable, the notation more recognizable, and I added six data tables from various cited sources (some of the data I collected myself), all in support of my conjecture. Similarly, through this revision process, we disproved many of my conjectures and theorems, and we sharpened and strengthened many of my proofs. However, the only conjecture that we were *unable to disprove* was my conjecture regarding Carmichael numbers. Furthermore, I discussed my paper with several mathematicians who are known for their work on Carmichael numbers and pseudoprimes (which are a superset of Carmichael numbers), all of whom agreed with the majority of my ideas. I also requested feedback from a mathematician who had not published any papers in this field, who stated: "I read through your paper on pseudoprimes, and while the subject is not my area of expertise, it is clear that you are familiar with the mathematical literature and are making a serious contribution."

After revising the paper thoroughly, I then submitted the paper to *The American Mathematical Monthly*. Although they were unable to find any mistakes (both mathematical and style-wise) and this time the paper received a good editorial review, the paper was rejected because "the Monthly tries to publish expositions of mathematics that are accessible to a broad mathematical audience. The material in your paper is rather technical, and we feel that many Monthly readers will find it forbidding. We will therefore not be able to accept it for publication. These are difficult decisions. The Monthly receives a large number of submissions each year, and we are able to publish only a small fraction of them."

I could not help but be amused by this rejection notice; however, I was somewhat flustered. Carmichael numbers are important in number theory because of their rarity (there are only 20138200 Carmichael numbers up to 10^{21}), and their existence demonstrates the ineffectiveness of the Fermat primality test. Furthermore, the fact that not too much is known about these numbers after almost a century of research, means that more work about them should be considered for inclusion within mathematical literature. Also, my paper is not forbidding as there are tables which present my assertions in non-verbiage form and these tables are even explained in detail. The notation is also entirely readable and widely-recognized.

As a final straw, I sent the paper to Carl Pomerance, in the hopes of a more extensive and in-depth peer review. At the time, Conjecture 1.0.4 (the second conjecture) had not been included in the manuscript; only Corollary 1.0.3 (the first conjecture) was presented as the main result. Although my correspondence with him was brief (parts of which I include in my paper), his advice was helpful. Pomerance's arguments in support of his conjecture compelled me to propose a second conjecture that was a refinement to his original 1981 one, mainly by utilizing finer estimates for the distribution of smooth numbers (a practice which he stated had not yet been done before). This conjecture, which later became Conjecture 1.0.4, gave extremely accurate counts for $C(x)$, the number of Carmichael numbers up to x, at least for smaller bounds (although asymptotically it is the same result as Pomerance's).

With these adjustments made, I submitted my manuscript to *Experimental Mathematics* as it is "a journal devoted to the experimental aspects of mathematics research." Unfortunately, two months later, they rejected the submission on the grounds that "the two conjectures presented by the author can each be substantially simplified by using known (or easily derived) asymptotics for the constituent parts....The first conjecture is extremely unlikely to be true, if only for the reason that it postulates an asymptotic formula for the number of Carmichael numbers up to x, while no other conjecture makes such a strong statement....Also, in the second conjecture, the author claims to be including more explicit secondary terms, but the $(1+o(1))$ factor just washes them out anyway. In short, the statements would need to be substantially simplified and polished to make this paper worth publishing in a strong journal such as EM." I agree with the referee that the statements would have to be simplified; a task which I had completed prior to submission, even going so far as to provide numerical estimates for the various constants used in the statement of Corollary 1.0.2. However, my points of contention with the referee are that the first conjecture cannot simply be disregarded as untrue due to the strength of its assertions (and in fact the numerical evidence compiled in my paper demonstrates its viability) and that the second conjecture must include secondary terms in it so that the discrepancies pointed out by Pinch will not occur.

If anything, the second conjecture appears to be more plausible than the first; however, both conjectures provide different and intriguing insights into the distribution of Carmichael numbers. The first conjecture asserts that an asymptotic formula for $C(x)$ easily follows based on the com-

Aran Nayebi
On the distribution of Carmichael numbers
Rejecta Mathematica, Vol. 2, No. 1, pp. 24–43, June 2011
2011 Rejecta Publications
math.rejecta.org

putation of numerical constants. The second conjecture indicates to us that if secondary terms exist, then the properties of smooth number counting functions *must* be examined further in order to effectively prove an equality for $C(x)$.

Frankly, submitting the paper to another peer-reviewed journal and waiting a few months to a year for a referee look over a paper which has already been examined by several mathematicians of the same expertise (if not more) is a waste of time. I have submitted my paper to *Rejecta Mathematica* in the hopes of advancing mathematics and the investigation of pseudoprimes and their variants.

On the distribution of Carmichael numbers

Aran Nayebi*

Abstract

Erdős conjectured in 1956 that there are $x^{1-o(1)}$ Carmichael numbers up to x. Pomerance made this conjecture more precise and proposed that there are $x^{1-\frac{\{1+o(1)\}\log\log\log x}{\log\log x}}$ Carmichael numbers up to x. At the time, his data tables up to $25 \cdot 10^9$ appeared to support his conjecture. However, Pinch extended this data and showed that up to 10^{21}, Pomerance's conjecture did not appear well-supported. Thus, the purpose of this paper is two-fold. First, we build upon the work of Pomerance and others to present an alternate conjecture regarding the distribution of Carmichael numbers that fits proven bounds and is better supported by Pinch's new data. Second, we provide another conjecture concerning the distribution of Carmichael numbers that sharpens Pomerance's heuristic arguments. We also extend and update counts pertaining to pseudoprimes and Carmichael numbers, and discuss the distribution of One-Parameter Quadratic-Base Test pseudoprimes.

1 Introduction

Fermat's "little" theorem states that if b is an integer prime to n, and if n is prime, then

$$b^n \equiv b \pmod{n}. \tag{1.0.1}$$

When $\gcd(b,n) = 1$, we can divide by b,

$$b^{n-1} \equiv 1 \pmod{n}. \tag{1.0.2}$$

A composite natural number n for which $b^{n-1} \equiv 1 \pmod{n}$ for any fixed integer $b \geq 2$ is a base b pseudoprime. A positive composite integer n is a Carmichael number if $b^{n-1} \equiv 1 \pmod{n}$ for all integers $b \geq 2$ with $\gcd(b,n) = 1$. The importance of Carmichael numbers is that they test the limits of the Fermat primality test, which ultimately led mathematicians to formulate more effective tests. Furthermore, there is little that is known about them; for instance, the infinitude of Carmichael numbers has only recently been proven by Alford, Granville, and Pomerance [3].

Let $\mathscr{P}_b(x)$ denote the number of base b pseudoprimes $\leq x$ and let $C(x)$ denote the number of Carmichael numbers $\leq x$. In 1899, Korselt [4] provided a method for identifying Carmichael numbers

Theorem 1.0.1. *An odd number n is a Carmichael number iff n is squarefree and $p-1 \mid n-1$ for all $p \mid n$, where p is a prime number.*

As a consequence of Theorem 1.0.1, it is easy to see that Carmichael numbers have at least three prime factors.

In 1910, Carmichael [24] found the smallest Carmichael number to be $561 = 3 \cdot 11 \cdot 17$.

*727 Moreno Avenue, Palo Alto, California, United States of America 94303-3618. Email: aran.nayebi@gmail.com

Table 1: Counts of k-prime Carmichael numbers

Bound	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$	$C_{11}(x)$	$C_{12}(x)$	$C(x)$
10^3	1	0	0	0	0	0	0	0	0	0	1
10^4	7	0	0	0	0	0	0	0	0	0	7
10^5	12	4	0	0	0	0	0	0	0	0	16
10^6	23	19	1	0	0	0	0	0	0	0	43
10^7	47	55	3	0	0	0	0	0	0	0	105
10^8	84	144	27	0	0	0	0	0	0	0	255
10^9	172	314	146	14	0	0	0	0	0	0	646
10^{10}	335	619	492	99	2	0	0	0	0	0	1547
10^{11}	590	1179	1336	459	41	0	0	0	0	0	3605
10^{12}	1000	2102	3156	1714	262	7	0	0	0	0	8241
10^{13}	1858	3639	7082	5270	1340	89	1	0	0	0	19279
10^{14}	3284	6042	14938	14401	5359	655	27	0	0	0	44706
10^{15}	6083	9938	29282	36907	19210	3622	170	0	0	0	105212
10^{16}	10816	16202	55012	86696	60150	16348	1436	23	0	0	246683
10^{17}	19539	25758	100707	194306	172234	63635	8835	240	1	0	585355
10^{18}	35586	40685	178063	414660	460553	223997	44993	3058	49	0	1401664
10^{19}	65309	63343	306310	849564	1159167	720406	196391	20738	576	2	3381806
10^{20}	120625	98253	514381	1681744	2774702	2148017	762963	114232	5804	56	8220777
10^{21}	224763	151566	846627	3230120	6363475	6015901	2714473	547528	42764	983	20138200

Based on Korselt's criterion, Erdős [21] formulated a method for constructing Carmichael numbers, which was the basis for the proof of Alford, Granville, and Pomerance [3]. His notion was to replace "$p-1 \mid n-1$ for all $p \mid n$" in Theorem 1.0.1 with $L \mid n-1$ for $L := \mathrm{lcm}_{p|n}(p-1)$. By focusing primarily on L, Erdős found every p for which $p-1 \mid L$ and then tried to find a product of those primes in which $\equiv 1 \pmod{L}$ [2]. His results heuristically suggested that for sufficiently large x,

$$C(x) = x^{1-o(1)}. \tag{1.0.3}$$

More convincingly, Theorem 4 of [3] shows that (1.0.3) holds if one assumes widely-believed assumptions regarding primes in arithmetic progressions. However, drawing upon available data at the time, Shanks [12] was skeptical of (1.0.3) because the counts of Carmichael numbers seemed to have noticeably fewer prime factors than those predicted by Erdős' heuristic.

Granville and Pomerance [2] conjectured that the reason for the difference between the computational evidence and the argument of (1.0.3) stems from a grouping of Carmichael numbers into two distinct classes, namely primitive and imprimitive. If we let $g = g(n) := \gcd(p_1-1, p_2-1, \cdots, p_k-1)$ for a squarefree integer $n = p_1p_2\cdots p_k$ and put $pa_i = p_i - 1$ for some integer a_i, then n is a primitive Carmichael number if $g(n) \leq [a_1, \cdots, a_k]$, and imprimitive if otherwise. Thus, since the observations of Shanks are more applicable to imprimitive Carmichael numbers and those of Erdős are more applicable to primitive Carmichael numbers, and most Carmichael numbers are in fact primitive whereas most Carmichael numbers with a fixed number of prime factors are imprimitive, then the two conjecturers easily reached different conclusions.

Interestingly, Pinch's counts of k-prime Carmichael numbers up to 10^{21} [25] reproduced in Table 1 imply that the number of prime factors of primitive Carmichael numbers tends to increase as x gets larger. As can be implied from Table 1, for the maximum number of distinct prime factors $k(x) \ll \frac{\log x}{\log^{(2)} x}$,

$$C(x) = C_3(x) + C_4(x) + C_5(x) + \cdots + C_{k(x)}(x), \tag{1.0.4}$$

where $\log^{(j)} x$ denotes the j-fold iteration of the natural logarithm for $j \geq 2$ (we shall use this notation from now on). Moreover, if we allow $C_k(x)$ to represent the number of Carmichael numbers $\leq x$ with precisely $k \geq 3$ prime factors, then it is conjectured that

$$C_k(x) = \Omega_k(x^{1/k}/\log^k x). \tag{1.0.5}$$

Table 2: Values of $h(x)$

Bound	$h(x)$
10^3	2.93319
10^4	2.19547
10^5	2.07632
10^6	1.97946
10^7	1.93388
10^8	1.90495
10^9	1.87989
10^{10}	1.86870
10^{11}	1.86421
10^{12}	1.86377
10^{13}	1.86240
10^{14}	1.86293
10^{15}	1.86301
10^{16}	1.86406
10^{17}	1.86472
10^{18}	1.86522
10^{19}	1.86565
10^{20}	1.86598
10^{21}	1.86619

Returning to the Erdős-Shanks controversy, Pomerance [2] sharpened the conjecture in (1.0.3) for all large x in order to be consistent with both Shanks' and Erdős' observations. Define the function $h(x)$ as

$$C(x) = x \cdot \exp\Big\{ -h(x)\frac{\log x \log^{(3)} x}{\log^{(2)} x}\Big\}. \tag{1.0.6}$$

According to Pomerance, distribution of Carmichael numbers is given by

$$C(x) = x^{1-\frac{\{1+o(1)\}\log^{(3)} x}{\log^{(2)} x}}, \tag{1.0.7}$$

for x sufficiently large. Unfortunately, according to Pinch [26], there appears to be no limiting value on h as indicated by the recent counts of Carmichael numbers up to 10^{21}. It is obvious that (1.0.7) holds iff $\lim h = 1$ in (1.0.6). However, Pinch [26] explains that the decrease in h is reversed between 10^{13} and 10^{14}, which is presented in Table 2. In fact, there is no clear evidence that suggests $\lim h = 1$.

As a result, we present an alternate conjecture

Conjecture 1.0.2.

$$C(x) \sim \frac{C_3(x)\mathscr{P}_b(x)}{\mathscr{P}_{b,2}(x)}, \tag{1.0.8}$$

where $\mathscr{P}_{b,2}(x)$ is the number of two-prime base b pseudoprimes $\leq x$ and $C_3(x)$ is the number of three-prime Carmichael numbers $\leq x$.

From Conjecture 1.0.2, we derive a corollary that the number of Carmichael numbers up to x sufficiently large is

Corollary 1.0.3.

$$C(x) \sim \frac{\psi' x^{\frac{5}{6}}}{\log x \cdot L(x)} \sim \frac{\psi_1' x^{\frac{1}{2}} \log^2 x \int_2^{x^{\frac{1}{3}}} \frac{dt}{\log^3 t}}{L(x)}, \tag{1.0.9}$$

where $L(x) = \exp\{\frac{\log x \log^{(3)} x}{\log^{(2)} x}\}$, $\psi' = \frac{\tau_3}{C}$, $\psi_1' = \frac{\tau_3}{27C}$. *If we let* p, q, *and* d *be odd primes, and we define* $\omega_{a,b,c}(p)$ *as the number of distinct residues modulo* p *represented by* a, b, c, *then the constants* C *and* τ_3 *are explicitly given as such,*

$$C = 4T \sum_{s\geq 1} \sum_{\substack{r>s \\ \gcd(r,s)=1}} \frac{\delta(rs)\rho(rs(r-s))}{(rs)^{\frac{3}{2}}} \tag{1.0.10}$$

$$T = 2\prod_d \frac{1-2/d}{(1-1/d)^2},$$

$$\rho(m) = \prod_{d|m} \frac{d-1}{d-2},$$

$$\delta(m) = \begin{cases} 2, & \text{if } 4 \mid m; \\ \\ 1, & \text{if otherwise.} \end{cases}$$

$$\tau_3 = \kappa_3 \lambda, \tag{1.0.11}$$

$$\lambda := 121.5 \prod_{p>3} \left(\frac{1-3/p}{(1-1/p)^3} \right),$$

$$\kappa_3 = \sum_{n\geq 1} \frac{\gcd(n,6)}{n^{4/3}} \prod_{\substack{p|n \\ p>3}} \frac{p}{p-3} \sum_{\substack{a<b<c,\, n=abc \\ a,b,c\, pairwise\, coprime}} \delta'(a,b,c) \prod_{\substack{p\nmid n \\ p>3}} \frac{p-\omega_{a,b,c}(p)}{p-3},$$

$$\delta'(a,b,c) = \begin{cases} 2, & \text{if } a \equiv b \equiv c \not\equiv 0 \pmod 3; \\ \\ 1, & \text{if otherwise.} \end{cases}.$$

Based upon the computation of C made by Galway [29] and the evaluation of κ_3 by Chick and Davies [16], we believe that ψ' will approach 69.51 and ψ_1' will approach 2.57; although these values are not yet borne out by the data. We also demonstrate that Corollary 1.0.3 fits the proven upper and lower bounds for $C(x)$, that ψ' and ψ_1' appear to approach constant values based upon Pinch's data, and we support Conjecture 1.0.2 through computational efforts.

In private communication [13], Pomerance suggests to us that the reason for $h(x)$ not approaching its conjectural limit of 1 is that "some secondary terms may be present. So, say in my conjecture, one replaces "$\log^{(3)} x$" with "$\log^{(3)} x + \log^{(4)} x$". It is the same conjecture, since the two are asymptotic...and so the Pinch phenomenon is banished". Hence, if secondary terms do indeed exist, then another conjecture regarding $C(x)$ would be to sharpen the heuristic arguments in [5] which, as a consequence, may better match the *actual* counts of Carmichael numbers. Since these heuristic arguments are dependent upon the number of y-smooth numbers up to x, represented by $\Psi(x, y)$, with y in the vicinity of $\exp\{(\log x)^{\frac{1}{2}}\}$, then it would suffice to utilize improvements concerning the asymptotic distribution of these numbers in the aforementioned region. As a result of these endeavors, we obtain the more precise heuristic:

Conjecture 1.0.4. *Let $\pi(x)$ be the prime counting function, for x sufficiently large $C(x)$ is*

$$x^{1-\frac{\{1+o(1)\}\log^{(3)} x+1}{\log^{(2)} x}}. \tag{1.0.12}$$

In Table 3, we define the function

$$a(x) := \left(\frac{(\log^{(2)} x)^2 \pi((\log x)^{\log^{(2)} x}) \exp\{-\{1+o(1)\}\log^{(2)} x \log^{(3)} x\}}{\log x} \right)^{\log x/(\log^{(2)} x)^2}.$$

Although Conjecture 1.0.4 states the same result and is a much more simplified version of $a(x)$, $a(x)$ is a slightly more precise version (for $x < 10^{100}$) of the conjecture and is thus used in the table instead of (1.0.12).

The reader may wonder why *two* conjectures are presented. The reason is that due to the lack of information regarding Carmichael numbers and their distribution. Corollary 1.0.3 asserts that if the values of ψ' and ψ_1' can be accurately determined then an asymptotic formula for $C(x)$ easily follows. Conjecture 1.0.4 indicates to us that if secondary terms exist, then the relation between the functions $\Psi(x, y)$ and $\Psi'(x, y)$ *must* be examined further (we explain this concept fully in §3.4) to effectively prove an equality for $C(x)$. We should note that the values of $C(x)$ predicted by Corollary 1.0.3 and Conjecture 1.0.4 appear to be closer to the actual values of $C(x)$ than Pomerance's conjecture in (2.2.2). Moreover, at least up to 10^{21}, it appears that Conjecture 1.0.4 is presenting more accurate values of $C(x)$ than Corollary 1.0.3; although, this may cease to be the case for larger bounds. In fact, (1.0.12) is asymptotically the same as (1.0.7); however, the usage of secondary terms in the former equation provides sharper estimates for smaller bounds than does (1.0.7).

Table 3: Comparisons between the actual and predicted Carmichael number counts

Bound	$C(x)$	$\frac{69.51x^{\frac{5}{6}}}{\log x \cdot L(x)}$	$\frac{2.57x^{\frac{1}{2}} \log^2 x \int_2^{x^{\frac{1}{3}}} \frac{dt}{\log^3 t}}{L(x)}$	$a(x)$	$x^{1-\frac{\{1+o(1)\}\log^{(3)} x}{\log^{(2)} x}}$
10^3	1	301.95	1092.82	3.50	94.89
10^4	7	594.43	2835.17	7.81	365.59
10^5	16	1316.29	6640.29	18.18	1485.33
10^6	43	3131.53	14806.24	43.43	6224.10
10^7	105	7826.17	32411.27	107.50	26636.80
10^8	255	20282.91	71150.56	274.074	115803.60
10^9	646	54070.80	159157.24	724.86	509769.35
10^{10}	1547	147451.71	367012.00	1926.56	2267174.18
10^{11}	3605	409716.38	878601.38	5245.56	10171329.99
10^{12}	8241	1156637.85	2188667.23	14488.22	45977679.09
10^{13}	19279	3309970.24	5664006.88	40424.93	209219668.02
10^{14}	44706	9585268.36	15162465.67	114558.014	957710051.36
10^{15}	105212	28049810.91	41763706.96	329251.92	4407472357.25
10^{16}	246683	82852448.55	117743387.56	955940.22	20382638275.29
10^{17}	585355	246785788.13	338238941.70	2796027.81	94682736406.04
10^{18}	1401644	740679196.52	986503770.93	8260103.95	441642695710.74
10^{19}	3381806	2238429061.23	2913197684.15	24637581.64	2067911761776.64
10^{20}	8220777	6807841639.58	8692508977.60	74026750.39	9717200728399.57
10^{21}	20138200	20826296835.28	26167265004.43	224193470.90	45814162191297.01

2 Preliminaries

Before delving into the main results of this paper, we shall first present results regarding pseudoprimes and Carmichael numbers that we will explicitly use later on in our derivations.

2.1 Pseudoprimes

Currently, the tightest bounds for pseudoprime distribution have been proven by Pomerance [27] [5].

Theorem 2.1.1 (R. A. Mollin 1989, Pomerance 1981)**.** *For the base* 2 *pseudoprime counting function,* $\exp\{(\log x)^{\frac{85}{207}}\} \leq \mathscr{P}_2(x) \leq x \cdot L(x)^{\frac{-1}{2}}$*, where* $L(x) = \exp\{\frac{\log x \log^{(3)} x}{\log^{(2)} x}\}$*. These bounds are applicable to* $\mathscr{P}_b(x)$ *for* $x \geq x_0(b)$*.*

Theorem 2.1.2 (Pomerance 1981)**.** *If we allow* $l_2(n)$ *to denote the exponent with multiplicative order of* 2 *modulo* n*, then* n *is a pseudoprime* (base 2) *iff* $l_2(n) \mid n-1$*.*

Conjecture 2.1.3 (Pomerance 1981)**.** *The number of solutions* w *for all* n *and* x *sufficiently large is,*

$$\#\{w \leq x : l_2(w) = n\} \leq x \cdot L(x)^{-1+\theta(x)}, \lim_{x\to\infty} \theta(x) = 0. \tag{2.1.1}$$

As a result, the number of base b *pseudoprimes for sufficiently large* $x \geq x_0(b)$ *is conjectured to be,*

$$\mathscr{P}_b(x) \sim x \cdot L(x)^{-1}. \tag{2.1.2}$$

Galway [29] has recently conjectured a formula for the distribution of pseudoprimes with two distinct prime factors, p and q, based on a longstanding conjecture of Hardy and Wright concerning the density of prime pairs. He noticed that a majority of these pseudoprimes satisfy the relation $\frac{p-1}{q-1} = \frac{r}{s}$, where r and s are small coprime integers. Thus, we heuristically have

Conjecture 2.1.4 (Galway 2004)**.** *Allow* p*,* q*, and* d *be odd primes, allow* $\mathscr{P}_{b,2}(x)$ *to represent the counting function for odd pseudoprimes with two distinct prime factors, and* $\mathscr{P}_{b,2}(x) := \#\{n \leq x : n = pq, p < q, \mathscr{P}_b(n)\}$*. Hence, as* $x \to \infty$*,*

$$\mathscr{P}_{b,2}(x) \sim \frac{Cx^{\frac{1}{2}}}{\log^2 x}, \tag{2.1.3}$$

where

$$C = 4T\sum_{s\geq 1} \sum_{\substack{r>s \\ \gcd(r,s)=1}} \frac{\delta(rs)\rho(rs(r-s))}{(rs)^{\frac{3}{2}}} \approx 30.03, \tag{2.1.4}$$

$$T = 2\prod_{d} \frac{1-2/d}{(1-1/d)^2} \approx 1.32, \tag{2.1.5}$$

$$\rho(m) = \prod_{d|m} \frac{d-1}{d-2}, \tag{2.1.6}$$

$$\delta(m) = \begin{cases} 2, & \textit{if } 4 \mid m; \\ \\ 1, & \textit{if otherwise.} \end{cases} \tag{2.1.7}$$

Table 4: Values of C

Bound	$\mathscr{P}_{b,2}(x)$	C
10^3	0	0
10^4	11	9.331
10^5	34	14.251
10^6	107	20.423
10^7	311	25.550
10^8	880	29.860
10^9	2455	33.340
10^{10}	6501	34.468
10^{11}	17207	34.908
10^{12}	46080	35.181
10^{13}	123877	35.100
10^{14}	334567	34.767
10^{15}	915443	34.534
10^{16}	2520503	34.210
10^{17}	7002043	33.928

Galway's conjecture is somewhat supported by Table 4 for it appears that C is slowly approaching its predicted constant value of 30.03:

Let $\omega(n)$ represent the number of different prime factors of n. Also, given an integer sequence $\{m_i\}_{i=1}^{\infty}$, note that a prime p is said to be a primitive prime factor of m_i if p divides m_i but does not divide any m_j for $j < i$.

Lemma 2.1.5 (Erdős 1949)**.** *Let n be a base 2 pseudoprime. For every k, there exist infinitely many squarefree base 2 pseudoprimes with $\omega(n) = k$* [20].

Theorem 2.1.6. *There exist infinitely many squarefree base b pseudoprimes n for any $b \geq 2$ with $\omega(n) = k$ distinct prime factors.*

Proof. Let $\{n_j\}_{j=1}^{\infty}$ be an integer sequence of base b pseudoprimes such that each term is greater than its preceding term, and $\omega(n_i) = k-1$, for any n_i in $\{n_j\}_{j=1}^{\infty}$. Let p_i be one of the primitive prime factors of $b^{n_i-1}-1$. Since $b^{n_i-1} \equiv 1 \pmod{p_i \cdot n_i}$ and $b^{p_i-1} \equiv 1 \pmod{p_i}$, $p_i \cdot n_i$ is a pseudoprime to base b. We observe that $b^{p_i-1} \equiv 1 \pmod{n_i}$ because $p_i - 1 \equiv 0 \pmod{(n_i-1)}$. As a result, it follows that $b^{n_i-1} \equiv 1 \pmod{n_i}$. Also, $b^{n_i p_i-1} \equiv 1 \pmod{p_i \cdot n_i}$ since $b^{n_i p_i - 1} = b^{(n_i-1)(p_i-1)} \cdot b^{n_i-1} \cdot b^{p_i-1}$. Hence, $p_i \cdot n_i$ is squarefree and $\omega(p_i \cdot n_i) = k$. Moreover, every integer satisfying $p_i \cdot n_i$ is different because n_i is composite, $p_i > n_i$, and $p_i \equiv 1 \pmod{(n_i - 1)}$. ■

Theorem 2.1.7. *For any base b pseudoprime, $b \geq 2$, having $k \geq 2$ distinct prime factors and for x sufficiently large,*

$$\mathscr{P}_{b,k+1}(x) \geq \mathscr{P}_{b,k}(\log_b x). \tag{2.1.8}$$

Proof. Let n be a pseudoprime with $k > 1$ distinct prime factors. Since $n-1$ is the smallest exponent ϵ such that $p \mid b^{\epsilon} - 1$ and ϵ divides an exponent h such that $p \mid b^h - 1$, it follows from Fermat's little theorem that $p \mid b^{p-1} - 1$. Thus, from Zsigmondy's theorem, there exists a prime $p > n$ for which $p \mid b^{n-1} - 1$ and $n - 1 \mid p - 1$ for $b \geq 2$. As a result,

$$np \mid b^{n-1} - 1. \tag{2.1.9}$$

On the other hand, since $np-1 = n(p-1)+n-1$ and $n-1 \mid p-1$, $n-1 \mid np-1$ and $np \mid b^{np-1}-1$. If we let $n, m \in \mathbb{N}^*$, the set of positive natural numbers, such that $n \neq m$ and $p > n$, $q > m$, then $np \neq mq$ for primes p and q. However, suppose we let $np = mq$ and $p > n$, then $m \mid p$. Hence, $m \geq p$ and $m > n$. Unfortunately, the latter statement is contradictory, and as a result $np \neq mq$. If n and m are two different base b pseudoprimes with $k \geq 2$ distinct prime factors, then np and mq are distinct pseudoprimes as well.

From (2.1.9),

$$p \mid (b^{\frac{n-1}{2}} - 1)(b^{\frac{n-1}{2}} + 1), \tag{2.1.10}$$

and

$$p \leq b^{\frac{n-1}{2}} + 1 < b^{\frac{n}{2}}. \tag{2.1.11}$$

If $n \leq \log_b x$, then $pn < x^{\frac{1}{2}} \log_b x < x$. It then follows that for every base b pseudoprime n with k distinct prime factors, $n = p_1 p_2 \cdots p_k \leq \log_b x$, there is at least one base b pseudoprime such that $p_1 p_2 \cdots p_k p < x$. ■

2.2 Carmichael Numbers

Improving upon Erdős' results in [21], Pomerance [5] sharpened the upper bound on $C(x)$.

Theorem 2.2.1 (Pomerance 1981)**.**

$$C(x) \leq x \cdot \exp\left\{ -\frac{\log x}{\log^{(2)} x} \left(\log^{(3)} x + \log^{(4)} x + \frac{\log^{(4)} x - 1}{\log^{(3)} x} + O\left(\left(\frac{\log^{(4)} x}{\log^{(3)} x} \right)^2 \right) \right) \right\}. \tag{2.2.1}$$

In the other direction, Alford, Granville, and Pomerance proved a lower bound for $C(x)$ for x sufficiently large [3].

Theorem 2.2.2 (Alford-Granville-Pomerance 1994)**.**

$$C(x) > x^{\frac{2}{7}}, \tag{2.2.2}$$

thus there are infinitely many Carmichael numbers.

Recently, Harman improved this lower bound [15].

Theorem 2.2.3 (Harman 2005)**.**

$$C(x) > x^{0.33336704}, \tag{2.2.3}$$

It is not yet even known if $C(x) > x^{\frac{1}{2}}$*.*

We provide in Table 5 a computation of the exponent β for which $C(x) = x^{\beta}$ for a sufficient value of x up to 10^{21}.

Conjecture 2.2.4 (Granville-Pomerance 2001)**.** *If we let* $C_3(x)$ *be the counting function for Carmichael numbers with* 3 *distinct prime factors, then*

$$C_3(x) \sim \tau_3 \frac{x^{\frac{1}{3}}}{\log^3 x} \sim \frac{\tau_3}{27} \int_2^{x^{\frac{1}{3}}} \frac{dt}{\log^3 t}, \tag{2.2.4}$$

where

$$\tau_3 = \kappa_3 \lambda \approx 2100, \tag{2.2.5}$$

Table 5: Values of β

Bound	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
$C(x)$	1	7	16	43	105	255	646	1547
β	0	0.21127	0.24082	0.27224	0.28874	0.30082	0.31225	0.31895

Bound	10^{11}	10^{12}	10^{13}	10^{14}	10^{15}	10^{16}
$C(x)$	3605	8241	19279	44706	105212	246683
β	0.32336	0.32633	0.32962	0.33217	0.33480	0.33700

Bound	10^{17}	10^{18}	10^{19}	10^{20}	10^{21}
$C(x)$	585355	1401644	3381806	8220777	20138200
β	0.33926	0.34148	0.34364	0.34575	0.34781

$$\lambda := 121.5\prod_{p>3}\left(\frac{1-3/p}{(1-1/p)^3}\right) \approx 77.1727, \tag{2.2.6}$$

$$\kappa_3 = \sum_{n\geq 1}\frac{\gcd(n,6)}{n^{4/3}}\prod_{\substack{p\mid n\\ p>3}}\frac{p}{p-3}\sum_{\substack{a<b<c,\, n=abc\\ a,b,c\, pairwise\, coprime}}\delta^{'}(a,b,c)\prod_{\substack{p\nmid n\\ p>3}}\frac{p-\omega_{a,b,c}(p)}{p-3}, \tag{2.2.7}$$

$$\delta^{'}(a,b,c) = \begin{cases} 2, & \textit{if } a\equiv b\equiv c \not\equiv 0 \pmod 3; \\ \\ 1, & \textit{if otherwise.}\end{cases}, \tag{2.2.8}$$

and $\omega_{a,b,c}(p)$ *is the number of distinct residues modulo* p *represented by* a, b, c.

Recent provisional estimates by Chick and Davies [16] of the slowly converging infinite series κ_3 suggest that $\kappa_3 = 27.05$ which gives $\tau_3 = 2087.5$.

3 On the Distribution of Carmichael Numbers

3.1 Two Conjectures Regarding k-prime Pseudoprimes and k-prime Carmichael numbers

We conjecture the following relations:

Conjecture 3.1.1. *For any fixed* $k \geq 2$*, let* $\mathscr{P}_{b,k}(x)$ *denote the counting function for base* b *pseudoprimes with* k *distinct prime factors, and let* $\mathscr{P}_b(x)$ *denote the counting function for base* b *pseudoprimes. Asymptotically,*

$$\frac{\mathscr{P}_{b,k}(x)}{\mathscr{P}_b(x)} = o(1). \tag{3.1.1}$$

In other terms, for any fixed base $b > 1$*, the* k*-prime base* b *pseudoprimes,* $\mathscr{P}_{b,k}(x)$*, form a set of relative density 0 in the set of* all *base* b *pseudoprimes,* $\mathscr{P}_b(x)$*, for that same value of* b*.*

We are only able to partially support Conjecture 3.1.1. First, we express the ratio $\frac{\mathscr{P}_{b,k}(x)}{\mathscr{P}_b(x)}$ as,

$$\frac{\mathscr{P}_{b,k}(x)}{\mathscr{P}_b(x)} = \frac{\mathscr{P}_{b,k}(x)}{\sum_{i=2}^{k(x)}\mathscr{P}_{b,i}(x)}, \tag{3.1.2}$$

where the maximum number of distinct prime factors, $k(x)$, of any integer $\leq x$ is $k(x) \ll \frac{\log x}{\log^{(2)} x}$. Let $\log_b^{(j)} x$ denote the the j-fold iteration of the base b logarithm. Thus,

$$\sum_{i=2}^{g(x)} \mathscr{P}_{b,i}(x) = \sum_{i=2}^{k-1} \mathscr{P}_{b,i}(x) + \mathscr{P}_{b,k}(x) + \sum_{i=k+1}^{k(x)} \mathscr{P}_{b,i}(x). \tag{3.1.3}$$

Due to Theorem 2.1.7, for any $h \leq k$ in (3.1.3), $\mathscr{P}_{b,k}(x) \geq \mathscr{P}_{b,h}(\log_b^{(k-h)} x)$, and for any $w \geq k$ in (3.1.3), $\mathscr{P}_{b,w}(x) \geq \mathscr{P}_{b,k}(\log_b^{(w-k)} x)$. Hence,

$$\sum_{i=2}^{k-1} \mathscr{P}_{b,i}(x) \leq \mathscr{P}_{b,2}(\log_b^{(k-2)} x) + \mathscr{P}_{b,3}(\log_b^{(k-3)} x) + \cdots + \mathscr{P}_{b,k-1}(\log_b x) \tag{3.1.4}$$

We cut off the terms from proceeding until $\frac{\log x}{\log^{(2)} x}$ because if such were the case, then no x could be sufficiently large to satisfy (3.1.5),

$$\sum_{i=k+1}^{k(x)} \mathscr{P}_{b,i}(x) \geq \mathscr{P}_{b,k+1}(\log_b x) + \cdots + \mathscr{P}_{b,r(x)}(\log_b^{(r(x)-k)} x), \tag{3.1.5}$$

where $r(x)$ is any function that grows slower than $\log^* x$, the iterated logarithm. We explicitly define $\log^* x$ as

$$\log^* x := \begin{cases} 0 & \text{if } x \leq 1; \\ 1 + \log^*(\log x) & \text{if } x > 1 \end{cases}. \tag{3.1.6}$$

Remark 3.1.2. We should note that the support for Conjecture 3.1.1 is rather weak. This is largely due to the weakness of Szymiczek's construction, $\mathscr{P}_{b,k+1}(x) \geq \mathscr{P}_{b,k}(\log_b x)$, in Theorem 2.1.7. We believe that the latter relation can be strengthened if a polynomial decrease can be proven. In other words, if $\mathscr{P}_{b,k+1}(x) \geq \mathscr{P}_{b,k}(x^c)$ for some $c \in (0,1)$. Similarly, in our support for Conjecture 3.1.1, we defined the function $r(x)$ as any function that grows slower than $\log^* x$, the iterated logarithm. Although it is not hard to see that any function growing faster than $\log^* x$ will fail, it is not obvious whether any function growing at the same rate as $\log^* x$ will succeed. However, we have several reasons to strongly believe that $r(x) = \log^* x$. First, for practical values of $x \leq 2^{65536}$ the iterated logarithm grows much more slowly than the logarithm. Second, the iterated logarithm's relation to the super-logarithm also supports its slow growth. Third, higher bases give smaller iterated logarithms, and $\log^* x$ is well defined for any base greater than $\exp\left\{\frac{1}{e}\right\}$. This implies that for any base $b \geq 2$, the iterated logarithm will grow even more slowly for higher pseudoprime bases.

Conjecture 3.1.3. *For any fixed $k \geq 3$, let $C_k(x)$ denote the number of k-prime Carmichael numbers up to x, and let $C(x)$ denote the Carmichael number counting function. Asymptotically,*

$$\frac{C_k(x)}{C(x)} = o(1). \tag{3.1.7}$$

3.2 Support for Conjecture 3.1.1 and Conjecture 3.1.3

So far, the claim established by Conjecture 3.1.1 is not yet borne out by the data in Table 6. We believe that the ratio $\frac{\mathscr{P}_{b,2}(x)}{\mathscr{P}_2(x)}$ will approach 0, but may do so slowly at first. On the other hand, it appears that the ratio $\frac{C_3(x)}{C(x)}$ in Table 7 rapidly approaches 0, thereby supporting Conjecture 3.1.3.

Table 6: Values of $\frac{\mathscr{P}_{b,2}(x)}{\mathscr{P}_2(x)}$

Bound	$\mathscr{P}_{b,2}(x)$	$\mathscr{P}_2(x)$	$\frac{\mathscr{P}_{b,2}(x)}{\mathscr{P}_2(x)}$
10^3	0	3	0.00
10^4	11	22	0.50
10^5	34	78	0.44
10^6	107	245	0.44
10^7	311	750	0.41
10^8	880	2057	0.43
10^9	2455	5597	0.44
10^{10}	6501	14884	0.44
10^{11}	17207	38975	0.44
10^{12}	46080	101629	0.45
10^{13}	123877	264239	0.47
10^{14}	334567	687007	0.49
10^{15}	915443	1801533	0.51
10^{16}	2520503	4744920	0.53
10^{17}	7002043	12604009	0.56

Furthermore, Pomerance, Selfridge, and Wagstaff's famous results [10] support both conjectures. In Conjecture 1 of their paper, they believe that for each $\epsilon > 0$, there is an $x_0(\epsilon)$ such that for all $x \geq x_0(\epsilon)$,

$$C(x) > x \cdot \exp\left\{\frac{-\{2+\epsilon\}\log x \cdot \log^{(3)} x}{\log^{(2)} x}\right\}. \tag{3.2.1}$$

Pomerance, Selfridge, and Wagstaff [10] show that $\mathscr{P}_{b,k}(x) \leq O_k(x^{2k/(2k+1)})$. If (3.2.1) is true, then the pseudoprimes "with exactly k prime factors form a set of relative density 0 in the set of all [pseudoprimes]" [10]. Similarly, in Theorem 7 of Granville and Pomerance [2], it is proven that $C_k(x) \leq x^{2/3+o_k(1)}$, and if (3.2.1) holds, "then for each k, $C_k(x) = o(C(x))$" [10].

Interestingly, we can also support the statements in Conjecture 3.1.1 and Conjecture 3.1.3 by relating them to their composite superset. Let the number of composites $\leq x$ with k distinct prime factors be denoted by $\pi_k(x)$ and let the number of composites $\leq x$ with k prime factors (not necessarily distinct) be represented by $\tau_k(x)$. Hence, we can prove upper and lower bounds for $\pi_k(x)$. In 22.18.2 of Hardy and Wright [14] for $k \geq 1$,

$$k!\pi_k(x) \leq \Pi_k(x) \leq k!\tau_k(x), \tag{3.2.2}$$

where $\Pi_k(x) = \frac{\vartheta_k(x)}{\log x} + O(\frac{x}{\log x})$ in 22.18.5. In 22.18.24, since $\vartheta_k(x) = \Pi_k(x)\log x - \displaystyle\int_2^x \frac{\Pi_k(x)}{t}dt \sim kx(\log^{(2)} x)^{k-1}$ for $k \geq 2$ and $\displaystyle\int_2^x \frac{\Pi_k(x)}{t}dt = O(x)$, $\Pi_k(x) \sim \frac{kx(\log^{(2)} x)^{k-1}}{\log x}$. As a result, it follows that

$$\pi_k(x) \leq (1+o(1))\frac{x(\log^{(2)} x)^{k-1}}{(k-1)!\log x}. \tag{3.2.3}$$

In the same respect, a lower bound for π_k can be formulated. In 22.18.3 it is proven that,

$$\tau_k(x) - \pi_k(x) \leq \sum_{p_1p_2\cdots p_{k-1}^2 \leq x} 1 \leq \sum_{p_1p_2\cdots p_{k-1} \leq x} 1 := \Pi_{k-1}(x). \tag{3.2.4}$$

Table 7: Values of $\frac{C_3(x)}{C(x)}$

Bound	$C_3(x)$	$C(x)$	$\frac{C_3(x)}{C(x)}$
10^3	1	1	1.00
10^4	7	7	1.00
10^5	12	16	0.75
10^6	23	43	0.53
10^7	47	105	0.45
10^8	84	255	0.33
10^9	172	646	0.27
10^{10}	335	1547	0.22
10^{11}	590	3605	0.16
10^{12}	1000	8241	0.12
10^{13}	1858	19279	0.096
10^{14}	3284	44706	0.073
10^{15}	6083	105212	0.058
10^{16}	10816	246683	0.044
10^{17}	19539	585355	0.033
10^{18}	35586	1401644	0.025
10^{19}	65309	3381806	0.019
10^{20}	120625	8220777	0.015
10^{21}	224763	20138200	0.011

Since $\pi_k(x) \geq \tau_k(x) - \Pi_{k-1}(x)$ and $\pi_k(x) \geq \frac{\Pi_k(x)}{k!} - \Pi_{k-1}(x)$,

$$\pi_k(x) \geq O\left(\frac{x(\log^{(2)} x)^{k-1}}{(k-1)!\log x}\right) - \frac{(k-1)x(\log^{(2)} x)^{k-2}}{\log x} + O\left(\frac{x}{\log x}\right).$$

We can improve the upper bound given in (3.2.3) to an equality,

$$\pi_k(x) \sim \frac{x(\log^{(2)} x)^{k-1}}{(k-1)!\log x}. \tag{3.2.5}$$

By the Erdős-Kac Theorem [22], we can formulate the probability that a number near x has k distinct prime factors using the fact that these numbers are distributed with a mean and variance of $\log^{(2)} x$. Hence, setting $\log^{(2)} x$ as the λ of the Poisson distribution $\mathrm{P}(k;\lambda)$ and taking its limit for any fixed k,

$$\lim_{x\to\infty} \mathrm{P}(k;\lambda) = \lim_{x\to\infty} \frac{(\log^{(2)} x)^{k-1}\exp\{-\log^{(2)} x\}}{(k-1)!} = 0, \tag{3.2.6}$$

where the asymptotic error bound is given by $O(\frac{1}{\log^{(2)} x})$ [17]. However, we caution the reader to consider that just because the probability of a general composite near x having k distinct prime factors goes to 0, does not necessarily fully prove that this probability will hold for either $\mathscr{P}_{b,k}(x)$ or $C_k(x)$.

3.3 An Alternate Conjecture

From Conjecture 3.1.1 and Conjecture 3.1.3, it is evident that the k-prime pseudoprimes and the k-prime Carmichael numbers are much more sparsely distributed than the set of all pseudoprimes and Carmichael numbers, respectively. We hypothesize that if k is minimized for both the k-prime pseudoprimes and the k-prime Carmichael numbers, then the ratios $\frac{\mathscr{P}_{b,2}(x)}{\mathscr{P}_b(x)}$ and $\frac{C_3(x)}{C(x)}$ will roughly achieve the same values for large enough x. We also recommend using the minimum number of distinct prime factors for both the pseudoprimes and the Carmichael numbers because first, there is no overlap between the three-prime Carmichael numbers and two-prime pseudoprimes and second, the distinct prime factors cannot be arbitrarily chosen. This idea leads us to believe that,

$$C(x) \sim \frac{C_3(x)\mathscr{P}_b(x)}{\mathscr{P}_{b,2}(x)}.$$

As a result, assuming Conjecture 2.1.3, Conjecture 2.1.4, Conjecture 2.2.4, and Conjecture 1.0.2, the amount of Carmichael numbers $\leq x$ given by the counting function $C(x)$ is conjectured to be for x sufficiently large,

$$C(x) \sim \frac{\psi' x^{\frac{5}{6}}}{\log x \cdot L(x)} \sim \frac{\psi_1' x^{\frac{1}{2}} \log^2 x \int_2^{x^{\frac{1}{3}}} \frac{dt}{\log^3 t}}{L(x)}, \tag{3.3.1}$$

where

$$\psi' = \frac{\tau_3}{C} \tag{3.3.2}$$

and

$$\psi_1' = \frac{\tau_3}{27C}. \tag{3.3.3}$$

In Table 8, the computed values of ψ' and ψ_1' up to 10^{21} are given. Hence, not only does Corol-

Table 8: Values of ψ' and ψ_1'

Bound	$C(x)$	$\frac{69.51x^{\frac{5}{6}}}{\log x \cdot L(x)}$	$\frac{2.57x^{\frac{1}{2}} \log^2 x \int_2^{x^{\frac{1}{3}}} \frac{dt}{\log^3 t}}{L(x)}$	ψ'	ψ_1'
10^3	1	301.95	1092.82	0.2302	0.0024
10^4	7	594.43	2835.17	0.8185	0.0063
10^5	16	1316.29	6640.29	0.8449	0.0062
10^6	43	3131.53	14806.24	0.9545	0.0075
10^7	105	7826.17	32411.27	0.9326	0.0083
10^8	255	20282.91	71150.56	0.8739	0.0092
10^9	646	54070.80	159157.24	0.8305	0.0104
10^{10}	1547	147451.71	367012.00	0.7293	0.0108
10^{11}	3605	409716.38	878601.38	0.6116	0.0105
10^{12}	8241	1156637.85	2188667.23	0.4953	0.0097
10^{13}	19279	3309970.24	5664006.88	0.4049	0.0087
10^{14}	44706	9585268.36	15162465.67	0.3242	0.0076
10^{15}	105212	28049810.91	41763706.96	0.2607	0.0065
10^{16}	246683	82852448.55	117743387.56	0.2070	0.0054
10^{17}	585355	246785788.13	338238941.70	0.1649	0.0044
10^{18}	1401644	740679196.52	986503770.93	0.1315	0.0037
10^{19}	3381806	2238429061.23	2913197684.15	0.1050	0.0030
10^{20}	8220777	6807841639.58	8692508977.60	0.0839	0.0024
10^{21}	20138200	20826296835.28	26167265004.43	0.0672	0.0020

lary 1.0.3 fit the proven bounds for $C(x)$ given in Theorem 2.2.1 and Theorem 2.2.3, but both ψ' and ψ_1' appear to be approaching constant values. However, there are several reasons as to why Corollary 1.0.3 may not be necessarily borne out by the data in the above table. For instance, the infinite series κ_3 is slowly convergent, and it is not until 10^{24} that κ_3 appears to approach its estimated value of 2087.5. However, the primary source of inaccuracy is due to Conjecture 2.1.3. Since Pomerance's conjecture for the distribution of pseudoprimes is applicable for sufficiently large x and pseudoprime counts have only recently been conducted to 10^{17} by Galway and Feitsma, we are not sure how "sufficiently large" x must be for Conjecture 2.1.3 to be an accurate model for pseudoprime distribution. Lastly, x must also be immensely large in order for $\frac{\mathscr{P}_{b,2}(x)}{\mathscr{P}_b(x)} = o(1)$.

3.4 An Improved Heuristic Argument

As mentioned before, Pomerance's heuristic arguments supporting his conjecture in (2.2.2) involve the distribution of smooth numbers. And, if secondary terms exist, then it would be worthwhile to sharpen these heuristics to produce a conjecture for $C(x)$. Let $\Psi(x, y)$ denote the number of y-smooth numbers $\leq x$ and let $\Psi'(x, y)$ denote the number of primes $p \leq x$ for which $p-1$ is squarefree and its prime factors are $\leq y$ [10]. It is conjectured in [5] that for $\exp\{\frac{1}{2}(\log x)^{\frac{1}{2}}\} \leq y \leq \exp\{(\log x)^{\frac{1}{2}}\}$,

$$\frac{1}{x}\Psi(x, y) \sim \frac{1}{\pi(x)}\Psi'(x, y). \tag{3.4.1}$$

If $0 < \alpha < 1$, it is well-known [1] that

$$\Psi\left(x, \exp\{c(\log x)^{\alpha}(\log^{(2)} x)^{\beta}\}\right) = x\exp\{-\{(1-\alpha)/c + o(1)\}(\log x)^{1-\alpha}(\log^{(2)} x)^{1-\beta}\}. \tag{3.4.2}$$

Concerning Carmichael numbers, we are interested in the case for which $\alpha = \frac{1}{2}$, $\beta = 0$, and $c = 1$. Hence,

$$\Psi\left(x, \exp\{(\log x)^{\frac{1}{2}}\}\right) = x\exp\{-\{1/2 + o(1)\}(\log x)^{\frac{1}{2}}(\log^{(2)} x)\}. \tag{3.4.3}$$

From (3.4.1) and (3.4.3), we make the following

Conjecture 3.4.1. *For* $\exp\{\frac{1}{2}(\log x)^{\frac{1}{2}}\} \leq y \leq \exp\{(\log x)^{\frac{1}{2}}\}$,

$$\Psi'(x, y) = \pi(x)\exp\{-\{1/2 + o(1)\}(\log x)^{\frac{1}{2}}(\log^{(2)} x)\}. \tag{3.4.4}$$

Let $A(x)$ denote the product of the primes $p \leq \log x/(\log^{(2)} x)^2$. Thus, $A(x) < x^{2/\log^{(2)} x}$ as in [10]. If we allow $r_1, \ldots, r_q$ to be the primes in the interval $\left(\log x/(\log^{(2)} x)^2, (\log x)^{\log^{(2)} x}\right)$ with $r_i - 1 \mid A(x)$. By Conjecture 3.4.1 we have for x sufficiently large,

$$q = \pi\left((\log x)^{\log^{(2)} x}\right)\exp\{-\{1 + o(1)\}\log^{(2)} x \log^{(3)} x\}. \tag{3.4.5}$$

Let $m_1, \ldots, m_N$ be the squarefree composite integers $\leq x$ composed of r_i and let

$$l = \left[\log x/(\log^{(2)} x)^2\right].$$

As discussed in [10], we have

$$N \geq \binom{q}{l} \geq \left(\frac{q}{l}\right)^l. \tag{3.4.6}$$

As a result,

$$N \geq \left(\frac{(\log^{(2)} x)^2 \pi((\log x)^{\log^{(2)} x}) \exp\{-\{1+o(1)\} \log^{(2)} x \log^{(3)} x\}}{\log x} \right)^{\log x / (\log^{(2)} x)^2}. \tag{3.4.7}$$

Since Euler's φ function and Carmichael's λ function are virtually the same, the lower bound in (3.4.7) should be applicable to $C(x)$. In fact, from the values of $a(x)$ in Table 3 and the precision of Conjecture 3.4.1, we have reason to believe that this result is asymptotically close to the actual value of $C(x)$.

4 One-Parameter Quadratic-Base Pseudoprimes: A Sidenote

As mentioned earlier, the discovery of Carmichael numbers demonstrated the fallability of Fermat's primality test therefore lending to the development of efficient probabilistic primality tests. Baillie, Pomerance, Selfridge, and Wagstaff [10] [23] have determined a primality test that is an amalgamation of the Miller-Rabin test and a Lucas test. However, even though Pomerance [7] presented a heuristic argument that the number of counter-examples up to x was $\gg x^{1-\epsilon}$ for $\epsilon > 0$, we have not been able to find any counter-examples up to 10^{17}. In fact, no precise probability of error has been given about this test either [30].

Grantham [18] has also provided a probable prime test known as the RQFT that has a *known* worst-case probability of error of 1/7710 per iteration.

An even stronger test known as the One-Parameter Quadratic-Base Test (OPQBT) has been given by Zhang [30], and is a version of the Baillie-PSW test that not only has a known probability of error but is more efficient than the RQFT except for a thin set of cases. We let $u (\neq \pm 2) \in \mathbb{Z}$, let $T_u = T \pmod{T^2 - uT + 1}$, and define the ring associated with parameter u as

$$R_u = \mathbb{Z}[T]/(T^2 - uT + 1) = \{a + bT_u : a, b \in \mathbb{Z}\}.$$

We then define an odd integer $n > 1$ as an OPQBT pseudoprime for $0 \leq u < n$ with

$$\epsilon = \left(\frac{u^2 - 4}{n} \right) \in \{-1, 1\},$$

where in the ring R_u, n must pass

$$T_u^{n-\epsilon} \equiv 1 \pmod{n}. \tag{4.0.8}$$

Moreover, n is defined as an OPQBT strong pseudoprime if for some $i = 0, 1, \cdots, k-1$, either

$$T_q^u \equiv 1 \pmod{n}, \tag{4.0.9}$$

or

$$T_u^{2^i q} \equiv -1 \pmod{n}, \tag{4.0.10}$$

in which for q odd, $n - \epsilon = 2^k q$ [30].

We have verified that there are no OPQBT pseudoprimes up to 10^{17}. Let the counting function $\mathscr{O}(x)$ denote that number of OPQBT pseudoprimes $\leq x$ and let $\mathscr{S}\mathscr{O}(x)$ denote the number of strong OPQBT pseudoprimes $\leq x$. The best upper bound we are able to prove is

$$\mathscr{S}\mathscr{O}(x) \leq \mathscr{O}(x) \leq x \cdot L(x)^{\frac{-1}{2}}, \tag{4.0.11}$$

since an upper bound on the pseudoprimes is applicable to an upper bound on the OPQBT pseudoprimes and strong OPQBT pseudoprimes.

Based upon Erdős' construction [21] and Pomerance's heuristics [7], in the interval $\left[H, H^j\right]$, for any fixed $j > 4$ and H sufficiently large, there are most likely $\exp\{H^2(1-4/j)\}$ counter-examples to Zhang's primality test, meaning that there are at least $x^{1-4/j}$ counter-examples below $x = \exp\{H^2\}$. Thus, for arbitrary j, the number of counter-examples to the OPQBT becomes generalized to $\gg x^{1-\epsilon}$ for $\epsilon > 0$. In other words, there are infinitely many counter-examples to Zhang's OPQBT.

Acknowledgements

I would like to express my gratitude to Charles R. Greathouse IV who provided invaluable guidance in the direction of this paper; Carl Pomerance who offered helpful comments on this paper in its initial stages; Harvey Dubner who provided guidance on the experimental aspects of this paper; Johan B. Henkens who helped me count the base 2 pseudoprimes and 2-strong pseudoprimes up to 10^{17} using Galway and Feitsma's data; Kazimierz Szymiczek who clarified Theorem 2.1.7; and David H. Low who assisted me with formatting and typesetting errors. I would also like to thank William F. Galway for sharing his viewpoints regarding the pseudoprimes with k distinct prime factors.

References

[1] A. Granville, *Smooth numbers: computational number theory and beyond*, Mathematical Sciences Research Institute Publications, **44** (2008) 1–58. `http://www.math.leidenuniv.nl/~psh/ANTproc/09andrew.pdf`.

[2] A. Granville and C. Pomerance, *Two Contradictory Conjectures Concerning Carmichael Numbers*, Math. Comp. **71** (2001): 883–908.

[3] A. Granville, C. Pomerance, and W. R. Alford, *There are Infinitely Many Carmichael Numbers*, Ann. of Math. **140** (1994): 703–722.

[4] A. Korselt, *Problème chinois*, L'intermèdiaire de mathématiciens **6** (1899): 142–143.

[5] C. Pomerance, *On the Distribution of Pseudoprimes*, Math. Comp. **37** (1981): 587–93.

[6] C. Pomerance, *A New Lower Bound for the Pseudoprime Counting Function*, Illinois J. Math. **26** (1982): 4–9.

[7] C. Pomerance, *Are there counter-examples to the Baillie-PSW primality test?*, Dopo Le Parole aangeboden aan Dr. A. K. Lenstra (H. W. Lenstra, jr., J. K. Lenstraand, P. Van Emde Boas, eds.), Amsterdam, 1984.

[8] C. Pomerance and R. Crandall, *Prime Numbers: A Computational Perspective*, New York: Springer, 2005.

[9] C. Pomerance and D. M. Gordon, *The Distribution of Lucas and Elliptic Pseudoprimes*, Math. Comp. **57** (1991): 825–38.

[10] C. Pomerance, J. L. Wagstaff, and S. S. Wagstaff, jr, *Pseudoprimes to* $25 \cdot 10^9$, Math. Comp. **35** (1980): 1003–026.

[12] D. Shanks, *Solved and unsolved problems in number theory*, 3rd ed., Chelsea, New York, 1985.

[13] Emails exchanged between A. Nayebi and C. Pomerance.

[14] E. M. Wright and G. H. Hardy, *An Introduction to the Theory of Numbers*, Oxford: Clarendon P, Oxford UP, 1998.

[15] G. Harman, *On the number of Carmichael numbers up to x*, Bull. Lond. Math. Soc. **37** (2005): 641–650.

[16] G. H. Davies and J. M. Chick, *The Evaluation of κ_3*, Math. Comp. **77** (2008): 547–550.

[17] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge: Cambridge UP, 1995.

[18] J. Grantham, *A probable prime test with high confidence*, J. Number Theory, **72** (1998) 32–47.

[19] K. Szymiczek, *On Pseudoprimes which are Products of Distinct Primes*, Amer. Math. Monthly **74** (1967): 35–37.

[20] P. Erdős, *On the Converse of Fermat's Theorem*, Amer. Math. Monthly, **56** (1949): 623–24.

[21] P. Erdős, *On pseudoprimes and Carmichael numbers*, Publ. Math. Debrecen **4** (1956): 201–206.

[22] P. Erdős and M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math **62** (1940): 738–742.

[23] R. Baillie and S. S. Wagstaff, jr., *Lucas pseudoprimes*, Math. Comp. **35** (1980): 1391–1417.

[24] R. D. Carmichael, *Note on a new number theory function*, Bull. Amer. Math. Soc. **16** (1910): 232–238.

[25] R. G.E. Pinch, *The Carmichael Numbers up to* 10^{21}, Proceedings Conference on Algorithmic Number Theory, Turku, May 2007. Turku Centre for Computer Science General Publications 46, edited by Anne-Maria Ernvall-Hytönen, Matti Jutila, Juhani Karhumäki and Arto Lepistö.

[26] R. G.E. Pinch, *The Carmichael Numbers up to* 10 *to the* 21, Eighth Algorithmic Number Theory Symposium ANTS-VIII May 17–22, 2008 Banff Centre, Banff, Alberta (Canada).

[27] W. F. Galway, *The Pseudoprimes below* 2^{64}, Simon Fraser University, 2002, `http://oldweb.cecm.sfu.ca/pseudoprime/psp-search-slides.pdf`.

[28] W. F. Galway, *Tables of pseudoprimes and related data*, 2002, `http://oldweb.cecm.sfu.ca/pseudoprime/psp1e15.gz`.

[29] W. F. Galway, *Research Statement*, 2004, `http://www.math.uiuc.edu/~galway/research-statement.pdf`.

[30] Z. Zhang, *A one-parameter quadratic-base version of the Baillie-PSW probable prime test*, Math. Comp. **71** (2002): 1699–1734.

ARTICLE

Satish Shirali
Extended real number system in measure theory
***Rejecta Mathematica*, Vol. 2, No. 1, pp. 44–48, June 2011**
2011 Rejecta Publications
math.rejecta.org

An open letter concerning

Extended real number system in measure theory

Satish Shirali

This article was originally submitted under the title "Extended real number system". Among the reasons given for its rejection was that there was undue focus on methods involved, such as consideration of separate cases, and that too little had been said about the relationship between what the review called "the proposed system" and other number systems that include infinity.

However, *no new system is being proposed* in the article and the very first sentence of the second paragraph includes the phrase "is usually defined as" in order to prevent any misinterpretation in this regard. But evidently to no avail.

Referring to the rule that infinity times zero should be zero as the rule "you propose," the review agreed with my observation that this does not work out well in many situations. However, the rule in question is quite standard in the extended real number system used for measure and integration, and nowhere does the article suggest that all conceivable systems are being studied under one roof.

The focus on multiplicity of cases has been argued for in the body of the article: Since multiplication in extended reals is defined by separating positive reals, negative reals, zero, positive infinity and negative infinity, verifying associativity alone requires an enormous number of dissimilar cases to be considered. The author feels that a construction procedure for the extended reals, involving only a manageable number of dissimilar cases—and this is what the article is mainly about, though not exclusively—is worth having on record.

Unfortunately, this is the only issue that is emphasized in the abstract. Within the article however, it has been pointed out that there is legitimate cause to question the consistency of a system having the usual properties which are assumed to hold for extended reals, and furthermore, that doubts arising on this score have been laid at rest in the rest of the discussion. The author feels that this is another feature that makes the exposition worth placing on record.

Besides a change in the title so as to include the phrase "in Measure Theory," there is an amendment in the abstract that reads "For the extended reals as used in measure theory (product $0 \cdot \infty$ is 0)" in place of "As an alternative." Also, the word "dissimilar" has been inserted.

Extended real number system in measure theory

Satish Shirali*

Abstract

The extended real number system is usually defined by appending two new elements and stating rules of addition and multiplication for them. The associative and distributive laws are then supposed to be verified case by case; however, the number of cases to be verified is well over sixty. For the extended reals as used in measure theory (product $0 \cdot \infty$ is 0), we offer a construction through equivalence classes, in which the number of dissimilar cases does not exceed five at any stage.

In any proof that requires consideration of separate cases, usually the number of cases is small and it is quite clear that all of them have been taken into account. However, when the number of cases is large, it may not be so clear that none have been left out. For example, verifying the associative law for a binary operation described in tabular form on a set of n elements is impractical when n exceeds 4. Another such instance is that of the associative and distributive laws in the extended real number system.

The extended real number system $\tilde{\mathbb{R}}$ is usually defined as the union of $\mathbb{R}$ with two elements, written ∞ and $-\infty$, and endowed with the structure described in (a)—(h) below, in addition to that already available on its subset $\mathbb{R}$:

(a) $-\infty < x < \infty$ for every $x \in \mathbb{R}$;
(b) $x + \infty = \infty + x = \infty$ and $x + (-\infty) = (-\infty) + x = -\infty$ for every $x \in \mathbb{R}$;
(c) $\infty + \infty = \infty$ and $(-\infty) + (-\infty) = -\infty$;
(d) $\infty \cdot \infty = (-\infty) \cdot (-\infty) = \infty$ and $(-\infty) \cdot \infty = \infty \cdot (-\infty) = -\infty$;
(e) $-(-\infty) = \infty$ and $-(\infty) = -\infty$;
(f) $x \cdot \infty = \infty \cdot x = \infty$ and $x \cdot (-\infty) = (-\infty) \cdot x = -\infty$
for every positive $x \in \mathbb{R}$;
(g) $x \cdot (\infty) = (\infty) \cdot x = -\infty$ and $x \cdot (-\infty) = (-\infty) \cdot x = \infty$
for every negative $x \in \mathbb{R}$;
(h) $(\pm\infty) \cdot 0 = 0 \cdot (\pm\infty) = 0$.

In contexts other than measure and integration, one may wish to omit (h) and take the products occurring in it as "undefined". For instance, it is omitted in [1, p.12] with the consequence on p.314 that the Lebesgue integral of the identically zero function on $\mathbb{R}$ is left undefined by (53), considering that (49) requires the function to be written as zero times the characteristic function of $\mathbb{R}$. However, the same author includes (h) in [2, p.19].

Without (e), there would seem to be no basis for the common practice of regarding $x - (-\infty)$ as meaning $x + \infty$ and $x - (\infty)$ as meaning $x + (-\infty)$. We have therefore chosen to state it explicitly although most authors prefer not to.

It is immediate from the properties (a)—(h) that addition and multiplication in $\tilde{\mathbb{R}}$ are commutative. However, the associative law of addition and the distributive law, which continue to be valid under the restriction that ∞ and $-\infty$ do not both appear in any of the sums involved, are supposed to be verified case by case. The associativity of multiplication can be verified, again case by case, to be valid without restriction.

*Email: satishshirali@usa.net, Address: House No.899, Sector 21, Panchkula, Haryana 134116, India

The multiplication described by (a)—(h) is, in effect, a binary operation on a set of five elements, namely, positive real, negative real, 0, ∞ and $-\infty$. In checking associativity therefore, one would have to consider 125 triplets (x, y, z); however, 27 of these involve only real numbers and need not be checked. The remaining 98 can be reduced to 68 by taking advantage of the obvious commutativity, but this is still an uncomfortably large number of cases to handle.

Consequently, in any effort at a case-by-case verification of the associative and distributive laws in the extended real number system, it would be a legitimate concern whether all cases have actually been taken into account or not.

We have avoided adding the requirement that $\frac{x}{\infty} = 0$ for $x \in \mathbb{R}$ so as to keep clear of the consequence that

$$0 = 0{\cdot}\infty = \tfrac{1}{\infty} \cdot \infty \text{ but } \tfrac{1\cdot\infty}{\infty} \text{ is undefined.}$$

However, this observation raises a second concern, namely, whether (a)—(h) already contain a contradiction, even without this requirement.

With a view to addressing both concerns, we outline a method of "constructing" $\tilde{\mathbb{R}}$ from $\mathbb{R}$, in which we describe ∞ and $-\infty$ set theoretically in terms of $\mathbb{R}$ rather than pull them out of the sky, and moreover, the associative and distributive laws become transparent with just two cases each. The definitions of addition and multiplication in $\tilde{\mathbb{R}}$ undoubtedly call for separate cases to be considered, but it is transparent that none are left out.

We begin by describing what the objects ∞ and $-\infty$ are.

Let ∞ denote the class of real sequences "diverging to ∞" in the usual sense (no circularity involved in this) and $-\infty$ denote the obvious analogous class. Furthermore, for each $x \in \mathbb{R}$, let $[\![x]\!]$ denote the class consisting of a single sequence, namely, the constant sequence with each term equal to x. Set $[\![\mathbb{R}]\!] = \{[\![x]\!] : x \in \mathbb{R}\}$ and $\tilde{\mathbb{R}} = [\![\mathbb{R}]\!] \cup \{-\infty,\infty\}$. Then each element of $\tilde{\mathbb{R}}$ is a class of sequences and the classes are disjoint.

For $\alpha,\beta \in \tilde{\mathbb{R}}$, define $\alpha < \beta$ to mean: for any sequences $\{a_n\} \in \alpha$ and $\{b_n\} \in \beta$, the inequality $a_n < b_n$ holds for all sufficiently large n. Then it is easy verify that

$$[\![x]\!] < [\![y]\!] \Leftrightarrow x < y \ \text{ if } x, y \in \mathbb{R} \tag{1}$$

Also, $-\infty < \alpha < \infty$ for all $\alpha \in [\![\mathbb{R}]\!]$, and $-\infty < \infty$. Thus (a) holds with x replaced by $[\![x]\!]$.

Suppose $\alpha,\beta \in \tilde{\mathbb{R}}$, and $\alpha \neq -\infty \neq \beta$. It is easy to see that the following four cases are exhaustive:

$$\text{(i)}\alpha, \beta \in [\![\mathbb{R}]\!] \quad \text{(ii)}\alpha \in [\![\mathbb{R}]\!] \text{ and } \beta = \infty \quad \text{(iii)}\alpha = \infty \text{ and } \beta \in [\![\mathbb{R}]\!] \quad \text{(iv)}\alpha, \beta = \infty.$$

It is equally straightforward to see in each of the four cases that, for any sequences $\{a_n\} \in \alpha$ and $\{b_n\} \in \beta$, the related sequence $\{a_n + b_n\}$ belongs to a unique $\gamma \in \tilde{\mathbb{R}}$. Therefore we may define $\alpha + \beta$ to be this unique $\gamma \in \tilde{\mathbb{R}}$.

In the course of arguing for the unique γ, it is also seen that

$$[\![x]\!] + \infty = \infty + [\![x]\!] = \infty \ \text{ if } \ x \in \mathbb{R},$$

$$\infty + \infty = \infty$$

and

$$[\![x]\!] + [\![y]\!] = [\![x + y]\!] \ \text{ if } x, y \in \mathbb{R} \tag{2}$$

One can proceed analogously when $\alpha,\beta \in \tilde{\mathbb{R}}$, and $\alpha \neq \infty \neq \beta$. All this establishes that $\alpha + \beta$ is uniquely defined except when one of them is ∞ and the other is $-\infty$ and that addition satisfies (c) as well as the properties claimed for it in (b), but with x replaced by $[\![x]\!]$.

Now consider $\alpha,\beta \in \tilde{\mathbb{R}}$. When $\alpha,\beta \in [\![\mathbb{R}]\!]$, any sequences $\{a_n\} \in \alpha$ and $\{b_n\} \in \beta$ must be constant sequences and it is immediate that the related (constant) sequence $\{a_nb_n\}$ belongs to a unique $\delta \in [\![\mathbb{R}]\!] \subseteq \tilde{\mathbb{R}}$. When $\alpha = \infty$, the following five cases for β are exhaustive:

$$[\![0]\!] < \beta \in [\![\mathbb{R}]\!], \quad [\![0]\!] > \beta \in [\![\mathbb{R}]\!], \quad [\![0]\!] = \beta, \quad \beta = \infty, \quad \beta = -\infty.$$

In each of the five cases, it is easy to arrive at the conclusion that: for any sequences $\{a_n\} \in \alpha$ and $\{b_n\} \in \beta$, the related sequence $\{a_n b_n\}$ belongs to a unique $\delta \in \tilde{\mathbb{R}}$. We note that it is essential here for $\beta = [\![0]\!]$ to consist of only the constant sequence $\{0,0,\dots\}$. Similarly when $\alpha = -\infty$. In view of the commutativity of multiplication in $\mathbb{R}$, the same conclusion can be drawn for the ten cases when $\beta = \pm\infty$. (It is inessential to the argument that the actual number of distinct cases is not 20 but 16, because the four cases when $\alpha = \pm\infty$ and $\beta = \pm\infty$ will occur twice among the 20.) Thus all cases have been covered and the aforementioned conclusion holds for all $\alpha,\beta \in \tilde{\mathbb{R}}$. Therefore we may define $\alpha\beta$ to be this unique $\delta \in \tilde{\mathbb{R}}$.

In the course of arguing for the unique δ, it is also seen that multiplication satisfies

$$[\![x]\!]\,[\![y]\!] = [\![xy]\!] \text{ if } x, y \in \mathbb{R} \tag{3}$$

as well as (d),(h), and that it further satisfies (f),(g), with x replaced by $[\![x]\!]$.

For any $\alpha \in \tilde{\mathbb{R}}$, there is a unique element $-\alpha \in \tilde{\mathbb{R}}$ such that

$$-[\![\alpha]\!] = [\![-\alpha]\!] \text{ if } \alpha \in \mathbb{R}$$

and $-(-\infty) = \infty$, $-(\infty) = -\infty$. Indeed, $-\alpha$ is the unique element of $\tilde{\mathbb{R}}$ such that whenever $\{a_n\} \in \alpha$, the sequence $\{-a_n\}$ belongs to the class $-\alpha$. This proves (e).

In view of (1), (2) and (3), the bijection $x \to [\![x]\!]$ is an isomorphism of ordered fields. Thus the subset $[\![\mathbb{R}]\!]$ of $\tilde{\mathbb{R}}$ is an isomorphic image of $\mathbb{R}$.

Having completed the construction of a system satisfying (a)—(h) and containing an isomorphic image of $\mathbb{R}$, we now turn our attention to the associative and distributive laws.

Suppose that either none among α,β,γ is ∞ or that none is $-\infty$. Let $\{a_n\} \in \alpha$, $\{b_n\} \in \beta$ and $\{c_n\} \in \gamma$. Then $(\alpha + \beta) + \gamma$ is the unique class in $\tilde{\mathbb{R}}$ containing the sequence $\{(a_n + b_n) + c_n\}$, while $\alpha + (\beta + \gamma)$ is the unique class in $\tilde{\mathbb{R}}$ containing the sequence $\{a_n + (b_n + c_n)\}$. By the associativity of addition in $\mathbb{R}$, it follows that the classes are the same. Thus the equality

$$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$$

holds provided that

either none among α, β, γ is ∞

or none among α, β, γ is $-\infty$.

Similarly, the equality

$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$$

holds provided that

either none among $\beta, \gamma, \alpha\beta, \alpha\gamma$ is ∞

or none among $\beta, \gamma, \alpha\beta, \alpha\gamma$ is $-\infty$.

In fact, both sides of the equality are the unique class containing the sequence $\{a_n(b_n + c_n)\}$, where $\{a_n\} \in \alpha$, $\{b_n\} \in \beta$ and $\{c_n\} \in \gamma$. It is left to the reader to formulate the corresponding statement regarding $(\alpha\beta)\gamma = \alpha(\beta\gamma)$, which is valid without any restrictions on α,β,γ.

We conclude with two remarks.

1. If $[\![x]\!]$ were enlarged to include all real sequences converging to x, then the properties (a)—(g) would follow in the same manner as above, but $\pm\infty \cdot 0$ and $0 \cdot (\pm\infty)$ would remain undefined.

2. If in the construction of $\mathbb{R}$ by Dedekind cuts as in [1, pp.17-21] or [3, pp.47-52], one includes the empty set and $\mathbb{Q}$ as cuts, then one gets $\tilde{\mathbb{R}}$, with the empty set serving as $-\infty$ and $\mathbb{Q}$ as ∞. The additional effort involved in checking this is minimal. However, the sum $-\infty + \infty$ needs to be specifically excluded in the general definition of sum, because otherwise it works out to be $-\infty$.

References

[1] Rudin, W., Principles of Mathematical Analysis, 3rd ed., McGraw-Hill, New York, 1976

[2] Rudin, W., Real and Complex Analysis, McGraw-Hill, New York, 1966

[3] Shirali, S. and Vasudeva, H.L., Mathematical Analysis, Alpha Science Publishers, Oxford, 2006.

ARTICLE

Hermann Bauer and Christoph Bauer
The problematic nature of Gödel's theorem
Rejecta Mathematica, Vol. 2, No. 1, pp. 49–58, June 2011
2011 Rejecta Publications
math.rejecta.org

An open letter concerning

The problematic nature of Gödel's theorem

Hermann Bauer and Christoph Bauer

Our paper "The problematic nature of Gödel's theorem" was rejected by MLQ (Mathematical Logic Quarterly). The managing editor (admittedly) has not read it, and no reviews were provided. His argument in favor of rejection was essentially that (a) "Gödel's results and techniques of proof are well-acknowledged by the scientific community ... since a lot of versed logicians have given detailed and well-understandable presentations and modifications of these results and proofs in a lot of frequently read textbooks." Additionally (b) the editor claims that in general texts like ours, which try to disprove well-accepted results, "... finally demonstrate only a lack of understanding by the authors. It cannot be the task of editors and referees to disprove all these 'disprovers'. In some cases, their errors are obvious, sometimes it takes a considerable amount of time to point out them. Who would spend this?"

In answer to (a) we would like to note that to our knowledge all secondary authors have directly assumed and sometimes even increased the problems of Gödel's original work we raise in our paper. As to (b) we would like to state that if indeed all or most such critics hitherto demonstrate only a lack of understanding by the authors, it may be improbable, but not impossible, that our criticism is nonetheless valid. As we are convinced that it actually is we would further like to point out that *mathematical truth is not a question of probability.* We recognize that many editors and referees, already over-burdened, must necessarily perform a certain degree of "triage" as papers are submitted—but does this not constitute a significant hole in the peer-review system? Is it possible to publish a mathematical paper which challenges the accepted orthodoxy?

We think that our paper could and should create a useful and necessary discussion about Gödel's theorem although, and in fact precisely because, it is a well established and unquestioned part in mathematical literature.

The problematic nature of Gödel's theorem

Hermann Bauer[1] and Christoph Bauer[2]

Abstract

In this paper we will show that some fundamental deductions Gödel's famous theorem is based upon, are highly problematic: Gödel begins the main part of his theorem by correlating to each formal expression (e.g. the variables) of a formal system a natural number. He calls the totality of these numbers x and he uses this x as a variable in recursive formulas. Then he correlates to each number x the formalized numeral $\mathbf{x}$. We will show that $\mathbf{x}$ can be nothing else than a variable in the formal system. In the well-known mapping ("Gödelization") he considers $\mathbf{x}$ as a predicate and correlates to it an expression, that is variable. This is contrary to the correlation of a fixed number to each variable, as the mapping is stated to be biunique. Gödel wanted the mapping itself not to be directly included in the proof, but in fact it is included and this is veiled by an incorrect predicate. In a further step (Corollary V) he effectively replaces $\mathbf{x}$ (that has a variable image) by the variable a (that has the image *17*). This problem shakes the foundation of Gödel's theorem and therefore its validity should be discussed again without any dogmatism.

1 Introduction

Gödel's theorem of incompleteness [3] is regarded as a fundamental part of metamathematics, and is not taken into question any more. One reason for this might be its age: One might believe that any mistake should have been discovered during more than seventy years. But one must consider the following: Gödel developed especially for his theorem complete new methods of proof, which could not easily be validated at the time. His sensational results themselves could be understood easily, the proof however was difficult to follow. Therefore many mathematicians accepted the proof without thorough examination. Though from time to time the theorem was criticized in manuscripts submitted to mathematical journals, but apparently none of these has been published.

Kleene in [4] performed a meticulous reworking of the proof. His intention was obviously to explain Gödel's theorem perfectly and to support its understanding, without however questioning the correctness of the proof. This might explain why such an outstanding mathematician took over uncritically a highly problematic *predicate* (see **p1** and **p2** in this paper) and made an incorrect deduction of it. Most later authors based their arguments on Kleene and did not correct this mistake either, using Gödel's methods without thorough examination to deduct other mathematical sentences. This of course does not demonstrate the correctness of the theorem, because one can deduct true as well as wrong sentences from a wrong sentence.

But the whole time it was problematic to understand, that the theorem can be correct, because the well known abridged version "a formula, that means its own improvability" is a circular statement. Such statements are highly problematic, for a statement on "something" must have a larger content than the mere naming of this "something", but both are identical here. But how can a statement be larger than itself? This becomes possible in Gödel's theorem, as we will show, because in the Gödelization he uses two essentially different mappings of the variables, one of them having a larger content than the other.

[1] Corresponding author. Vinkelgasse 21, D 53332 Bornheim, hebbauer@freenet.de
[2] University of Geneva, 30. Q. E. Ansermet, 1205 Geneva, Christoph.Bauer@unige.ch

On the other hand abridged versions were for many mathematicians a reason to accept the theorem, as the contents of the proof seems to be quite plausible there. In the first part of our paper (Sections **2.1** and **2.2**) we will show the reader that this is problematic, as the abridged versions are by nature not usable to take a final decision upon the truth of a theorem (see also discussion in [1]).

In Section **3** we will support our arguments by going back to Gödel's original paper. The crucial problematic of the theorem will be raised in Section **4**: In the proof the term **x**, that represents the "numerals" (formalized numbers), is a variable, the Gödelizing of which is highly problematic.

In the last section we will finally extend these realizations to a more general level and invite the readers to discuss its implications.

2 The problematic nature

2.1 The problematic nature of the first abridged version of Gödel's theorem

As is generally known the goal of Gödel's theorem is to produce an explicit number-theoretic formula, which is undecidable i.e. neither provable nor refutable. The existence of such a formula had not yet been proven and according to Hilbert's program was not possible at all.

The best known abridged version was formulated by Gödel himself: *"Wir haben also einen Satz vor uns, der seine eigene Unbeweisbarkeit behauptet."* ([3], p.175.) *Kleene* translates/formulates it as follows ([4], p. 205): "A means that A is unprovable." The word "mean" implies "is equivalent" i.e. the formula is then and only then unprovable, if it is true. If "true" would be equivalent to "provable" then the formula would be provable then an only then, if it is not provable. That would be a contradiction against formal logic. To avoid this contradiction one must assume that there exists a true and unprovable formula. This anticipates the result of the proof (because a true formula cannot be refused, if arithmetic is consistent).

2.2 The problematic nature of the second abridged version of the theorem

A second version that has been published in the journal *"Spektrum der Wissenschaft"* ([4], p.51) seems at first sight again plausible, but as we will show is wrong: The set of all proofs is enumerable and also the set of all formulas. It is possible to find a formula $Dem(a;b)$ that is provable then and only then, if b is the number of a proof and a is the number of the formula proved by this proof. Now one can build the formula:

$$f(a) \equiv \forall b\, (\neg\, Dem(a;b)), \tag{1a}$$

which means, that the formula with number a is not provable. The number of formula (1a) now is named g, and then it is stated, that the formula

$$f(g) \equiv \forall b\, (\neg\, Dem(g;b)) \tag{1b}$$

built by substituting g for a in formula (1a) means her own improvability. That is incorrect. In fact follows to the provability of formula (1b) that formula (1a) and not formula (1b) itself is unprovable.[3]

3 Main features of Gödel's theorem

Gödel begins his theorem with a formal mathematical system (NL) that comprises formal number theory and logic. (We call it upper level). The object language of this system has for its base seven[4] formal sym-

[3] (1a) is obviously refutable, because it contains a free variable and is refuted by any proof. Therefore (1b) is provable.

[4] It is useful (we think necessary) to add an eighth zero entity for "equal to" as Kleene did (We choose *15*). He also added zeroes for "plus" and "times" and "successor variable" and uses another sort of Gödelization.

bols and moreover an infinite number of symbols for variables, as presents the following table (Gödel calls them *"Grundzeichen"*, Kleene calls them "zero entities" or shortly "zeroes").

Zero entities in upper level NL	Name	Correlated numbers in lower level cn
0 (in bold)	zero	*1*
′	successor	*3*
¬	not	*5*
∨	or	*7*
∀	for all	*9*
(	opening bracket	*11*
)	closing bracket	*13*
=	equal to	*15*
a	first variable	*17*
b	second variable	*19*
x	third variable	*23*
y	fourth variable	*29*
etc.	etc.	etc.

All formal objects of the System NL are finite sequences of zero entities (*"Grundzeichenreihen"* by Gödel). E.g. the natural numbers greater than ***0*** are represented by *´**0***, *´´**0***, *´´´**0*** … .[5] i.e. by two zero entities. We write *N1*, *N2*, *N3* … *N*(*n*) for that. They are called "numerals" (*"Zahlzeichen"* by Gödel).

Formal expressions that are identical are connected by "≡". At cn (see below) this is correlated to "=".

Now distinct odd numbers are correlated to the zero entities (*"Zahlengrundzeichen"* by Gödel—see the third column of the table). To each formal object is in this way correlated a sequence of odd numbers[6]:

$$X \equiv n_1\, n_2\, n_3 \,....\, n_k$$

By the well-known Gödelization now to each formal object is correlated by a bi-unique mapping a natural number $\tilde{x}$ (Gödel number)[7]:

$$X \to \tilde{x}\text{ , where}^{8} \tag{2a}$$

$$\tilde{x} = G[X] = G[n_1\, n_2\, n_3 \,...\, n_k] = 2^{n_1} \bullet 3^{n_2} \bullet 5^{n_3} \bullet ... \bullet p_k^{n_k}\, , \tag{2b}$$

where p_k is the *k*-th of the prime numbers in order of magnitude. For example is:

$$G[N2] = G[\acute{}\acute{}\mathbf{0}\,] = 2^3 \cdot 3^3 \cdot 5^1 = 1080$$

By this mapping and the theory of primitive recursive functions it becomes possible to develop for metamathematical predicates PR(*X*) in NL equivalent formulas pr($\tilde{x}$) in classical number-theory cn (we

[5] Gödel places the successor symbols before the zero symbol.
[6] *X is a variable Expression* of course.
[7] The tilde above *x* is placed to distinguish it from the variable *x* in NL. Gödel wrote *x* only. We later on will realize that $\tilde{x}$ is in fact a variable in cn. The exact formulation for "the number $\tilde{x}$ " therefore is : "a value of the variable $\tilde{x}$."
[8] *G*[*X*] is only a short expression für the Gödel number of *X*.

call it "lower level"). For example "X is a *numeral*", "X is a *variable*", "X is a *formula*", "X is an *axiom*" etc. and finally (for two formal expressions X and Y) "Y is a proof for X" ([4], p. 253 f). For example the equivalent for the predicate $V(X) \equiv$ "X is a variable" is[9]:

$$v(\tilde{x}): \exists z\,(\tilde{x} = 2^z \wedge prim\ z \wedge z > 13). \tag{3}$$

It means, that the number correlations z to the variables are the primes greater 13 and their Gödel numbers are 2^z.

The decisive step of deduction in Gödel's proof (that was ignored in the abridged version **2.2)** is, that in each formula $X(a)$ with the single free variable a the "Gödel numeral" of $X(a)$ is substituted.

The Gödel numeral is defined as the formalized Gödel number i.e. the zero ($\boldsymbol{0}$) with leading successor signs (´) their number being $\tilde{x}$.

We write according to Kleene **x** for the Gödel numeral[10] of the Gödel number $\tilde{x}$ (and **n** for the numeral of the number n).

The Gödel number of **x** is:

$$G[\mathbf{x}] = 2^3 \cdot 3^3 \cdot 5^3 \cdot \ldots \mathrm{p}^3_{\tilde{x}} \cdot \mathrm{p}^1_{\tilde{x}+1}. \tag{4}$$

The result of the substitution one can write $R \equiv X_{\mathbf{x}}(\mathbf{x})$. We call it the "Richard formula" of the formula $X(a)$.

Let be e.g. $X(a) \equiv a = a$. Then is $G[X(a)] = \tilde{x} = 2^{17} \cdot 3^{15} \cdot 5^{17}$, and $\mathbf{x} = N\,(2^{17} \cdot 3^{15} \cdot 5^{17})$ (for the meaning of "N" see above) and the Richard formula of $X(a)$ is:

$$\mathbf{x} = \mathbf{x}, \text{ i.e.,}$$
$$N\,(2^{17} \cdot 3^{15} \cdot 5^{17}) = N\,(2^{17} \cdot 3^{15} \cdot 5^{17}).$$

We can shorten the substitution- predicate as follows (cf. ([4], p. 253 Dn5):

$$R \equiv X_{\mathbf{x}}(\mathbf{x}) \equiv SB(X,a,\mathbf{x}); \tag{5a}$$

Its equivalent in cn is[11]:

$$\tilde{r} = sb\,(\tilde{x}, 17, \mathrm{G}\,[\mathbf{x}]), \tag{5b}$$

that means in essence, to get $\tilde{r}$ one must replace on the right side of (2b) the potencies with exponent 17 by G [**x**] of (4) and then increase[12] the primes of G [**x**] and the following primes in this way that all primes are positioned in order of magnitude again (see [3], p. 184).

In order to carry out the substitution (5a) Gödel establishes a predicate that relates $\tilde{x}$ to **x,** the problematic connected with that will be shown in section **4**.

Ahead of this we will present the remaining part of the theorem. Gödel now creates in NL[13] the predicate "Y is a proof for $X_{\mathbf{x}}(\mathbf{x})$"[14]. The equivalent of this in cn can by recursions be formulated as

$$dem'(\tilde{x}\,;\tilde{y}), \tag{6a}$$

where $\tilde{x}$ is the Gödel number of $X(a)$ and $\tilde{y}$ is the Gödel number of Y. Formula (6a) now is "formally expressed" i.e. an equivalent formula is created in NL[15] :

[9] according to [3], p. 182, no. 11 for $n = 1$.

[10] I will show later on **x** to be a variable in NL. Today $\ulcorner x \urcorner$ is written for **x**.

[11] See in [3], p. 188 formula (8.1) the expression on the right side in square brackets, where stands y instead of x (and 19 instead of 17) . Gödel writes $Z(x)$ instead of $G[\mathbf{x}]$ and sets the number 17 above instead of ahead of it. Let it be noted that $X_{\mathbf{x}}(\mathbf{x})$ is not a formal Expression in NL, but its equivalent in cn is a term there.

[12] Only if 17 is the first exponent $G[\mathbf{x}]$ remains unchanged.

[13] For *Gödel´s* mode of formulation see section **4.1**.

[14] The exact predicate is "X is a proof of Y_{y}" – X and Y are exchanged.

$$DEM'(\mathbf{x}; \mathbf{y}), \tag{6b}$$

where **x** it the numeral of $\tilde{x}$ and **y** that of $\tilde{y}$.

This is combined with the second problematic step, which effectively comes to replace **x** by the variable *a* and **y** by the variable *b*. The result of this step is (cf. [4], p. 206, LEMMA 21):

$$DEM'(a; b). \tag{6c}$$

Formula (6c) is (if we accept the deduction) provable then and only then, if there is substituted for *a* the Gödel numeral **x** of a formula $X(a)$ and for *b* is substituted the Gödel numeral **y** of a proof of the Richard formula $X_{\mathbf{x}}(\mathbf{x})$ of $X(a)$.

The next formula created is:

$$f(a) \equiv \forall b\,(\neg\, DEM'(a;b)). \tag{1a'}$$

In this formula finally its Gödel numeral $\boldsymbol{g}$ is substituted for the variable *a*:

$$f(\boldsymbol{g}) \equiv \forall b\,(\neg\, DEM'(\mathbf{g};b)). \tag{1b'}$$

(1b') means, that the Richard formula of the formula with the Gödel numeral **g** is unprovable. The formula with the Gödel numeral **g** is formula (1a'); its Richard formula is formula (1b') itself. It means, that it is unprovable. Now it is easy to demonstrate, that formula (1b') is formally undecidable in a consistent (ω-consistent respectively) arithmetic system.

4 The fundamental problematic nature of the theorem

The key problematic of the theorem has to do with the meaning of **x**. The question is, whether in the proof **x** can be anything else than a variable in NL. We deny this and will justify our conclusion soon. At first we would like to explain the consequences for Gödel's proof: The Gödelization correlates to **x** the variable expression $G[\mathbf{x}]$ of formula (4). If **x** is a variable it must be correlated a definite natural number to it. There results the contradiction $G[\mathbf{x}]$ = const (see **4.3.3**).

4.1 The problematic relation and its correction

We will now show how Gödel establishes the predicate concerning the relation between $\tilde{x}$ and **x,** earlier referred to be problematic. Since Gödel's mode of formulation is very unusual, we will equally refer to Kleene´s reworking afterwards.

As is generally known, statements on itself can produce antinomies by their *"Zirkelhaftigkeit"* (basing on a circular argument). The abridged version above (Section **2.1**) is such a statement. Gödel comments on this problem with the words:

> "Ein solcher Satz hat entgegen dem Anschein nichts Zirkelhaftes, denn er behauptet zunächst die Unbeweisbarkeit einer ganz bestimmten Formel und erst nachträglich (gewissermaßen zufällig) stellt sich heraus, daß diese Formel gerade die ist, in der er selbst ausgedrückt wurde." ([3], p. 175, footnote 15).
>
> Such a theorem against appearance has nothing to do with a circular argument, because it first of all states the improvability of a definite formula and later only (so to speak coincidentally) it emerges, that this formula is just the one, that expresses the theorem itself.

In order to avoid antinomy on any rate, Gödel intends, to carry out all calculations on the lower level cn, but to name there the terms and formulas after their meaning in the upper level NL *written in italics* ([3], p. 179, line 20f).

[15] This step is left out in the referee of *Nagel* and *Newman* [6], p. 82 – 89, therefore it is not correct.

For examples: The sentence in the metalanguage of NL "*a* is a variable" Gödel formulates[16]: *"17* is a *variable* (italic)". The sentence "*´0* is a numeral" Gödel would formulate: "$2^3 \cdot 3^1$ is a *numeral* (italic)" that means: "$2^3 \cdot 3^1$ is correlated (by Gödelization) to a numeral."

If that procedure is carried out consequently, any antinomy can be avoided in fact. But Gödel is inconsequent, for he defines the following predicate ([3] p.183, nr. 16 and 17): "G[**n**] is the *numeral* (italics) for the natural number (not italics!) *n*". One can write for this:

$$nu(\mathrm{G}[\mathbf{n}];\, n). \qquad (7a)$$

The inconsequence is, that "natural number" is not written in italics. It is an expression of the lower level cn and is named at this level. Gödel here contradicts his declared intention, to name only after the meaning in the upper level NL. The consequent formulation of the meaning of (7a) becomes obvious, if one pursue, how Gödel uses it further: He replaces *n* by $\tilde{x}$ and G[**n**] by G[**x**] (i.e. *Z*(*n*) by *Z*(*x*) in his formulation) without mentioning that particularly[17]. We formulate the result analogous to the predicate and (7a):

"G[**x**] is the *numeral* for the natural number $\tilde{x}$", and **(p1)**

$$nu(\mathrm{G}[\mathbf{x}];\, \tilde{x}). \qquad (7b)$$

(7b) is a relation at the lower level and corresponds with formula (4) in this paper. A consequent formulation of its meaning (according to Gödel´s declared intention) one finds by considering, that $\tilde{x}$ is the Gödel number, correlated to the formal expressions *X*(*a*) of NL[18]. The correct formulation in Gödel´s diction therefore is: "G[**x**] is the *numeral* (italics) correlated to the *formal Expression* (italics!) *X*(*a*)." It means at the upper level "**x** is the numeral (not italics) correlated to the formal expression (not italics) *X*(*a*)." The consequences of this we will clarify after having referred its reworking by Kleene.

4.2 Reworking by Kleene

Gödel simultaneously works in both levels, as we saw. Kleene does not adopt this. He formulates the metatheory at the upper level NL and wants to separate strictly the number-theoretic equivalents at the lower level cn. But he is equally inconsequent in connection with the problematic predicate (**p1**). He formulates at the upper level the predicate:

"**x** is the numeral for the natural number $\tilde{x}$"[19] (abbreviation *Nu*(**x**; $\tilde{x}$). **(p2)**

That is inadmissible, for $\tilde{x}$ does not exist at this level. (**p2**) is a **mixed predicate.** The equivalent in cn Kleene formulates effectively[20]: *nu*(G[**x**]; $\tilde{x}$), i.e., he correlates $\tilde{x}$ to $\tilde{x}$, which is wrong.

The correction is only possible at the upper level. As we already have seen, $\tilde{x}$ must be replaced there by the formal object *X*(*a*):

$$NU(\mathbf{x};\, X(a)). \qquad (8)$$

Therefore it must be possible to relate the formal objects directly at the upper level to their Gödel numeral **x** without the detour via $\tilde{x}$. We can write:

$$X(a) \rightarrow \mathbf{x}. \qquad (9)$$

16 He writes *17* instead of 2^{17}.

17 See [3], p. 188 formula (8.1). It is notable, that *Gödel* names *n* a natural number, where *x* is an argument of the recursive relation *Q* (*x*;*y*), i.e. it is a variable (see section **4. 3. 3**).

18 The restriction to *X*(*a*) instead of *X* is insignificant of course.

19 [4] p.254, DN11 and [7] p. 38 V_{11}

20 according to [4], p. 258, example 2 and explicit [7] p. 42, line 14

This way the mapping of Gödelization is directly included in the proof, although Gödel wanted to avoid this.

4.3 Consequence for the meaning of x

Predicate (8) requires **x** to be defined immediately in NL. Therefore we have to search the meaning of **x** in NL.

4.3.1 Metamathematical symbol

Can one interpret **x** as a metamathematical symbol for a numeral, which can be replaced by a numeral? (cf. [4], p. 82f.)

Such a metamathematical symbol however is not a formal expression in NL and therefore has not a number-theoretic image in cn, while **x** has the image G[**x**] there.

4.3.2 Infinite set

Another possible interpretation would be to interpret **x** (and **y** also) as a symbol for an infinite set of numerals in NL. According to that *DEM´* (**x**; **y**) would be an infinite set of formulas in NL. But Gödel´s proof requires drawing a conclusion from this infinite set to a formal expression and this is not expressible in a formal deduction that must be finite.

The problem will be clarified by giving the gist of the central corollary V at Gödel (([3] p.186) applied to the step of deduction from formula (6a) to formula (6b) in this paper respectively (6c) (we refer to it in a diction according to the one used in this paper hitherto):

For the recursive relation *dem´*($\tilde{x}$; $\tilde{y}$) there exists a relation (*"Relatiosszeichen"* by Gödel) *DEM´*(*a*; *b*) with the free variables *a* and *b* and for all pairs of numbers $\tilde{x}$; $\tilde{y}$ the following is valid: From *dem´*($\tilde{x}$; $\tilde{y}$) follows that *DEM´* (**x**; **y**) is provable.

The proof of this corollary (not given by Gödel, but by Kleene in ([4] p. 238 to p.245) represents an infinite set of proofs, each having the result "*DEM´* (**x**; **y**) is provable", where **x** and **y** are elements of infinite sets of numerals corresponding to the numbers $\tilde{x}$ and $\tilde{y}$. But these sets are not expressible in the formal system NL and therefore a conclusion to a relation *DEM´*(*a*; *b*) is not possible there.

This objection is also valid for the mathematical sketch of the proof, that Gödel sets ahead of it ([3], p. 174f, where the "class *K*" is an infinite set of natural numbers), but for which he however did not demand exactness.

4.3.3 Variable

In Gödel's proof $\tilde{x}$ and $\tilde{y}$ are variables, for he uses them as arguments of recursive relations basing on recursive functions, whose arguments are variables of course. Using (7a) in the formulation of (7b), he defines a recursive relation *Q*($\tilde{x}$; $\tilde{y}$), where $\tilde{x}$ and $\tilde{y}$ are variables.[21]

The correct interpretation of (2b) therefore is: By the Gödelization the variable $\tilde{x}$ is correlated to the variable formal expression *X* in this sense, that to each formal expression is correlated a value of the variable. The equation of this correlation can be expressed by (2b).

According to that **x** is correlated to *X* in NL (8 and 9) and therefore **x** can be nothing else than a variable in NL e.g. *x*.[22] This result moreover can be realized by the following deduction:

The mixed predicate *Nu*(**x**; $\tilde{x}$) (**p2**) according to Kleene ([4], p. 254) is defined as follows[23]:

$$Nu(\boldsymbol{0};\, 0) \wedge ((Nu(\mathbf{x};\tilde{x}) \Rightarrow Nu(\acute{}\mathbf{x};\tilde{x}+1)).$$

By induction proof results:

[21] cf. [4] p. 254 – 258
[22] We can write X_x, where *x* is a parameter.
[23] I use the formal symbols, though the predicates are mixed ones.

$$\forall n(\tilde{x} = n \Rightarrow \mathbf{x} = \mathbf{n}).$$

The correct interpretation of this predicate is: "**x** is a corresponding variable in NL to the variable $\tilde{x}$ in cn". Therefore **x** is identical with a variable in NL (e.g. $\mathbf{x} \equiv x$).

But now in cn is correlated to this the contradiction G[**x**] = const (e.g. G[**x**] = 2^{17}) for all **x**. Therefore the nature of **x** is highly problematic.

If we could accept **x** and **y** to be variables, the last statement of corollary V could be formulated very simply:

… for all pairs of number values of (the pair of variables) $\tilde{x}\,;\tilde{y}$ the following is valid: From *dem´*($\tilde{x}\,;\tilde{y}$) follows that *DEM´* (**x**; **y**) for the corresponding pair of numeral values (of the pair of variables **x**; **y**) is provable. The transition to *DEM´*(*a*; *b*) then only changes the names of variables.

In Gödel's theorem the problematic nature of **x** and **y** is transmitted to all other variables e.g. to *a* and *b*. Only if one ignores this, one can correlate to formula (1a') a fixed Gödel numeral **g**. Only then one can substitute this numeral in (1a') for the variable *a* and create the formula (1b'), that states something about itself.

5 The depth of the problem

The key problematic represented in this article concerns the problem of mapping the formal expressions of NL into the natural numbers of cn. For such a mapping the basic elements need to be independent of each other. This is not the case here. The signs correlated to "zero", "successor" and "variable" are not independent. We can define the recursive predicate "$S(x)$ is the x-th successor of $\boldsymbol{0}$" as follows. $S(\boldsymbol{0}) = \boldsymbol{0}$ and $S(´x) = ´S(x)$. Then is valid: $S(x) = x \Rightarrow$ $S(´x) = ´S(x) = ´x$. Therefore it results (by induction proof):

$$\forall x(S(x) = x), \text{ i.e. } S(x) \equiv x$$

Therefore two different images (G[**x**] and a constant) of a variable are possible, although the mapping is stated to be unique and therefore circular deductions and finally contradictions result.

The problem is connected to the notion of variables in general. Are variables entities for themselves or vacant places only for concrete numbers (respective numerals)? Both opinions are possible, but both are one-sided. Their combination only makes the notion of variables comprehensible. At the Gödelization however it is impossible to combine both opinions. An entity for itself should have a definite Gödel number, whilst to a vacant place the Gödelization should correlate a vacant place.[24] Therefore a unique mapping of the full notion of variables is impossible. A discussion of this problem is necessary and welcome.

The existence of undecidable formulas has tried to be proven with the aid of Turing machines. If such proofs use a predicate like (**p2**) they are equally problematic.[25]

However there are proofs of this subject, which are based on the theory of machine-numbers. These are not discussed here.[26]

6 Summary

As we have shown, Gödel's theorem has key problematic that asks for further discussions. As we pointed out these problems are not obvious, as the formulas of the lower level as well as the predicates of the higher level are by themselves correct. If Gödelization is accepted as a fact one can then indeed deduct

[24] The problem has relations even to the discussion about formalistic and realistic conception of ideas (cf. [1], p. 16).
[25] *Kleen* formulates such a proof in [4], p. 376-386.
[26] cf [2], Kap. 10.

without any further contradiction formulas, that state something about themselves, e.g. that a formula states itself to be provable, decidable, refutable, not existent"(!) etc. But the mapping itself and only this is contradictory: The expression **x** is a variable and must have a fixed natural number as an image and not the variable expression G[**x**]. As a consequence the deduction of Gödels theorem becomes impossible, as the variable a has then a variable image and the image of $X(a)$ for a fixed X is variable too. Only by using two essentially different images of the variables it becomes possible to make statements about themselves. The incorrect deductions (**p1** in this paper and corollary V in [3]) veil the notion of **x** as a variable.

References

[1] H. Bauer, Kritik des Gödel'schen Unvollständigkeitsbeweises, in Mathematisch-Physikalische Korrespondenz nr. 222, p. 13 –17 (Dornach, 2005).
[2] H.D. Ebbinghaus, et al., Einführung in die mathematische Logik (Darmstadt, 1978).
[3] K Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, in Monatshefte für Mathematik und Physik nr. 38, p. 173 – 198 (1931).
[4] S. C Kleene, Introduction to Metamathematics (Amsterdam, 1952).
[5] G. Guerrerio, Kurt Gödel, Biographie, in Spektrum der Wissenschaften. (1,2002)
[6] E. Nagel and J.R Newman, Der Gödelsche Beweis (Wien – München, 1964).
[7] W. Stegmüller, Unvollständigkeit und Unentscheidbarkeit (Wien – New York 1970).

ARTICLE

Jean-François Burnol
Scattering, determinants, hyperfunctions in relation to $\frac{\Gamma(1-s)}{\Gamma(s)}$
***Rejecta Mathematica*, Vol. 2, No. 1, pp. 59–118, June 2011**
2011 Rejecta Publications
math.rejecta.org

An open letter concerning

Scattering, determinants, hyperfunctions in relation to $\frac{\Gamma(1-s)}{\Gamma(s)}$

Jean-François Burnol

I wrote this paper in 2006, and submitted it to a journal specializing in integral equations and operator theory. After circa 14 months I received a report which I reproduce in full here (I allow myself to correct the spelling of a mathematician's name cited in the report):

"In spite of desperate efforts, the referee has failed to understand what the paper is about. Apparently it does not have a definite goal but consists of miscellaneous remarks to the papers by de Branges and Rovnyak. It is practically impossible to distinguish original results in this jumble. Actually, the text does not look as a mathematical article but rather as some notes for personal use. In the referee's opinion, the paper should be rewritten according to conventional rules and its volume should be divided by the factor 5-10. The author should try to formulate the results which he considers to be new."

Let me explain why I consider the publication of the paper important. First of all the referee's report only serves to demonstrate that the referee did not read the manuscript. I tried to point this out to the editor in chief, to no avail:

" *Dear Professor Burnol,*
I read all your letters to us. I am not changing my mind! Your paper is not accepted for publication. This decision is final and the discussions about this paper this time I consider finished. Sincerely, XXX "

I think this illustrates nicely how dysfunctional the peer-review process may be, at times. Regarding the paper itself, it is well structured, and its goal was to prove new mathematical theorems (!), a goal which was achieved (!). I corrected a typo in 2008 (there was a superfluous imaginary i in some equations, see the footnote on page 1), this is the only change to the 2006 version.

The referee asked me to divide the "volume" by between five and ten, a request which at that time particularly infuriated me. In fact, a more acceptable comment would have been to point out that the paper contained material for between 3 and 5 reasonably sized quasi-independent publications (of reasonable, but obviously not earth-shaking interest!), but I wanted to make a common exposition with in particular a common introduction. What would be the point of repeating 5 times the same introduction? An introduction is made necessary by the fact that my perspective is unique and links together a priori disjoint topics, the reader needs some help in entering this framework.

Another difficulty is that in 2008, during a stay at Institut des Hautes Études Scientifiques (IHES), I made very significant advances (establishing links with domains apparently completely unrelated, and which moreover have been of great interest for the last thirty years to large communities of researchers), on which I have had opportunities to give lectures at IHES, at the European Conference of Mathematics (ECM) at Amsterdam, and at a workshop at the Independent University of Moscow (Conference Zeta functions II). I have circulated a hand-written manuscript of about 80 pages, and prior to publishing this novel material in peer-reviewed journals, I need to make my earlier work available to the mathematical community.

I did sufficiently serious and dedicated work on this in 2006 resulting in a paper of about 65 pages. It would be all too easy, and far more beneficial to my career, to instead divide the paper into at least 3 publications, but I just don't see the point. If one is not sufficiently committed to mathematics to place great importance on the form one gives to one's own contributions, if one is ready to obey arbitrary diktats, if all that matters is adding lines of publications to a CV, then one practices a job and not a passion and one does not care about his/her legacy, one lives amidst superficial illusions and pleasures.

This paper will be necessary reading to get a full understanding of my earlier as well as of my future works.

Scattering, determinants, hyperfunctions in relation to $\frac{\Gamma(1-s)}{\Gamma(s)}$

Jean-François Burnol*

Abstract

The method of realizing certain self-reciprocal transforms as (absolute) scattering, previously presented in summarized form in the case of the Fourier cosine and sine transforms, is here applied to the self-reciprocal transform $f(y) \mapsto \mathcal{H}(f)(x) = \int_0^\infty J_0(2\sqrt{xy})f(y)\,dy$, which is isometrically equivalent to the Hankel transform of order zero and is related to the functional equations of the Dedekind zeta functions of imaginary quadratic fields. This also allows to re-prove and to extend theorems of de Branges and V. Rovnyak regarding square integrable functions which are self-or-skew reciprocal under the Hankel transform of order zero. Related integral formulae involving various Bessel functions are all established internally to the method. Fredholm determinants of the kernel $J_0(2\sqrt{xy})$ restricted to finite intervals $(0, a)$ give the coefficients of first and second order differential equations whose associated scattering is (isometrically) the self-reciprocal transform $\mathcal{H}$, closely related to the function $\frac{\Gamma(1-s)}{\Gamma(s)}$. Remarkable distributions involved in this analysis are seen to have most natural expressions as (difference of) boundary values (*i.e.* hyperfunctions.) The present work is completely independent from the previous study by the author on the same transform $\mathcal{H}$, which centered around the Klein-Gordon equation and relativistic causality. In an appendix, we make a simple-minded observation regarding the resolvent of the Dirichlet kernel as a Hilbert space reproducing kernel.

1 Introduction (The idea of co-Poisson)

We explain the underlying framework and the general contours of this work. Throughout the paper, we have tried to formulate the theorems in such a form that one can, for most of them, read their statements without having studied the preceeding material in its entirety, so a sufficiently clear idea of the results and methods is easily accessible. Setting up here all notations and necessary preliminaries for stating the results would have taken up too much space.

The Riemann zeta function $\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \dots$ is a meromorphic function in the entire complex plane with a simple pole at $s = 1$, residue 1. Its functional equation is usually written in one of the following two forms:

$$\pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})\zeta(s) = \pi^{-\frac{1-s}{2}}\Gamma(\frac{1-s}{2})\zeta(1-s) \tag{1a}$$

$$\zeta(s) = \chi_0(s)\zeta(1-s) \qquad \chi_0(s) = \pi^{s-\frac{1}{2}}\frac{\Gamma(\frac{1-s}{2})}{\Gamma(\frac{s}{2})} \tag{1b}$$

The former is related to the expression of $\pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})\zeta(s)$ as a left Mellin transform[1] and to the Jacobi

*E-mail: burnol@math.univ-lille1.fr, Address: Université Lille 1, UFR de Mathématiques, Cité scientifique M2, F-59655 Villeneuve d'Ascq, France

[1]in the left Mellin transform we use $s-1$, in the right Mellin transform we use $-s$.

identity:

$$\pi^{-\frac{s}{2}}\Gamma(\frac{s}{2})\zeta(s) = \frac{1}{2}\int_0^\infty (\theta(t)-1)t^{\frac{s}{2}-1}\,dt \qquad (\Re(s)>1) \tag{2a}$$

$$= \frac{1}{2}\int_0^\infty (\theta(t)-1-\frac{1}{\sqrt{t}})t^{\frac{s}{2}-1}\,dt \qquad (0<\Re(s)<1) \tag{2b}$$

$$\theta(t) = 1 + 2\sum_{n\geq 1} q^{n^2} \qquad q = e^{-\pi t} \qquad \theta(t) = \frac{1}{\sqrt{t}}\theta(\frac{1}{t}) \tag{2c}$$

The latter form of the functional equation is related to the expression of $\zeta(s)$ as the right Mellin transform of a tempered distribution with support in $[0,+\infty)$, which is self-reciprocal under the Fourier cosine transform:[2]

$$\zeta(s) = \int_0^\infty (\sum_{m\geq 1}\delta_m(x) - 1)x^{-s}dx \tag{3a}$$

$$\int_0^\infty 2\cos(2\pi xy)(\sum_{n\geq 1}\delta_n(y)-1)\,dy = \sum_{m\geq 1}\delta_m(x) - 1 \qquad (x>0) \tag{3b}$$

This last identity may be written in the more familiar form:

$$\int_{\mathbb{R}} e^{2\pi i xy}\sum_{n\in\mathbb{Z}}\delta_n(y)dy = \sum_{m\in\mathbb{Z}}\delta_m(x) \tag{4}$$

which expresses the invariance of the "Dirac comb" distribution $\sum_{m\in\mathbb{Z}}\delta_m(x)$ under the Fourier transform. As a linear functional on Schwartz functions ϕ , the invariance of $\sum_{m\in\mathbb{Z}}\delta_m(x)$ under Fourier is expressed as the Poisson summation formula:

$$\sum_{n\in\mathbb{Z}}\widetilde{\phi}(n) = \sum_{m\in\mathbb{Z}}\phi(m) \qquad \widetilde{\phi}(y) = \int_{\mathbb{R}} e^{2\pi i xy}\phi(x)\,dx \tag{5}$$

The Jacobi identity is the special instance with $\phi(x) = \exp(-\pi t x^2)$, and conversely the validity of (5) for Schwartz functions (and more) may be seen as a corollary to the Jacobi identity.

The *idea of co-Poisson* [4] leads to *another* formulation of the functional equation as an identity involving functions. The co-Poisson identity ((10) below) appeared in the work of Duffin and Weinberger [13]. In one of the approaches to this identity, we start with a function g on the positive half-line such that both $\int_0^\infty g(t)\,dt$ and $\int_0^\infty g(t)t^{-1}\,dt$ are absolutely convergent. Then we consider the averaged distribution $g*D(x) = \int_0^\infty g(t)D(\frac{x}{t})\frac{dt}{t}$ where $D(x) = \sum_{n\geq 1}\delta_n(x) - \mathbf{1}_{x>0}(x)$. This gives (for $x>0$):

$$g*D(x) = \sum_{n=1}^\infty \frac{g(x/n)}{n} - \int_0^\infty \frac{g(1/t)}{t}\,dt \tag{6}$$

If g is smooth with support in $[a,A]$, $0<a<A<\infty$, then the co-Poisson sum $g*D$ has Schwartz decrease at $+\infty$ (easy from applying the Poisson formula to $\frac{g(1/t)}{t}$; *cf.* [8, 4.29] for a general statement). The right Mellin transform $\widehat{g*D}(s)$ is related to the right Mellin transform $\widehat{g}(s)$ of g via the identity:

$$\widehat{g*D}(s) = \int_0^\infty (g*D)(x)x^{-s}\,dx = \zeta(s)\int_0^\infty g(x)x^{-s}\,dx = \zeta(s)\widehat{g}(s) \tag{7}$$

[2] of course, $\delta_m(x) = \delta(x-m)$.

This is because the right Mellin transform of a multiplicative convolution is the product of the right Mellin transforms. The necessary calculus of tempered distributions needed for this and other statements in this paragraph is detailed in [8]. The functional equation in the form of (1b) gives:[3]

$$\widehat{g * D}(s) = \chi_0(s)\zeta(1-s)\widehat{g}(s) = \chi_0(s)\widehat{I(g) * D}(1-s) \qquad I(g)(t) = \frac{g(1/t)}{t} \tag{8}$$

One may reinterpret this in a manner involving the cosine transform $\mathcal{C}$ acting on $L^2(0,+\infty;dx)$. The Mellin transform of a function $f(x)$ in $L^2(0,\infty;dx)$ is a function $\widehat{f}(s)$ on $\Re(s)=\frac{1}{2}$ which is nothing else than the Plancherel Fourier transform of $e^{\frac{1}{2}u}f(e^u)$: $\widehat{f}(\frac{1}{2}+i\gamma) = \int_0^\infty f(x)x^{-\frac{1}{2}-i\gamma}\,dx = \int_{-\infty}^{\infty} f(e^u)e^{\frac{u}{2}}e^{-i\gamma u}\,du$, $\int_0^\infty |f(x)|^2\,dx = \int_{-\infty}^\infty |f(e^u)e^{\frac{u}{2}}|^2\,du = \frac{1}{2\pi}\int_{\Re(s)=\frac{1}{2}} |\widehat{f}(s)|^2|ds|$. The unitary operator $\mathcal{C}I$ is scale invariant hence it is diagonalized by the Mellin transform: $\widehat{\mathcal{C}I(f)}(s) = \chi_0(s)\widehat{f}(s)$, $\widehat{\mathcal{C}(f)}(s) = \chi_0(s)\widehat{f}(1-s)$, where $\chi_0(s)$ is obtained for example using $f(x)=e^{-\pi x^2}$ and coincides with the chi-function defined in (1b). It has modulus 1 on the critical line as $\mathcal{C}$ is unitary. So (8) says that the *co-Poisson intertwining identity* holds:

$$\mathcal{C}(g * D) = I(g) * D \tag{9}$$

The co-Poisson intertwining (9) or explicitely:

$$\int_0^\infty 2\cos(2\pi xy)\left(\sum_{m=1}^\infty \frac{g(x/m)}{m} - \int_0^\infty \frac{g(1/t)}{t}\,dt\right)dx = \sum_{n=1}^\infty \frac{g(n/y)}{y} - \int_0^\infty g(t)\,dt \qquad (y>0) \tag{10}$$

is, when g is smooth with support in $[a,A]$, $0<a<A<\infty$, an identity of (even) Schwartz functions. If g is only supposed to be such that $\int_0^\infty |g(t)|(1+\frac{1}{t})\,dt < \infty$ then the co-Poisson intertwining $\mathcal{C}(g*D) = I(g)*D$ holds as an identity of distributions (either considered even or with support in $[0,\infty)$). Sufficient conditions for pointwise validity have been established [8]. The general statement of the intertwining is $\mathcal{C}(g*E) = I(g)*\mathcal{C}(E)$ where E is an arbitrary tempered distribution with support on $[0,\infty)$ (see footnote[4]) and it is proven directly. The co-Poisson identity (10) is another manner, not identical with the Poisson summation formula, to express the invariance of D under the cosine transform, or the invariance of the Dirac comb under the Fourier transform.

If the integrable function g has its support in $[a,A]$, $0<a<A<\infty$, then $g*D$ is constant in $(0,a)$ and its cosine transform (thanks to the co-Poisson intertwining) is constant in $(0,A^{-1})$. Up to a rescaling we may take $A=a^{-1}$, and then $a<1$ (if a non zero example is wanted.) Appropriate modifications allow to construct non zero even Schwartz functions constant in $(-a,a)$ and with Fourier transform again constant in $(-a,a)$ where $a>0$ is arbitrary [8].

Schwartz functions are square-integrable so here we have made contact with the investigation of de Branges [1], V Rovnyak [28] and J. and V. Rovnyak [29, 30] of square integrable functions on $(0,\infty)$ vanishing on $(0,a)$ and with Hankel transform of order ν vanishing on $(0,a)$. For $\nu=-\frac{1}{2}$ the Hankel transform of order ν is $f(y)\mapsto\sqrt{\frac{2}{\pi}}\int_0^\infty \cos(xy)f(y)\,dy$ and up to a scale change this is the cosine transform considered above. The co-Poisson idea allows to attach the zeta function to, among the spaces defined by de Branges [1], the spaces associated with the cosine transform: it has allowed the definition of some novel Hilbert spaces [3] of entire functions in relation with the Riemann zeta function and Dirichlet L-functions (the co-Poisson idea is in [4] on the adeles of an

[3] one observes that $\widehat{I(g)}(s) = \widehat{g}(1-s)$.

[4] both sides in fact depend only on $E(x)+E(-x)$ as a distribution on the line, which may be identically 0, and this happens exactly when E is a linear combination of odd derivatives of the delta function.

arbitrary algebraic number field K; then, the study of the related Hilbert spaces was begun for $K = \mathbb{Q}$. Further results were obtained in [7].)

The study of the function $\chi_0(s) = \pi^{s-\frac{1}{2}}\frac{\Gamma(\frac{1-s}{2})}{\Gamma(\frac{s}{2})}$, of unit modulus on the critical line, is interesting. We proposed to realize the χ_0 function as a "scattering matrix". This is indeed possible and has been achieved in [6]. The distributions, functions, and differential equations involved are all related to, or expressed by, the Fredholm determinants of the finite cosine transform, which in turn are related to the Fredholm determinants of the finite Dirichlet kernels $\frac{\sin(t(x-y))}{\pi(x-y)}$ on $[-1, 1]$. The study of the Dirichlet kernels is a topic with a vast literature. A minor remark will be made in an appendix.

We mentioned the Riemann zeta function and how it relates to $\chi_0(s) = \pi^{s-\frac{1}{2}}\frac{\Gamma(\frac{1-s}{2})}{\Gamma(\frac{s}{2})}$ and to the cosine transform. Let us now briefly consider the Dedekind zeta function of the Gaussian number field $\mathbb{Q}(i)$ and how it relates to $\chi(s) = \frac{\Gamma(1-s)}{\Gamma(s)}$ and to the $\mathcal{H}$ transform. The $\mathcal{H}$ transform is

$$\mathcal{H}(g)(y) = \int_0^\infty J_0(2\sqrt{xy})g(x)\,dx \qquad J_0(2\sqrt{xy}) = \sum_{n=0}^\infty (-1)^n \frac{x^n y^n}{n!^2} \tag{11}$$

Up to the unitary transformation $g(x) = (2x)^{-\frac{1}{4}}f(\sqrt{2x})$, $\mathcal{H}(g)(y) = (2y)^{-\frac{1}{4}}k(\sqrt{2y})$, it becomes the Hankel transform of order zero $k(y) = \int_0^\infty \sqrt{xy}J_0(xy)f(x)\,dx$. It is a self-reciprocal, unitary, scale reversing operator $(\mathcal{H}(g(\lambda x))(y) = \frac{1}{\lambda}\mathcal{H}(g)(\frac{y}{\lambda}))$. We shall also extend its action to tempered distributions on $\mathbb{R}$ with support in $[0, +\infty)$. At the level of right Mellin transforms of elements of $L^2(0, \infty; dx)$ it acts as:

$$\widehat{\mathcal{H}(g)}(s) = \chi(s)\widehat{g}(1-s) \qquad \chi(s) = \frac{\Gamma(1-s)}{\Gamma(s)} \qquad \Re(s) = \frac{1}{2} \tag{12}$$

It has $e^{-x}\mathbf{1}_{x\geq 0}(x)$ as one among its self-reciprocal functions, as is verified directly by series expansion $\int_0^\infty J_0(2\sqrt{xy})e^{-y}\,dy = \sum_{n=0}^\infty \frac{(-1)^n}{n!^2}x^n\int_0^\infty y^n e^{-y}\,dy = e^{-x}$. The identity

$$\int_0^\infty J_0(2\sqrt{t})t^{-s}\,dt = \chi(s) = \frac{\Gamma(1-s)}{\Gamma(s)} \tag{13}$$

is equivalent to a special case of well-known formulas of Weber, Sonine and Schafheitlin [33, 13.24.(1)]. Here we have an absolutely convergent integral for $\frac{3}{4} < \Re(s) < 1$ and in that range the identity may be proven as in: $e^{-x} = \int_0^\infty J_0(2\sqrt{xy})e^{-y}\,dy = \int_0^\infty J_0(2\sqrt{y})\frac{1}{x}e^{-\frac{y}{x}}\,dy$, $\Gamma(1-s) = \int_0^\infty J_0(2\sqrt{y})(\int_0^\infty x^{-s-1}e^{-\frac{y}{x}}\,dx)\,dy = \Gamma(s)\int_0^\infty J_0(2\sqrt{y})y^{-s}\,dy$. The integral is semi-convergent for $\Re(s) > \frac{1}{4}$, and of course (13) still holds. In particular on the critical line and writing $t = e^u$, $s = \frac{1}{2} + i\gamma$, we obtain the identities of tempered distributions $\int_\mathbb{R} e^{\frac{1}{2}u}J_0(2e^{\frac{1}{2}u})e^{-i\gamma u}\,du = \chi(\frac{1}{2} + i\gamma)$, $e^{\frac{1}{2}u}J_0(2e^{\frac{1}{2}u}) = \frac{1}{2\pi}\int_\mathbb{R}\chi(\frac{1}{2} + i\gamma)e^{+i\gamma u}\,du$.

We have $\zeta_{\mathbb{Q}(i)}(s) = \frac{1}{4}\sum_{(n,m)\neq(0,0)}\frac{1}{(n^2+m^2)^s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{4^s} + \frac{2}{5^s} + \frac{1}{8^s} + \cdots = \sum_{n\geq 1}\frac{c_n}{n^s}$ and it is a meromorphic function in the entire complex plane with a simple pole at $s = 1$, residue $\frac{\pi}{4}$. Its functional equation assumes at least two convenient well-known forms:

$$(\sqrt{4})^s(2\pi)^{-s}\Gamma(s)\zeta_{\mathbb{Q}(i)}(s) = (\sqrt{4})^{1-s}(2\pi)^{-(1-s)}\Gamma(1-s)\zeta_{\mathbb{Q}(i)}(1-s) \tag{14a}$$

$$(\frac{1}{\pi})^s\zeta_{\mathbb{Q}(i)}(s) = \chi(s)(\frac{1}{\pi})^{1-s}\zeta_{\mathbb{Q}(i)}(1-s) \qquad \chi(s) = \frac{\Gamma(1-s)}{\Gamma(s)} \tag{14b}$$

The former is related to the expression of $\pi^{-s}\Gamma(s)\zeta_{\mathbb{Q}(i)}(s)$ as a left Mellin transform:

$$\pi^{-s}\Gamma(s)\zeta_{\mathbb{Q}(i)}(s) = \frac{1}{4}\int_0^\infty (\theta(t)^2 - 1)t^{s-1}\,dt \qquad (\Re(s) > 1) \tag{15a}$$

$$= \frac{1}{4}\int_0^\infty (\theta(t)^2 - 1 - \frac{1}{t})t^{s-1}\,dt \qquad (0 < \Re(s) < 1) \tag{15b}$$

$$\theta(t)^2 = \frac{1}{t}\theta(\frac{1}{t})^2 \tag{15c}$$

The latter form of the functional equation is related to the expression of $(\frac{1}{\pi})^s\zeta_{\mathbb{Q}(i)}(s)$ as the right Mellin transform of a tempered distribution which is supported in $[0,\infty)$ and which is self-reciprocal under the $\mathcal{H}$-transform:

$$(\frac{1}{\pi})^s\zeta_{\mathbb{Q}(i)}(s) = \int_0^\infty (\sum_{m\geq 1} c_m\delta_{\pi m}(x) - \frac{1}{4})x^{-s}dx \tag{16a}$$

$$\int_0^\infty J_0(2\sqrt{xy})(\sum_{n\geq 1} c_n\delta_{\pi n}(y) - \frac{1}{4})\,dy = \sum_{m\geq 1} c_m\delta_{\pi m}(x) - \frac{1}{4}\mathbf{1}_{x>0}(x) = E(x) \qquad (x > 0) \tag{16b}$$

The invariance of E under the $\mathcal{H}$-transform is equivalent to the validity of the functional equation of $(\frac{1}{\pi})^s\zeta_{\mathbb{Q}(i)}(s)$ and it having a pole with residue $\frac{1}{4}$ at $s = 1$. The co-Poisson intertwining becomes the assertion:

$$y > 0 \implies \int_0^\infty J_0(2\sqrt{xy})\left(\sum_{m=1}^\infty c_m\frac{g(x/\pi m)}{\pi m} - \frac{1}{4}\int_0^\infty g(\frac{1}{t})\,\frac{dt}{t}\right)dx = \sum_{n=1}^\infty c_n\frac{g(\pi n/y)}{y} - \frac{1}{4}\int_0^\infty g(t)\,dt \tag{17}$$

If g is smooth with support in $[b,B]$, $0 < b < B < \infty$, then we have on the right hand side a function of Schwartz decrease at $+\infty$ (compare to Theorem 3), and its $\mathcal{H}$-transform is also of Schwartz decrease at $+\infty$. The former is constant for $0 < y < \pi B^{-1}$ and the latter is constant for $0 < x < \pi b$. The supremum of the values obtainable for the product of the lengths of the intervals of constancy is π^2. But, as for the cosine and sine transforms, there does exist smooth functions which are constant on a given $(0,a)$ for arbitrary $a > 0$ with an $\mathcal{H}$ transform again constant on $(0,a)$ and have Schwartz decrease at $+\infty$ (the two constants being arbitrarily prescribed.)

De Branges and V. Rovnyak have obtained [1, 28] rather complete results in the study of the Hankel transform of order zero $f(x) \mapsto g(y) = \int_0^\infty \sqrt{xy}J_0(xy)f(x)\,dx$ from the point of view of understanding the support property of being zero and with transform again zero in a given interval $(0,b)$. They obtained an isometric expansion (Theorem 1 of section 2) and also the detailed description of the related spaces of entire functions ([1]). The more complicated case of the Hankel transforms of non-zero integer orders was treated by J. and V. Rovnyak [29, 30]. These, rather complete, results are an indication that the Hankel transform of order zero or of integer order is easier to understand than the cosine or sine transforms, and that doing so thoroughly could be useful to better understand how to try to understand the cosine and sine transforms.

The kernel $J_0(2\sqrt{uv})$ of the $\mathcal{H}$-transform satisfies the Klein-Gordon equation in the variables $x = v - u$, $t = v + u$:

$$(\frac{\partial^2}{\partial u\partial v} + 1)J_0(2\sqrt{uv}) = (\Box + 1)J_0(2\sqrt{uv}) = (\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} + 1)J_0(\sqrt{t^2 - x^2}) = 0 \tag{18}$$

It is a noteworthy fact that the support condition, initially considered by de Branges and V. Rovnyak, and which, nowadays, is also seen to be in relation with the co-Poisson identities, has turned out to be related to the relativistic causality governing the propagation of solutions to the Klein-Gordon equation. This has been established in [9] where we obtained as an application of this idea the isometric expansion of [1, 28] in a novel manner. It was furthermore proven in [9] that the $\mathcal{H}$ transform is indeed an (absolute) scattering, in fact the scattering from the past boundary to the future boundary of the Rindler wedge $0 < |t| < x$ for solutions of a first order, two-component ("Dirac"), form of the KG equation.

In the present paper, which is completely independent from [9], we shall again study the $\mathcal{H}$-transform and show in particular how to recover in yet a different way the earlier results of [1, 28] and also we shall extend them. This will be based on the methods from [5, 6], and uses the techniques motivated by the study of the co-Poisson idea [8]. Our exposition will thus give a fully detailed account of the material available in summarized form in [5, 6]. Then we proceed with a development of these methods to provide the elucidation of the (two dimensions bigger) spaces of functions constant in $(0, a)$ and with $\mathcal{H}$-transforms constant in $(0, a)$.

The use of tempered distributions is an important point of our approach[5]; also one may envision the co-Poisson idea as asking not to completely identify a distribution with the linear functional it "is". In this regard it is of note that the distributions which arise following the method of [5] are seen in the present case of the study of the $\mathcal{H}$-transform to have a very natural formulation as differences of boundary values of analytic functions, that is, as hyperfunctions [23]. We do not use the theory of hyperfunctions as such, but could not see how not to mention that this is what these distributions seem to be in a natural manner.

The paper contains no number theory. And, the reader will need no prior knowledge of [2]; some familiarity with the m-function of Hermann Weyl [10, 21, 26] is necessary at one stage of the discussion (there is much common ground, in fact, between the properties of the m-function and the axioms of [2]). The reproducing kernel in any space with the axioms of [2] has a specific appearance (equation (109) below) which has been used as a guide to what we should be looking for. The validity of the formula is re-proven in the specific instance considered here[6]. Regarding the differential equations governing the deformation, with respect to the parameter $a > 0$ [7], of the Hilbert spaces, we depart from the general formalism of [2] and obtain them in a canonical form, as defined in [21, §3]. Interestingly this is related to the fact that the A and B functions (connected to the reproducing kernel, equation (109)) which are obtained by the method of [5] turn out not to be normalized according to the rule in general use in [2]. Each rule of normalization has its own advantages; here the equations are obtained in the Schrödinger and Dirac forms familiar from the spectral theory of linear second order differential equations [10, 21, 26]. This allows to make reference to the well-known Weyl-Stone-Titchmarsh-Kodaira theory [10, 21, 26], and to understand $\mathcal{H}$ as a scattering. Regarding spaces with the axioms of [2], the articles of Dym [14] and Remling [27] will be useful to the reader interested in second order linear differential equations. And we refer the reader with number theoretical interests to the recent papers of Lagarias [18, 19].

The author has been confronted with a dilemma: a substantial portion of the paper (most of chapters 5, 6, 8) has a general validity for operators having a kernel of the multiplicative type $k(xy)$ possessing certain properties in common with the cosine, sine or $\mathcal{H}$ transforms. But on the

[5] at the bottom of page 456 of [1] the formulas given for $A(a, z)$ and $B(a, z)$ as completed Mellin transforms are lacking terms which would correspond to Dirac distributions; possibly related to this, the isometric expansion as presented in Theorem II of [1] is lacking corresponding terms. The exact isometric expansion appears in [28] and the exact formulas for $A(a, z)$ and $B(a, z)$ as completed Mellin transforms appear, in an equivalent form, in [30, eq.(37)].

[6] the critical line here plays the rôle of the real axis in [2], s is $\frac{1}{2} - iz$ and the use of the variable s is most useful in distinguishing the right Mellin transforms which need to be completed by a Gamma factor from the left Mellin transforms of "theta"-like functions.

[7] the a here corresponds to $\frac{1}{2}a^2$ in [1].

other hand the (essentially) unique example where all quantities arising may be computed is the $\mathcal{H}$ transform (and transforms derived from it, or closely related to it, as the Hankel transforms of integer orders). We have tried to give proofs whose generality is obvious, but we also made full use of distributions, as this allows to give to the quantities arising very natural expressions. Also we never hesitate using arguments of analyticity although for some topics (for example, some aspects involving certain integral equations and Fredholm determinants) this is certainly not really needed.

2 Hardy spaces and the de Branges-Rovnyak isometric expansion

Let us state the isometric expansion of [1, 28] regarding the square integrable Hankel transforms of order zero. We reformulate the theorem to express it with the $\mathcal{H}$ transform (11) rather than the Hankel transform of order zero.

Theorem 1 ([1], [28])**.** *: Let $k \in L^2(0,\infty;dx)$. The functions f_1 and g_1, defined as the following integrals:*

$$f_1(y) = \int_y^\infty J_0(2\sqrt{y(x-y)})k(x)\,dx\ , \tag{19a}$$

$$g_1(y) = k(y) - \int_y^\infty \sqrt{\frac{y}{x-y}}J_1(2\sqrt{y(x-y)})k(x)\,dx\ , \tag{19b}$$

exist in L^2 in the sense of mean-square convergence, and they verify:

$$\int_0^\infty |f_1(y)|^2 + |g_1(y)|^2\,dy = \int_0^\infty |k(x)|^2\,dx\ . \tag{19c}$$

The function k is given in terms of the pair (f_1, g_1) as:

$$k(x) = g_1(x) + \int_0^x J_0(2\sqrt{y(x-y)})f_1(y)\,dy - \int_0^x \sqrt{\frac{y}{x-y}}J_1(2\sqrt{y(x-y)})g_1(y)\,dy \tag{19d}$$

The assignment $k \mapsto (f_1, g_1)$ is a unitary equivalence of $L^2(0,\infty;dx)$ with $L^2(0,\infty;dy) \oplus L^2(0,\infty;dy)$ such that the $\mathcal{H}$-transform acts as $(f_1, g_1) \mapsto (g_1, f_1)$. Furthermore k and $\mathcal{H}(k)$ both identically vanish in $(0,a)$ if and only if f_1 and g_1 both identically vanish in $(0,a)$.

Let us mention the following (which follows from the proof we have given of Thm. 1 in [9]): *if f_1, f_1', g_1, g_1' are in L^2 then k, k' and $\mathcal{H}(k)'$ are in L^2. Conversely if k, k' and $\mathcal{H}(k)'$ are in L^2 then the integrals defining $f_1(y)$ and $g_1(y)$ are convergent for each $y > 0$ as improper Riemann integrals, and f_1' and g_1' are in L^2.*

It will prove convenient to work with $(f(x), g(x)) = \frac{1}{2}(g_1(\frac{x}{2}) + f_1(\frac{x}{2}), g_1(\frac{x}{2}) - f_1(\frac{x}{2}))$:

$$f(y) = \frac{1}{2}k(\frac{y}{2}) + \frac{1}{2}\int_{y/2}^\infty \left(J_0(\sqrt{y(2x-y)}) - \sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\right)k(x)\,dx \tag{20a}$$

$$g(y) = \frac{1}{2}k(\frac{y}{2}) - \frac{1}{2}\int_{y/2}^\infty \left(J_0(\sqrt{y(2x-y)}) + \sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\right)k(x)\,dx \tag{20b}$$

$$\begin{aligned} k(x) = f(2x) &+ \frac{1}{2}\int_0^{2x} \left(J_0(\sqrt{y(2x-y)}) - \sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\right)f(y)\,dy \\ &+ g(2x) - \frac{1}{2}\int_0^{2x} \left(J_0(\sqrt{y(2x-y)}) + \sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\right)g(y)\,dy \end{aligned} \tag{20c}$$

$$\int_0^\infty |k(x)|^2\,dx = \int_0^\infty |f(y)|^2 + |g(y)|^2\,dy \tag{20d}$$

The $\mathcal{H}$ transform on k acts as $(f,g) \mapsto (f,-g)$. The pair $(k, \mathcal{H}(k))$ identically vanishes on $(0,a)$ if and only if the pair (f,g) identically vanishes on $(0,2a)$. The structure of the formulas is more apparent after observing $(x,y>0)$:

$$\frac{\partial}{\partial x}\Big(\frac{1}{2}\,J_0(\sqrt{y(2x-y)})\mathbf{1}_{0<y<2x}(y)\Big) = \delta_{2x}(y) - \frac{1}{2}\sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\mathbf{1}_{0<y<2x}(y) \qquad (21)$$

In this section I shall prove the existence of an isometric expansion $k \leftrightarrow (f,g)$ having the stated support properties and relation to the $\mathcal{H}$-transform; that this construction does give the equations (20a), (20b), (20c), will only be established in the last section (9) of the paper. The method followed in this section coincides partly with the one of V. Rovnyak [28]; we try to produce the most direct arguments, using the commonly known facts on Hardy spaces. The reader only interested in Theorem 1 is invited after having read the present section to then jump directly to section 9 for the conclusion of the proof.

To a function $k \in L^2(0,\infty;dx)$ we associate the analytic function

$$\widetilde{k}(\lambda) = \int_0^\infty e^{i\lambda x}k(x)\,dx \qquad (\Im(\lambda) > 0) \qquad (22)$$

with boundary values for $\lambda \in \mathbb{R}$ again written $\widetilde{k}(\lambda)$, which defines an element of $L^2(\mathbb{R}, \frac{d\lambda}{2\pi})$, the assignment $k \mapsto \widetilde{k}$ being unitary from $L^2(0,\infty;dx)$ onto $\mathbb{H}^2(\Im(\lambda>0), \frac{d\lambda}{2\pi})$. Next we have the conformal equivalence and its associated unitary map from $\mathbb{H}^2(\Im(\lambda>0), \frac{d\lambda}{2\pi})$ to $\mathbb{H}^2(|w|<1, \frac{d\theta}{2\pi})$:

$$w = \frac{\lambda - i}{\lambda + i} \qquad K(w) = \frac{1}{\sqrt{2}}\frac{\lambda+i}{i}\,\widetilde{k}(\lambda) \qquad (23)$$

It is well known that this indeed unitarily identifies the two Hardy spaces. With $k_0(x) = e^{-x}$, $\widetilde{k}_0(\lambda) = \frac{i}{\lambda+i}$, $K_0(w) = \frac{1}{\sqrt{2}}$, and $\|k_0\|^2 = \int_0^\infty e^{-2x}dx = \frac{1}{2} = \|K_0\|^2$. The functions $\widetilde{k}_n(\lambda) = (\frac{\lambda-i}{\lambda+i})^n\frac{i}{\lambda+i}$ correspond to $K_n(w) = \frac{1}{\sqrt{2}}w^n$. To obtain explicitely the orthogonal basis $(k_n)_{n\geq 0}$, we first observe that $w = 1 - 2\frac{i}{\lambda+i}$, so as a unitary operator it acts as:

$$w \cdot k(x) = k(x) - 2\int_0^x e^{-(x-y)}k(y)\,dy = k(x) - e^{-x}2\int_0^x e^y k(y)\,dy \qquad (24)$$

Writing $k_n(x) = P_n(x)e^{-x}$ we thus obtain $P_{n+1}(x) = P_n(x) - 2\int_0^x P_n(y)\,dy$:

$$P_n(x) = \left(1 - 2\int_0^x\right)^n \cdot 1 = \sum_{j=0}^n \binom{n}{j}\frac{(-2x)^j}{j!} \qquad (25)$$

So as is well-known $P_n(x) = L_n^{(0)}(2x)$ (in the notation of [31, §5]) where the Laguerre polynomials $L_n^{(0)}(x)$ are an orthonormal system for the weight $e^{-x}dx$ on $(0,\infty)$.

One of the most common manner to be led to the $\mathcal{H}$-transform is to define it from the two-dimensional Fourier transform as:

$$\begin{gathered}\mathcal{H}(f)(\frac{1}{2}r^2) = \frac{1}{2\pi}\iint e^{i(x_1y_1+x_2y_2)}f(\frac{y_1^2+y_2^2}{2})dy_1dy_2 = \int_0^\infty\left(\int_0^{2\pi}e^{irs\cos\theta}\frac{d\theta}{2\pi}\right)f(\frac{1}{2}s^2)sds \\ \mathcal{H}(f)(\frac{1}{2}r^2) = \int_0^\infty J_0(rs)f(\frac{1}{2}s^2)\,sds \qquad r^2 = x_1^2+x_2^2, s^2 = y_1^2+y_2^2\end{gathered} \qquad (26)$$

which proves its unitarity, self-adjointness, and self-reciprocal character and the fact that it has e^{-x} has self-reciprocal function. The direct verification of $\mathcal{H}(k_0) = k_0$ is immediate: $\mathcal{H}(k_0)(x) = \int_0^\infty J_0(2\sqrt{xy})e^{-y}\,dy = \sum_{n=0}^\infty \frac{(-1)^n}{n!^2} x^n \int_0^\infty y^n e^{-y}\,dy = e^{-x}$. Then, $\mathcal{H}(e^{-tx}) = t^{-1}e^{-\frac{x}{t}}$ for each $t > 0$. So $\int_0^\infty e^{-tx}\mathcal{H}(k)(x)\,dx = t^{-1}\int_0^\infty e^{-\frac{1}{t}x}k(x)\,dx$ hence:

$$\forall k \in L^2(0,\infty;dx) \qquad \widetilde{\mathcal{H}(k)}(\lambda) = \frac{i}{\lambda}\widetilde{k}(\frac{-1}{\lambda}) \tag{27}$$

With the notation $\mathcal{H}(K)$ for the function in $\mathbb{H}^2(|w|<1)$ corresponding to $\mathcal{H}(k)$, we obtain from (23), (27), an extremely simple result:[8]

$$\mathcal{H}(K)(w) = K(-w) \tag{28}$$

This obviously leads us to associate to $K(w) = \sum_{n=0}^\infty c_n w^n$ the functions:

$$F(w) := \sum_{n=0}^\infty c_{2n}w^n \tag{29a}$$

$$G(w) := \sum_{n=0}^\infty c_{2n+1}w^n \tag{29b}$$

$$K(w) = F(w^2) + wG(w^2) \tag{29c}$$

and to k the functions f and g in $L^2(0,+\infty;dx)$ corresponding to F and G. Certainly, $\|k\|^2 = \|f\|^2 + \|g\|^2$, and the assignment of (f,g) to k is an isometric identification. Furthermore, certainly the $\mathcal{H}$ transform acts in this picture as $(f,g) \mapsto (f,-g)$. Let us now check the support properties. Let $\alpha(m)$ be the leftmost point of the (essential) support of a given $m \in L^2(0,\infty;dx)$. As is well-known,

$$-\alpha(m) = \limsup_{t\to+\infty} \frac{1}{t}\log|\widetilde{m}(it)|\,, \tag{30}$$

If w corresponds to λ via (23) then w^2 corresponds to $\frac{1}{2}(\lambda - \frac{1}{\lambda})$, so if to a function f with corresponding $F(w)$ we associate the function $\psi(f) \in L^2(0,\infty;dx)$ which corresponds to $F(w^2)$,

$$(t+1)\widetilde{\psi(f)}(i\,t) = (\frac{t+\frac{1}{t}}{2}+1)\widetilde{f}(i\,\frac{t+\frac{1}{t}}{2})\,, \tag{31}$$

then we have the identity:

$$\alpha(\psi(f)) = \frac{1}{2}\alpha(f) \tag{32}$$

Returning to F (resp. f) and G (resp. g) associated via (29a), (29b), to K (resp. k) we thus have $k = \psi(f) + w\cdot\psi(g)$, $\mathcal{H}(k) = \psi(f) - w\cdot\psi(g)$, hence if the pair (f,g) vanishes on $(0,2a)$ then the pair $(k,\mathcal{H}(k))$ vanishes on $(0,a)$ (clearly the unitary operator of multiplication by $w = \frac{\lambda-i}{\lambda+i}$ does not affect $\alpha(m)$.) Conversely, as $\alpha(f) = 2\alpha(k+\mathcal{H}(k))$ and $\alpha(g) = 2\alpha(k-\mathcal{H}(k))$, if the pair $(k,\mathcal{H}(k))$ vanishes on $(0,a)$ then the pair (f,g) vanishes on $(0,2a)$.

We have thus established the existence of an isometric expansion, its support properties, and its relation to the $\mathcal{H}$-transform. That there is indeed compatibility of (20a) and (20b) with (29a) and (29b), and with (20c), will be established in the last section (9) of the paper with a direct study of (31). In the meantime equations (20a), (20b), (20c) and (20d) will have been confirmed in another manner. Yet another proof of the isometric expansion has been given in [9].

[8] we also take note of the operator identity $\mathcal{H}\cdot w = -w\cdot\mathcal{H}$.

3 Tempered distributions and their $\mathcal{H}$ and Mellin transforms

Any distribution D on $\mathbb{R}$ has a primitive. If the closed support of D is included in $[0,+\infty)$, then it has a unique primitive, which we will denote $\int_0^x D(x)\,dx$, or, more safely, $D^{(-1)}$, which also has its support in $[0,+\infty)$. The temperedness of such a D is equivalent to the fact that $D^{(-N)}$ for $N \gg 0$ is a continuous function with polynomial growth. With $D^{(-N)}(x) = (1+x^2)^M g_{(N,M)}(x)$, $M \gg 0$, we can express D as $P(x, \frac{d}{dx})(g)$ where P is a polynomial and $g \in L^2(0,\infty;dx)$. Conversely any such expression is a tempered distribution vanishing in $(-\infty, 0)$. The Fourier transforms of such tempered distributions $\widetilde{D}(\lambda)$ appear thus as the boundary values of $Q(\frac{d}{d\lambda}, \lambda)f(\lambda)$ where Q are polynomials and the f's belong to $\mathbb{H}^2(\Im(\lambda) > 0)$. As taking primitives is allowed we know without further ado that this class of analytic functions is the same thing as the space of functions $g(\lambda) = R(\frac{d}{d\lambda}, \lambda, \lambda^{-1})f(\lambda)$, R a polynomial and $f \in \mathbb{H}^2$. It is thus clearly left stable by the operation:

$$g \mapsto \mathcal{H}(g)(\lambda) := \frac{i}{\lambda} g(\frac{-1}{\lambda}) \qquad (\Im(\lambda) > 0) \tag{33}$$

which will serve to define the action of $\mathcal{H}$ on tempered distributions with support in $[0,+\infty)$.

Let us also use (33), where now $\lambda \in \mathbb{R}$, to define $\mathcal{H}$ as a unitary operator on $L^2(-\infty,+\infty;dx)$. It will anti-commute with $f(x) \to f(-x)$ so:

$$\mathcal{H}(f)(x) = \int_{-\infty}^{\infty} (J_0(2\sqrt{xy})\mathbf{1}_{x>0}(x)\mathbf{1}_{y>0}(y) - J_0(2\sqrt{xy})\mathbf{1}_{x<0}(x)\mathbf{1}_{y<0}(y))f(y)\,dy \tag{34}$$

Useful operator identities are easily established from (33):

$$x\frac{d}{dx} \cdot \mathcal{H} = -\mathcal{H} \cdot \frac{d}{dx}x \qquad \text{and} \qquad \frac{d}{dx}x \cdot \mathcal{H} = -\mathcal{H} \cdot x\frac{d}{dx} \tag{35a}$$

$$\frac{d}{dx} \cdot \mathcal{H} = \mathcal{H} \cdot \int_0^x \qquad \text{and} \qquad \int_0^x \cdot \mathcal{H} = \mathcal{H} \cdot \frac{d}{dx} \tag{35b}$$

$$x \cdot \mathcal{H} = -\mathcal{H} \cdot \frac{d}{dx}\, x\, \frac{d}{dx} \qquad \text{and} \qquad \mathcal{H} \cdot x = -\frac{d}{dx}\, x\, \frac{d}{dx} \cdot \mathcal{H} \tag{35c}$$

It is important that $\frac{d}{dx}$ is always taken in the distribution sense. It would actually be possible to define the action of $\mathcal{H}$ on distributions supported in $[0,+\infty)$ without mention of the Fourier transform, because these identities uniquely determine $\mathcal{H}(D)$ if D is written $(\frac{d}{dx})^N(1+x)^M g_{N,M}(x)$ with $g_{N,M} \in L^2(0,\infty;dx)$. But the proof needs some organizing then as it is necessary to check independence from the choice of N and M, and also to establish afterwards all identities above. So (33) provides the easiest road. Still, in this context, let us mention the following which relates to the restriction of $\mathcal{H}(D)$ to $(0,+\infty)$:

Lemma 2. *Let k be smooth on $\mathbb{R}$ with compact support in $[0,+\infty)$. Then $\mathcal{H}(k)$ is the restriction to $[0,+\infty)$ of an entire function γ which has Schwartz decrease as $x \to +\infty$. For any tempered distribution D with support in $[0,+\infty)$, there holds*

$$\int_0^\infty \mathcal{H}(D)(x)k(x)\,dx = \int_0^\infty D(x)\gamma(x)\,dx\ , \tag{36}$$

where in the right hand side in fact one has $\int_{-\epsilon}^\infty D(x)\gamma(x)\theta(x)\,dx$ where the smooth function θ is 1 for $x \geq -\frac{\epsilon}{3}$ and 0 for $x \leq -\frac{\epsilon}{2}$ and is otherwise arbitrary (as is ϵ).

Let us suppose $k = 0$ for $x > B$. Defining:

$$\gamma(x) = \int_0^B J_0(2\sqrt{xy})\, k(y)\, dy \tag{37}$$

we obtain an entire function and, according to our definitions, $\mathcal{H}(k)(x) = \gamma(x)\mathbf{1}_{x>0}(x)$ as a distribution or a square-integrable function. Using (35c) ($\mathcal{H} = -\frac{1}{x}\mathcal{H}\cdot \frac{d}{dx}\, x\, \frac{d}{dx}$ for $x > 0$) and bounding J_0 by 1 we see (induction) that γ is $O(x^{-N})$ for any N as $x \to +\infty$, and using (35a) ($\frac{d}{dx}\cdot \mathcal{H} = -\frac{1}{x}\mathcal{H}\cdot \frac{d}{dx}x$ for $x > 0$) the same applies to its derivative and also to its higher derivatives. So it is of the Schwartz class for $x \to +\infty$.

Replacing D by $\mathcal{H}(D)$ in (36) it will be more convenient to prove:

$$\int_0^\infty D(x)k(x)\, dx = \int_0^\infty \mathcal{H}(D)(x)\gamma(x)\, dx \tag{38}$$

If (38) holds for D (and all k's) then $< D', k >= - < D, k' >= - < \mathcal{H}(D), -\theta(x) \int_x^\infty \gamma(y)dy >$ (observe that $\int_0^x \gamma(y)\, dy = \mathcal{H}(k')(x)$ vanishes at $+\infty$) so $< D', k >= + < \int_0^x \mathcal{H}(D), \theta\gamma >=< \mathcal{H}(D'), \theta\gamma >$ hence (38) holds as well for D' (and all k's). So we may assume D to be a continuous function of polynomial growth. It is also checked using (35c) that if (38) holds for D it holds for xD. So we may reduce to D being square-integrable, and the statement then follows from the self-adjointness of $\mathcal{H}$ on L^2 (or we reduce to Fubini).

The behavior of $\mathcal{H}$ with respect to the translations $\tau_a : f(x) \mapsto f(x - a)$ is important. For $f \in L^2(\mathbb{R}; dx)$ the value of a is arbitrary and we can define

$$\tau_a^\# := \mathcal{H}\, \tau_a\, \mathcal{H} \tag{39a}$$

$$\widetilde{\tau_a(f)}(\lambda) = e^{ia\lambda}\widetilde{f}(\lambda) \tag{39b}$$

$$\widetilde{\tau_a^\#(f)}(\lambda) = e^{ia\frac{-1}{\lambda}}\widetilde{f}(\lambda) \tag{39c}$$

We observe the remarkable commutation relations (which would fail for the cosine or sine transforms):

$$\forall a, b \qquad \tau_a\tau_b^\# = \tau_b^\#\tau_a \tag{40}$$

For a distribution D the action of $\tau_a^\#$ is here defined only for $a \geq -\alpha(\mathcal{H}(D))$, where $\alpha(E)$ is the leftmost point of the closed support of the distribution E. On this topic from the validity of (30) when $f \in L^2(0, \infty; dx)$, and invariance of α under derivation[9], integration, and multiplication by x, one has:

$$-\alpha(E) = \limsup_{t\to+\infty} \frac{1}{t}\, \log|\widetilde{E}(it)| \tag{41}$$

We thus have the property, not shared by the cosine or sine transforms:

$$a \geq -\alpha(\mathcal{H}(D)) \implies \alpha(\tau_a^\#(D)) = \alpha(D) \tag{42}$$

We now consider D with $\alpha(D) > 0$ and $\alpha(\mathcal{H}(D)) > 0$ and prove that its Mellin transform is an entire function with trivial zeros at $0, -1, -2, \ldots$, following the method of regularization by multiplicative convolution and co-Poisson intertwining from [8]. The other, very classical in spirit, proof shall be presented later. The latter method is shorter but the former provides complementary information.

[9] It is important in order to avoid a possible confusion to insist on the fact that $\frac{d}{dx}$ is always taken in the distribution sense so for example $\frac{d}{dx}\mathbf{1}_{x>0} = \delta(x)$ indeed has the leftmost point of its support not affected by $\frac{d}{dx}$.

In [8, §4.A] the detailed explanations relative to the notion of multiplicative convolution are given:

$$(g * D)(x) \text{ “ = ” } \int_{\mathbb{R}} g(t)D(\frac{x}{t})\,\frac{dt}{|t|}\;, \tag{43}$$

where we will in fact always take g to have compact support in $(0, +\infty)$. It is observed that

$$g * xD = x(\frac{g}{x} * D) \qquad (g * D)' = \frac{g}{x} * D' \tag{44}$$

The notion of right Mellin transform $\int_0^\infty D(x)x^{-s}dx$ is developed in [8, §4.C], for D with support in $[a, +\infty)$, $a > 0$:

$$\widehat{D}(s) = s(s+1)\cdots(s+N-1)\widehat{D^{(-N)}}(s+N)\;, \tag{45}$$

where $N \gg 0$. The meaning of $\widehat{D}$ is as the maximal possible analytic continuation to a half-plane $\Re(s) > \sigma$, where σ is as to the left as is possible. The notion is extended[10] in [8, §4.F] to the case where the restriction of D to $(-a, a)$ is “quasi-homogeneous”. For example, if $D|_{(-a,a)} = \mathbf{1}_{0<x<a}$ (resp. δ), then $\widehat{D}$ is defined as $\widehat{D}_1$ with $D_1 = D - \mathbf{1}_{0<x<\infty}$ (resp. $D - \delta$.) Then, also in the extended case, the following holds:

$$\widehat{g * D}(s) = \widehat{g}(s)\widehat{D}(s) \tag{46}$$

where g in an integrable function with compact support in $(0, \infty)$ and $\widehat{g}(s)$ is the entire function $\int_0^\infty g(t)t^{-s}\,dt$. We then have the following theorem:

Theorem 3. *Let D a tempered distribution with support in $[a, +\infty)$, $a > 0$ and such that $\mathcal{H}(D)$ also has a positive leftmost point of support. Let g be a smooth function with compact support in $(0, \infty)$. Then the multiplicative convolution $g * D$ belongs to the Schwartz class.*

This is the analog of [8, Thm 4.29]. The function $k(t) = (Ig)(t) = \frac{g(1/t)}{t}$ is defined and it is written as $k = \mathcal{H}(\gamma \mathbf{1}_{x>0})$ where γ is the entire function, of Schwartz decrease at $+\infty$ such that $\mathcal{H}(k) = \gamma \cdot \mathbf{1}_{x>0}$. Then it is observed that

$$t > 0 \implies \quad (g * D)(t) = \int_0^\infty D(x)\frac{k(x/t)}{t}\,dx = \int_0^\infty \mathcal{H}(D)(x)\gamma(tx)\,dx \tag{47}$$

We have used Lemma 2. Then the Schwartz decrease of $\int_0^\infty \mathcal{H}(D)(x)\gamma(tx)\,dx$ as $t \to +\infty$ is established as is done at the end of the proof of [8, Thm 4.29], integrating by parts enough times to transform $\mathcal{H}(D)$ into a continuous function of polynomial growth, identically zero on $[0, c]$, $c > 0$.

Theorem 4. *Let D a tempered distribution with a positive lefmost point of support and such that $\mathcal{H}(D)$ also has a positive leftmost point of support. Then $\widehat{D}(s)$ and $\Gamma(s)\widehat{D}(s)$ are entire functions and:*

$$\Gamma(s)\widehat{D}(s) = \Gamma(1-s)\widehat{\mathcal{H}(D)}(1-s) \tag{48}$$

We first establish:

Theorem 5 (“co-Poisson intertwining”)**.** *Let D be a tempered distribution supported in $[0, +\infty)$ and let g be an integrable function with compact support in $(0, \infty)$. Then, with $(Ig)(t) = \frac{g(1/t)}{t}$:*

$$\mathcal{H}(g * D) = (Ig) * \mathcal{H}(D) \tag{49}$$

[10] if D is near the origin a function with an analytic character, then straightforward elementary arguments allow a complementary discussion. However if D is just an element of $L^2(0, \infty; dx)$ then $\widehat{D}$ is a square-integrable function on the critical line, and nothing more nor less.

Let us first suppose that D is an L^2 function. In that case, we will use the Mellin-Plancherel transform $f \mapsto \widehat{f}(s) = \int_0^\infty f(t)t^{-s}\,dt$, for f square integrable and $\Re(s) = \frac{1}{2}$. Then $\widehat{g * f}$ is, changing variables, the Fourier transform of an additive convolution where one of the two has compact support, well known to be the product $\widehat{g}\cdot\widehat{f}$. We need also to understand the Mellin transform of $\mathcal{H}(f)$. Let us suppose $f_t(x) = \exp(-tx)$. Then $\mathcal{H}(f_t) = \frac{1}{t} f_{\frac{1}{t}}$ has Mellin transform $\widehat{\mathcal{H}(f_t)}(s) = t^{-s}\Gamma(1-s)$ and $\widehat{f_t}(s) = t^{s-1}\Gamma(1-s)$, so we have the identity for such f's:

$$\widehat{\mathcal{H}(f)}(s) = \frac{\Gamma(1-s)}{\Gamma(s)}\widehat{f}(1-s) \tag{50}$$

The linear combinations of the f_t's are dense in L^2, so (49) holds for all f's as an identity of square integrable functions on the critical line. We are now in a position to check the intertwining: $\widehat{\mathcal{H}(g * f)}(s) = \frac{\Gamma(1-s)}{\Gamma(s)}\widehat{g}(1-s)\widehat{f}(1-s) = \widehat{Ig}(s)\widehat{\mathcal{H}(f)}(s) = \widehat{Ig * \mathcal{H}(f)}(s)$.

For the case of an arbitrary distribution it will then be sufficient to check that if (49) holds for D it holds for xD and for D'. This is easily done using (44). We have $g * (D') = (xg * D)'$, so $\mathcal{H}(g * D') = \int_0^x \mathcal{H}(xg * D) = \int_0^x (\frac{Ig}{x} * \mathcal{H}(D)) = Ig * (\int_0^x \mathcal{H}(D)) = Ig * \mathcal{H}(D')$. A similar proof is done for xD. This completes the proof of the intertwining.

The theorem 4 is then established as is [8, Thm 4.30]. We pick an arbitrary g smooth with compact support in $(0, \infty)$. We know by theorem 3 that $g * D$ is a Schwartz function as $x \to +\infty$, and certainly it vanishes identically in a neighborhood of the origin, so $\widehat{g * D}(s) = \widehat{g}(s)\widehat{D}(s)$ is an entire function. So $\widehat{D}(s)$ is a meromorphic function in the entire complex plane, in fact an entire function as g is arbitrary. We then use the intertwining and (50) for square integrable functions. This gives $\widehat{g}(1-s)\widehat{\mathcal{H}(D)}(s) = \widehat{Ig * \mathcal{H}(D)}(s) = \widehat{\mathcal{H}(g * D)}(s) = \frac{\Gamma(1-s)}{\Gamma(s)}\widehat{g}(1-s)\widehat{D}(1-s)$. Hence, indeed, after replacing s by $1-s$:

$$\Gamma(s)\widehat{D}(s) = \Gamma(1-s)\widehat{\mathcal{H}(D)}(1-s) \tag{51}$$

The left-hand side may have poles only at $0, -1, \ldots$, and the right-hand side only at $1, 2, \ldots$. So both sides are entire functions and $\widehat{D}(s)$ has trivial zeros at $0, -1, -2, \ldots$

We now give another proof of Theorem 4, which is more classical, as it is the descendant of the second of Riemann's proof, and is the familiar one from the theory of theory of L-series and modular functions. The existence of two complementary proofs is instructive, as it helps to better understand the rôle of the right Mellin transform $\int_0^\infty f(x)x^{-s}\,dx$ vs. the left Mellin transform $\int_0^\infty \theta(it)t^{s-1}\,dt$.

To the distribution D we associate its "theta" function[11] $\theta_D(\lambda) = \widetilde{D}(\lambda) = \int_0^\infty e^{i\lambda x}D(x)\,dx$, which is an analytic function for $\Im(\lambda) > 0$ [12]. Right from the beginning we have:

$$\theta_{\mathcal{H}(D)}(it) = \frac{1}{t}\theta_D(\frac{i}{t}) \tag{52}$$

If the leftmost point of the support of D is positive then $\theta_D(it)$ has exponential decrease as $t \to +\infty$ and $\int_1^\infty \theta_D(it)t^{s-1}\,dt$ is an entire function. If also the leftmost point of support of $\mathcal{H}(D)$ is positive then $\theta_{\mathcal{H}(D)}(it)$ has exponential decrease as $t \to +\infty$ and $\int_0^1 \theta_D(it)t^{s-1}\,dt = \int_1^\infty \theta_{\mathcal{H}(D)}(it)t^{-s}\,dt$ is an entire function. So, under the support property considered in Theorem 4 $\mathcal{D}(s) := \int_0^\infty \theta_D(it)t^{s-1}\,dt$ is indeed an entire function, and the functional equation is

$$\mathcal{D}(s) = \mathcal{D}^*(1-s) \tag{53}$$

[11] the author hopes to be forgiven this temporary terminology in a situation where only the behavior under $\lambda \mapsto \frac{-1}{\lambda}$ is at work.

[12] we adopt the usual notation, and consider θ_D as a function of it rather than t.

with $\mathcal{D}^*(s) = \int_0^\infty \theta_{\mathcal{H}(D)}(t)t^{s-1}\,dt$.

To conclude we also need to establish:

$$\mathcal{D}(s) = \Gamma(s)\widehat{D}(s) \tag{54}$$

We shall prove this for $\Re(s) \gg 0$ under the hypothesis that D has support in $[a, +\infty)$, $a > 0$ (no hypothesis on $\mathcal{H}(D)$). In that case, as $\theta_D(it)$ is $O(t^{-N})$ for a certain N as $t \to 0$ ($t > 0$), and is of exponential decrease as $t \to +\infty$, we can define $\mathcal{D}(s) = \int_0^\infty \theta_D(it)t^{s-1}\,dt$ as an analytic function for $\Re(s) \gg 0$. Let us suppose that D is a continuous function which is $O(x^{-2})$ as $x \to +\infty$. Then, for, $\Re(s) > 0$, the identity (54) holds as an application of the Fubini theorem. We then apply our usual method to check that if (54) holds for D it also holds for xD and for D'. For this, obviously we need things such as $\widehat{D'}(s) = s\widehat{D}(s+1)$ [8, 4.15] and $\widehat{xD}(s) = \widehat{D}(s-1)$, the formulas $\theta_{D'} = -i\lambda\theta_D$, $\theta_{xD} = -i\frac{\partial}{\partial\lambda}\theta_D$, and $\Gamma(s+1) = s\Gamma(s)$. The verifications are then straightforward.

In summary we have seen how the support property for D and $\mathcal{H}(D)$ is related in two complementary manners to the functional equation, one using the right Mellin transform $\widehat{D}(s)$ of D and the idea of co-Poisson, the other using the left Mellin transform $\mathcal{D}(s)$ of the "theta" function θ_D associated to D as an analytic function on the upper half-plane and the behavior of $\theta_D(it)$ under $t \mapsto \frac{1}{t}$. It is possible to push further the analysis and to characterize the class of entire functions $\mathcal{D}(s) = \Gamma(s)\widehat{D}(s)$, as has been done in [8] in the case of the cosine and sine transforms. It is also explained in [8] how the discussion extends to allow finitely many poles. The proofs and statements given there are easily adapted to the case of the $\mathcal{H}$ transform. Only the case of poles at 1 and 0 will be needed here and this corresponds, either to the condition that D and $\mathcal{H}(D)$ both restrict in $(-a, a)$ for some $a > 0$ to multiples of the Dirac delta function, or, that they are both constant in $[0, a)$ for some $a > 0$. We recall that the Mellin transform $\widehat{D}(s)$ is defined in such a manner, that it is not affected from either substracting δ or $\mathbf{1}_{x>0}$ from D.

4 A group of distributions and related integral formulas

We now derive some integral identities which will prove central. The identities will be re-obtained later as the outcome of a less direct path. We are interested in the tempered distribution $g_a(x)$ whose Fourier transform is $\exp(ia\frac{-1}{\lambda})$. Indeed $\tau_a^\#(f)$ (equation (39a)) is the additive convolution of f with g_a: we note that g_a differs from $\delta(x)$ by a square integrable function as $1 - \exp(-ia\lambda^{-1}) = O_{|\lambda|\to\infty}(|\lambda|^{-1})$; so there is a convolution formula $\tau_a^\#(f) = f - f_a * f$ for a certain square integrable function f_a. For $f \in L^2$, the convolution $f_a * f$ as the Fourier transform of an L^1-function is continuous on $\mathbb{R}$. Starting from the identity $\exp(ia\frac{-1}{\lambda}) = -i\lambda\frac{i}{\lambda}\exp(ia\frac{-1}{\lambda})$ we identify g_a for $a \geq 0$ as $\frac{\partial}{\partial x}\mathcal{H}\delta_a$. It is important that $\frac{\partial}{\partial x}$ is taken in the distribution sense. So we have, simply:

$$g_a(x) = \delta(x) - \frac{aJ_1(2\sqrt{ax})}{\sqrt{ax}}\mathbf{1}_{x>0}(x) \qquad (a \geq 0) \tag{55}$$

If $a < 0$ then $g_a(x) = g_{-a}(-x)$, $f_a(x) = f_{-a}(-x)$. So:

$$g_{-a}(x) = \delta(x) - \frac{aJ_1(2\sqrt{-ax})}{\sqrt{-ax}}\mathbf{1}_{x<0}(x) \qquad (-a \leq 0) \tag{56}$$

The group property under the additive convolution $g_a * g_b = g_{a+b}$ leads to remarkable integral identities $f_{a+b} = f_a + f_b - f_a * f_b$ involving the Bessel functions. The pointwise validity is guaranteed by continuity; the Plancherel identity confirms the identity, where $f_a(x) = \frac{aJ_1(2\sqrt{ax})}{\sqrt{ax}}\mathbf{1}_{x>0}(x)$ for $a \geq 0$ and $f_{-a}(x) = f_a(-x)$:

$$f_{a+b} = f_a + f_b - f_a * f_b \tag{57}$$

At $x = 0$ the pointwise identity is obtained by continuity from either $x > 0$ or $x < 0$. We have essentially two cases: $g_a * g_b$ for $a, b \geq 0$ and $g_a * g_{-b}$ for $a \geq b \geq 0$. The following is obtained:

Proposition 6. *Let $a \geq b \geq 0$ and $x \geq 0$. There holds:*

$$\frac{(a+b)J_1(2\sqrt{(a+b)x})}{\sqrt{(a+b)x}} = \frac{aJ_1(2\sqrt{ax})}{\sqrt{ax}} + \frac{bJ_1(2\sqrt{bx})}{\sqrt{bx}} - \int_0^x \frac{aJ_1(2\sqrt{ay})}{\sqrt{ay}}\ \frac{bJ_1(2\sqrt{b(x-y)})}{\sqrt{b(x-y)}}\,dy \quad \text{(58a)}$$

$$\frac{(a-b)J_1(2\sqrt{(a-b)x})}{\sqrt{(a-b)x}} = \frac{aJ_1(2\sqrt{ax})}{\sqrt{ax}} - \int_x^\infty \frac{aJ_1(2\sqrt{ay})}{\sqrt{ay}}\ \frac{bJ_1(2\sqrt{b(y-x)})}{\sqrt{b(y-x)}}\,dy \quad \text{(58b)}$$

$$0 = \frac{bJ_1(2\sqrt{bx})}{\sqrt{bx}} - \int_0^\infty \frac{aJ_1(2\sqrt{ay})}{\sqrt{ay}}\ \frac{bJ_1(2\sqrt{b(y+x)})}{\sqrt{b(y+x)}}\,dy \quad \text{(58c)}$$

Exchanging a and b and changing variables we combine (58b) and (58c) into one single equation for $x \geq 0$ and $a, b \geq 0$:

$$\frac{(a-b)J_1(2\sqrt{(a-b)x})}{\sqrt{(a-b)x}}\mathbf{1}_{a-b\geq 0}(a-b) = \frac{aJ_1(2\sqrt{ax})}{\sqrt{ax}} - \int_0^\infty \frac{aJ_1(2\sqrt{a(y+x)})}{\sqrt{a(y+x)}}\ \frac{bJ_1(2\sqrt{by})}{\sqrt{by}}\,dy \quad (59)$$

The formula for $x = 0$ in (59) is obtained by continuity. It is equivalent to

$$\int_0^\infty J_1(u)J_1(cu)\,\frac{du}{u} = \frac{1}{2}\min(c, \frac{1}{c}) \qquad (c > 0) \quad (60)$$

which is a very special case of formulas of Weber, Sonine and Schafheitlin ([33, 13.42.(1)]). Another interesting special case of (59) is for $a = b$. The formula becomes

$$\frac{J_1(2\sqrt{x})}{\sqrt{x}} = \int_0^\infty \frac{J_1(2\sqrt{y})}{\sqrt{y}}\,\frac{J_1(2\sqrt{x+y})}{\sqrt{x+y}}\,dy \quad (61)$$

which is equivalent to a special case of a formula of Sonine ([33, 13.48.(12)]).

We already mentioned the equation $\frac{\partial^2}{\partial u\partial v}J_0(2\sqrt{uv}) = -J_0(2\sqrt{uv})$. New identities are obtained from (59) or (58a) after taking either the a or the b derivative. We investigate no further (59) as the corresponding semi-convergent integrals, in a form or another, are certainly among the formulas of [33, §13]. Let us rather focus more closely on the case $a, b \geq 0$ ((58a).) We have a function which is entire in a, b, and x and the identity holds for all complex values of a, b, and x. Let us take the derivative with respect to a:

$$J_0(2\sqrt{(a+b)x}) = J_0(2\sqrt{ax}) - \int_0^x J_0(2\sqrt{ay})\,\frac{bJ_1(2\sqrt{b(x-y)})}{\sqrt{b(x-y)}}\,dy \quad (62)$$

We replace b by $-b$ and then set $x = b$. This gives:

$$I_0(2\sqrt{b(b-a)}) = J_0(2\sqrt{ba}) + \int_0^b J_0(2\sqrt{ay})\,\frac{bI_1(2\sqrt{b(b-y)})}{\sqrt{b(b-y)}}\,dy \quad (63)$$

We take the derivative of (62) with respect to b:

$$-\frac{xJ_1(2\sqrt{(a+b)x})}{\sqrt{(a+b)x}} = -\int_0^x J_0(2\sqrt{ay})J_0(2\sqrt{b(x-y)})\,dy \quad (64)$$

Then we replace b by $-b$ and set $x=b$:

$$\frac{bI_1(2\sqrt{b(b-a)})}{\sqrt{b(b-a)}} = \int_0^b J_0(2\sqrt{ay})I_0(2\sqrt{b(b-y)})\,dy \tag{65}$$

Combining (63) and (65) by addition and substraction we discover that we have solved certain integral equations:

$$\phi_b^+(x) = I_0(2\sqrt{b(b-x)}) - \frac{bI_1(2\sqrt{b(b-x)})}{\sqrt{b(b-x)}} = (1+\frac{\partial}{\partial x})I_0(2\sqrt{b(b-x)}) \tag{66a}$$

$$\phi_b^-(x) = I_0(2\sqrt{b(b-x)}) + \frac{bI_1(2\sqrt{b(b-x)})}{\sqrt{b(b-x)}} = (1-\frac{\partial}{\partial x})I_0(2\sqrt{b(b-x)}) \tag{66b}$$

$$\phi_b^+(x) + \int_0^b J_0(2\sqrt{xy})\phi_b^+(y)\,dy = J_0(2\sqrt{bx}) \tag{66c}$$

$$\phi_b^-(x) - \int_0^b J_0(2\sqrt{xy})\phi_b^-(y)\,dy = J_0(2\sqrt{bx}) \tag{66d}$$

The significance will appear later in the paper and we leave the matter here. The method was devised after the importance of solving equations (66c) and (66d) had emerged and after the solutions (66a) and (66b) had been obtained as the outcome of a more indirect path. Of course, direct verification by replacement of the Bessel functions by their series expansions is possible and easy.

5 Orthogonal projections and Hilbert space evaluators

Let $a>0$ and let P_a be the orthogonal projection on $L^2(0,a;dx)$ and $Q_a = \mathcal{H}P_a\mathcal{H}$ the orthogonal projection on $\mathcal{H}(L^2(0,a;dx))$ and let $K_a \subset L^2(0,\infty;dx)$ be the Hilbert space of square integrable functions f such that both f and $\mathcal{H}(f)$ have their supports in $[a,\infty)$. Also we shall write $H_a = P_a\mathcal{H}P_a$. Also we shall very often use $D_a = H_a^2 = P_a\mathcal{H}P_a\mathcal{H}P_a$. Using:

$$J_0(2\sqrt{xy}) = \sum_{n=0}^{\infty}(-1)^n\frac{x^ny^n}{n!^2}\ , \tag{67}$$

we exhibit $H_a = P_a\mathcal{H}P_a$ as a limit in operator norm of finite rank operators so $P_a\mathcal{H}P_a$ is a compact (self-adjoint) operator. It is not possible for a non zero $f\in L^2(0,a;dx)$ to be such that $\|H_a(f)\| = \|f\|$, as this would imply that $H_a(f)$ vanishes identically for $x>a$, but $H_a(f)$ is an entire function. So the operator norm of H_a is strictly less than one, and $1\pm H_a$ as well as $1-D_a$ are invertible. We consider the equation

$$\phi = u + \mathcal{H}(v) \qquad u,v \in L^2(0,a;dx) \tag{68}$$

Hence:

$$u + H_a(v) = P_a(\phi) \tag{69a}$$

$$H_a(u) + v = P_a(\mathcal{H}(\phi)) \tag{69b}$$

$$u = (1-D_a)^{-1}(P_a(\phi) - H_aP_a\mathcal{H}(\phi)) \tag{69c}$$

$$v = (1-D_a)^{-1}(-H_aP_a(\phi) + P_a\mathcal{H}(\phi)) \tag{69d}$$

Then if $\phi_n = u_n + \mathcal{H}(v_n)$ is L^2-convergent, (u_n) and (v_n) will be convergent, and the vector space sum $L^2(0,a;dx)+\mathcal{H}(L^2(0,a;dx))$ is closed. Its elements are analytic functions for $x > a$ so certainly this is a proper subspace of L^2. Hence we obtain that each K_a is not reduced to $\{0\}$ and

$$K_a^\perp = L^2(0,a;dx) + \mathcal{H}(L^2(0,a;dx)) \tag{70}$$

We also mention that $\cup_{a>0} K_a$ is dense in but not equal to $L^2(0,\infty;dx)$, more generally that $\cup_{a>b} K_a$ is dense in but not equal to K_b, and also obviously $\cap_{a<\infty} K_a = \{0\}$, $\cap_{a<b} K_a = K_b$.

In this section $a > 0$ will be fixed (all defined quantities and functions will depend on a, but this will not always be explicitely indicated.) We shall be occupied with understanding the vectors $X_s^a \in K_a$ such that

$$\forall f \in K_a \qquad \int_a^\infty f(x) X_s^a(x)\,dx = \widehat{f}(s) = \int_a^\infty f(x) x^{-s}\,dx \tag{71}$$

and in particular we are interested in computing

$$X_a(s,z) = \int_a^\infty X_s^a(x) X_z^a(x)\,dx \tag{72}$$

As a is fixed here, we shall drop the superscript a to lighten the notation. For the time being we shall restrict to $\Re(s) > \frac{1}{2}$ and we define X_s to be the orthogonal projection to K_a of $\mathbf{1}_{x>a}(x)x^{-s}$. As a preliminary to this study we need to say a few words regarding:

$$g_s(x) := \mathcal{H}(\mathbf{1}_{x>a}(x)x^{-s}) = \int_a^\infty J_0(2\sqrt{xy})y^{-s}\,dy \tag{73}$$

The integral is absolutely convergent for $\Re(s) > \frac{3}{4}$, semi-convergent for $\Re(s) > \frac{1}{4}$, and g_s is defined by the equation as an L^2 function for $\Re(s) > \frac{1}{2}$ (it will prove to be entire in s for each $x > 0$). We need the following identity, which shows also that $g_s(x)$ is analytic in $x > 0$:

$$g_s(x) = \chi(s)x^{s-1} - \int_0^a J_0(2\sqrt{xy})y^{-s}\,dy = \chi(s)x^{s-1} - \sum_{n=0}^\infty (-1)^n \frac{x^n\, a^{n+1-s}}{n!^2(n+1-s)} \tag{74}$$

This is obtained first in the range $\frac{3}{4} < \Re(s) < 1$: $\int_a^\infty J_0(2\sqrt{xy})y^{-s}\,dy = x^{s-1}\int_{ax}^\infty J_0(2\sqrt{y})y^{-s}\,dy = x^{s-1}\left(\chi(s) - \int_0^{ax} J_0(2\sqrt{y})y^{-s}\,dy\right) = x^{s-1}\chi(s) - \int_0^a J_0(2\sqrt{xy})y^{-s}\,dy$. The poles at $s = 1$, $s = 2$, $\dots$ are only apparent. The identity is valid by analytic continuation in the entire plane $\Re(s) > \frac{1}{2}$. For each given $x > 0$ we have in fact an entire function of $s \in \mathbb{C}$. But we are here more interested in g_s as a function of x and we indeed see that it is analytic in $\mathbb{C}\setminus]-\infty, 0]$ (it is an entire function of x if $s \in -\mathbb{N}$). [13]

There are unique vectors u_s, v_s in $L^2(0,a;dx)$ such that

$$\mathbf{1}_{x>a}(x)x^{-s} = X_s(x) + u_s(x) + \mathcal{H}(v_s)(x) \tag{75}$$

and they are the solutions to the system of equations:

$$u_s + H_a(v_s) = 0 \tag{76a}$$
$$H_a(u_s) + v_s = P_a(g_s) \tag{76b}$$

[13] For some other transforms $k(xy)$, such as the cosine transform, the argument must be slightly modified in order to accomodate the fact $\int_0^\infty k(y)y^{-s}\,dy$ has no range of absolute convergence.

From (76a) we see that u_s is in fact the restriction to $(0,a)$ of an entire funtion, and from (76b) that v_s is the restriction to $(0,a)$ of a function which is analytic in $\mathbb{C}\setminus]-\infty,0]$. Redefining u_s and v_s to now refer to these analytic functions their defining equations become (on $(0,+\infty)$):

$$u_s + \mathcal{H}P_a(v_s) = 0 \tag{77a}$$
$$\mathcal{H}P_a(u_s) + v_s = g_s \tag{77b}$$

and (75) becomes (we set $X_s(a) = X_s(a+)$):

$$\mathbf{1}_{x\geq a}(x)x^{-s} = X_s(x) + \mathbf{1}_{0<x<a}(x)u_s(x) + \mathcal{H}P_a(v_s)(x) \tag{78a}$$
$$\mathbf{1}_{x\geq a}(x)x^{-s} = X_s(x) - \mathbf{1}_{x\geq a}(x)u_s(x) \tag{78b}$$
$$X_s(x) = \mathbf{1}_{x\geq a}(x)(x^{-s} + u_s(x)) \tag{78c}$$

The key to the next steps will be the idea to investigate the distribution $(x\frac{d}{dx} + s)X_s$ on the (positive) real line. Let D_s be $x\frac{d}{dx} + s$. There holds:

$$D_s\mathcal{H} = -\mathcal{H}D_{1-s} \tag{79}$$

To compute $\frac{d}{dx}P_a(v_s)$ we first suppose $\Re(s) > 1$, so (we know the behavior as $x \to 0$ from (74)) $\frac{d}{dx}P_a(v_s) = P_a(v_s') - v_s(a)\delta_a(x)$ and $x\frac{d}{dx}P_a(v_s) = P_a(xv_s') - av_s(a)\delta_a(x)$. This remains true for $\Re(s) > \frac{1}{2}$. Applying D_s to (77a) thus gives $D_s(u_s) - \mathcal{H}\left(P_aD_{1-s}(v_s) - av_s(a)\delta_a(x)\right)$. We similarly apply D_{1-s} to (77b) and obtain the following system:

$$D_s(u_s)(x) - (\mathcal{H}P_aD_{1-s}v_s)(x) = -av_s(a)J_0(2\sqrt{ax}) \tag{80a}$$
$$-(\mathcal{H}P_aD_su_s)(x) + D_{1-s}(v_s)(x) = (D_{1-s}g_s)(x) - au_s(a)J_0(2\sqrt{ax}) \tag{80b}$$

From (73), we have $D_{1-s}g_s = -\mathcal{H}D_s(\mathbf{1}_{x>a}x^{-s}) = -\mathcal{H}(a^{1-s}\delta_a(x)) = -a^{1-s}J_0(2\sqrt{ax})$. Let us define

$$J_0^a(x) = J_0(2\sqrt{ax}) \tag{81}$$

We have proven:

$$+D_su_s - \mathcal{H}P_aD_{1-s}v_s = -av_s(a)J_0^a \tag{82a}$$
$$-\mathcal{H}P_aD_su_s + D_{1-s}v_s = -a(a^{-s} + u_s(a))J_0^a \tag{82b}$$

Restricting to the interval $(0,a)$ and solving, we find:

$$P_aD_su_s = -a(1-D_a)^{-1}(v_s(a)J_0^a + (a^{-s} + u_s(a))H_aJ_0^a) \tag{83a}$$
$$P_aD_{1-s}v_s = -a(1-D_a)^{-1}((a^{-s} + u_s(a))J_0^a + v_s(a)H_aJ_0^a) \tag{83b}$$

It is advantageous at this stage to define ϕ_a^+ and ϕ_a^- to be the solutions of the equations (in $L^2(0,a;dx)$):

$$\phi_a^+ + H_a\phi_a^+ = J_0^a \tag{84a}$$
$$\phi_a^- - H_a\phi_a^- = J_0^a \tag{84b}$$

We already know from (66a) and (66b) exactly what ϕ_a^+ and ϕ_a^- are (in this special case of the $\mathcal{H}$ transform), but we shall proceed as if we didn't. We see from (84a), (84b) that ϕ_a^+ and ϕ_a^- are entire functions, and we can rewrite the system as:[14]

$$\phi_a^+ + \mathcal{H}P_a\phi_a^+ = J_0^a \tag{85a}$$
$$\phi_a^- - \mathcal{H}P_a\phi_a^- = J_0^a \tag{85b}$$

[14] in conformity with our conventions, these are identities on $(0,\infty)$; to see them as identities on $\mathbb{C}$ one must read $\int_0^a J_0(2\sqrt{xy})\phi_a^+(y)\,dy$ rather than $(\mathcal{H}P_a\phi_a^+)(x)$.

We observe the identities:

$$(1-D_a)^{-1}J_0^a = P_a\frac{\phi_a^+ + \phi_a^-}{2} \tag{86a}$$

$$(1-D_a)^{-1}H_aJ_0^a = P_a\frac{-\phi_a^+ + \phi_a^-}{2} \tag{86b}$$

So (83a) and (83b) become

$$D_su_s = +a\frac{a^{-s}+u_s(a)-v_s(a)}{2}\phi_a^+ - a\frac{a^{-s}+u_s(a)+v_s(a)}{2}\phi_a^- \tag{87a}$$

$$D_{1-s}v_s = -a\frac{a^{-s}+u_s(a)-v_s(a)}{2}\phi_a^+ - a\frac{a^{-s}+u_s(a)+v_s(a)}{2}\phi_a^- \tag{87b}$$

From (87a) we compute successively (again, these are identities on $(0,+\infty)$):

$$\mathcal{H}P_aD_su_s = a\frac{a^{-s}+u_s(a)-v_s(a)}{2}(J_0^a-\phi_a^+) - a\frac{a^{-s}+u_s(a)+v_s(a)}{2}(-J_0^a+\phi_a^-) \tag{88}$$

$$P_aD_su_s = a\frac{a^{-s}+u_s(a)-v_s(a)}{2}(\delta_a-\mathcal{H}\phi_a^+) - a\frac{a^{-s}+u_s(a)+v_s(a)}{2}(-\delta^a+\mathcal{H}\phi_a^-) \tag{89}$$

In (89), $\mathcal{H}\phi_a^+$ should perhaps be more precisely written as $\mathcal{H}(\phi_a^+\mathbf{1}_{x>0})$. From (85a) we know that $\phi_a^+\mathbf{1}_{x>0}$ is tempered as a distribution. From (78c) we compute $D_sX_s = \mathbf{1}_{x>a}D_s(u_s) + a(a^{-s} + u_s(a))\delta_a(x) = D_su_s - P_aD_su_s + a(a^{-s}+u_s(a))\delta_a(x)$. From (87a) and(89) then follows:

$$\begin{aligned} D_sX_s = +a\frac{a^{-s}+u_s(a)-v_s(a)}{2}(\phi_a^+ + \mathcal{H}\phi_a^+ - \delta_a) - a\frac{a^{-s}+u_s(a)+v_s(a)}{2}(\phi_a^- - \mathcal{H}\phi_a^- + \delta^a) \\ +a(a^{-s}+u_s(a))\delta_a(x) \end{aligned} \tag{90}$$

And the result of the computation is:

$$D_sX_s = +a\frac{a^{-s}+u_s(a)-v_s(a)}{2}(\phi_a^+ + \mathcal{H}\phi_a^+) + a\frac{a^{-s}+u_s(a)+v_s(a)}{2}(-\phi_a^- + \mathcal{H}\phi_a^-) \tag{91}$$

We then define the remarkable distributions:

$$A_a = \frac{\sqrt{a}}{2}(\phi_a^+ + \mathcal{H}\phi_a^+) \tag{92a}$$

$$-iB_a = \frac{\sqrt{a}}{2}(-\phi_a^- + \mathcal{H}\phi_a^-) \tag{92b}$$

$$E_a = A_a - iB_a \tag{92c}$$

From (84a) we observe that A_a has its support in $[a,\infty)$. Furthermore it is $\mathcal{H}$ invariant. Similarly, $-iB_a$, which is $\mathcal{H}$ anti invariant, also has its support in $[a,+\infty)$. We recover A_a and $-iB_a$ from E_a through taking the invariant and anti-invariant parts. We may also rewrite D_sX_s as:

$$D_sX_s = \sqrt{a}(a^{-s}+u_s(a))E_a - \sqrt{a}\,v_s(a)\mathcal{H}E_a \tag{93}$$

Some other manners of writing A_a and $-iB_a$ are useful: from (85a) $\mathcal{H}\phi_a^+ = \delta_a - P_a\phi_a^+$ and from (85b) $\mathcal{H}\phi_a^- = \delta_a + P_a\phi_a^-$, so:

$$A_a = \frac{\sqrt{a}}{2}(\delta_a + \phi_a^+\mathbf{1}_{x>a}) \tag{94a}$$

$$-iB_a = \frac{\sqrt{a}}{2}(\delta_a - \phi_a^-\mathbf{1}_{x>a}) \tag{94b}$$

And, we take also notice of the following definitions and identities:

$$j_a = \sqrt{a}(\delta_a - \phi_a^+ \mathbf{1}_{0<x<a}) \qquad j_a = \sqrt{a}\mathcal{H}\phi_a^+ \qquad A_a = \frac{1}{2}(j_a + \mathcal{H}j_a) \tag{95a}$$

$$-ik_a = \sqrt{a}(\delta_a + \phi_a^- \mathbf{1}_{0<x<a}) \qquad -ik_a = \sqrt{a}\mathcal{H}\phi_a^- \qquad B_a = \frac{1}{2}(k_a - \mathcal{H}k_a) \tag{95b}$$

From (85a) and (85b) we know that ϕ_a^+ and ϕ_a^- are bounded, so the right Mellin transforms are defined directly for $\Re(s) > 1$ by:[15]

$$\widehat{A_a}(s) = \frac{\sqrt{a}}{2}\left(a^{-s} + \int_a^\infty \phi_a^+(x)x^{-s}\,dx\right) \tag{96a}$$

$$-i\widehat{B_a}(s) = \frac{\sqrt{a}}{2}\left(a^{-s} - \int_a^\infty \phi_a^-(x)x^{-s}\,dx\right) \tag{96b}$$

$$\widehat{E_a}(s) = \sqrt{a}\left(a^{-s} + \frac{1}{2}\int_a^\infty (\phi_a^+(x) - \phi_a^-(x))x^{-s}\,dx\right) \tag{96c}$$

$$\widehat{\mathcal{H}(E_a)}(s) = \sqrt{a}\,\frac{1}{2}\int_a^\infty (\phi_a^+(x) + \phi_a^-(x))x^{-s}\,dx \tag{96d}$$

From (85a) and (85b) we know that $\phi_a^+ - \phi_a^-$ is square-integrable at $+\infty$, so, using $\mathcal{H}\phi_a^+ = \delta_a - P_a\phi_a^+$ and $\mathcal{H}\phi_a^- = \delta_a + P_a\phi_a^-$ we compute:

$$\int_a^\infty (\phi_a^+(x) - \phi_a^-(x))x^{-s}\,dx = \int_0^\infty (\mathcal{H}\phi_a^+ - \mathcal{H}\phi_a^-)g_s(x)\,dx = -\int_0^a (\phi_a^+(x) + \phi_a^-(x))g_s(x)\,dx \tag{97}$$

Then using (86a):

$$\int_0^a \frac{\phi_a^+(x) + \phi_a^-(x)}{2}g_s(x)\,dx = \int_0^a (1-D_a)^{-1}(J_0^a)(x)g_s(x)\,dx = \int_0^a J_0^a(x)((1-D_a)^{-1}(g_s))(x)\,dx \tag{98}$$

Comparing with (76a) and (76b) the right-most term of (98) may be written as $\int_0^a J_0(2\sqrt{ax})v_s(x)\,dx$ which in turn we recognize from (77a) to be $-u_s(a)$. We have thus proven the identity:

$$\widehat{E_a}(s) = \sqrt{a}(a^{-s} + u_s(a)) \tag{99}$$

In a similar manner we have:

$$\int_a^\infty \frac{\phi_a^+(x) + \phi_a^-(x)}{2}x^{-s}\,dx = \int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx + \int_0^a \frac{-\phi_a^+(x) + \phi_a^-(x)}{2}g_s(x)\,dx \tag{100}$$

$$\begin{aligned}\int_0^a \frac{-\phi_a^+(x) + \phi_a^-(x)}{2}g_s(x)\,dx &= \int_0^a ((1-D_a)^{-1}H_aJ_0^a)(x)g_s(x)\,dx = -\int_0^a J_0(2\sqrt{ax})u_s(x)\,dx \\ &= v_s(a) - g_s(a) = v_s(a) - \int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx\end{aligned} \tag{101}$$

[15] the integral for $\widehat{E_a}(s)$ is certainly absolutely convergent for $\Re(s) > \frac{1}{2}$ as $\phi_a^+ - \phi_a^-$ is square integrable on $(0,\infty)$, and in fact it is absolutely convergent for $\Re(s) > \frac{1}{4}$. As we know already completely explicitely ϕ_a^+ and ϕ_a^-, we do not pause on this here. A general argument suitable to establish in more general cases absolute convergence for $\Re(s) > \sigma$ for some $\sigma < \frac{1}{2}$ will be given later.

$$\widehat{A_a}(s) + i\widehat{B_a}(s) = \widehat{\mathcal{H}(E_a)}(s) = \sqrt{a}\int_a^\infty \frac{\phi_a^+(x)+\phi_a^-(x)}{2} x^{-s}\,dx = \sqrt{a}v_s(a) \tag{102}$$

Then, we obtain the reformulation of (93) as:

$$D_s X_s = \widehat{E_a}(s)E_a - \widehat{\mathcal{H}(E_a)}(s)\mathcal{H}E_a \tag{103}$$

And, noting $\widehat{D_sX_s}(z) = (s+z-1)\widehat{X_s}(z) = (s+z-1)\int_a^\infty X_s(x)X_z(x)\,dx$ we are finally led to the remarkable result:

$$X_a(s,z) = \int_a^\infty X_s^a(x)X_z^a(x)\,dx = \frac{\widehat{E_a}(s)\widehat{E_a}(z) - \widehat{\mathcal{H}(E_a)}(s)\widehat{\mathcal{H}E_a}(z)}{s+z-1} \tag{104}$$

This equation has been proven under the assumption $\Re(s) > 1$, and $\Re(z) > \frac{1}{2}$. To complete the discussion we need to know that the evaluators $f \mapsto \widehat{f}(s)$, $s \in \mathbb{C}$ are indeed continuous linear forms on K_a. For $\Re(s) > \frac{1}{2}$, we have $\widehat{f}(s) = \int_a^\infty f(x)x^{-s}\,dx$. For $\Re(s) < \frac{1}{2}$ we have $\widehat{f}(s) = \frac{\Gamma(1-s)}{\Gamma(s)}\widehat{\mathcal{H}(f)}(1-s)$. For $\Re(s) = \frac{1}{2}$ continuity follows by the Banach-Steinhaus theorem, and of course more elementary proofs exist (as in [3] for the cosine or sine transform). So we do have unique Hilbert space vectors $X_s^a \in K_a$ such that $\forall f \in K_a \forall s \in \mathbb{C}$ $\widehat{f}(s) = \int_a^\infty X_s^a(x)f(x)\,dx$. Then (104) holds throughout $\mathbb{C}\times\mathbb{C}$ by analytic continuation.

The vectors X_s^a are zero for $s \in -\mathbb{N}$, and it is more precise to use vectors $\mathcal{X}_s^a = \Gamma(s)X_s^a$ which are non-zero for all $s \in \mathbb{C}$. These vectors are the evaluators[16] for $f \mapsto \mathcal{F}(s)$, $\mathcal{F}(s) = \Gamma(s)\widehat{f}(s)$. We recapitulate some of the results in the following theorem, whose analog for the cosine (or sine) transform was given in [5] (up to changes of variables and notations, the first paragraph as well as equation (108) are theorems from [1]; the equations (105), (106), (107) are our contributions. In this specific case of $\mathcal{H}$ we shall later identify exactly ϕ_a^+ and ϕ_a^- and E_a and $\mathcal{E}_a$. As we shall explain the analog of the $\mathcal{E}_a$-function in [1] has value 1 at $s = \frac{1}{2}$, and is not identical with the $\mathcal{E}_a$ here):

Theorem 7. *For a given $a > 0$ let K_a be the Hilbert space of square integrable functions $f(x)$ on $[a,+\infty)$ whose $\mathcal{H}$-transforms $\int_0^\infty J_0(2\sqrt{xy})f(y)\,dy$ (in the L^2-sense) again vanish for $0 < x < a$. The completed right Mellin transforms $\Gamma(s)\widehat{f}(s) = \Gamma(s)\int_a^\infty f(x)x^{-s}\,dx$ are entire functions and evaluations at $s \in \mathbb{C}$ are continuous linear forms.*
Let $\mathcal{X}_s^a$ for each $s \in \mathbb{C}$ be the unique vector in K_a such that $\forall f \in K_a$ $\Gamma(s)\widehat{f}(s) = \int_a^\infty f(x)\mathcal{X}_s^a(x)\,dx$.
Let ϕ_a^+ and ϕ_a^- be the entire functions which are the solutions to:

$$\phi_a^+(x) + \int_0^a J_0(2\sqrt{xy})\phi_a^+(y)\,dy = J_0(2\sqrt{ax}) \tag{105}$$

$$\phi_a^-(x) - \int_0^a J_0(2\sqrt{xy})\phi_a^-(y)\,dy = J_0(2\sqrt{ax}) \tag{106}$$

Then

$$\widehat{E_a}(s) = \sqrt{a}\Big(a^{-s} + \frac{1}{2}\int_a^\infty (\phi_a^+(x) - \phi_a^-(x))x^{-s}\,dx\Big) \tag{107}$$

is an entire function with trivial zeros at $-\mathbb{N}$ and, defining $\mathcal{E}_a(s) = \Gamma(s)\widehat{E_a}(s)$, we have:

$$\forall s,z \in \mathbb{C} \quad \int_a^\infty \mathcal{X}_s^a(x)\mathcal{X}_z^a(x)\,dx = \frac{\mathcal{E}_a(s)\mathcal{E}_a(z) - \mathcal{E}_a(1-s)\mathcal{E}_a(1-z)}{s+z-1} \tag{108}$$

[16] evaluators for the “euclidean” product $\int fg\,dx$, not the “hilbertian” $\int f\overline{g}\,dx$.

We knew in advance that we had to end up with a formula such as (108) (with a $\mathcal{E}$ function to be discovered[17]), and this is why we started investigating $(x\frac{d}{dx}+s)X_s^a(x)$ in the first place! The reason is this: the Hilbert space of the entire functions $\mathcal{F}(s)$, $f\in K_a$ (the Hilbert structure is the one from K_a, or $(\mathcal{F}_1,\mathcal{F}_2)=\frac{1}{2\pi}\int_{\Re(s)=\frac12}\mathcal{F}_1(s)\overline{\mathcal{F}_2(s)}\frac{|ds|}{|\Gamma(s)|^2}$) verifies the de Branges axioms [2], up to the change of variable $s=\frac12-iz$. Let us recall the axioms of [2] for a (non-zero) Hilbert space of entire functions $F(z)$:

(H1) for each z, evalution at z is a continuous linear form,

(H2) for each F, $z\mapsto\overline{F(\overline{z})}$ belongs to the Hilbert space and has the same norm as F,

(H3) if $F(w)=0$ then $G(z)=\frac{z-\overline{w}}{z-w}F(z)$ belongs to the space and has the same norm as F

Let $K(z,w)$ be defined as the evaluator at z: $\forall F\ F(z)=(F,K(z,\cdot))$. It is anti-analytic in z and analytic in w (the scalar product is complex linear in its first entry, and conjugate linear in its second entry). It is a reproducing kernel: $K(z,w)=(K(z,\cdot),K(w,\cdot))$. It is proven in [2] that (H1), (H2), (H3) entail the existence of an entire function $E(z)$ with $|E(z)|>|E(\overline{z})|$ for $\Im(z)>0$, such that the space is exactly the set of entire functions $F(z)$ such that both $\frac{F(z)}{E(z)}$ and $\frac{\overline{F(\overline{z})}}{E(z)}$ belong to $\mathbb{H}^2(\Im(z)>0)$, and the Hilbert space norm of F is $\frac{1}{2\pi}\int_{\mathbb{R}}|F(t)|^2\frac{dt}{|E(t)|^2}$.[18] We have incorporated a 2π for easier comparison with our conventions. Then the reproducing kernel is expressed as:

$$K(z,w)=\frac{\overline{E(z)}E(w)-E(\overline{z})\overline{E(\overline{w})}}{i(\overline{z}-w)}\tag{109}$$

The function E is not unique; if the space has the isometric symmetry $F(z)\mapsto F(-z)$, a function E exists which is real on the imaginary axis and writing $E=A-iB$ where A and B are real on the real axis, the pair (A,B) is unique up to $A\mapsto kA$, $B\mapsto k^{-1}B$, A is even and B is odd. If $A(0)\neq 0$ (this happens exactly when the space contains at least one element not vanishing at 0) then it may be uniquely normalized so that $A(0)=1$. Then E is uniquely determined.

Model examples are the Paley-Wiener spaces of entire functions $F(z)$ of exponential type at most τ with $\|F\|^2=\frac{1}{2\pi}\int_{\mathbb{R}}|F(t)|^2\,dt<\infty$. Then $E(z)=e^{-i\tau z}$ is a possible E function. The Paley-Wiener spaces are related to the study of the differential operator $-\frac{d^2}{dx^2}$ on the positive half-line, and an important class of spaces verifying the axioms of [2] is associated with the theory of the eigenfunction expansions for Schrödinger operators $-\frac{d^2}{dx^2}+V(x)$ ([27]). In these examples the spaces are indexed by a parameter τ (the Schrödinger operator is first studied on a finite interval $(0,\tau)$) and they are ordered by isometric inclusions (the E-function of a bigger space may be used in the computation of the norm of an element of a smaller space). Typically indeed, de Branges spaces are studied included in one fixed space $L^2(\mathbb{R},\frac{1}{2\pi}d\nu)$, are ordered by isometric inclusion and indexed by a parameter[19]. Obviously this theory is intimately related with the Weyl-Stone-Titchmarsh-Kodaira theory of the spectral measure. The articles of Dym [14] and Remling [27], the book of Dym and McKean [15], will be useful to the interested reader. In the case of the study of $\mathcal{H}$ we will have $d\nu(\gamma)=|\Gamma(\frac12+i\gamma)|^{-2}d\gamma$. It is an important flexibility of the axioms not to be limited to functions of finite exponential type, and also the spectral measures are not necessarily such that $(1+\gamma^2)^{-1}$

[17] the method was initially developed by the author for the cosine and sine transforms [5, 6] and leads for them to the only known "explicit" formulas for $\mathcal{E}$; for the zero order Hankel transform the problem of computing the reproducing kernel had been already solved by de Branges [1].

[18] the conditions on $F(z)$ are not formulated in [2] as Hardy space conditions, but they are exactly equivalent.

[19] the axioms allow for "jumps" in the isometric chain of inclusions, as occur in the theory of the Krein strings [15], discrete Schrödinger equations being special cases.

is integrable. It has turned out in our study of the spaces associated with the $\mathcal{H}$-transform that the naturally occurring E function is not the one normalized to take value 1 at $z = 0$. Rather the normalization will prove to be $\lim_{\sigma\to+\infty} \frac{-iB(i\sigma)}{A(i\sigma)} = 1$. This has an important impact on the aspect of the differential equations which will govern the deformation of the K_a's with respect to a: they will take the form of a first order linear differential system in canonical form (as generally studied in [21, §3].)

The space of the functions $\mathcal{F}(s) = \Gamma(s)\widehat{f}(s)$, $f \in K_a$ verify (this is easy) the de Branges axioms, with $s = \frac{1}{2} - iz$ and they were defined[20] in [1]. The spaces K_a have the real structure, which is manifest in the s variable through the isometry $\mathcal{F}(s) \mapsto \overline{\mathcal{F}(\overline{s})}$. Rather than with the reproducing kernel $K(z_1, z_2)$ we work mainly with $\mathcal{X}(s_1, s_2) = K(-\overline{z_1}, z_2)$ which is analytic in both variables. Of course then it is $\mathcal{X}(\overline{s}, s)$ which gives the squared norm of the evaluator at s. Writing $\mathcal{E}(s) = E(z)$ we obtain from (109):

$$\mathcal{X}(s_1, s_2) = \frac{\mathcal{E}(s_1)\mathcal{E}(s_2) - \mathcal{E}(1-s_1)\mathcal{E}(1-s_2)}{s_1 + s_2 - 1} \tag{110}$$

which is indeed what has appeared on the right hand side of (108). With $\mathcal{E}(s) = \mathcal{A}(s) - i\mathcal{B}(s)$, $\mathcal{A}$ (resp. $\mathcal{B}$) even (resp. odd) under $s \mapsto 1 - s$, this is also:

$$\mathcal{X}(s_1, s_2) = 2\frac{-i\mathcal{B}(s_1)\mathcal{A}(s_2) + \mathcal{A}(s_1)(-i\mathcal{B}(s_2))}{s_1 + s_2 - 1} \tag{111}$$

and for $\Re(s) \neq \frac{1}{2}$, $0 < \mathcal{X}(\overline{s}, s) = 2\frac{\Im(\mathcal{B}(s)\overline{\mathcal{A}(s)})}{\Re(s)-\frac{1}{2}}$ so both $\mathcal{A}$ and $\mathcal{B}$ have all their zeros on the critical line.[21]

The method in this chapter has been developed in [5, 8, 6] for the case of the cosine and sine transforms, and it leads to the currently only known "explicit" formulae[22] for the structural elements $\mathcal{E}$, $\mathcal{A}$, $\mathcal{B}$ and reproducing kernels for the spaces for the cosine and sine transforms. So far, almost nothing very specific to $\mathcal{H}$ has been used apart from it being self-adjoint self-reciprocal with an entire multiplicative kernel $k(xy)$. The next section is still of a very general validity.

As was mentioned in the Introduction the realization of the structural elements of the spaces as right Mellin transforms of distributions is a characteristic aspect of the method; the Dirac delta's in the expressions for $A_a(x)$ and $-iB_a(x)$ could have been overlooked if we had only been prepared to use functions, and the whole development was based on the computation of $(x\frac{d}{dx} + s)X_s(x)$ as a distribution. This aspect will be further reinforced in the concluding chapter of the paper (section 9) where it will be seen that the distributions $A_a(x)$ and $-iB_a(x)$ are very naturally differences of boundary values of analytic functions, so they are hyperfunctions [23] in a natural manner.

Let us consider the behavior of $\widehat{A_a}(s)$, $\widehat{B_a}(s)$, $\widehat{E_a}(s)$ and $\widehat{\mathcal{H}(E_a)}(s)$ for $\Re(s) \geq \frac{1}{2}$. Let us first look at $\widehat{E_a}(s) = \sqrt{a}\Big(a^{-s} + \frac{1}{2}\int_a^\infty (\phi_a^+(x) - \phi_a^-(x))x^{-s}\,dx\Big)$. We remark that $\phi_a^+(x) - \phi_a^-(x)$ is the $\mathcal{H}$-transform of $-(\phi_a^+(x) + \phi_a^-(x))\mathbf{1}_{0<x<a}(x)$.

Lemma 8. *Let $k(x)$ a continuous function on $[0, +\infty)$ and $A \in [0, 1]$ be such that $k_1(x) = \int_0^x k(t)\,dt = O(x^A)$ as $x \to \infty$. Let $a > 0$ and let $f(x)$ be an absolutely continuous function on $[0, a]$. Then $\int_0^a k(xy)f(y)\,dy = O(x^{A-1})$ as $x \to +\infty$.*

There exists $C < \infty$ such that $\forall x > 0$ $|k_1(x)| \leq C\,x^A$. Then $\int_0^a k(xy)f(y)\,dy = \frac{1}{x}k_1(xa)f(a) - \frac{1}{x}\int_0^a k_1(xy)f'(y)\,dy$, and $|\int_0^a k_1(xy)f'(y)\,dy| \leq Cx^A \int_0^a y^A|f'(y)|\,dy$. This was easy...

[20] in the variable z, and associated with the Hankel transform of order zero, rather than with the $\mathcal{H}$ transform.

[21] this is also seen from $2\mathcal{A}(s) = \mathcal{E}(s) + \mathcal{E}(1-s)$ as $|\mathcal{E}(s)| > |\mathcal{E}(1-s)|$ for $\Re(s) > \frac{1}{2}$. As $\mathcal{X}(\overline{s}, s) = \frac{|\mathcal{E}(s)|^2 - |\mathcal{E}(1-s)|^2}{2\Re(s)-1}$ this is in fact the same argument.

[22] as "explicit" as the Fredholm determinants of the finite Dirichlet kernels are "explicit".

With $k(x) = J_0(2\sqrt{x})$, one has $k_1(x) = \sqrt{x}J_1(2\sqrt{x}) = O(x^{\frac{1}{4}})$. We have $\phi_a^+(x) - \phi_a^-(x) = -\int_0^a J_0(2\sqrt{xy})(\phi_a^+(x) + \phi_a^-(x))\,dy$ and from Lemma 8 this is $O(x^{-\frac{3}{4}})$. So the integral in the expression for $\widehat{E_a}(s)$ is absolutely convergent for $\Re(s) > \frac{1}{4}$. In particular $\widehat{E_a}$ is bounded on the critical line. But then $\widehat{\mathcal{H}(E_a)}(s) = \chi(s)\widehat{E_a}(1-s)$ is also bounded. Hence:

Proposition 9. *The functions $\widehat{A_a}$ and $\widehat{B_a}$ are bounded on the critical line.*

Let us turn to the situation regarding $\Re(s) = \sigma \to +\infty$.

Let $f(x)$ be a function of class C^2 on $[0,a]$ and $e(x) = \int_0^a J_0(2\sqrt{xy})f(y)\,dy$. It is $O(x^{-\frac{3}{4}})$. There holds $\frac{d}{dx}xe(x) = \int_0^a(\frac{d}{dy}yJ_0(2\sqrt{xy}))f(y)\,dy = af(a)J_0(2\sqrt{ax}) - \int_0^a J_0(2\sqrt{xy})\,yf'(y)\,dy$. Let $k(x) = \int_0^a J_0(2\sqrt{xy})\,yf'(y)\,dy$. By the Lemma 8 it is $O(x^{-\frac{3}{4}})$. For $\Re(s) > \frac{3}{4}$, with absolutely convergent integrals:

$$af(a)\int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx - \int_a^\infty k(x)x^{-s}\,dx = -ae(a)a^{-s} + s\int_a^\infty e(x)x^{-s}\,dx \tag{112}$$

We show that the left hand side of (112) is $O(a^{-s}\frac{1}{s})$ for $\Re(s) > \frac{5}{4}$. We apply to k what we did for e, $\frac{d}{dx}xk(x) = a^2f'(a)J_0(2\sqrt{ax}) - \int_0^a J_0(2\sqrt{xy})\,y(yf')'(y)\,dy$. This is $O(1)$ (using $|J_0| \leq 1$). So for $\Re(s) > 1$, we can compute $\int_0^\infty(\frac{d}{dx}xk(x))x^{-s}\,dx$ by integration by parts, this gives $-ak(a)a^{-s} + s\int_a^\infty k(x)x^{-s}\,dx$. So for $\Re(s) \geq 1+\epsilon$ we have $\int_a^\infty k(x)x^{-s}\,dx = O(a^{-s}\frac{1}{s})$. Then regarding $\int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx$ we note that $\frac{d}{dx}xJ_0(2\sqrt{ax}) = J_0(2\sqrt{ax}) - \sqrt{ax}J_1(2\sqrt{ax})$, so for $\Re(s) \geq \frac{5}{4}+\epsilon$ we can apply the same method of integration by parts, and prove that $\int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx = O(a^{-s}\frac{1}{s})$. So the left hand side of (112) is indeed $O(a^{-s}\frac{1}{s})$ for $\Re(s) \geq \frac{3}{2}$ and we have:

Lemma 10. *Let $f(x)$ be a function of class C^2 on $[0,a]$ and let $e(x) = \int_0^a J_0(2\sqrt{xy})f(y)\,dy$. One has*

$$\int_a^\infty e(x)x^{-s}\,dx = \frac{a^{-s}}{s}\big(ae(a) + O(\frac{1}{s})\big) \qquad (\Re(s) \geq \frac{3}{2}) \tag{113}$$

Let us return to $\int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx = \frac{1}{s}\big(J_0(2a)a^{1-s} + \int_a^\infty(J_0(2\sqrt{ax}) - \sqrt{ax}J_1(2\sqrt{ax}))x^{-s}\,dx\big)$. We want to iterate so we also need $x\frac{d}{dx}\sqrt{ax}J_1(2\sqrt{ax}) = \frac{1}{4}2\sqrt{ax}\frac{d}{d2\sqrt{ax}}2\sqrt{ax}J_1(2\sqrt{ax}) = axJ_0(2\sqrt{ax})$. So we can integrate by parts and obtain that the last Mellin integral is $O(a^{-s}\frac{1}{s})$ for $\Re(s) \geq \frac{7}{4}+\epsilon$. So, certainly:

$$\int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx = \frac{a^{-s}}{s}\big(aJ_0(2a) + O(\frac{1}{s})\big) \qquad (\Re(s) \geq \frac{5}{2}) \tag{114}$$

Using $\phi_a^+ = J_0^a - \mathcal{H}P_a\phi_a^+$ and $\phi_a^- = J_0^a + \mathcal{H}P_a\phi_a^-$ and combining (113) and (114) we obtain:

Proposition 11. *One has for $\Re(s) \geq \sigma_0$ (here $\sigma_0 = \frac{5}{2}$ for example):*

$$\widehat{E_a}(s) = a^{\frac{1}{2}-s}(1 + \frac{a\phi^+(a) - a\phi^-(a)}{2s} + O(\frac{1}{s^2})) \qquad \widehat{A_a}(s) = \frac{\sqrt{a}}{2}a^{-s}(1 + \frac{a\phi_a^+(a)}{s} + O(\frac{1}{s^2})) \tag{115a}$$

$$\widehat{\mathcal{H}(E_a)}(s) = a^{\frac{1}{2}-s}(\frac{a\phi^+(a) + a\phi^-(a)}{2s} + O(\frac{1}{s^2})) \qquad -i\widehat{B_a}(s) = \frac{\sqrt{a}}{2}a^{-s}(1 - \frac{a\phi_a^-(a)}{s} + O(\frac{1}{s^2})) \tag{115b}$$

Theorem 12. *One has*

$$\lim_{\sigma\to+\infty}\frac{-i\mathcal{B}_a(\sigma)}{\mathcal{A}_a(\sigma)} = 1 \qquad \textit{and} \qquad \frac{\mathcal{E}_a(1-\sigma)}{\mathcal{E}_a(\sigma)} \sim_{\sigma\to+\infty} \frac{a\phi_a^+(a) + a\phi_a^-(a)}{2\sigma} \tag{116}$$

So the functions $\mathcal{A}_a$ and $\mathcal{B}_a$ are not normalized as is usually done in [2] which is to impose (when possible) to the E function to have value 1 at the origin (which for us is $s = \frac{1}{2}$; the exact value of $\mathcal{A}_a(\frac{1}{2})$ will be obtained later.) This difference in normalization is related to the realization of the differential equations governing the deformation of the spaces K_a as a first order differential system in "canonical" form, as in the classical spectral theory of linear differential equations ([21, 10].) This allows to realize the self-reciprocal scale reversing operator as a scattering [6].

6 Fredholm determinants, the first order differential system, and scattering

Let us return to the defining equations for the entire functions ϕ_a^+ and ϕ_a^-:

$$\phi_a^+ + \mathcal{H}P_a\phi_a^+ = J_0^a \tag{117a}$$
$$\phi_a^- - \mathcal{H}P_a\phi_a^- = J_0^a \tag{117b}$$

Either we read these equations as identities on $(0,\infty)$, or we decide that $\mathcal{H}P_a\phi_a^\pm$ in fact stands for $\int_0^a J_0(2\sqrt{xy})\phi_a^\pm(y)\,dy$, and the equation holds for $x \in \mathbb{C}$; the latter option slightly conflicts with our earlier definition of $\mathcal{H}$ as an operator on functions or distributions. But whatever choice is made this has no impact on what comes next. We shall apply to the equations the operators $a\frac{\partial}{\partial a}$ and $x\frac{\partial}{\partial x}$. As $J_0^a(x) = J_0(2\sqrt{ax})$ we have $a\frac{\partial}{\partial a}J_0^a = x\frac{\partial}{\partial x}J_0^a$. We write $\delta_x = x\frac{\partial}{\partial x} + \frac{1}{2} = \frac{\partial}{\partial x}x - \frac{1}{2}$. First we have:

$$a\frac{\partial}{\partial a}\phi_a^+ + \mathcal{H}P_a a\frac{\partial}{\partial a}\phi_a^+ = -a\phi_a^+(a)J_0^a + a\frac{\partial}{\partial a}J_0^a \tag{118a}$$
$$a\frac{\partial}{\partial a}\phi_a^- - \mathcal{H}P_a a\frac{\partial}{\partial a}\phi_a^- = +a\phi_a^-(a)J_0^a + a\frac{\partial}{\partial a}J_0^a \tag{118b}$$

Then, as $x\frac{\partial}{\partial x}\mathcal{H} = -\mathcal{H}\frac{\partial}{\partial x}x$, $\delta_x\mathcal{H} = -\mathcal{H}\delta_x$, $\delta_x P_a f = (P_a\delta_x f) - af(a)\delta_a(x)$, $\delta_x J_0^a = a\frac{\partial}{\partial a}J_0^a + \frac{1}{2}J_0^a$:

$$\delta_x\phi_a^+ - \mathcal{H}P_a\delta_x\phi_a^+ = (\frac{1}{2} - a\phi_a^+(a))J_0^a + a\frac{\partial}{\partial a}J_0^a \tag{119a}$$
$$\delta_x\phi_a^- + \mathcal{H}P_a\delta_x\phi_a^- = (\frac{1}{2} + a\phi_a^-(a))J_0^a + a\frac{\partial}{\partial a}J_0^a \tag{119b}$$

Combining we obtain:

$$a\frac{\partial}{\partial a}\phi_a^+ - \delta_x\phi_a^- + \mathcal{H}P_a(a\frac{\partial}{\partial a}\phi_a^+ - \delta_x\phi_a^-) = -(a\phi_a^+(a) + a\phi_a^-(a) + \frac{1}{2})J_0^a \tag{120a}$$
$$a\frac{\partial}{\partial a}\phi_a^- - \delta_x\phi_a^+ - \mathcal{H}P_a(a\frac{\partial}{\partial a}\phi_a^- - \delta_x\phi_a^+) = +(a\phi_a^+(a) + a\phi_a^-(a) - \frac{1}{2})J_0^a \tag{120b}$$

Comparing with (117a) and (117b), and as there is uniqueness:

$$a\frac{\partial}{\partial a}\phi_a^+ - \delta_x\phi_a^- = -(a\phi_a^+(a) + a\phi_a^-(a) + \frac{1}{2})\phi_a^+ \tag{121a}$$
$$a\frac{\partial}{\partial a}\phi_a^- - \delta_x\phi_a^+ = +(a\phi_a^+(a) + a\phi_a^-(a) - \frac{1}{2})\phi_a^- \tag{121b}$$

The quantity $a\phi_a^+(a) + a\phi_a^-(a)$ will play a fundamental rôle and we shall denote it by $\mu(a)$.[23] So:

$$(a\frac{\partial}{\partial a} + \frac{1}{2} + \mu(a))\phi_a^+ = \delta_x \phi_a^- \tag{122a}$$

$$(a\frac{\partial}{\partial a} + \frac{1}{2} - \mu(a))\phi_a^- = \delta_x \phi_a^+ \tag{122b}$$

It follows easily from this that $a\frac{\partial}{\partial a}(\phi_a^+\phi_a^-) = -\phi_a^+\phi_a^- + \frac{1}{2}\frac{\partial}{\partial x}x((\phi_a^+)^2 + (\phi_a^-)^2)$. So

$$a\frac{d}{da}\int_0^a \phi_a^+(x)\phi_a^-(x)\,dx = a\phi_a^+(a)\phi_a^-(a) - \int_0^a \phi_a^+(x)\phi_a^-(x)\,dx + \frac{1}{2}a(\phi_a^+(a)^2 + \phi_a^-(a)^2) \tag{123}$$

$$a\frac{d}{da}\,a\int_0^a \phi_a^+(x)\phi_a^-(x)\,dx = \frac{1}{2}\mu(a)^2 \tag{124}$$

We then compute:

$$\int_0^a \phi_a^+(x)\phi_a^-(x)\,dx = \int_0^a ((1-D_a)^{-1}J_0^a)(x)J_0^a(x)\,dx\ , \tag{125}$$

where we recall $\phi_a^+ = (1+H_a)^{-1}J_0^a$, $\phi_a^- = (1-H_a)^{-1}J_0^a$, $D_a = H_a^2$. The operator D_a acts on $L^2(0,a;dx)$ with kernel $D_a(x,z) = \int_0^a J_0(2\sqrt{xy})J_0(2\sqrt{yz})\,dy$. After the change of variables $x = at$, $y = au$, $z = av$ this becomes the operator d_a on $L^1(0,1;dt)$ with kernel $d_a(t,v) = \int_0^1 aJ_0(2a\sqrt{tu})\,aJ_0(2a\sqrt{uv})\,du$. We compute the derivative with respect to a:

$$\frac{\partial}{\partial a}\int_0^1 aJ_0(2a\sqrt{tu})\,aJ_0(2a\sqrt{uv})\,du \tag{126a}$$

$$= \int_0^1 ((2u\frac{\partial}{\partial u}+1)J_0(2a\sqrt{tu}))\,aJ_0(2a\sqrt{uv})\,du + \int_0^1 aJ_0(2a\sqrt{tu}))((2\frac{\partial}{\partial u}u - 1)J_0(2a\sqrt{uv}))\,du \tag{126b}$$

$$= 2aJ_0(2a\sqrt{t})J_0(2a\sqrt{v}) \tag{126c}$$

So $\frac{d}{da}d_a$ is a rank one operator, with range $\mathbb{C}J_0(2a\sqrt{t})\mathbf{1}_{0<t<1}(t)$. We now use the well-known formula

$$\frac{d}{da}\log\det(1-d_a) = -\mathrm{Tr}((1-d_a)^{-1}\frac{d}{da}d_a) \tag{127}$$

The rank one operator $(1-d_a)^{-1}\frac{d}{da}d_a$ has the function $(1-d_a)^{-1}2aJ_0(2a\sqrt{t})$ as eigenvector and the eigenvalue is $\int_0^1 J_0(2a\sqrt{t})((1-d_a)^{-1}2aJ_0(2a\sqrt{v}))(t)\,dt$. Going back to $(0,a)$ we obtain $2\int_0^a J_0(2\sqrt{ax})((1-D_a)^{-1}J_0(2\sqrt{az}))(x)\,dx$ and in view of (125) we have proven:

$$\frac{d}{da}\log\det(1-D_a) = -2\int_0^a \phi_a^+(x)\phi_a^-(x)\,dx \tag{128}$$

Then, using (124), we have the important formula:

$$\mu(a)^2 = -a\frac{d}{da}\,a\frac{d}{da}\,\log\det(1-D_a) \tag{129}$$

[23] maybe it would be unfair to hide the fact that $\mu(a) = 2a$, in this study of $\mathcal{H}$! In a later section a further mu function, associated with a variant of $\mathcal{H}$, will also be found explicitely and it will be quite more complicated.

We shall now also relate $\phi_a^+(a)$ and $\phi_a^-(a)$ to Fredholm determinants. In fact the following holds:

$$a\phi_a^+(a) = +a\frac{d}{da}\log\det(1+H_a) \tag{130a}$$

$$a\phi_a^-(a) = -a\frac{d}{da}\log\det(1-H_a) \tag{130b}$$

This is the application of a well-known general theorem, for any continuous kernel $k(x,y)$: if $w(x)+\int_0^a k(x,y)w(y)\,dy = k(x,a)$ for $0\le x\le a$ then $w(a) = +\frac{d}{da}\log\det_{(0,a)}(\delta(x-y)+k(x,y))$. A proof may be given which is of a somewhat similar kind as the one given above for (128), or one may more directly use the Fredholm's formulas for the determinant and the resolvent.[24] The theorem is proven in the book of P. Lax [20], Theorem 12 of Chapter 24 (Lax treats the case of a kernel on $(a,+\infty)$, here we have the simpler case of a finite interval $(0,a)$.) This means that $\mu(a)$ has another expression in terms of Fredholm determinants:

$$\mu(a) = a\frac{d}{da}\log\frac{\det(1+H_a)}{\det(1-H_a)} \tag{131}$$

Combining (129) and (131) we obtain:

$$-2a\frac{d}{da}\,a\frac{d}{da}\log\det(1+H_a) = \left(a\frac{d}{da}\log\frac{\det(1+H_a)}{\det(1-H_a)}\right)^2 - a\frac{d}{da}a\frac{d}{da}\log\frac{\det(1+H_a)}{\det(1-H_a)} \tag{132a}$$

$$-2a\frac{d}{da}\,a\frac{d}{da}\log\det(1-H_a) = \left(a\frac{d}{da}\log\frac{\det(1+H_a)}{\det(1-H_a)}\right)^2 + a\frac{d}{da}a\frac{d}{da}\log\frac{\det(1+H_a)}{\det(1-H_a)} \tag{132b}$$

$$2a\frac{d}{da}a\phi_a^+(a) = -\mu(a)^2 + a\mu'(a) \tag{132c}$$

$$2a\frac{d}{da}a\phi_a^-(a) = +\mu(a)^2 + a\mu'(a) \tag{132d}$$

$$\frac{d}{da}a(\phi_a^-(a)-\phi_a^+(a)) = a(\phi_a^+(a)+\phi_a^-(a))^2 \tag{132e}$$

These Fredholm determinants identities are reminiscent of certain well-known Gaudin identities [22, App. A16], which apply to the even and odd parts of an additive (Toeplitz) convolution kernel on an interval $(-a,a)$; here the situation is with kernels $k(xy)$ which have a multiplicative look, and reduction to the additive case would give $g(t+u)$ type kernels on semi-infinite intervals.

We have defined $A_a = \frac{\sqrt{a}}{2}(\phi_a^+ + \mathcal{H}\phi_a^+)$ and $-iB_a = \frac{\sqrt{a}}{2}(-\phi_a^- + \mathcal{H}\phi_a^-)$. Let us recall that here $\phi_a^\pm$ is restricted to $[0,+\infty)$ and is then tempered as a distribution. Using the differential equations (122a) and (122b) and the commutation property $\delta_x\mathcal{H} = -\mathcal{H}\delta_x$, $\delta_x = x\frac{\partial}{\partial x}+\frac12$, we have $\delta_x A_a = \frac{\sqrt{a}}{2}(\delta_x\phi_a^+ - \mathcal{H}\delta_x\phi_a^+) = \frac{\sqrt{a}}{2}(a\frac{\partial}{\partial a}+\frac12-\mu(a))(\phi_a^- - \mathcal{H}\phi_a^-) = -(a\frac{\partial}{\partial a}-\mu(a))(-iB_a)$ and $\delta_x(-iB_a) = \frac{\sqrt{a}}{2}(-(a\frac{\partial}{\partial a}+\frac12+\mu(a))\phi_a^+ - (a\frac{\partial}{\partial a}+\frac12+\mu(a))\mathcal{H}\phi_a^+) = -(a\frac{\partial}{\partial a}+\mu(a))A_a$. The following first order system of differential equations therefore holds:

$$a\frac{\partial}{\partial a}A_a = -\mu(a)A_a - \delta_x(-iB_a) \tag{133a}$$

$$a\frac{\partial}{\partial a}(-iB_a) = +\mu(a)(-iB_a) - \delta_x A_a \tag{133b}$$

[24] let us recall that for a continuous kernel on a finite interval, the formula of Fredholm for a determinant as a convergent series always applies, even if the operator given by the kernel is not trace class, which may happen.

Then we also have the second order differential equations $(a\frac{\partial}{\partial a} - \mu(a))(a\frac{\partial}{\partial a} + \mu(a))A_a = +\delta_x^2 A_a$ and $(a\frac{\partial}{\partial a} + \mu(a))(a\frac{\partial}{\partial a} - \mu(a))(-iB_a) = +\delta_x^2(-iB_a)$, or, taking the right Mellin transforms, and writing $s = \frac{1}{2} + i\gamma$, $\delta_x = i\gamma$:

$$-a\frac{\partial}{\partial a}a\frac{\partial}{\partial a}\widehat{A_a} + (\mu(a)^2 - a\mu'(a))\widehat{A_a} = \gamma^2\widehat{A_a} \tag{134a}$$

$$-a\frac{\partial}{\partial a}a\frac{\partial}{\partial a}(-i\widehat{B_a}) + (\mu(a)^2 + a\mu'(a))(-i\widehat{B_a}) = \gamma^2(-i\widehat{B_a}) \tag{134b}$$

With the new variable $u = \log(a)$ we obtain Dirac and Schrödinger equations which are associated with this study of $\mathcal{H}$, modeled on the study of the cosine and sine transforms summarized in [5, 6]. All quantities in the statement of the theorem will be completely explicited later in terms of Bessel functions, but we keep the notation sufficiently general to allow, if an interesting other case arises, to write down the identical results:

Theorem 13. *For each $a > 0$ let ϕ_a^+ and ϕ_a^- be the entire functions which are the solutions to:*

$$\phi_a^+(x) + \int_0^a J_0(2\sqrt{xy})\phi_a^+(y)\,dy = J_0(2\sqrt{ax}) \tag{135a}$$

$$\phi_a^-(x) - \int_0^a J_0(2\sqrt{xy})\phi_a^-(y)\,dy = J_0(2\sqrt{ax}) \tag{135b}$$

Let H_a be the integral operator on $L^2(0, a; dx)$ with kernel $J_0(2\sqrt{xy})$. There holds:

$$\phi_a^+(a) = +\frac{d}{da}\log\det(1 + H_a) \tag{135c}$$

$$\phi_a^-(a) = -\frac{d}{da}\log\det(1 - H_a)\ . \tag{135d}$$

The tempered distributions $A_a = \frac{\sqrt{a}}{2}(1 + \mathcal{H})(\phi_a^+\mathbf{1}_{0<x<\infty})$ and $B_a = i\frac{\sqrt{a}}{2}(-1 + \mathcal{H})(\phi_a^-\mathbf{1}_{0<x<\infty})$ vanish on $(-\infty, a)$ and are respectively self-reciprocal and skew-reciprocal under $\mathcal{H}$. Their completed right Mellin transforms $\mathcal{A}_a(s) = \Gamma(s)\widehat{A_a}(s)$ and $\mathcal{B}_a(s) = \Gamma(s)\widehat{B_a}(s)$ are entire functions with all their zeros on the critical line, they are respectively even and odd for $s \leftrightarrow 1 - s$, and they verify the following Dirac and Schrödinger types of differential equations in the variable $u = \log(a)$, $-\infty < u < +\infty$,

$$\frac{d}{du}\mathcal{A}_a = -\mu(a)\mathcal{A}_a - \gamma\mathcal{B}_a \tag{135e}$$

$$\frac{d}{du}\mathcal{B}_a = +\mu(a)\mathcal{B}_a + \gamma\mathcal{A}_a \tag{135f}$$

$$\gamma^2\mathcal{A}_a = \left(-\frac{d^2}{du^2} + V_+(u)\right)\mathcal{A}_a \tag{135g}$$

$$\gamma^2\mathcal{B}_a = \left(-\frac{d^2}{du^2} + V_-(u)\right)\mathcal{B}_a \tag{135h}$$

$$V_+(\log a) = \mu(a)^2 - \frac{d\,\mu(a)}{du} = -2\frac{d^2\log\det(1 + H_a)}{du^2} \tag{135i}$$

$$V_-(\log a) = \mu(a)^2 + \frac{d\,\mu(a)}{du} = -2\frac{d^2\log\det(1 - H_a)}{du^2} \tag{135j}$$

$$\mu(a) = \frac{d}{du}\log\frac{\det(1 + H_a)}{\det(1 - H_a)} = a\phi_a^+(a) + a\phi_a^-(a) \tag{135k}$$

where $s = \frac{1}{2} + i\gamma$.

Let us consider the Hilbert space of pairs $\begin{bmatrix}\alpha(u)\\ \beta(u)\end{bmatrix}$ on $\mathbb{R}$ with squared norms $\int_{-\infty}^{\infty} |\alpha(u)|^2 + |\beta(u)|^2 \frac{du}{2}$, and the two equivalent differential systems in canonical forms:

$$\left(\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\frac{d}{du} - \begin{bmatrix}0 & \mu(e^u)\\ \mu(e^u) & 0\end{bmatrix}\right)\begin{bmatrix}\alpha(u)\\ \beta(u)\end{bmatrix} = \gamma\begin{bmatrix}\alpha(u)\\ \beta(u)\end{bmatrix} \tag{136}$$

$$\left(\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\frac{d}{du} + \begin{bmatrix}-\mu(e^u) & 0\\ 0 & \mu(e^u)\end{bmatrix}\right)\begin{bmatrix}\alpha(u)+\beta(u)\\ -\alpha(u)+\beta(u)\end{bmatrix} = \gamma\begin{bmatrix}\alpha(u)+\beta(u)\\ -\alpha(u)+\beta(u)\end{bmatrix} \tag{137}$$

The components obey the corresponding Schrödinger equations:

$$-\alpha''(u) + V_+(u)\alpha(u) = \gamma^2\alpha \qquad V_+(u) = \mu(e^u)^2 - \frac{d\,\mu(e^u)}{du} \tag{138a}$$

$$-\beta''(u) + V_-(u)\beta(u) = \gamma^2\beta \qquad V_-(u) = \mu(e^u)^2 + \frac{d\,\mu(e^u)}{du} \tag{138b}$$

Regarding the behavior at $-\infty$, we are in the limit-point case for each of the Schrödinger equations (138a) and (138b) because clearly (say, from the defining integral equations for ϕ_a^+ and ϕ_a^-) one has $\phi_a^+(a) \to_{a\to 0} J_0(0) = 1$, $\phi_a^-(a) \to_{a\to 0} 1$, $\mu(a) \sim_{a\to 0} 2a$, so the potentials are exponentially vanishing as $u \to -\infty$. Perhaps we should reveal that one has exactly $\mu(a) = 2a = 2e^u$ so we are dealing here with quite concrete Schrödinger equations and Dirac systems whose exact solutions will later be written explicitely in terms of modified Bessel functions, but we delay using any information which would be too specific of the $\mathcal{H}$-transform.

For each $\gamma \in \mathbb{C}$

$$u \mapsto \begin{bmatrix}\mathcal{A}_{\exp(u)}(\frac12 + i\gamma)\\ \mathcal{B}_{\exp(u)}(\frac12 + i\gamma)\end{bmatrix} \tag{139}$$

is a (non-zero) solution of the system (136), and we now show that it is square-integrable (with respect to $du = d\log(a)$) at $+\infty$. Let us recall the equation (111) $(s+z-1)\mathcal{X}_a(s,z) = -2i\mathcal{B}_a(s)\mathcal{A}_a(z) - 2i\mathcal{A}_a(s)\mathcal{B}_a(z)$, from which we deduce

$$a\frac{\partial}{\partial a}\mathcal{X}_a(s,z) = -2\mathcal{A}_a(s)\mathcal{A}_a(z) - 2(i\mathcal{B}_a(s))(i\mathcal{B}_a(z)) \tag{140}$$

We have[25] $\|\mathcal{X}_s^a\|^2 = \mathcal{X}_a(s,\overline{s})$, $\mathcal{A}_a(\overline{s}) = \overline{\mathcal{A}_a(s)}$, $i\mathcal{B}_a(\overline{s}) = \overline{i\mathcal{B}_a(s)}$, so

$$a\frac{\partial}{\partial a}\|\mathcal{X}_s^a\|^2 = -2|\mathcal{A}_a(s)|^2 - 2|\mathcal{B}_a(s)|^2 \tag{141}$$

and as of course $\lim_{a\to+\infty}\|\mathcal{X}_s^a\|^2 = 0$ (we have $\|\mathcal{X}_s^a\|^2 \le \int_a^\infty |\mathcal{X}_s^1|^2(x)\,dx$ for $a \ge 1$) we obtain:

$$\forall s \in \mathbb{C} \quad \|\mathcal{X}_s^a\|^2 = 2\int_a^\infty (|\mathcal{A}_a(s)|^2 + |\mathcal{B}_a(s)|^2)\frac{da}{a} \tag{142}$$

This establishes the square-integrability at $+\infty$ of $\begin{bmatrix}\mathcal{A}_{\exp(u)}(s)\\ \mathcal{B}_{\exp(u)}(s)\end{bmatrix}$, for any $s \in \mathbb{C}$.

The solutions of (136) with eigenvalue $\gamma = 0$ are $\begin{bmatrix}\mathcal{A}_a(\frac12)\\ 0\end{bmatrix}$ and $\begin{bmatrix}0\\ \mathcal{A}_a(\frac12)^{-1}\end{bmatrix}$. The former is square-integrable, so from $2 \le t + t^{-1}$ the latter then necessarily is not. This confirms that the Dirac system (136) is in the limit point case at $+\infty$ (according to a general theorem of Levitan [21, §13,

[25] let us recall the notation $\mathcal{X}_s^a = \Gamma(s)X_s^a \in L^2(a,+\infty;dx)$.

Thm 7.1] any first order differential operator $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right] \frac{d}{du} + \left[\begin{smallmatrix} a(u) & b(u) \\ c(u) & d(u) \end{smallmatrix}\right]$ with continuous coefficients is in the limit point case at infinity). So the pair (139) is in fact, for any $\gamma \in \mathbb{C}$, the unique solution of (136) which is square-integrable at $+\infty$. Also the Schrödinger equation (138b) is in the limit point case as not all of its solutions are square integrable at $+\infty$. Whether the limit-point case at $+\infty$ holds for equation (138a) is less evident. Let us recall from [10, §9, Thm 2.4] and [26, §X, Thm X.8] that a sufficient condition for this is the existence of a lowerbound $\liminf_{u\to+\infty} V_+(u)/u^2 > -\infty$. We will prove in the next chapter that $\mu(a) = 2a = 2e^u$ so this is certainly the case here. In the present chapter only the fact that the Dirac system is known to be in the limit-point case will be used.

We now take $u_0 = \log a_0$ and apply on (u_0, ∞) the Weyl-Stone-Titchmarsh-Kodaira theory ([10, §9], [21, §3]). Let $\psi(u,s)$ be the unique solution of the system (136) for the eigenvalue γ, $s = \frac{1}{2} + i\gamma$, and with the initial condition $\psi(u_0, s) = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ and let $\phi(u,s)$ be the unique solution with the initial condition $\phi(u_0, s) = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. Let $m(\gamma)\psi(u,s) + \phi(u,s)$ for $\Im\gamma > 0$ be the unique solution which is square-integrable on $(u_0, +\infty)$. So $m(\gamma) = \frac{\mathcal{A}_{a_0}(s)}{\mathcal{B}_{a_0}(s)}$, $s = \frac{1}{2} + i\gamma$, $\Re(s) < \frac{1}{2}$. It is a fundamental general property of the m function from Hermann Weyl's theory that $\Im(m(\gamma)) > 0$ (for $\Im(\gamma) > 0$.) Here, we have a case where the m-function is found to be meromorphic on all of $\mathbb{C}$; so we see that its poles and zeros on $\mathbb{R}$ are simple. Furthermore, the spectral measure ν is obtained via the formula $\nu(a,b) = \lim_{\epsilon\to 0^+} \frac{1}{\pi}\int_a^b \Im m(\gamma + i\epsilon)\, d\gamma$ (under the condition $\nu\{a,b\} = 0$). We obtain:

$$d\nu(\gamma) = \sum_{\mathcal{B}_{a_0}(\rho)=0} \frac{\mathcal{A}_{a_0}(\rho)}{-i\mathcal{B}'_{a_0}(\rho)} \delta(\gamma - \Im\rho) \tag{143}$$

The spectrum is thus purely discrete and the general theory tells us further that the finite linear combinations $\sum_\rho c_\rho \frac{\mathcal{A}_{a_0}(\rho)}{-i\mathcal{B}'_{a_0}(\rho)} \psi(u,\rho)$ have squared norms $\sum_\rho \frac{\mathcal{A}_{a_0}(\rho)}{-i\mathcal{B}'_{a_0}(\rho)} |c_\rho|^2$ and also that they are dense in $L^2((u_0,\infty) \to \mathbb{C}^2; du)$. For $\mathcal{B}_{a_0}(\rho) = 0$, $\psi(u,\rho) = \mathcal{A}_{a_0}(\rho)^{-1} \left[\begin{smallmatrix} \mathcal{A}_{\exp(u)}(\rho) \\ \mathcal{B}_{\exp(u)}(\rho) \end{smallmatrix}\right] \mathbf{1}_{u \geq u_0}(u)$, so the vectors $Z_\rho^{a_0} = \left[\begin{smallmatrix} 2\mathcal{A}_{\exp(u)}(\rho) \\ 2\,\mathcal{B}_{\exp(u)}(\rho) \end{smallmatrix}\right] \mathbf{1}_{u \geq u_0}(u)$ are an orthogonal basis of $L^2((u_0,\infty) \to \mathbb{C}^2; \frac{1}{2}du)$ and they satisfy $\|Z_\rho^{a_0}\|^2 = -2\,\mathcal{A}_{a_0}(\rho)\, i\, \mathcal{B}'_{a_0}(\rho)$. Similarly a spectral interpretation is given to the zeros of $\mathcal{A}_{a_0}$ if one looks at the initial condition $\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. The factors of 2 and $\frac{1}{2}$, have been incorporated so that the statement may be translated (taking into account results established later) into the fact that the evaluators $K_{a_0}(\rho, z)$, for $\mathcal{B}_{a_0}(\rho) = 0$, are an orthogonal basis of the Hilbert space of the functions $\Gamma(z)\widehat{f}(z)$, $f \in K_{a_0}$. This last statement is a general theorem (under a certain condition) for spaces with the de Branges axioms [2, §22].

To discuss in a self-contained manner the generalized Parseval identity which is associated with the differential system on the full line, it is convenient to make a preliminary majoration of $\|X_s^a\|^2$, $\Re(s) = \frac{1}{2}$. From (108) we have, for $\Re(s) = \frac{1}{2}$: $\|\mathcal{X}_s^a\|^2 = 2\Re(\mathcal{E}_a(s)\overline{\mathcal{E}'_a(s)})$. And $\mathcal{E}_a(s) = \Gamma(s)\widehat{E_a}(s)$. And $\widehat{E_a}(s) = \sqrt{a}\Big(a^{-s} + \frac{1}{2}\int_a^\infty (\phi_a^+(x) - \phi_a^-(x)) x^{-s}\, dx\Big)$. We know from the discussion of Lemma 9 that the integral in the expression for $\widehat{E_a}(s)$ is absolutely convergent for $\Re(s) > \frac{1}{4}$. Hence by the Riemann-Lebesgue lemma $\widehat{E_a}(\frac{1}{2} + i\gamma) \sim a^{-i\gamma}$ as $|\gamma| \to \infty$, $\gamma \in \mathbb{R}$. Similarly, $\widehat{E_a}'(\frac{1}{2} + i\gamma) \sim -\log(a) a^{-i\gamma}$. So, with $\|\mathcal{X}_s^a\|^2 = |\Gamma(s)|^2 \|X_s^a\|^2$ and using Stirling's formula we obtain:

Lemma 14. *For each given $a > 0$ one has $\|X_s^a\|^2 \sim 2\log|s|$ as $|s| \to \infty$, $\Re(s) = \frac{1}{2}$.*

From (104) expressed using A_a and B_a we see that $\frac{\widehat{B_a}(s)}{s - \frac{1}{2}}$ is square integrable, so $s^{-1}\widehat{B_a}(s)$ is square integrable on the critical line (with respect to $|ds|$). Then using again (104) we see that

$s^{-1}\widehat{A_a}(s)$ is also square integrable on the critical line.[26] Let us pick a function $F(s)$ on the critical line which is such that $sF(s)$ is square integrable. Then $F(s)\overline{\widehat{A_a}(s)}$ and $F(s)\overline{\widehat{B_a}(s)}$ are absolutely integrable on the critical line and $(\int_{\Re(s)=\frac{1}{2}}|F(s)\overline{\widehat{A_a}(s)}|\frac{|ds|}{2\pi})^2 \leq C\int_{\Re(s)=\frac{1}{2}}\frac{|\widehat{A_a}(s)|^2}{|s|^2}\frac{|ds|}{2\pi}$ and similarly with B_a. If we define $\alpha_F(u) = 2\int_{\Re(s)=\frac{1}{2}}F(s)\overline{\widehat{A_a}(s)}\frac{|ds|}{2\pi}$ and $\beta_F(u) = 2\int_{\Re(s)=\frac{1}{2}}F(s)\overline{\widehat{B_a}(s)}\frac{|ds|}{2\pi}$ we then compute:

$$\int_{u_0}^{\infty}|\alpha_F(u)|^2 + |\beta_F(u)|^2\,du \leq C\int_{\Re(s)=\frac{1}{2}}\frac{\int_{u_0}^{\infty}(|\widehat{A_a}(s)|^2 + |\widehat{B_a}(s)|^2)\,du}{|s|^2}\ \frac{|ds|}{2\pi}$$
$$= \frac{C}{2}\int_{\Re(s)=\frac{1}{2}}\frac{\|X_s^a\|^2}{|s|^2}\frac{|ds|}{2\pi} < \infty \quad (144)$$

So $\alpha_F(u)$ and $\beta_F(u)$ are square integrable at $+\infty$. More precisely the above upper bound holds as well for $\int_{\Re(s)=\frac{1}{2}}|F(s)\widehat{A_a}(s)|\frac{|ds|}{2\pi}$ and $\int_{\Re(s)=\frac{1}{2}}|F(s)\widehat{B_a}(s)|\frac{|ds|}{2\pi}$. So the double integrals

$$\iint_{u_0<u<\infty,\Re(s)=\frac{1}{2}}\mathcal{A}_{\exp(u)}(z)\mathcal{A}_{\exp(u)}(s)\mathcal{F}(s)\ \frac{|ds|}{2\pi|\Gamma(s)|^2}\ du \quad (145a)$$
$$\iint_{u_0<u<\infty,\Re(s)=\frac{1}{2}}\mathcal{B}_{\exp(u)}(z)\mathcal{B}_{\exp(u)}(s)\mathcal{F}(s)\ \frac{|ds|}{2\pi|\Gamma(s)|^2}\ du \quad (145b)$$

where $z\in\mathbb{C}$ is arbitrary, and $\mathcal{F}(s) = \Gamma(s)F(s)$, are absolutely convergent and Fubini may be employed. Using (140):

$$\mathcal{X}_{\exp(u_0)}(z,\overline{s}) = 2\int_{u_0}^{\infty}(\mathcal{A}_{\exp(u)}(z)\overline{\mathcal{A}_{\exp(u)}(s)} + \mathcal{B}_{\exp(u)}(z)\overline{\mathcal{B}_{\exp(u)}(s)})\,du \quad (146)$$

And we obtain the following identity of absolutely convergent integrals, for any $\mathcal{F}(s) = \Gamma(s)F(s)$ with $s\,F(s)\in L^2(\Re(s)=\frac{1}{2};\frac{|ds|}{2\pi})$:

$$\int_{\Re(s)=\frac{1}{2}}\mathcal{X}_{\exp(u_0)}(z,\overline{s})\mathcal{F}(s)\,\frac{|ds|}{2\pi|\Gamma(s)|^2} = \int_{u_0}^{\infty}(\mathcal{A}_{\exp(u)}(z)\alpha_F(u) + \mathcal{B}_{\exp(u)}(z)\beta_F(u))\,du \quad (147)$$

We shall prove that this identity holds under the weaker hypothesis $F(s)\in L^2(\Re(s)=\frac{1}{2};\frac{|ds|}{2\pi})$. First, still with $s\,F(s)$ square integrable we suppose additionally that $F=\widehat{f}$ with $f\in K_{\exp(u_0)}$[27]. The hilbertian kernel $K_{\exp(u_0)}(z,s)$ is $\mathcal{X}_{\exp(u_0)}(\overline{z},s)$ so $\overline{K_{\exp(u_0)}(z,s)} = \mathcal{X}_{\exp(u_0)}(z,\overline{s})$. The equations give then:

$$\mathcal{F}(z) = \int_{u_0}^{\infty}(\mathcal{A}_{\exp(u)}(z)\alpha_F(u) + \mathcal{B}_{\exp(u)}(z)\beta_F(u))\,du \quad (148a)$$
$$\alpha_F(u) = 2\int_{\Re(s)=\frac{1}{2}}\mathcal{F}(s)\mathcal{A}_a(s)\frac{|ds|}{2\pi\,|\Gamma(s)|^2} \quad (148b)$$
$$\beta_F(u) = 2\int_{\Re(s)=\frac{1}{2}}\mathcal{F}(s)\mathcal{B}_a(s)\frac{|ds|}{2\pi\,|\Gamma(s)|^2} \quad (148c)$$

[26] we know in fact according to proposition 11 that $\widehat{A_a}$ and $\widehat{B_a}$ are bounded on the critical line.
[27] this is certainly possible as we know that the $f(x)$ which are smooth, vanishing on $(0,a)$ and of Schwartz decrease as $x\to+\infty$ are dense in K_a.

We have worked under the hypothesis that $sF(s)$ is square integrable. To show that the formulae extend in the L^2 sense, we first examine:

$$|\alpha_F(u)|^2 = 4\int_{\Re(s_1)=\frac{1}{2}} \mathcal{F}(s_1)\mathcal{A}_a(s_1)\frac{|ds_1|}{2\pi\,|\Gamma(s_1)|^2}\int_{\Re(s_2)=\frac{1}{2}} \overline{\mathcal{F}(s_2)}\mathcal{A}_a(s_2)\frac{|ds_2|}{2\pi\,|\Gamma(s_2)|^2} \tag{149}$$

$$\int_{u_0}^{\infty} |\alpha_F(u)|^2+|\beta_F(u)|^2\,\frac{du}{2} = \iint_{\Re(s_i)=\frac{1}{2}} \mathcal{F}(s_1)\overline{\mathcal{F}(s_2)}\mathcal{X}_{\exp(u_0)}(s_1,\overline{s}_2)\frac{|ds_1|}{2\pi\,|\Gamma(s_1)|^2}\frac{|ds_2|}{2\pi\,|\Gamma(s_2)|^2} \tag{150}$$

There was absolute convergence in the triple integral used as an intermediate. Also $\mathcal{X}_{\exp(u_0)}(s_1,\overline{s}_2) = \mathcal{X}_{\exp(u_0)}(s_2,\overline{s}_1)$ and $\int_{\Re(s_1)=\frac{1}{2}} \mathcal{F}(s_1)\mathcal{X}_{\exp(u_0)}(s_2,\overline{s}_1)\frac{|ds_1|}{2\pi\,|\Gamma(s_1)|^2} = \mathcal{F}(s_2)$. Hence:

$$\int_{u_0}^{\infty} (|\alpha_F(u)|^2+|\beta_F(u)|^2)\,\frac{du}{2} = \int_{\Re(s)=\frac{1}{2}} |F(s)|^2\frac{|ds|}{2\pi} = \int_{\exp(u_0)}^{\infty} |f(x)|^2\,dx \tag{151}$$

So with an arbitrary $f\in K_a$, $F=\widehat{f}$, $\mathcal{F}(s)=\Gamma(s)\widehat{f}(s)$, the assignment $f\mapsto(\alpha_F,\beta_F)$ exists in the sense of L^2 convergence when one approximates f by a sequence f_n in K_a such that $s\widehat{f_n}(s)$ is in $L^2(\Re(s)=\frac{1}{2};\frac{|ds|}{2\pi})$, and $f\mapsto(\alpha_F,\beta_F)$ is linear and isometric. We check that its range is all of $L^2(u_0,\infty;\frac{du}{2})\oplus L^2(u_0,\infty;\frac{du}{2})$. For this let us identify the functions $\alpha_w(u)$ and $\beta_w(u)$ which will correspond to $\mathcal{F}(s)=\mathcal{X}_{a_0}(w,s)$ ($a_0=\exp(u_0)$.) On one hand from (147) it must be the case that:

$$\forall z\in\mathbb{C}\quad \int_{\Re(s)=\frac{1}{2}} \mathcal{X}_{a_0}(z,\overline{s})\mathcal{X}_{a_0}(w,s)\,\frac{|ds|}{2\pi|\Gamma(s)|^2} = \int_{u_0}^{\infty}(\mathcal{A}_{\exp(u)}(z)\alpha_w(u)+\mathcal{B}_{\exp(u)}(z)\beta_w(u))\,du \tag{152}$$

The left hand side is $\mathcal{X}_{a_0}(w,z)$ which on the other hand is given by the formula $2\int_{u_0}^{\infty}(\mathcal{A}_a(z)\mathcal{A}_a(w)-\mathcal{B}_a(z)\mathcal{B}_a(w))\,du$. The functions $u\mapsto\begin{bmatrix}\mathcal{A}_a(z)\\ \mathcal{B}_a(z)\end{bmatrix}$, $z\in\mathbb{C}$ are certainly dense in $L^2((u_0,\infty)\to\mathbb{C}^2;\frac{du}{2})$ as we know in particular that the pairs for the ρ's such that $\mathcal{B}_{a_0}(\rho)=0$ give an orthogonal basis. So we have the identification on $(u_0,+\infty)$:

$$\alpha_w(u) = 2\,\mathcal{A}_{\exp(u)}(w) \qquad \beta_w(u) = -2\mathcal{B}_{\exp(u)}(w) \tag{153}$$

This proves that the range is all of $L^2((u_0,\infty)\to\mathbb{C}^2;\frac{du}{2})$. Let us note that in this identification the hilbertian evaluator $K_{a_0}(w,\cdot)$ is sent to the pair $u\mapsto 2\,\mathbf{1}_{u>u_0}(u)(\overline{\mathcal{A}_a(w)},\overline{\mathcal{B}_a(w)})$. To check if all is coherent we compute the hilbertian scalar product $(K_{a_0}(w,\cdot),K_{a_0}(z,\cdot))$. We obtain $4\int_{u_0}^{\infty}\overline{\mathcal{A}_a(w)}\mathcal{A}_a(z)+\overline{\mathcal{B}_a(w)}\mathcal{B}_a(z)\frac{du}{2} = 2\int_{u_0}^{\infty}\mathcal{A}_a(\overline{w})\mathcal{A}_a(z)-\mathcal{B}_a(\overline{w})\mathcal{B}_a(z)\,du = \mathcal{X}_{exp(u_0)}(\overline{w},z)$, which is indeed $K_{a_0}(w,z)$.

Let us return to the consideration of a general $F(s)\in L^2(\Re(s)=\frac{1}{2};\frac{|ds|}{2\pi})$. Under the hypothesis that $s\,F(s)$ is square integrable we have assigned to F the functions

$$\alpha_F(u) = 2\int_{\Re(s)=\frac{1}{2}} \mathcal{F}(s)\mathcal{A}_a(s)\frac{|ds|}{2\pi\,|\Gamma(s)|^2} = \int_{\Re(s)=\frac{1}{2}} F(s)\overline{2\widehat{A_a}(s)}\frac{|ds|}{2\pi} \tag{154a}$$

$$\beta_F(u) = 2\int_{\Re(s)=\frac{1}{2}} \mathcal{F}(s)\mathcal{B}_a(s)\frac{|ds|}{2\pi\,|\Gamma(s)|^2} = \int_{\Re(s)=\frac{1}{2}} F(s)\overline{2\widehat{\mathcal{B}_a}(s)}\frac{|ds|}{2\pi} \tag{154b}$$

which are square-integrable at $+\infty$. From (147) there holds, for any $a_0=\exp(u_0)$:

$$\int_{\Re(s)=\frac{1}{2}} X_{\exp(u_0)}(z,\overline{s})F(s)\,\frac{|ds|}{2\pi} = \int_{u_0}^{\infty}(2\widehat{A_{\exp(u)}}(z)\alpha_F(u)+2\widehat{B_{\exp(u)}}(z)\beta_F(u))\,\frac{du}{2} \tag{155}$$

The function of z on the left side is the orthogonal projection F_{a_0} of F to the space $\widehat{K_{a_0}}$. So, we deduce by unicity $\alpha_F(u)\mathbf{1}_{u\geq u_0}(u) = \alpha_{F_{u_0}}(u)$ and $\beta_F(u)\mathbf{1}_{u\geq u_0}(u) = \beta_{F_{u_0}}(u)$. We then obtain $\|F_{a_0}\|^2 = \int_{u_0}^\infty (|\alpha_F(u)|^2 + |\beta_F(u)|^2)\frac{du}{2}$ so α_F and β_F are square-integrable on $(-\infty,+\infty)$, and as $\cup_a K_a$ is dense in $L^2(0,\infty;dx)$ the assignment $F\mapsto(\alpha_F,\beta_F)$ is isometric, and also it has a dense range in $L^2(\mathbb{R}\to\mathbb{C}^2;\frac{du}{2})$. We can then remove the hypothesis that $s\,F(s)$ is square integrable and define the functions α_F and β_F to be the limit in the L^2 sense of functions α_n and β_n associated with F_n's such that $\|F-F_n\|\to 0$ and the $s\,F_n$ are square-integrable. Summing up:

Theorem 15. *There are unitary identifications $L^2(0,\infty;dx)\tilde{\to}L^2(\Re(s)=\frac12;\frac{|ds|}{2\pi})\tilde{\to}L^2(\mathbb{R}\to\mathbb{C}^2;\frac{du}{2})$ given in the L^2 sense by the formulas, where $\Re(s)=\frac12$:*

$$F(s) = \widehat{f}(s) = \int_0^\infty f(x)x^{-s}\,dx \qquad\qquad f(x) = \int_{\Re(s)=\frac12} F(s)x^{s-1}\,\frac{|ds|}{2\pi} \tag{156a}$$

$$\alpha(u) = \lim_{n\to\infty}\int_{\Re(s)=\frac12} F_n(s)\,\overline{2\widehat{A_{\exp(u)}}(s)}\,\frac{|ds|}{2\pi} \qquad (F_n\to_{L^2} F;\qquad s\,F_n(s)\in L^2) \tag{156b}$$

$$\beta(u) = \lim_{n\to\infty}\int_{\Re(s)=\frac12} F_n(s)\,\overline{2\widehat{B_{\exp(u)}}(s)}\,\frac{|ds|}{2\pi} \tag{156c}$$

$$F(s) = \lim_{a_0\to 0}\int_{\log(a_0)}^\infty (\alpha(u)\,2\widehat{A_{\exp(u)}}(s) + \beta(u)\,2\widehat{B_{\exp(u)}}(s))\,\frac{du}{2} \tag{156d}$$

The orthogonal projection of f to K_{a_0} corresponds to the replacement of $\alpha(u)$ by $\alpha(u)\mathbf{1}_{u>u_0}(u)$ and of $\beta(u)$ by $\beta(u)\mathbf{1}_{u>u_0}(u)$ ($u_0=\log(a_0)$.). The unitary operators $f\mapsto\mathcal{H}(f)$, $F(s)\mapsto\chi(s)F(1-s)$, correspond to $(\alpha,\beta)\mapsto(\alpha,-\beta)$. For $f=X_z^{a_0}$ one has $\alpha(u)=2\widehat{A_{\exp(u)}}(z)\mathbf{1}_{u>\log(a_0)}(u)$ and $\beta(u)=-2\widehat{B_{\exp(u)}}(z)\mathbf{1}_{u>\log(a_0)}(u)$. The self-adjoint operator $F(s)\mapsto\gamma F(s)$ ($s=\frac12+i\gamma$) corresponds to the canonical operator:

$$H = \begin{bmatrix} 0 & \frac{d}{du} \\ -\frac{d}{du} & 0\end{bmatrix} - \begin{bmatrix} 0 & \mu(e^u) \\ \mu(e^u) & 0\end{bmatrix} \tag{156e}$$

which, in $L^2(\mathbb{R}\to\mathbb{C}^2;\frac{du}{2})$, is essentially self-adjoint when defined on the domain of the functions of class C^1 (or even C^∞) with compact support. The unitary operator $e^{i\tau H}$ acts on $L^2(0,\infty;dx)$ as $f(x)\mapsto e^{\frac12\tau}f(e^\tau x)$.

For the statement of self-adjointness we start with α and β of class C^1 with compact support, define F by (156d) and integrate by parts to confirm that $\gamma F(s)$ corresponds to $H\left(\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]\right)$. We know by Hermann Weyl's theorem that in the limit point case the pairs $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ of class C^1 with compact support are a core of self-adjointness (*cf.* [21, §13].) On the other hand we know that multiplication by γ on $L^2(\Re(s)=\frac12;\frac{|ds|}{2\pi})$ with maximal domain is a self-adjoint operator. So the two self-adjoint operators are the same.

Having discussed the matter from the point of view of the isometric expansion we now turn to another topic, the topic of the scattering, or rather total reflection against the potential barrier at $+\infty$. Another pair of solutions of the first order system (136) (hence also of the second order differential equations) is known. Let us recall from equations (95a), (95b) that we defined $j_a = \sqrt{a}(\delta_a - \phi_a^+\mathbf{1}_{0<x<a}) = \sqrt{a}\mathcal{H}\phi_a^+$ and $-ik_a = \sqrt{a}(\delta_a + \phi_a^-\mathbf{1}_{0<x<a}) = \sqrt{a}\mathcal{H}\phi_a^-$. Using again (122a) and (122b) it is checked that j_a and k_a verify the exact same differential system as A_a and B_a:

$$a\frac{\partial}{\partial a}j_a = -\mu(a)j_a + i\delta_x k_a \tag{157a}$$

$$a\frac{\partial}{\partial a}k_a = +\mu(a)k_a - i\delta_x j_a \tag{157b}$$

The right Mellin transforms $\widehat{j_a}(s)$ and $-i\widehat{k_a}(s)$ are defined as

$$\widehat{j_a}(s) = a^{\frac{1}{2}-s} - \sqrt{a}\int_0^a \phi_a^+(x)x^{-s}\,dx \tag{158a}$$

$$-i\widehat{k_a}(s) = a^{\frac{1}{2}-s} + \sqrt{a}\int_0^a \phi_a^-(x)x^{-s}\,dx \tag{158b}$$

As ϕ_a^+ and ϕ_a^- are analytic these are meromorphic functions in $\mathbb{C}$ with possible[28] pole locations at $s = 1, 2, \ldots$ From the point of view of the Schrödinger equation (138a) and as $u = \log(a) \to -\infty$, $a \to 0$, we thus see that, for $s = \frac{1}{2}+i\gamma$, $\gamma \in \mathbb{R}$, $\widehat{j_a}(\frac{1}{2}+i\gamma)$ and $\widehat{j_a}(\frac{1}{2}-i\gamma)$ are two (linearly independent for $\gamma \neq 0$) solutions, differing from $e^{-i\gamma u}$ and $e^{i\gamma u}$ by an exponentially small (in $u = \log(a)$) quantity (and similarly with $-i\widehat{k_a}$ with respect to the Schrödinger equation (135h)). So we have identified the unique solutions which verify the Jost conditions at $-\infty$.

As $P_a\phi_a^+$ is square integrable, and also using (85a), we have on the critical line:

$$\Gamma(s)\widehat{j_a}(s) + \Gamma(1-s)\widehat{j_a}(1-s) \tag{159a}$$

$$= \Gamma(s)\widehat{j_a}(s) + \Gamma(1-s)a^{s-\frac{1}{2}} - \Gamma(1-s)\sqrt{a}\int_0^a \phi_a^+(x)x^{s-1}\,dx \tag{159b}$$

$$= \Gamma(s)\widehat{j_a}(s) + \Gamma(1-s)a^{s-\frac{1}{2}} - \Gamma(s)\sqrt{a}\int_0^\infty (\mathcal{H}P_a\phi_a^+)(x)x^{-s}\,dx \tag{159c}$$

$$= \Gamma(s)\widehat{j_a}(s) + \Gamma(1-s)a^{s-\frac{1}{2}} + \Gamma(s)\sqrt{a}\int_0^\infty (\phi_a^+(x) - J_0(2\sqrt{ax}))x^{-s}\,dx \tag{159d}$$

$$\begin{aligned} &= \Gamma(s)a^{\frac{1}{2}-s} + \Gamma(1-s)a^{s-\frac{1}{2}} - \Gamma(s)\sqrt{a}\int_0^a J_0(2\sqrt{ax})x^{-s}\,dx \\ &\qquad + \Gamma(s)\sqrt{a}\int_a^\infty (\phi_a^+(x) - J_0(2\sqrt{ax}))x^{-s}\,dx \end{aligned} \tag{159e}$$

As $J_0(2\sqrt{ax}) - \phi_a^+(x)$ is square integrable both integrals are simultaneously absolutely convergent at least for $\frac{1}{2} < \Re(s) < 1$ (the $\frac{1}{2}$ can be improved, but this does not matter). As the boundary values on the critical line coincide we have an identity of analytic functions. We recognize in $\int_a^\infty J_0(2\sqrt{ax})x^{-s}\,dx$, which is absolutely convergent for $\Re(s) > \frac{3}{4}$, the quantity $g_s(a)$ (equation (73)). And from equation (74) we know $g_s(a) = \chi(s)a^{s-1} - \int_0^a J_0(2\sqrt{ax})x^{-s}\,dx$. So

$$\Gamma(s)\widehat{j_a}(s) + \Gamma(1-s)\widehat{j_a}(1-s) = \Gamma(s)a^{\frac{1}{2}-s} + \Gamma(s)\sqrt{a}\int_a^\infty \phi_a^+(x)x^{-s}\,dx\ , \tag{160}$$

which is indeed $2\mathcal{A}_a(s)$. From the equation (158a) $\widehat{j_a}(s) = a^{\frac{1}{2}}(a^{-s} - \int_0^a \phi_a^+(x)x^{-s}\,dx)$ (valid as is for $\Re(s) < 1$) the function $u \mapsto \widehat{j_a}(s)$ differs from $u \mapsto e^{-i\gamma u}$ by an error which is relatively exponentially smaller (we write $s = \frac{1}{2} + i\gamma$, $\Im(\gamma) > -\frac{1}{2}$). So $\widehat{j_a}$ is the Jost solution at $-\infty$ of the Schrödinger equation (135g). The identity relating $\mathcal{B}_a(s)$ and $\widehat{k_a}(s) = i\,a^{\frac{1}{2}}(a^{-s} + \int_0^a \phi_a^-(x)x^{-s}\,dx)$ is proven similarly.

Theorem 16. *The unique[29] solution, square integrable at $u = +\infty$, of the Schrödinger equation* (135g) *(resp.* (135h)*; $\gamma \neq 0$) is expressed in terms of the functions $\widehat{j_a}(\frac{1}{2}+i\gamma)$ (resp. $-i\,\widehat{k_a}(\frac{1}{2}+i\gamma)$)*

[28] the poles do exist.

[29] here we make use of the fact that (135g) is in the limit point case at $+\infty$, because it is proven in the next chapter, or known from (66a), (66b), that $\mu(a) = 2a = 2\,e^u$, in this study of the $\mathcal{H}$ transform.

satisfying at $-\infty$ the Jost condition $\widehat{j_a} \sim_{u\to-\infty} e^{-i\gamma u}$ (resp. $-i\widehat{k_a} \sim_{u\to-\infty} e^{-i\gamma u}$) as:

$$\mathcal{A}_a(s) = \frac{1}{2}\Big(\Gamma(s)\widehat{j_a}(s) + \Gamma(1-s)\widehat{j_a}(1-s)\Big) \tag{161a}$$

$$\mathcal{B}_a(s) = \frac{1}{2}\Big(\Gamma(s)\widehat{k_a}(s) - \Gamma(1-s)\widehat{k_a}(1-s)\Big) \tag{161b}$$

Let us add a time parameter t and consider the wave equation:

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial u^2} + \mu^2 - \frac{d\mu}{du}\right)\Phi(t,u) = 0 \tag{162}$$

Then $\Phi(t,u) = e^{i\gamma t}\widehat{j_{\exp(u)}}(\frac{1}{2}+i\gamma)$ is a solution which behaves as $e^{i\gamma(t-u)}$ as $u \to -\infty$. This wave is thus right-moving, it is an incoming wave from $u=-\infty$ at $t=-\infty$. For a given frequency γ there is a unique, up to multiplicative factor, wave which respects the condition of being at each time square integrable at $u=+\infty$. This wave is $e^{i\gamma t}\mathcal{A}_{\exp(u)}(\frac{1}{2}+i\gamma)$. So the equation (161a) represents the decomposition in incoming and reflected components. There is in the reflected component a phase shift $\theta_\gamma = \arg\chi(s)$, the solution behaving approximatively at $u\to-\infty$ as $C(\gamma)\cos(\gamma u + \frac{1}{2}\theta_\gamma)$. This is an absolute scattering, as there is nothing a priori to compare it too. We will thus declare that equation (162) has realized $\chi(s)$ as an (absolute) scattering. Similarly the Schrödinger equation (135h) realizes $-\chi(s)$ as an absolute scattering.

We have $2\mathcal{A}_a(\frac{1}{2}) = 2\Gamma(\frac{1}{2})\widehat{j_a}(\frac{1}{2})$ and $\widehat{j_a}(\frac{1}{2}) = 1 - a^{\frac{1}{2}}\int_0^a \phi_a^+(x)x^{-\frac{1}{2}}\,dx$. So $\lim_{a\to0}\mathcal{A}_a(\frac{1}{2}) = \Gamma(\frac{1}{2}) = \sqrt{\pi}$. On the other hand $a\frac{d}{da}\mathcal{A}_a(\frac{1}{2}) = -\mu(a)\mathcal{A}_a(\frac{1}{2})$ and $\mu(a) = a\frac{d}{da}\log\frac{\det(1+H_a)}{\det(1-H_a)}$. so:

$$\mathcal{A}_a(\frac{1}{2}) = \sqrt{\pi}\,\widehat{A_a}(\frac{1}{2}) = \sqrt{\pi}\,\frac{\det(1-H_a)}{\det(1+H_a)} \tag{163}$$

We have $a\frac{d}{da}\|\mathcal{X}^a_{\frac{1}{2}}\|^2 = -2\mathcal{A}_a(\frac{1}{2})^2$. And $\mathcal{X}^a_{\frac{1}{2}} = \Gamma(\frac{1}{2})X^a_{\frac{1}{2}}$. So:

Theorem 17. *The squared-norm of the evaluator $f \mapsto X^a_{\frac{1}{2}}(f) = \int_a^\infty \frac{f(x)}{\sqrt{x}}\,dx$ on the Hilbert space K_a of square integrable functions vanishing on $(0,a)$ and with $\mathcal{H}$ transforms again vanishing on $(0,a)$ is:*

$$\|X^a_{\frac{1}{2}}\|^2 = 2\int_a^\infty \left(\det\frac{1-H_b}{1+H_b}\right)^2\frac{db}{b} \tag{164}$$

where H_a is the restriction of $\mathcal{H}$ to $L^2(0,a;dx)$.

It will be seen that $\det(1+H_a) = e^{a-\frac{1}{2}a^2}$ and $\det(1-H_a) = e^{-a-\frac{1}{2}a^2}$. Having spent a long time in the general set-up we now turn to determine explicitely what the functions ϕ_a^+, ϕ_a^-, etc... are.

7 The K-Bessel function in the theory of the $\mathcal{H}$ transform

Let us recall that we may define the $\mathcal{H}$ transform on all of $L^2(\mathbb{R};dx)$ through the formula $\widetilde{\mathcal{H}(f)}(\lambda) = \frac{i}{\lambda}\widetilde{f}(\frac{-1}{\lambda})$. This anticommutes with $f(x)\mapsto f(-x)$, and $\mathcal{H}$ leaves separately invariant $L^2(0,+\infty;dx)$ and $L^2(-\infty,0;dx)$. We defined the groups $\tau_a : f(x)\mapsto f(x-a)$ and $\tau_a^\# = \mathcal{H}\tau_a\mathcal{H}$. We observed that the two groups are mutually commuting, and that if the leftmost point of the support of f is at $\alpha(f)\geq 0$ then the leftmost point of the support of $\tau_b^\#(f)$, for any $b\geq 0$, more precisely for any $b \geq -\alpha(\mathcal{H}(f))$, is still exactly at $\alpha(f)$. From this we obtain the exact description of K_a:

Lemma 18. *One has* $K_a = \tau_a\tau_a^{\#}L^2(0,+\infty;dx)$.

Let now Q be the orthogonal projection $L^2(\mathbb{R};dx) \mapsto L^2(0,+\infty;dx)$. The orthogonal projection Q_a from $L^2(0,\infty;dx)$ to K_a is thus exactly $\tau_a\tau_a^{\#}Q\tau_{-a}\tau_{-a}^{\#}$. It will be easier to work with $R_a = Q\tau_{-a}\tau_{-a}^{\#}$, especially as we are interested in scalar products so we can skip the $\tau_a\tau_a^{\#}$ isometry. First, we obtain $g_t^a(x) = R_a(f_t)(x)$, for $f_t(x) = e^{-tx}$. The part of $\tau_{-a}(f_t)$ supported in $x<0$ will be sent by $\tau_{-a}^{\#}$ to a function supported again in $x<0$. We can forget about it and we have thus first $e^{-ta}e^{-tx}\mathbf{1}_{x>0}(x)$, whose $\mathcal{H}$ transform is $e^{-ta}\frac{1}{t}\exp(-\frac{x}{t})$, which we translate to the left, again we cut the part in $x<0$, and we reapply $\mathcal{H}$, this gives $g_t^a(x) = e^{-a(t+\frac{1}{t})}e^{-tx}\mathbf{1}_{x>0}(x)$. In other words we have used in this computation:

$$Q\tau_{-a}\tau_{-a}^{\#} = \mathcal{H}Q\tau_{-a}\mathcal{H}Q\tau_{-a} \qquad (a\geq 0) \tag{165}$$

The orthogonal projection $f_t^a := Q_a(f_t)$ of $f_t(x) = e^{-tx}\mathbf{1}_{x>0}(x)$ to K_a is thus $\tau_a\tau_a^{\#}(g_t^a)$. We can then compute exactly the Fourier transform of f_t^a as $f_t^a(i\tau) = (e^{-\tau x}, \tau_a\tau_a^{\#}(g_t^a))_{L^2(\mathbb{R})}$ which is also $(\tau_{-a}^{\#}\tau_{-a}e^{-\tau x}, g_t^a)_{L^2(\mathbb{R})} = (g_\tau, g_t^a) = e^{-a(t+\frac{1}{t})}e^{-a(\tau+\frac{1}{\tau})}\frac{1}{t+\tau}$. Hence:

Lemma 19. *The orthogonal projection* f_t^a *to* K_a *of* $e^{-tx}\mathbf{1}_{x>0}(x)$ *has its Fourier transform* $\widetilde{f_t^a}(\lambda)$ *which is given as:*

$$\widetilde{f_t^a}(i\tau) = e^{-a(t+\frac{1}{t}+\tau+\frac{1}{\tau})}\frac{1}{t+\tau} \tag{166}$$

The Gamma completed right Mellin transform $\mathcal{F}_t^a(s)$ of f_t^a is the left Mellin transform of $\widetilde{f_t^a}(i\tau)$.

$$\int_a^\infty f_t^a(x)\mathcal{X}_s^a(x)\,dx = \mathcal{F}_t^a(s) = e^{-a(t+\frac{1}{t})}\int_0^\infty e^{-a(\tau+\frac{1}{\tau})}\frac{\tau^{s-1}}{t+\tau}\,d\tau \tag{167}$$

Let us write W_s^a for the element of $L^2(0,+\infty;dx)$ such that $\tau_a\tau_a^{\#}W_s^a = \mathcal{X}_s^a$. We have $\mathcal{F}_t^a(s) = (\mathcal{X}_s^a, f_t^a) = (W_s^a, g_t^a) = e^{-a(t+\frac{1}{t})}\int_0^\infty W_s^a(x)e^{-tx}\,dx$. So the Laplace transform of $W_s^a(x)$ is exactly:

$$\int_0^\infty W_s^a(x)e^{-tx}\,dx = \int_0^\infty e^{-a(\tau+\frac{1}{\tau})}\frac{\tau^{s-1}}{t+\tau}\,d\tau \tag{168}$$

Writing $\frac{1}{t+\tau} = \int_0^\infty e^{-(t+\tau)x}\,dx$, we recover $W_s^a(x)$ as:

$$W_s^a(x) = \int_0^\infty e^{-a(\tau+\frac{1}{\tau})}\tau^{s-1}e^{-\tau x}\,d\tau \tag{169}$$

Then we obtain $\int_0^\infty W_s^a(x)W_z^a(x)\,dx$ which is nothing else than $\mathcal{X}_a(s,z)$:

Theorem 20. *The (analytic) reproducing kernel associated with the space of the completed right Mellin transforms of the elements of* K_a *is*

$$\mathcal{X}_a(s,z) = \iint_{[0,+\infty)^2} e^{-a(t+\frac{1}{t}+u+\frac{1}{u})}\frac{t^{s-1}u^{z-1}}{t+u}\,dtdu \tag{170}$$

Here is a shortened argument: the analytic reproducing kernel $\mathcal{X}_a(s,z)$ is the completed right Mellin transform of $\mathcal{X}_s^a(x)$, so this is $\int_0^\infty(\mathcal{X}_s^a, e^{-tx})t^{s-1}\,dt$. But for $\Re(s) > \frac{1}{2}$, $(\mathcal{X}_s^a, e^{-tx}) = \Gamma(s)(Q_a(x^{-s}\mathbf{1}_{x>a}), e^{-tx}) = \Gamma(s)(x^{-s}\mathbf{1}_{x>a}, f_t^a) = \mathcal{F}_t^a(s)$ (Q_a is the orthogonal projection to K_a). This gives again (170).

To proceed further, we compute $(s+z-1)\mathcal{X}_a(s,z)$. Using integration by parts, multiplication by s (resp. z) is converted into $-t\frac{d}{dt}$ (resp. $-u\frac{d}{du}$; there are no boundary terms.)

$$s\mathcal{X}_a(s,z) = \iint_{[0,+\infty)^2} (a(t-\frac{1}{t}) + \frac{t}{t+u})e^{-a(t+\frac{1}{t}+u+\frac{1}{u})}\frac{t^{s-1}u^{z-1}}{t+u}\,dtdu \tag{171a}$$

$$z\mathcal{X}_a(s,z) = \iint_{[0,+\infty)^2} (a(u-\frac{1}{u}) + \frac{u}{t+u})e^{-a(t+\frac{1}{t}+u+\frac{1}{u})}\frac{t^{s-1}u^{z-1}}{t+u}\,dtdu \tag{171b}$$

$$\begin{aligned}(s+z-1)\mathcal{X}_a(s,z) &= a\iint_{[0,+\infty)^2} (t-\frac{1}{t}+u-\frac{1}{u})e^{-a(t+\frac{1}{t}+u+\frac{1}{u})}\frac{t^{s-1}u^{z-1}}{t+u}\,dtdu \\ &= a\int_0^\infty e^{-a(t+\frac{1}{t})}t^{s-1}\,dt\int_0^\infty e^{-a(u+\frac{1}{u})}u^{z-1}\,du \\ &\quad - a\int_0^\infty e^{-a(t+\frac{1}{t})}t^{s-2}\,dt\int_0^\infty e^{-a(u+\frac{1}{u})}u^{z-2}\,du\end{aligned} \tag{171c}$$

The K-Bessel function is $K_s(x) = \frac{1}{2}\int_0^\infty e^{-x\frac{1}{2}(t+\frac{1}{t})}t^{s-1}\,dt = \int_0^\infty e^{-x\cosh u}\cosh(su)\,du$. It is an even function of s. It has, for each $x>0$, all its zeros on the imaginary axis, and was used by Pólya in a famous work on functions inspired by the Riemann ξ-function and for which he proved the validity of the Riemann hypothesis [24, 25]. We have obtained the formula

$$\mathcal{X}_a(s,z) = \frac{E(s)E(z) - E(1-s)E(1-z)}{s+z-1} \qquad E(s) = 2\sqrt{a}K_s(2a) \tag{172}$$

To confirm $\mathcal{E}_a(s) = 2\sqrt{a}K_s(2a)$, let us define temporarily $A(s) = \frac{1}{2}\sqrt{a}\int_0^\infty e^{-a(t+\frac{1}{t})}(1+\frac{1}{t})t^{s-1}\,dt$ and $-iB(s) = \frac{1}{2}\sqrt{a}\int_0^\infty e^{-a(t+\frac{1}{t})}(1-\frac{1}{t})t^{s-1}\,dt$ which are respectively even and odd under $s\mapsto 1-s$ and are such that $E(z) = A(z) - iB(z)$. We have $\forall s,z\in\mathbb{C}\quad -iB(s)A(z) + A(s)(-iB(z)) = -i\mathcal{B}_a(s)\mathcal{A}_a(z) + \mathcal{A}_a(s)(-i\mathcal{B}_a(z))$ and considering separately the even and odd parts in z, we find that there exists a constant $k(a)$ such that $A(s) = k(a)\mathcal{A}_a(s)$ and $B(s) = k(a)^{-1}\mathcal{B}_a(s)$. Let us check that $\lim_{\sigma\to\infty}\frac{-iB(\sigma)}{A(\sigma)} = 1$. It is a corollary to $\lim_{\sigma\to\infty}K_\sigma(x)/K_{\sigma+1}(x) = 0$ which is elementary: $\int_{-\infty}^0 \exp(-x\cosh u)e^{\sigma u}\,du = O(1)$ $(\sigma\to+\infty)$, and for each $T>0$, $\int_T^\infty \exp(-x\cosh u)e^{\sigma u}\,du \geq T\exp(-x\cosh 3T)e^{2\sigma T}$, $\int_0^T \exp(-x\cosh u)e^{\sigma u}\,du \leq Te^{\sigma T}$, and combining we get $K_\sigma(x) = (1+o(1))\frac{1}{2}\int_T^\infty \exp(-x\cosh u)e^{\sigma u}\,du$. So $\limsup_{\sigma\to+\infty}\frac{K_\sigma(x)}{K_{\sigma+1}(x)} \leq e^{-T}$ for each $T>0$. Using (116), we then conclude $k(a) = 1$.

Let us examine the equality $\mathcal{E}_a(s) = 2\sqrt{a}K_s(2a) = \sqrt{a}\int_0^\infty \exp(-a(t+\frac{1}{t}))t^{s-1}\,dt$. It exhibits $\mathcal{E}_a$ as the left Mellin transform of $\sqrt{a}\exp(-a(t+\frac{1}{t}))$, so the distribution E_a is determined as the distribution whose Fourier transform is $\sqrt{a}\exp(i\,a(\lambda-\lambda^{-1}))$. Using τ_a and $\tau_a^\#$, this means exactly:

$$E_a = \sqrt{a}\,\tau_a^\#\,\tau_a\delta = \sqrt{a}\,\mathcal{H}\tau_a\mathcal{H}\delta(x-a) \tag{173}$$

We exploit the symmetry $K_s = K_{-s}$, which corresponds to $\widetilde{E_a}(\lambda) = \widetilde{E_a}(-\lambda^{-1}) = -i\lambda\widetilde{\mathcal{H}E_a}(\lambda)$, so the unexpected identity appears:

$$E_a = \frac{d}{dx}\mathcal{H}E_a \tag{174}$$

From (173) we read $\mathcal{H}E_a = \sqrt{a}\tau_a J_0(2\sqrt{ax}) = \sqrt{a}J_0(2\sqrt{a(x-a)})\mathbf{1}_{x>a}(x)$. Using (174), and recalling equations (96c) and (96d) we deduce:

$$\frac{\phi_a^+(x)+\phi_a^-(x)}{2} = J_0(2\sqrt{a(x-a)}) = I_0(2\sqrt{a(a-x})) \tag{175a}$$

$$\frac{\phi_a^+(x)-\phi_a^-(x)}{2} = \frac{\partial}{\partial x}J_0(2\sqrt{a(x-a)}) = \frac{\partial}{\partial x}I_0(2\sqrt{a(a-x)}) \tag{175b}$$

We knew already from equations (66a), (66b)! Summing up we have proven:

Theorem 21. *Let $\mathcal{H}$ be the self-reciprocal operator with kernel $J_0(2\sqrt{xy})$ on $L^2(0,\infty;dx)$. Let H_a be the restriction of $\mathcal{H}$ to $L^2(0,a;dx)$. The solutions to the integral equations $\phi_a^+ + H_a\phi_a^+ = J_0(2\sqrt{ax})$ and $\phi_a^- - H_a\phi_a^- = J_0(2\sqrt{ax})$ are the entire functions:*

$$\phi_a^+(x) = I_0(2\sqrt{a(a-x)}) + \frac{\partial}{\partial x} I_0(2\sqrt{a(a-x)}) \tag{176a}$$

$$\phi_a^-(x) = I_0(2\sqrt{a(a-x)}) - \frac{\partial}{\partial x} I_0(2\sqrt{a(a-x)}) \tag{176b}$$

One has $1-a = \phi_a^+(a) = \frac{d}{da}\log\det(1+H_a)$ and $1+a = \phi_a^-(a) = -\frac{d}{da}\log\det(1-H_a)$, and

$$\det(1+H_a) = e^{+a-\frac{1}{2}a^2} \tag{176c}$$

$$\det(1-H_a) = e^{-a-\frac{1}{2}a^2} \tag{176d}$$

The tempered distributions $A_a = \frac{\sqrt{a}}{2}(1+\mathcal{H})\phi_a^+$ and $-i\,B_a = \frac{\sqrt{a}}{2}(-1+\mathcal{H})\phi_a^-$, respectively invariant and anti-invariant under $\mathcal{H}$, are also given as:

$$A_a(x) = \frac{\sqrt{a}}{2}\left(\delta_a(x) + \mathbf{1}_{x>a}(x)\big(J_0(2\sqrt{a(x-a)}) + \frac{\partial}{\partial x}J_0(2\sqrt{a(x-a)})\big)\right) \tag{176e}$$

$$-iB_a(x) = \frac{\sqrt{a}}{2}\left(\delta_a(x) - \mathbf{1}_{x>a}(x)\big(J_0(2\sqrt{a(x-a)}) - \frac{\partial}{\partial x}J_0(2\sqrt{a(x-a)})\big)\right) \tag{176f}$$

Their Fourier transforms are $\int_{\mathbb{R}} e^{i\lambda x}A_a(x)\,dx = \frac{\sqrt{a}}{2}(1+\frac{i}{\lambda})\exp(ia(\lambda-\frac{1}{\lambda}))$ and $-i\int_{\mathbb{R}} e^{i\lambda x}B_a(x)\,dx = \frac{\sqrt{a}}{2}(1-\frac{i}{\lambda})\exp(ia(\lambda-\frac{1}{\lambda}))$. The Gamma completed right Mellin transforms are:

$$\Gamma(s)\widehat{A_a}(s) = \mathcal{A}_a(s) = \sqrt{a}(K_s(2a) + K_{s-1}(2a)) \tag{176g}$$

$$-i\Gamma(s)\widehat{B_a}(s) = -i\mathcal{B}_a(s) = \sqrt{a}(K_s(2a) - K_{s-1}(2a)) \tag{176h}$$

$$\mathcal{A}_a(s) - i\mathcal{B}_a(s) = \mathcal{E}_a(s) = 2\sqrt{a}\,K_s(2a) = \sqrt{a}\int_0^\infty e^{-a(t+\frac{1}{t})}t^{s-1}\,dt \tag{176i}$$

They verify the first order system, where $\mu(a) = a\phi^+(a) + a\phi^-(a) = 2a$:

$$\left(\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix} a\frac{d}{da} - \begin{bmatrix}0 & \mu(a)\\ \mu(a) & 0\end{bmatrix}\right)\begin{bmatrix}\mathcal{A}_a(s)\\ \mathcal{B}_a(s)\end{bmatrix} = -i(s-\frac{1}{2})\begin{bmatrix}\mathcal{A}_a(s)\\ \mathcal{B}_a(s)\end{bmatrix} \tag{176j}$$

The pair $\left[\begin{smallmatrix}\mathcal{A}_a(s)\\ \mathcal{B}_a(s)\end{smallmatrix}\right]$ is the unique solution of the first order system which is square-integrable with respect to $d\log(a)$ at $+\infty$. The total reflection against the exponential barriers at $\log(a)\to+\infty$ of the associated Schrödinger equations realizes $+\frac{\Gamma(1-s)}{\Gamma(s)}$ and $-\frac{\Gamma(1-s)}{\Gamma(s)}$ ($\Re(s) = \frac{1}{2}$) as scattering matrices.

From (163) we have $\mathcal{A}_a(\frac{1}{2}) = \sqrt{\pi}\,e^{-2a}$. To normalize $\mathcal{A}_a$ according to $\mathcal{A}_a(\frac{1}{2}) = 1$, we would have to make the replacement $\mathcal{A}_a \to \pi^{-\frac{1}{2}}e^{2a}\mathcal{A}_a$ and $\mathcal{B}_a \to \sqrt{\pi}\,e^{-2a}\mathcal{B}_a$ and the expression of $\mathcal{E}_a$ in terms of the K-Bessel function would be less simple. Let us also note that according to (116) we must have $\frac{K_{\sigma-1}(2a)}{K_\sigma(2a)} \sim_{\sigma\to+\infty} \frac{a}{\sigma}$.

Regarding the isometric expansion, as given in theorem 15, we apply it to a function $F(s) = \int_0^\infty k(x)x^{-s}\,dx$ such that $\frac{d}{dx}x\,k(x)$ as a distribution on $\mathbb{R}$ is in L^2. Using the L^2-function $\frac{1}{s}\widehat{A_a}(s)$, which is the Mellin transform of the function $C_a(x) = \frac{1}{x}\int_0^x A_a(x)\,dx$, and the Parseval identity we obtain $\alpha(u) = 2\int_0^\infty(-x\frac{d}{dx}k(x))C_a(x)\,dx$ as an absolutely convergent integral. The square integrable function $C_a(x)$ is explicitely:

$$C_a(x) = \frac{\sqrt{a}}{2x}\mathbf{1}_{x>a}(x)\Big(J_0(2\sqrt{a(x-a)}) + \sqrt{\frac{x-a}{a}}J_1(2\sqrt{a(x-a)})\Big) \tag{177}$$

And under the hypothesis made on $x\frac{d}{dx}k(x)$ we obtain the existence of:

$$\alpha(u) = \lim_{X\to\infty} \sqrt{a}k(a) - 2Xk(X)C_a(X) + \sqrt{a}\int_a^X k(x)(1+\frac{\partial}{\partial x})J_0(2\sqrt{a(x-a)})\,dx \tag{178}$$

Let us observe that $k(x) = \int_x^\infty \frac{l(y)}{y}\,dy$ with $l(y)\in L^2$, so $|k(x)|^2 \le \frac{C}{x}$ and $Xk(X)C_a(X) = O(X^{-\frac14})$. Hence:

$$\alpha(u) = \sqrt{a}k(a) + \sqrt{a}\int_a^{\to\infty} k(x)(1+\frac{\partial}{\partial x})J_0(2\sqrt{a(x-a)})\,dx \tag{179}$$

Comparing with equation (20a) we see that the $f(y)$ defined there is related to $\alpha(u)$, $u = \log(a)$ by the formula $f(y) = \frac{1}{2\sqrt{a}}\alpha(\log(a))$, $a = \frac{y}{2}$, so $|f(y)|^2\,dy = \frac{1}{4a}|\alpha(\log(a))|^2 2\,da = |\alpha(\log(a))|^2\frac12 d\log(a)$. Similarly we obtain $\beta(u)$:

$$\beta(u) = \sqrt{a}k(a) + \sqrt{a}\int_a^{\to\infty} k(x)(-1+\frac{\partial}{\partial x})J_0(2\sqrt{a(x-a)})\,dx \tag{180}$$

and comparing with (20b), $g(y) = \frac{1}{2\sqrt{a}}\beta(\log(a))$, $|g(y)|^2\,dy = |\beta(\log(a))|^2\frac12 d\log(a)$. So according to theorem 15 we do have equation (20d):

$$\int_0^\infty (|f(y)|^2 + |g(y)|^2)\,dy = \int_0^\infty |k(x)|^2\,dx \tag{181}$$

From 15 the assignment $k\to(\alpha,\beta)$ extends to a unitary identification $L^2(\Re(s)=\frac12;\frac{|ds|}{2\pi})\widetilde{\to}L^2(\mathbb{R}\to\mathbb{C}^2;\frac{du}{2})$, which has the property $\mathcal{H}(k)\to(\alpha,-\beta)$. In order to complete the proof of the isometric expansion, it remains to check the equation (20c) which expresses k in terms of f and g. According to 15 we recover $k(x)$ has the inverse Mellin transform of $\int_{\mathbb{R}}(\alpha(u)\,2\widehat{A_a}(s) + \beta(u)\,2(-i\widehat{B_a}(s)))\,\frac{du}{2}$. Expressing this in terms of $f(y)$ and $g(y)$, $y = 2a$, $u = \log(a)$, this means the identity of distributions, where we suppose for simplicity that $f(y)$ and $g(y)$ have compact support in $(0,+\infty)$ (as usual, this means having support away from 0 as well as ∞.):

$$\begin{aligned} k(x) &= \int_0^\infty \left(2\sqrt{\frac{y}{2}}f(y)\,2A_{\frac{y}{2}}(x) + 2\sqrt{\frac{y}{2}}g(y)\,2(-iB_{\frac{y}{2}}(x))\right)\frac{dy}{2y} \\ &= 2\int_0^\infty (\sqrt{y}f(2y)\,A_y(x) + \sqrt{y}g(2y)\,2(-iB_y(x)))\,\frac{dy}{y} \end{aligned} \tag{182}$$

Then imagining that we are integrating against a test function $\psi(x)$ and using Fubini we obtain:

$$\begin{aligned} &2\int_0^\infty \sqrt{y}f(2y)\,A_y(x)\frac{dy}{y} \\ &= \int_0^\infty f(2y)\left(\delta(x-y) + \mathbf{1}_{x>y}\Big(J_0(2\sqrt{y(x-y)}) - \sqrt{\frac{y}{x-y}}J_1(2\sqrt{y(x-y)})\Big)\right)\,dy \end{aligned} \tag{183}$$

$$\begin{aligned}&= f(2x) + \int_0^x \Big(J_0(2\sqrt{y(x-y)}) - \sqrt{\frac{y}{x-y}}J_1(2\sqrt{y(x-y)})\Big)f(y)\,dy\\ &= f(2x) + \frac{1}{2}\int_0^{2x} \Big(J_0(\sqrt{y(2x-y)}) - \sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\Big)f(y)\,dy\end{aligned} \tag{184}$$

Proceeding similarly with $g(y)$ one obtains for $2\int_0^\infty \sqrt{y}g(2y)(-i\,B_y(x))\frac{dy}{y}$:

$$g(2x) - \frac{1}{2}\int_0^{2x} \Big(J_0(\sqrt{y(2x-y)}) + \sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\Big)f(y)\,dy \tag{185}$$

Combining (183) and (185) in the formula (182) for $k(x)$ we obtain equation (20c).

8 The reproducing kernel and differential equations for the extended spaces

Let $L_a \subset L^2(0,\infty;dx)$ be the Hilbert space of square integrable functions f which are constant in $(0,a)$ and with their $\mathcal{H}$-transforms again constant in $(0,a)$. The distribution $x\frac{d}{dx}\frac{d}{dx}x\,f = \frac{d}{dx}x\,x\frac{d}{dx}f$ vanishes in $(0,a)$ and its $\mathcal{H}$ transform does too. So $s(s-1)\widehat{f}(s)$ is an entire function with trivial zeros at $-\mathbb{N}$. The Hilbert space of the functions $s(s-1)\Gamma(s)\widehat{f}(s)$ satisfies the axioms of [2]; we prove everything according to the methods developed in the earlier chapters. Our goal is to determine the evaluators and reproducing kernel for the spaces L_a.

For $f \in L_a$, $\widehat{f}(s)$ is a meromorphic function with at most a pole at $s=1$ and also $\widehat{f}(0)$ does not necessarily vanish. The Mellin-Plancherel transform $\int_0^\infty f(x)x^{-s}\,dx = \int_0^a c(f)x^{-s}\,dx + \int_a^\infty f(x)x^{-s}\,dx$ has polar part $-\frac{c(f)}{s-1}$. Let us write $(f,Y_1) = -c(f) = \mathrm{Res}(\widehat{f}(s), s=1) = s(s-1)\Gamma(s)\widehat{f}(s)|_{s=1}$. This defines an element $Y_1 \in L_a$. We define also $\mathcal{Y}_1^a = \Gamma(1)Y_1 = Y_1$. Then $(f,\mathcal{Y}_1^a) = s(s-1)\Gamma(s)\widehat{f}(s)|_{s=1}$. We also define $\mathcal{Y}_0^a$ as the vector such that $(f,\mathcal{Y}_0^a) = s(s-1)\Gamma(s)\widehat{f}(s)|_{s=0} = -\widehat{f}(0)$. One observes $(f,\mathcal{H}(Y_1)) = (\mathcal{H}(f),Y_1) = s(s-1)\Gamma(s)\widehat{\mathcal{H}(f)}(s)|_{s=1} = s(s-1)\Gamma(1-s)\widehat{f}(1-s)|_{s=1} = -\widehat{f}(0) = (f,\mathcal{Y}_0^a)$ so $\mathcal{Y}_0^a = \mathcal{H}(\mathcal{Y}_1^a)$. To lighten the notation we sometimes write $\mathcal{Y}_1$ and $\mathcal{Y}_0$ instead $\mathcal{Y}_1^a$ and $\mathcal{Y}_0^a$ when no confusion can arise.

We will also consider the vectors $X_s^\times \in L_a$ such that $\forall f \in L_a\ \widehat{f}(s) = (X_s^\times, f)$.[30] The orthogonal projection of $X_s^\times$ to $K_a \subset L_a$ is X_s. Let us look more closely at this orthogonal projection. First let N_a be the (closed) vector space sum $L^2(0,a;dx) + \mathcal{H}L^2(0,a;dx)$. Inside N_a we have the codimension two space M_a defined as the sum of $(\mathbf{1}_{0<x<a})^\perp \cap L^2(0,a;dx)$ and of its image under $\mathcal{H}$. Finally, let R_a be the orthogonal complement in N_a of M_a, which has dimension two. For a function f to belong to L_a it is necessary and sufficient that its orthogonal projection to N_a be perpendicular to the functions in $L^2(0,a;dx)$ which are perpendicular to $\mathbf{1}_{0<x<a}$, and the same for the $\mathcal{H}$-transform, so this means exactly that its orthogonal projection to N_a belongs to R_a. So we have the orthogonal decomposition of $L^2(0,\infty;dx)$ into the sum of the three spaces K_a, R_a and M_a and $L_a = K_a \oplus R_a$. For $f \in L_a$ to be in K_a it is necessary and sufficient that $c(f) = -(f,\mathcal{Y}_1^a) = 0$ and the same for $c(\mathcal{H}(f))$, so this means that $\{\mathcal{Y}_1^a, \mathcal{Y}_0^a\}$ is a basis of R_a. The function $\mathcal{Y}_1^a$ belongs to $N_a = L^2(0,a;dx) + \mathcal{H}L^2(0,a;dx)$ and as such is uniquely written as $u_1 + \mathcal{H}v_1$. As $\mathcal{Y}_1^a \in L_a$ we have constants $\alpha,\beta \in \mathbb{C}$ such that:

$$u_1 + H_a v_1 = -\alpha \tag{186a}$$
$$H_a u_1 + v_1 = -\beta \tag{186b}$$

[30] sometimes written $X_s^{a\times}$.

where we recall that P_a is the restriction to $(0,a)$ and $H_a = P_a\mathcal{H}P_a$, $D_a = H_a^2$. From what was said previously $\alpha = (\mathcal{Y}_1^a, \mathcal{Y}_1^a)$ and $\beta = (\mathcal{H}(\mathcal{Y}_1^a), \mathcal{Y}_1^a) = (\mathcal{Y}_0^a, \mathcal{Y}_1^a)$. We have thus:

$$u_1 = (1-D_a)^{-1}(-\alpha\mathbf{1}_{0<x<a} + \beta H_a(\mathbf{1}_{0<x<a})) \tag{186c}$$
$$v_1 = (1-D_a)^{-1}(+\alpha H_a(\mathbf{1}_{0<x<a}) - \beta\mathbf{1}_{0<x<a}) \tag{186d}$$

Also the function $\mathcal{Y}_1^a$ may be obtained as the orthogonal projection to L_a of $-\frac{1}{a}\mathbf{1}_{0<x<a}$. Indeed it follows from what has been seen above that for any element $f \in L_a$, $(f, \mathcal{Y}_1^a) = -\frac{1}{a}\int_0^a f(x)\,dx$. As the function $-\frac{1}{a}\mathbf{1}_{0<x<a}$ already belongs to N_a, we have

$$-\frac{1}{a}\mathbf{1}_{0<x<a} = u_1 + \mathcal{H}v_1 + u_2 + \mathcal{H}v_2 \tag{187}$$

where $u_2 + \mathcal{H}v_2$ belongs to M_a, which means that $u_2 \in L^2(0,a;dx)$ verifies $\int_0^a u_2(x)\,dx = 0$ and $v_2 \in L^2(0,a;dx)$ verifies $\int_0^a v_2(x)\,dx = 0$. But there is unicity so we have exactly

$$u_1 + u_2 = -\frac{1}{a}\mathbf{1}_{0<x<a} \qquad v_1 + v_2 = 0 \tag{188}$$

And we deduce:

$$\int_0^a u_1(x)\,dx = -1 \qquad \int_0^a v_1(x)\,dx = 0 \tag{189}$$

So α and β are determined as the solutions of the system:

$$\alpha(\mathbf{1}_{0<x<a}, (1-D_a)^{-1}\mathbf{1}_{0<x<a}) - \beta(\mathbf{1}_{0<x<a}, (1-D_a)^{-1}H_a\mathbf{1}_{0<x<a}) = 1 \tag{190a}$$
$$\alpha(\mathbf{1}_{0<x<a}, (1-D_a)^{-1}H_a\mathbf{1}_{0<x<a}) - \beta(\mathbf{1}_{0<x<a}, (1-D_a)^{-1}\mathbf{1}_{0<x<a}) = 0 \tag{190b}$$

We thus have:

Proposition 22. *Let $p(a)$ and $q(a)$ be defined as*

$$p(a) = \int_0^a (1-D_a)^{-1}(\mathbf{1}_{0<x<a})(x)\,dx \tag{191a}$$
$$q(a) = \int_0^a (1-D_a)^{-1}H_a(\mathbf{1}_{0<x<a})(x)\,dx \tag{191b}$$

then:

$$\begin{bmatrix} p(a) & -q(a) \\ -q(a) & p(a) \end{bmatrix} \begin{bmatrix} (\mathcal{Y}_1,\mathcal{Y}_1) & (\mathcal{Y}_1,\mathcal{Y}_0) \\ (\mathcal{Y}_0,\mathcal{Y}_1) & (\mathcal{Y}_0,\mathcal{Y}_0) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{191c}$$

The evaluators $\mathcal{Y}_1^a$ and $\mathcal{Y}_0^a = \mathcal{H}(\mathcal{Y}_1^a)$ are given as $u_1 + \mathcal{H}v_1$ and $\mathcal{H}u_1 + v_1$ with:

$$u_1 = -(\mathcal{Y}_1,\mathcal{Y}_1)(1-D_a)^{-1}(\mathbf{1}_{0<x<a}) + (\mathcal{Y}_0,\mathcal{Y}_1)(1-D_a)^{-1}H_a(\mathbf{1}_{0<x<a}) \tag{191d}$$
$$v_1 = -(\mathcal{Y}_0,\mathcal{Y}_1)(1-D_a)^{-1}(\mathbf{1}_{0<x<a}) + (\mathcal{Y}_0,\mathcal{Y}_0)(1-D_a)^{-1}H_a(\mathbf{1}_{0<x<a}) \tag{191e}$$

We have introduced, for $s \neq 0,1$, $X_s^\times$ as the evaluator $\widehat{f}(s)$ for functions in L_a. We shall write $\mathcal{X}_s^\times = \Gamma(s)X_s^\times$ and then $\mathcal{Y}_s = s(s-1)\mathcal{X}_s^\times$. This is compatible with our previous definitions of $\mathcal{Y}_1^a$ and $\mathcal{Y}_0^a$. We note that the orthogonal projection of $\mathcal{X}_s^\times$ to K_a is $\mathcal{X}_s$. So we may write

$$\mathcal{X}_s^\times = \mathcal{X}_s + \lambda(s)\mathcal{Y}_1^a + \mu(s)\mathcal{Y}_0^a \tag{192}$$

We shall also write $\mathcal{Y}_a(s,z) = \int_0^\infty \mathcal{Y}_s^a(x)\mathcal{Y}_z^a(x)\,dx = z(z-1)\Gamma(z)\widehat{\mathcal{Y}_s^a}(z)$. One has $\mathcal{H}(\mathcal{Y}_s^a) = \mathcal{Y}_{1-s}^a$ and $\mathcal{Y}_a(s,z) = \mathcal{Y}_a(1-s,1-z) = \mathcal{Y}_a(z,s)$. Taking the scalar products with $\mathcal{Y}_1^a$ and $\mathcal{Y}_0^a$ in (192) we obtain

$$\frac{1}{s(s-1)}\mathcal{Y}_a(1,s) = \lambda(s)\alpha + \mu(s)\beta \tag{193a}$$

$$\frac{1}{s(s-1)}\mathcal{Y}_a(1,1-s) = \lambda(s)\beta + \mu(s)\alpha \tag{193b}$$

$$\lambda(s) = \frac{1}{s(s-1)}(p\mathcal{Y}_a(1,s) - q\mathcal{Y}_a(1,1-s)) \tag{193c}$$

$$\mu(s) = \frac{1}{s(s-1)}(-q\mathcal{Y}_a(1,s) + p\mathcal{Y}_a(1,1-s)) = \lambda(1-s) \tag{193d}$$

Combining with (192), this gives:

$$\mathcal{Y}_s = s(s-1)\mathcal{X}_s + \mathcal{Y}_a(1,s)(p\mathcal{Y}_1^a - q\mathcal{Y}_0^a) + \mathcal{Y}_a(1,1-s)(-q\mathcal{Y}_1^a + p\mathcal{Y}_0^a) \tag{194}$$

$$\text{Let}\quad T_a(s) = p(a)\mathcal{Y}_a(1,s) - q(a)\mathcal{Y}_a(1,1-s) \tag{195}$$

Proposition 23. *The (analytic) reproducing kernel $\mathcal{Y}_a(s,z)$ of the extended space L_a is given by each of the following expressions:*

$$s(s-1)z(z-1)\mathcal{X}_a(s,z) + \begin{bmatrix}\mathcal{Y}_a(1,s) & \mathcal{Y}_a(1,1-s)\end{bmatrix}\begin{bmatrix}p(a) & -q(a)\\ -q(a) & p(a)\end{bmatrix}\begin{bmatrix}\mathcal{Y}_a(1,z)\\ \mathcal{Y}_a(1,1-z)\end{bmatrix} \tag{196a}$$

$$= s(s-1)z(z-1)\mathcal{X}_a(s,z) + \begin{bmatrix}T_a(s) & T_a(1-s)\end{bmatrix}\begin{bmatrix}\alpha(a) & \beta(a)\\ \beta(a) & \alpha(a)\end{bmatrix}\begin{bmatrix}T_a(z)\\ T_a(1-z)\end{bmatrix} \tag{196b}$$

$$= s(s-1)z(z-1)\mathcal{X}_a(s,z) + T_a(s)\mathcal{Y}_a(1,z) + T_a(1-s)\mathcal{Y}_a(1-z) \tag{196c}$$

A very important observation, before turning to the determination of the quantities $p(a)$ and $q(a)$ shall now be made. Let L be the unitary operator:

$$L(f)(x) = f(x) - \frac{1}{x}\int_0^x f(y)\,dy \tag{197}$$

It is the operator of multiplication by $\frac{s-1}{s}$ at the level of right Mellin transforms. Obviously it converts functions constant on $(0,a)$ into functions vanishing on $(0,a)$. Let us now consider the operator

$$\mathcal{H}^\diamond = L\,\mathcal{H}\,L^{-1} = L^2\,\mathcal{H} = \mathcal{H}L^{-2} \tag{198}$$

It is a unitary, self-adjoint, self-reciprocal, scale reversing operator whose kernel is easily computed to be

$$k^\diamond(xy) = J_0(2\sqrt{xy}) - 2\frac{J_1(2\sqrt{xy})}{\sqrt{xy}} + \frac{1-J_0(2\sqrt{xy})}{xy} = \sum_{n=0}^\infty (-1)^n\frac{n^2x^ny^n}{(n+1)!^2} \tag{199}$$

It has $L(e^{-x}) = (1+\frac{1}{x})e^{-x} - \frac{1}{x}$ as self-reciprocal function; the Mellin transform is $\frac{s-1}{s}\Gamma(1-s)$ which, multiplied by $s(s-1)$ gives $(1-s)^2\Gamma(1-s)$ which is the Mellin transform of a more convenient invariant function for $\mathcal{H}^\diamond$, the function $x(x-1)e^{-x}$. This function is the analog for $\mathcal{H}^\diamond$ of e^{-x} for $\mathcal{H}$. Let us now consider the space $L(L_a)$. It consists of the square integrable functions vanishing identically on $(0,a)$ and having $\mathcal{H}^\diamond$ transforms also identically zero on $(0,a)$. But then the entire theory applies to $\mathcal{H}^\diamond$ exactly as it did for $\mathcal{H}$, up to some minor details in the proofs where the function J_0 was really used like in Lemma 9 or proposition 11. We have for $\mathcal{H}^\diamond$ versions of all quantities previously considered for $\mathcal{H}$. To check that the proof of 11 may be adapted, we need to look at $k_1^\diamond(x) = \int_0^x k^\diamond(t)\,dt = \sqrt{x}J_1(2\sqrt{x}) + 2(J_0(2\sqrt{x})-1) + 2\int_0^{2\sqrt{x}} \frac{1-J_0(u)}{u}\,du = O_{x\to+\infty}(x^{\frac14})$. So we may employ lemma 8 as was done for $\mathcal{H}$. Proposition 11 and Theorem 12 thus hold. We must be careful that the operator L^{-1} is always involved when comparing functions or distributions related to $\mathcal{H}^\diamond$ with those related to $\mathcal{H}$. For example, one has $X_s^\times = \frac{s}{s-1}L^{-1}X_s^\diamond$ and $\mathcal{Y}_s = s(s-1)\mathcal{X}_s^\times = L^{-1}\mathcal{X}_s^\diamond$. The two types of Gamma completed Mellin transforms differ: for $\mathcal{H}$ we consider $\Gamma(s)\widehat{f}(s)$ while for $\mathcal{H}^\diamond$ we consider $s^2\Gamma(s)\widehat{g}(s)$. Indeed this is quite the coherent thing to do in order that:

$$s^2\Gamma(s)\widehat{g}(s) = s(s-1)\Gamma(s)\widehat{f}(s) \tag{200}$$

for $g = L(f)$. The bare Mellin transforms of elements of spaces $K_a^\diamond$ are not always entire in the complex plane: they may have a pole at $s=0$. After multiplying by $s^2\Gamma(s)$ which is the *left* Mellin transform of the self-invariant function $x(x-1)e^{-x}$, as $\Gamma(s)$ is the *left* Mellin transform of e^{-x}, we do obtain entire functions, whose trivial zeros are at -1, -2, $\dots$ (0 is not a trivial zero anymore.) From equation (200) we see that the (analytic) reproducing kernel $\mathcal{X}_a^\diamond(s,z)$ exactly coincides with the function $\mathcal{Y}_a(s,z)$ whose initial computation has been given in Proposition 23. Also the Schrödinger equations will realize $\pm\left(\frac{1-s}{s}\right)^2\frac{\Gamma(1-s)}{\Gamma(s)}$ as scattering matrices, and there will be an isometric expansion generalizing the de Branges-Rovnyak expansion to the spaces L_a. We will determine exactly the functions $\mathcal{A}_a^\diamond(s)$, $\mathcal{B}_a^\diamond(s)$, $\mathcal{E}_a^\diamond(s)$ and especially the function $\mu^\diamond(a)$. It will be seen that this is a more complicated function than the simple-minded $\mu(a) = 2a$...

The key now is to obtain the functions $p(a)$ and $q(a)$ defined in Proposition 22, and the function $\mathcal{Y}_a(1,s)$. It turns out that their computation also involves the quantities (we recall that $J_0^a(x) = J_0(2\sqrt{ax})$):

$$r(a) = 1 + \int_0^a ((1-D_a)^{-1}H_a\cdot J_0^a)(x)\,dx \tag{201a}$$

$$s(a) = \int_0^a ((1-D_a)^{-1}\cdot J_0^a)(x)\,dx \tag{201b}$$

In order to compute r, s, p, q we shall need the already defined functions ϕ_a^+ ($=(1+H_a)^{-1}J_0^a$ on $(0,a)$), ϕ_a^-, ($=(1-H_a)^{-1}J_0^a$) as well as the entire functions ψ_a^+ and ψ_a^- verifying:

$$\psi_a^+ + \mathcal{H}P_a\psi_a^+ = 1 \tag{202a}$$

$$\psi_a^- - \mathcal{H}P_a\psi_a^- = 1 \tag{202b}$$

We have $r(a) = 1 + \frac12\int_0^a(-\phi_a^+(x)+\phi_a^-(x))\,dx$, and we know explicitely $\phi_a^\pm$. But, we shall proceed in a more general manner. First we recall the differential equations (121a), (121b) which are verified by $\phi_a^\pm$ (where $\delta_x = x\frac{\partial}{\partial x}+\frac12$):

$$a\frac{\partial}{\partial a}\phi_a^+ = +\delta_x\phi_a^- - (\mu(a)+\frac12)\phi_a^+ \tag{203a}$$

$$a\frac{\partial}{\partial a}\phi_a^- = +\delta_x\phi_a^+ + (\mu(a)-\frac12)\phi_a^- \tag{203b}$$

We compute $ar'(a) = a\frac{-\phi_a^+(a)+\phi_a^-(a)}{2} + \frac12\int_0^a(x\frac{\partial}{\partial x}+1)(\phi_a^+(x)-\phi_a^-(x)) + \mu(a)(\phi_a^+(x)+\phi_a^-(x))\,dx$ and simplifying this gives exactly $ar'(a) = \mu(a)\frac12\int_0^a(\phi_a^+(x)+\phi_a^-(x))\,dx = s(a)$. Similarly starting with

$s(a) = \frac{1}{2}\int_0^a(\phi_a^+(x)+\phi_a^-(x))\,dx$ we obtain $as'(a) = \frac{1}{2}\mu(a)+\frac{1}{2}\int_0^a(x\frac{\partial}{\partial x}(\phi_a^+(x)+\phi_a^-(x))+\mu(a)(-\phi_a^+(x)+\phi_a^-(x)))\,dx$ which gives $\mu(a)r(a)-s(a)$. So the quantities r and s verify the system:

$$ar'(a) = \mu(a)s(a) \tag{204a}$$
$$(as)'(a) = \mu(a)r(a) \tag{204b}$$

Either solving the system taking into account the behavior as $a \to 0$ or using the explicit formulas for $\phi_a^\pm$ we obtain in this specific instance of the study of $\mathcal{H}$, for which $\mu(a) = 2a$, that $r(a) = I_0(2a)$ and $s(a) = I_1(2a)$.

From (202a) and (202b) we obtain two types of differential equations, either involving $x\frac{\partial}{\partial x}$ or $a\frac{\partial}{\partial a}$. From $\psi_a^+(x) + \int_0^a J_0(2\sqrt{xy})\psi_a^+(x)\,dx = 1$, we obtain $(1+\mathcal{H}P_a)a\frac{\partial}{\partial a}\psi_a^+(x) = -a\psi_a^+(a)J_0^a$. We do similarly with ψ_a^- and deduce:

$$a\frac{\partial}{\partial a}\psi_a^+(x) = -a\psi_a^+(a)\phi_a^+(x) \tag{205a}$$
$$a\frac{\partial}{\partial a}\psi_a^-(x) = +a\psi_a^-(a)\phi_a^-(x) \tag{205b}$$

Regarding the differential equations with $x\frac{\partial}{\partial x}$, which we shall actually not use, the computation is done using only the fact that the kernel is a function of xy so $x\frac{\partial}{\partial x}J_0(2\sqrt{xy}) = y\frac{\partial}{\partial y}J_0(2\sqrt{xy})$. We only state the result:

$$(x\frac{\partial}{\partial x}+\frac{1}{2})\psi_a^+(x) = \frac{1}{2}\psi_a^-(x) - a\psi_a^+(a)\phi_a^-(x) \tag{206a}$$
$$(x\frac{\partial}{\partial x}+\frac{1}{2})\psi_a^-(x) = \frac{1}{2}\psi_a^+(x) + a\psi_a^-(a)\phi_a^+(x) \tag{206b}$$

Let us now turn to the quantities $p(a)$ and $q(a)$. We have $p(a) = \int_0^a(1-D_a)^{-1}(\mathbf{1}_{0<x<a})(x)\,dx = \frac{1}{2}\int_0^a(\psi_a^+(x)+\psi_a^-(x))\,dx$. So $p'(a) = \frac{1}{2}(\psi_a^+(a)+\psi_a^-(a)) - \frac{1}{2}\psi_a^+(a)\int_0^a\phi_a^+(x)\,dx + \frac{1}{2}\psi_a^-(a)\int_0^a\phi_a^-(x)\,dx$. Reorganizing this gives:

$$p'(a) = \frac{\psi_a^+(a)+\psi_a^-(a)}{2}(1+\int_0^a\frac{-\phi_a^+(x)+\phi_a^-(x)}{2}\,dx) + \frac{-\psi_a^+(a)+\psi_a^-(a)}{2}\int_0^a\frac{+\phi_a^+(x)+\phi_a^-(x)}{2}\,dx \tag{207}$$

We remark that from the integral equations defining $\psi_a^\pm$ we have $\psi_a^+(a) = 1-\int_0^a J_0(2\sqrt{ax})\psi_a^+(x)\,dx = 1-\int_0^a\phi_a^+(x)\,dx$ and $\psi_a^-(a) = 1+\int_0^a J_0(2\sqrt{ax})\psi_a^-(x)\,dx = 1+\int_0^a\phi_a^-(x)\,dx$. So $\frac{\psi_a^+(a)+\psi_a^-(a)}{2} = r(a)$ and $\frac{-\psi_a^+(a)+\psi_a^-(a)}{2} = s(a)$. Hence the quantity $p(a)$ verifies:

$$p'(a) = r(a)^2 + s(a)^2 \tag{208}$$

With exactly the same method one obtains:

$$q'(a) = 2r(a)s(a) \tag{209}$$

Let us observe that $q(a) = \frac{1}{2}\int_0^a((1-H_a)^{-1}-(1+H_a)^{-1})(1)\,dx = O(a^2)$ and $p(a) = \frac{1}{2}\int_0^a((1+H_a)^{-1}+(1-H_a)^{-1})(1)\,dx \sim_{a\to 0} a$. So $(p\pm q) \sim_{a\to 0} a$. Also $r(a) \sim_{a\to 0} 1$ and $s(a) \sim_{a\to 0} a$. The equation for $p(a)$ can be integrated:

$$p(a) = a(r(a)^2 - s(a)^2) \tag{210}$$

Indeed this has the correct derivative. Regarding $q(a)$ the situation is different, one has $q' = 2rs = \frac{2a}{\mu}rr'$ so in the special case considered here, and only in that case we have $q(a) = \frac{1}{2}(r(a)^2-1)$. Summing up:

Proposition 24. *The quantities $r(a)$, $s(a)$, $p(a)$ and $q(a)$ verify the differential equations $ar'(a) = \mu(a)s(a)$, $as'(a)+s(a) = \mu(a)r(a)$, $p'(a) = r(a)^2+s(a)^2$, $q'(a) = 2r(a)s(a)$, $p(a) = a(r(a)^2-s(a)^2)$. In the special case of the $\mathcal{H}$ transform one has:*

$$r(a) = I_0(2a) \tag{211a}$$
$$s(a) = I_1(2a) \tag{211b}$$
$$p(a) = a(I_0^2(2a) - I_1^2(2a)) \tag{211c}$$
$$q(a) = \frac{1}{2}(I_0^2(2a) - 1) \tag{211d}$$

We now need to determine $\mathcal{Y}_a(1,s) = s(s-1)\Gamma(s)\widehat{Y_1}(s)$. There holds $Y_1 = u_1 + \mathcal{H}v_1 = -\alpha\mathbf{1}_{0<x<a} + \mathbf{1}_{x>a}\mathcal{H}v_1$. So $\widehat{Y_1}(s) = \alpha\frac{a^{1-s}}{s-1} + \int_a^\infty(\mathcal{H}v_1)(x)x^{-s}\,dx$. Then $\int_a^\infty(\mathcal{H}v_1)(x)x^{-s}\,dx = \int_0^\infty v_1(x)g_s(x)\,dx = \int_0^a v_1(x)g_s(x)\,dx$, where the function g_s from (73) has been used. Recalling from (76a), (76b) the analytic functions u_s, equal to $-(1-D_a)^{-1}H_a(g_s)$ on $(0,a)$, and v_s, equal to $(1-D_a)^{-1}P_a(g_s)$ on $(0,a)$, and using (186d) and self-adjointness we obtain

$$\int_0^a v_1(x)g_s(x)\,dx = -\alpha\int_0^a u_s(x)\,dx - \beta\int_0^a v_s(x)\,dx \tag{212}$$

Let us now recall that we computed ((83a)) $(x\frac{\partial}{\partial x}+s)u_s$ and found it to be on the interval $(0,a)$ given as $-av_s(a)(1-D_a)^{-1}(J_0^a) - a(a^{-s}+u_s(a))(1-D_a)^{-1}H_a(J_0^a)$. Integrating and also using equations (99) and (102) we obtain

$$\sqrt{a}\widehat{E_a}(s) - a^{1-s} + (s-1)\int_0^a u_s(x)\,dx = -\sqrt{a}\widehat{\mathcal{H}(E_a)}(s)s(a) - \sqrt{a}\widehat{E_a}(s)(r(a)-1) \tag{213}$$

$$\int_0^a u_s(x)\,dx = \sqrt{a}\,\frac{a^{\frac{1}{2}-s} - \widehat{E_a}(s)r(a) - \widehat{\mathcal{H}(E_a)}(s)s(a)}{s-1} \tag{214}$$

We have similarly ((83b)) $(x\frac{\partial}{\partial x}+1-s)v_s = -\sqrt{a}\widehat{E_a}(s)(1-D_a)^{-1}(J_0^a) - \sqrt{a}\widehat{\mathcal{H}(E_a)}(s)(1-D_a)^{-1}H_a(J_0^a)$ so integration gives $av_s(a) - s\int_0^a v_s(x)\,dx = -\sqrt{a}\widehat{E_a}(s)s(a) - \sqrt{a}\widehat{\mathcal{H}(E_a)}(s)(r(a)-1)$ hence

$$\int_0^a v_s(x)\,dx = \sqrt{a}\,\frac{\widehat{E_a}(s)s(a) + \widehat{\mathcal{H}(E_a)}(s)r(a)}{s} \tag{215}$$

Combining (214), (215) with (212), and using $\mathcal{Y}_a(1,s) = s(s-1)\Gamma(s)\widehat{Y_1^a}(s)$:

$$\widehat{Y_1}(s) = \sqrt{a}\,\widehat{E_a}(s)\Big(\frac{\alpha(a)r(a)}{s-1} - \frac{\beta(a)s(a)}{s}\Big) + \sqrt{a}\,\widehat{\mathcal{H}(E_a)}(s)\Big(\frac{\alpha(a)s(a)}{s-1} - \frac{\beta(a)r(a)}{s}\Big) \tag{216a}$$

$$\mathcal{Y}_a(1,s) = \sqrt{a}\,\Big(s\alpha(a)(\mathcal{E}_a(s)r(a) + \mathcal{E}_a(1-s)s(a)) + (1-s)\beta(a)(\mathcal{E}_a(s)s(a) + \mathcal{E}_a(1-s)r(a))\Big) \tag{216b}$$

Proposition 25. *The functions $\mathcal{Y}_a(1,s)$ and $\mathcal{Y}_a(1,1-s)$ verify*

$$\begin{bmatrix}\mathcal{Y}_a(1,s)\\ \mathcal{Y}_a(1,1-s)\end{bmatrix} = \sqrt{a}\,\begin{bmatrix}\alpha(a) & \beta(a)\\ \beta(a) & \alpha(a)\end{bmatrix}\begin{bmatrix}s(\mathcal{E}_a(s)r(a) + \mathcal{E}_a(1-s)s(a))\\ (1-s)(\mathcal{E}_a(s)s(a) + \mathcal{E}_a(1-s)r(a))\end{bmatrix} \tag{217}$$

Comparing with equation (195) we get: $T_a(s) = \sqrt{a}s(\mathcal{E}_a(s)r(a) + \mathcal{E}_a(1-s)s(a))$. So:

Theorem 26. *The analytic reproducing kernel $\mathcal{Y}_a(s,z)$ associated with the extended spaces L_a is:*

$$\mathcal{Y}_a(s,z) = s(s-1)z(z-1)\mathcal{X}_a(s,z) + \begin{bmatrix} T_a(s) & T_a(1-s)\end{bmatrix}\begin{bmatrix}\alpha(a) & \beta(a)\\ \beta(a) & \alpha(a)\end{bmatrix}\begin{bmatrix}T_a(z)\\ T_a(1-z)\end{bmatrix} \tag{218a}$$

$$\mathcal{X}_a(s,z) = \frac{\mathcal{E}_a(s)\mathcal{E}_a(z) - \mathcal{E}_a(1-s)\mathcal{E}_a(1-z)}{s+z-1} \qquad \mathcal{E}_a(s) = 2\sqrt{a}K_s(2a) \tag{218b}$$

$$T_a(s) = \sqrt{a}\, s\,(\mathcal{E}_a(s)r(a) + \mathcal{E}_a(1-s)s(a)) \qquad r(a) = I_0(2a) \quad s(a) = I_1(2a) \tag{218c}$$

$$\alpha(a) = \frac{p(a)}{p(a)^2 - q(a)^2} \qquad p(a) = a(I_0^2(2a) - I_1^2(2a)) \tag{218d}$$

$$\beta(a) = \frac{q(a)}{p(a)^2 - q(a)^2} \qquad q(a) = \frac{1}{2}(I_0^2(2a) - 1) \tag{218e}$$

We proceed now to the determination of $\mathcal{A}_a^\diamond$, $\mathcal{B}_a^\diamond$ and $\mathcal{E}_a^\diamond = \mathcal{A}_a^\diamond(s) - i\mathcal{B}_a^\diamond(s)$. The function $\mathcal{A}_a^\diamond(s)$ is even under $s \to 1-s$ and $\mathcal{B}_a^\diamond(s)$ is odd. We must have:

$$z\mathcal{Y}_a(1,z) = 2(-i\mathcal{B}_a^\diamond(1))\mathcal{A}_a^\diamond(z) + 2\mathcal{A}_a^\diamond(1)(-i\mathcal{B}_a^\diamond(z)) \tag{219}$$

On the other hand from (195) we have $\mathcal{Y}_a(1,z) = \alpha T_a(z) + \beta T_a(1-z)$. Let us write

$$\begin{aligned} z\,T_a(z) &= \sqrt{a}(z(z-1)r(a)\mathcal{E}_a(z) + zr(a)\mathcal{E}_a(z) \\ &\qquad + z(z-1)s(a)\mathcal{E}_a(1-z) + zs(a)\mathcal{E}_a(1-z)) \end{aligned} \tag{220}$$

$$z\,T_a(1-z) = \sqrt{a}(-z(z-1)s(a)\mathcal{E}_a(z) - z(z-1)r(a)\mathcal{E}_a(1-z)) \tag{221}$$

$$\begin{aligned} z\mathcal{Y}_a(1,z) = \sqrt{a}\Big(&z(z-1)((\alpha r - \beta s)\mathcal{E}_a(z) + (\alpha s - \beta r)\mathcal{E}_a(1-z)) \\ &+ z\alpha(r\,\mathcal{E}_a(z) + s\mathcal{E}_a(1-z))\Big) \end{aligned} \tag{222}$$

Extracting the even part $(z\mathcal{Y}_a(1,z))^+$ and the odd part $(z\mathcal{Y}_a(1,z))^-$:

$$(z\mathcal{Y}_a(1,z))^+ = \sqrt{a}\Big(z(z-1)(\alpha-\beta)(r+s)\mathcal{A}_a + (z-\frac{1}{2})\alpha(r-s)(-i\mathcal{B}_a) + \frac{1}{2}\alpha(r+s)\mathcal{A}_a\Big) \tag{223}$$

$$(z\mathcal{Y}_a(1,z))^- = \sqrt{a}\Big(z(z-1)(\alpha+\beta)(r-s)(-i\mathcal{B}_a) + (z-\frac{1}{2})\alpha(r+s)\mathcal{A}_a + \frac{1}{2}\alpha(r-s)(-i\mathcal{B}_a)\Big) \tag{224}$$

We have $(z\mathcal{Y}_a(1,z))^+ = 2(-i\mathcal{B}_a^\diamond(1))\mathcal{A}_a^\diamond(z)$ and $(z\mathcal{Y}_a(1,z))^- = 2\mathcal{A}_a^\diamond(1)(-i\mathcal{B}_a^\diamond(z))$. Let us define $K(a) = (2(-i\mathcal{B}_a^\diamond(1)))^{-1}$ and $L(a) = (2\mathcal{A}_a^\diamond(1))^{-1}$. We know that:

$$\lim_{\sigma\to+\infty} \frac{-i\mathcal{B}_a^\diamond(\sigma)}{\mathcal{A}_a^\diamond(\sigma)} = 1 \tag{225}$$

So it must be that

$$K(a)(\alpha-\beta)(r+s) = L(a)(\alpha+\beta)(r-s) \tag{226}$$

Also, taking $z=1$ in (223) we have $\frac{1}{KL} = 2\sqrt{a}\frac{1}{2}\alpha\,(r\mathcal{E}_a(1) + s\mathcal{E}_a(0)) = \alpha T_a(1)$. But referring to (195) one has $T_a(1) = p\alpha - q\beta = 1$. So:

$$K(a)L(a) = \frac{1}{\alpha(a)} \tag{227}$$

Then:

$$K(a)^2 = \frac{1}{\alpha(a)}\frac{(\alpha+\beta)(r-s)}{(\alpha-\beta)(r+s)} = \frac{\alpha^2-\beta^2}{\alpha(a)}\frac{r^2-s^2}{(\alpha-\beta)^2(r+s)^2} = \frac{1}{p}\frac{p}{a}\frac{1}{(\alpha-\beta)^2(r+s)^2} \tag{228}$$

$$(\alpha-\beta)(r+s)K(a) = a^{-\frac{1}{2}} \tag{229}$$

We conclude:

$$\mathcal{A}_a^\diamond = z(z-1)\mathcal{A}_a + (z-\frac{1}{2})\frac{\alpha(r-s)}{(\alpha-\beta)(r+s)}(-i\mathcal{B}_a) + \frac{\alpha}{2(\alpha-\beta)}\mathcal{A}_a \tag{230a}$$

$$-i\mathcal{B}_a^\diamond = z(z-1)(-i\mathcal{B}_a) + (z-\frac{1}{2})\frac{\alpha(r+s)}{(\alpha+\beta)(r-s)}\mathcal{A}_a + \frac{\alpha}{2(\alpha+\beta)}(-i\mathcal{B}_a) \tag{230b}$$

Let us now observe that $\frac{\alpha}{\alpha\pm\beta} = \frac{p}{p\pm q}$ and further:

$$\frac{\alpha}{\alpha-\beta}\frac{r-s}{r+s} = \frac{p}{p-q}\frac{a(r-s)^2}{p} = a\frac{p'-q'}{p-q} = a\frac{d}{da}\log(p-q) \tag{231a}$$

$$\frac{\alpha}{\alpha+\beta}\frac{r+s}{r-s} = \frac{p}{p+q}\frac{a(r+s)^2}{p} = a\frac{p'+q'}{p+q} = a\frac{d}{da}\log(p+q) \tag{231b}$$

$$\mathcal{A}_a^\diamond(z) = (z(z-1)+\frac{1}{2}\frac{p}{p-q})\mathcal{A}_a(z) + a\frac{d}{da}\log(p-q)(z-\frac{1}{2})(-i\mathcal{B}_a(z)) \tag{232a}$$

$$-i\mathcal{B}_a^\diamond(z) = (z(z-1)+\frac{1}{2}\frac{p}{p+q})(-i\mathcal{B}_a(z)) + a\frac{d}{da}\log(p+q)(z-\frac{1}{2})\mathcal{A}_a(z) \tag{232b}$$

Combining we get finally:

Theorem 27. *The E function associated with the entire functions $s(s-1)\Gamma(s)\widehat{f}(s)$, $f \in L_a$ is:*

$$\begin{aligned}\mathcal{E}_a^\diamond(z) = \Big(z(z-1) + \frac{1}{2}a\frac{d}{da}\log(p(a)^2-q(a)^2)(z-\frac{1}{2}) + \frac{1}{2}p(a)\alpha(a)\Big)\mathcal{E}_a(z)\\ + \Big(\frac{1}{2}a\frac{d}{da}\log\frac{p(a)+q(a)}{p(a)-q(a)}(z-\frac{1}{2}) + \frac{1}{2}p(a)\beta(a)\Big)\mathcal{E}_a(1-z)\end{aligned} \tag{233}$$

where $p(a) = a(I_0^2(2a)-I_1^2(2a))$, $q(a) = \frac{1}{2}(I_0^2(2a)-1)$, $\alpha(a) = \frac{p(a)}{p(a)^2-q(a)^2}$, $\beta(a) = \frac{q(a)}{p(a)^2-q(a)^2}$, and $\mathcal{E}_a(z) = 2\sqrt{a}K_z(2a)$.

We shall now obtain by two methods the function $\mu^\diamond(a)$. First, we compute $\mathcal{E}_a^\diamond(\frac{1}{2}) = (-\frac{1}{4}+\frac{1}{2}p(\alpha+\beta))\mathcal{E}_a(\frac{1}{2}) = \frac{1}{4}\frac{p+q}{p-q}\mathcal{E}_a(\frac{1}{2})$ and invoke $a\frac{d}{da}\mathcal{E}_a^\diamond(\frac{1}{2}) = -\mu^\diamond(a)\mathcal{E}_a^\diamond(\frac{1}{2})$. We thus have:

Theorem 28. *The mu function for the chain of spaces L_a, $0 < a < \infty$ is*

$$\mu^\diamond(a) = \mu(a) + a\frac{d}{da}\log\frac{p-q}{p+q} \tag{234}$$

$$= 2a + a\frac{d}{da}\log\frac{(2a-1)I_0^2(2a)-2aI_1^2(2a)+1}{(2a+1)I_0^2(2a)-2aI_1^2(2a)-1} \tag{235}$$

$$= 2a - 2 + o(1) \qquad (a\to+\infty) \tag{236}$$

The asymptotic behavior is a corollary to $\lim_{a\to\infty} 2a\frac{-pq'+qp'}{p^2-q^2} = -2$ which itself follows from $p^2-q^2 \sim \frac{1}{4}I_0(2a)^4\frac{1}{16a^2}$ and $(\frac{q}{p})' \sim +\frac{1}{16a^3}$ which are easily deduced from the asymptotic expansion $I_0(x) = \frac{e^x}{\sqrt{2\pi x}}(1+\frac{1}{8x}+\frac{9}{128x^2}+\dots)$ ([33]). Of course the $o(1)$ is in fact an $O(a^{-1})$.

The second method to obtain $\mu^\diamond(a)$ relies on $\frac{\mathcal{E}_a^\diamond(1-\sigma)}{\mathcal{E}_a^\diamond(\sigma)} \sim_{\sigma\to+\infty} \frac{\mu^\diamond(a)}{2\sigma}$ ((116)). We have:

$$\frac{\mathcal{E}_a^\diamond(\sigma)}{\sigma^2\mathcal{E}_a(\sigma)} = 1 + \frac{\frac{1}{2}a\frac{d}{da}\log(p^2-q^2)-1}{\sigma} + O(\frac{1}{\sigma^2}) \tag{237a}$$

$$\frac{\mathcal{E}_a^\diamond(1-\sigma)}{\sigma^2\mathcal{E}_a(1-\sigma)} \to_{\sigma\to\infty} 1 - \frac{1}{2}(a\frac{d}{da}\log\frac{p+q}{p-q})\frac{2}{\mu(a)} \tag{237b}$$

so $\frac{\mu^\diamond(a)}{\mu(a)} = 1 - \frac{1}{\mu(a)}a\frac{d}{da}\log\frac{p+q}{p-q}$ and (234) is confirmed. We can use this method to gather more information. From (115a) we have, as $\Re(s)\to+\infty$:

$$\widehat{E_a}(s) = a^{\frac{1}{2}-s}(1+\frac{a\phi^+(a)-a\phi^-(a)}{2s}+O(\frac{1}{s^2})) \tag{238a}$$

$$\widehat{E_a^\diamond}(s) = a^{\frac{1}{2}-s}(1+\frac{a\phi^{\diamond+}(a)-a\phi^{\diamond-}(a)}{2s}+O(\frac{1}{s^2})) \tag{238b}$$

Let us be careful that $\mathcal{E}_a(s) = \Gamma(s)\widehat{E_a}(s)$ while $\mathcal{E}_a^\diamond(s) = s^2\Gamma(s)\widehat{E_a^\diamond}(s)$. We obtain:

$$a\phi^{\diamond+}(a) - a\phi^{\diamond-}(a) = a\phi^+(a) - a\phi^-(a) + a\frac{d}{da}\log(p^2-q^2) - 2 \tag{239a}$$

$$a\phi^{\diamond+}(a) = a\phi^+(a) + a\frac{d}{da}\log\frac{p-q}{a} \tag{239b}$$

$$a\phi^{\diamond-}(a) = a\phi^-(a) - a\frac{d}{da}\log\frac{p+q}{a} \tag{239c}$$

We recall that $(p\pm q)\sim_{a\to0} a$. We integrate (239b) and (239c) using (130a), (130b) and this gives $\det(1+H_a^\diamond) = \frac{p-q}{a}\det(1+H_a)$ and $\det(1-H_a^\diamond) = \frac{p+q}{a}\det(1-H_a)$.

$$\det(1+H_a^\diamond) = \frac{p-q}{a}\det(1+H_a) = \det(1+H_a)\,\frac{1}{a}\int_0^a (r-s)^2\,da \tag{240a}$$

$$\det(1-H_a^\diamond) = \frac{p+q}{a}\det(1-H_a) = \det(1-H_a)\,\frac{1}{a}\int_0^a (r+s)^2\,da \tag{240b}$$

Theorem 29. *Let $\mathcal{H}^\diamond = L\,\mathcal{H}\,L^{-1}$ be the self-reciprocal operator on $L^2(0,\infty;dx)$ with kernel:*

$$J_0(2\sqrt{xy}) - 2\frac{J_1(2\sqrt{xy})}{\sqrt{xy}} + \frac{1-J_0(2\sqrt{xy})}{xy} = \sum_{n=0}^\infty (-1)^n\frac{n^2x^ny^n}{(n+1)!^2} \tag{241a}$$

and let $H_a^\diamond$ be the restriction to $L^2(0,a;dx)$. Then:

$$\det(1+H_a^\diamond) = e^{+a-\frac{1}{2}a^2}\,\frac{1}{a}\int_0^a (I_0(2a)-I_1(2a))^2\,da = e^{+a-\frac{1}{2}a^2}\big(I_0^2(2a)-I_1^2(2a)-\frac{I_0^2(2a)-1}{2a}\big) \tag{241b}$$

$$\det(1-H_a^\diamond) = e^{-a-\frac{1}{2}a^2}\,\frac{1}{a}\int_0^a (I_0(2a)+I_1(2a))^2\,da = e^{-a-\frac{1}{2}a^2}\big(I_0^2(2a)-I_1^2(2a)+\frac{I_0^2(2a)-1}{2a}\big) \tag{241c}$$

From theorem 26 $\|\mathcal{Y}_{\frac12}\|^2 = \frac{1}{16}\|\mathcal{X}_{\frac12}\|^2 + 2(\alpha+\beta)T_a(\frac12)^2$ and $T_a(\frac12) = \frac12\sqrt{a}\,(r+s)\mathcal{E}_a(\frac12)$. Also, $(\alpha+\beta)(r+s)^2 = \frac{1}{p-q}(r+s)^2$. Furthermore $\|\mathcal{Y}_{\frac12}\|^2 = \frac{1}{16}\Gamma(\frac12)^2\|X^\diamond_{\frac12}\|^2 = \frac{\pi}{16}\|X^\diamond_{\frac12}\|^2$ and $\|\mathcal{X}_{\frac12}\|^2 = \pi\|X_{\frac12}\|^2$. And also from (163) $\mathcal{E}_a(\frac12) = \sqrt{\pi}\ \frac{\det(1-H_a)}{\det(1+H_a)}$ and from theorem 17 one has $\|X^a_{\frac12}\|^2 = 2\int_a^\infty \left(\det\frac{1-H_b}{1+H_b}\right)^2 \frac{db}{b}$ and the analog holds for $X^{\diamond a}_{\frac12}$. Let us observe that $X^{\diamond a}_{\frac12} = -LX^\times_{\frac12}$ so $\|X^{\diamond a}_{\frac12}\| = \|X^{a\times}_{\frac12}\|$. So

$$\|X^{a\times}_{\frac12}\|^2 = 2\int_a^\infty \left(\det\frac{1-H_b^\diamond}{1+H_b^\diamond}\right)^2\frac{db}{b} = 2\int_a^\infty \left(\det\frac{1-H_b}{1+H_b}\right)^2\frac{db}{b} + 8a\frac{p'+q'}{p-q}\left(\det\frac{1-H_a}{1+H_a}\right)^2 \quad (242)$$

Theorem 30. *Let L_a be the Hilbert space of square integrable functions on $f\in L^2(0,\infty;dx)$ such that both f and $\mathcal{H}(f) = \int_0^\infty J_0(2\sqrt{xy})f(y)\,dy$ are constant on $(0,a)$. Then the squared norm of the linear form $f\mapsto\int_0^\infty \frac{f(x)}{\sqrt{x}}\,dx$ is given by either one of the following two expressions:*

$$2\int_a^\infty\left(\frac{(2b+1)I_0^2(2b)-2bI_1^2(2b)-1}{(2b-1)I_0^2(2b)-2bI_1^2(2b)+1}\right)^2\frac{e^{-4b}}{b}\,db \quad (243a)$$

$$= 2\int_a^\infty\frac{e^{-4b}}{b}\,db + 2\frac{8a(I_0(2a)+I_1(2a))^2}{(2a-1)I_0^2(2a)-2aI_1^2(2a)+1}e^{-4a} \quad (243b)$$

The squared norm of the restriction of the linear form to the subspace K_a of functions vanishing on $(0,a)$ and with $\mathcal{H}(f)$ also vanishing on $(0,a)$ is $2\int_a^\infty\frac{e^{-4b}}{b}\,db$.

One may express the wish to verify explicitely from equations (230a) and (230b), or in the equivalent form

$$\mathcal{A}^\diamond_a(z) = \Big((z-\frac12)^2+\frac14\frac{p+q}{p-q}\Big)\mathcal{A}_a(z) + (a\frac{d}{da}\log(p-q))(z-\frac12)(-i\mathcal{B}_a(z)) \quad (244a)$$

$$-i\mathcal{B}^\diamond_a(z) = \Big((z-\frac12)^2+\frac14\frac{p-q}{p+q}\Big)(-i\mathcal{B}_a(z)) + (a\frac{d}{da}\log(p+q))(z-\frac12)\mathcal{A}_a(z) \quad (244b)$$

the differential system:

$$a\frac{\partial}{\partial a}\mathcal{A}^\diamond_a(z) = -\mu^\diamond(a)\mathcal{A}^\diamond_a(z) - (z-\frac12)(-i\mathcal{B}^\diamond_a(z)) \quad (245a)$$

$$a\frac{\partial}{\partial a}(-i\mathcal{B}^\diamond_a(z)) = +\mu^\diamond(a)(-i\mathcal{B}^\diamond_a(z)) - (z-\frac12)\mathcal{A}^\diamond_a(z) \quad (245b)$$

and also to verify explicitely the reproducing kernel formula

$$\mathcal{Y}_a(s,z) = \frac{\mathcal{E}^\diamond_a(s)\mathcal{E}^\diamond_a(z)-\mathcal{E}^\diamond_a(1-s)\mathcal{E}^\diamond_a(1-z)}{s+z-1} \quad (246)$$

The interested reader will see that the algebra has a tendency to become slightly involved if one does not benefit from the following preliminary observations: using $p'=r^2+s^2$, $q'=2rs$, $ar'=\mu r$, $as'=\mu r-s$, $p=a(r^2-s^2)$ one first establishes $aq''+q'=2\mu p'$, $ap''+p'-\frac{p}{a}=2\mu q'$. Using this one checks easily:

$$\left(\frac{p'+q'}{p+q}\right)' + \left(\frac{p'+q'}{p+q}\right)^2 = \frac{1}{a^2}\frac{p}{p+q} + \frac{2\mu-1}{a}\frac{p'+q'}{p+q} \quad (247a)$$

$$\left(\frac{p'-q'}{p-q}\right)' + \left(\frac{p'-q'}{p-q}\right)^2 = \frac{1}{a^2}\frac{p}{p-q} - \frac{2\mu+1}{a}\frac{p'-q'}{p-q} \quad (247b)$$

Also the identity

$$\frac{p^2}{p^2-q^2} = a\frac{p'+q'}{p+q}a\frac{p'-q'}{p-q} = p\alpha \tag{247c}$$

is useful. The verifications may then be done.

9 Hyperfunctions in the study of the $\mathcal{H}$ transform

In this final section we return to the equation (31):

$$\widetilde{\psi(f)}(i\,t) = \frac{t+1}{2t}\widetilde{f}(i\,\frac{t+\frac{1}{t}}{2})\;, \tag{248}$$

Let us recall that $f \in L^2(0,\infty;dx)$ and $\psi : L^2(0,\infty;dx) \to L^2(0,\infty;dx)$ is the isometry which corresponds to $F(w) \mapsto F(w^2)$, where $F(w) = \sum_{n=0}^\infty c_n w^n$, $f(x) = \sum_{n=0}^\infty c_n P_n(x)e^{-x}$, $P_n(x) = L_n^{(0)}(2x)$. Let $g = \psi(f)$. Using $\lambda = it$, in the L^2 sense:

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\lambda+i}{2\lambda}\widetilde{f}(\frac{\lambda-\frac{1}{\lambda}}{2})e^{-i\lambda x}\,d\lambda \tag{249}$$

It is natural to consider separately $\lambda > 0$ and $\lambda < 0$. So let us define:

$$G_+(x) = \frac{1}{2\pi}\int_{-\infty}^{0} \frac{\lambda+i}{2\lambda}\widetilde{f}(\frac{\lambda-\frac{1}{\lambda}}{2})e^{-i\lambda x}\,d\lambda \tag{250a}$$

$$G_-(x) = -\frac{1}{2\pi}\int_{0}^{\infty} \frac{\lambda+i}{2\lambda}\widetilde{f}(\frac{\lambda-\frac{1}{\lambda}}{2})e^{-i\lambda x}\,d\lambda \tag{250b}$$

We observe that G_+ is in the Hardy space of $\Im(x) > 0$ and G_- is in the Hardy space of $\Im(x) < 0$. Their boundary values must coincide on $(-\infty, 0)$ as $g \in L^2(0,+\infty;dx)$. So we have a single analytic function $G(z)$ on $\mathbb{C} \setminus [0,+\infty)$ with $G = G_+$ for $\Im(x) > 0$ and $G = G_-$ for $\Im(x) < 0$. Then $g = \psi(f) = G_+ - G_-$ is computed as

$$g(x) = G(x+i0) - G(x-i0) \tag{251}$$

In other words g is most naturally seen as a hyperfunction [23], as a difference of boundary values of analytic functions. We shall now compute it explicitely, and also we will show later that this observation extends to the distributions $A_a(x)$, $-iB_a(x)$, $E_a(x)$ which are associated with the study of the $\mathcal{H}$ transform. The point of course is that the corresponding functions G will for them have a simple natural expression.

We have, for $\Im(z) > 0$:

$$G(z) = \frac{1}{2\pi}\int_{0}^{\infty} \frac{\lambda-i}{2\lambda}\widetilde{f}(\frac{-\lambda+\frac{1}{\lambda}}{2})e^{+i\lambda z}\,d\lambda \tag{252a}$$

$$G(z) = \frac{1}{2\pi}\int_{0}^{\infty} \frac{\lambda-i}{2\lambda}\left(\int_0^\infty e^{i\frac{1}{2}x(-\lambda+\frac{1}{\lambda})}f(x)\,dx\right)e^{+i\lambda z}\,d\lambda \tag{252b}$$

Let $\mu = \frac{1}{2}(\lambda - \frac{1}{\lambda})$, $\lambda = \mu + \sqrt{1+\mu^2}$, $\frac{\lambda-i}{2\lambda}\,d\lambda = \frac{\lambda}{\lambda+i}\,d\mu$, with, for $0 < \lambda < \infty$, $-\infty < \mu < \infty$.

$$G(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\left(\int_0^\infty e^{-i\mu x}f(x)\,dx\right)e^{+i\lambda z}\,d\mu \tag{252c}$$

For $\mu \to +\infty$, $\lambda = 2\mu + \frac{1}{2\mu} + \dots$, and for $\mu \to -\infty$, $\lambda = -\frac{1}{2\mu} + \dots$, and $\frac{\lambda}{\lambda+i} \sim \frac{i}{2\mu}$ as $\mu \to -\infty$. So far the inner integral is in the L^2 sense. We shall now suppose that f and f' are in L^1 (so $\lim_{x\to\infty} f(x) = 0$) and write $\int_0^\infty e^{-i\mu x} f(x)\,dx = \int_0^\infty e^{-i\mu x - x} e^x f(x)\,dx = \frac{f(0)}{i\mu+1} + \frac{1}{i\mu+1}\int_0^\infty e^{-i\mu x}(f(x)+f'(x))\,dx$.

$$G(z) = \frac{1}{2\pi}\left(f(0)\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{+i\lambda z}}{i\mu+1}\,d\mu + \int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{+i\lambda z}}{i\mu+1}\left(\int_0^\infty e^{-i\mu x}(f(x)+f'(x))\,dx\right) d\mu\right) \tag{252d}$$

In this manner, with $f \in L^1$, $f' \in L^1$, $\Im(z) > 0$, we have an absolutely convergent double integral.

$$G(z) = \frac{1}{2\pi}\left(f(0)\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{+i\lambda z}}{i\mu+1}\,d\mu + \int_0^\infty \left(\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{-i\mu x + i\lambda z}}{i\mu+1}\,d\mu\right)(f(x)+f'(x))\,dx\right) \tag{252e}$$

Observing $\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{+i\lambda z}}{i\mu+1}\,d\mu = \frac{1}{2\pi}\int_0^\infty \frac{\lambda-i}{2\lambda}\frac{1}{i\mu+1}e^{+i\lambda z}\,d\lambda = \frac{1}{2\pi i}\int_0^\infty \frac{1}{\lambda-i}e^{+i\lambda z}\,d\lambda$, we then suppose $\Re(z) < 0$, $\Im(z) > 0$ (or $\Im(z) \geq 0$) so that we may rotate the contour to $\lambda = -it$, $0 \leq t < \infty$. This procedure gives thus:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{+i\lambda z}}{i\mu+1}\,d\mu = \frac{1}{2\pi i}\int_0^\infty \frac{e^{tz}}{1+t}\,dt \tag{252f}$$

Also:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\lambda}{\lambda+i}\frac{e^{-i\mu x + i\lambda z}}{i\mu+1}\,d\mu = \frac{1}{2\pi i}\int_0^\infty \frac{1}{\lambda-i}e^{-i\frac{x}{2}(\lambda-\frac{1}{\lambda})+i\lambda z}\,d\lambda \tag{252g}$$

We rotate the contour to $\lambda \in i[0, -\infty)$, which is licit as $x \geq 0$ and, for $\Re(z) < 0$, $x \geq 0$, we obtain:

$$\frac{1}{2\pi i}\int_0^\infty \frac{e^{zt - \frac{x}{2}(t+\frac{1}{t})}}{1+t}\,dt \tag{252h}$$

Going back this allows to write (252e), for $\Re(z) < 0$, $\Im(z) > 0$ as:

$$G(z) = \frac{1}{2\pi i}\left(f(0)\int_0^\infty \frac{e^{zt}}{1+t}\,dt + \int_0^\infty \left(\int_0^\infty \frac{e^{zt-\frac{x}{2}(t+\frac{1}{t})}}{1+t}\,dt\right)(f(x)+f'(x))\,dx\right) \tag{252i}$$

and finally, after integrating by parts:

$$G(z) = \frac{1}{2\pi i}\int_0^\infty \left(\int_0^\infty \frac{1}{2}(1+\frac{1}{t})e^{zt-\frac{1}{2}y(t+\frac{1}{t})}\,dt\right) f(y)\,dy \tag{252j}$$

This last expression (still temporarily under the hypothesis $f, f' \in L^1$) is certainly a priori absolutely convergent for $\Re(z) < 0$ and gives $G(z)$ in this half-plane.

We are led to the study of:

$$a(z,y) = \frac{1}{2\pi i}\int_0^\infty \frac{1}{2}(1+\frac{1}{t})e^{zt-\frac{1}{2}y(t+\frac{1}{t})}\,dt \tag{253a}$$

We still temporarily assume $\Re(z) < 0$. We even suppose $z < 0$ and make a change of variable:

$$a(z,y) = \frac{1}{2\pi i}\left(\sqrt{\frac{y}{y-2z}}\,\frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\sqrt{y(y-2z)}(u+\frac{1}{u})}\,du + \frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\sqrt{y(y-2z)}(u+\frac{1}{u})}\frac{1}{u}\,du\right) \tag{253b}$$

$$a(z,y) = \frac{1}{2\pi i}\left(\sqrt{\frac{1}{y-2z}}\,\frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\sqrt{y-2z}(v+y\frac{1}{v})}\,dv + \frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\sqrt{y-2z}(v+y\frac{1}{v})}\frac{1}{v}\,dv\right) \tag{253c}$$

$$a(z,y) = \frac{1}{2\pi i}\left(\sqrt{\frac{y}{y-2z}}\,K_1(\sqrt{y(y-2z)}) + K_0(\sqrt{y(y-2z)})\right) \tag{253d}$$

For any $z \in \mathbb{C}\setminus[0,+\infty)$ and any $y \geq 0$ the integrals in (253c) converge absolutely and define an analytic function of z. Furthermore the K Bessel functions decrease exponentially as $y \to +\infty$ in (253d). For fixed z, $a(z,y)$ is certainly a square-integrable function of y (also at the origin), locally uniformly in z so the equation (252j) defines G as an analytic function on the entire domain $\mathbb{C}\setminus[0,+\infty)$. Then by an approximation argument (252j) applies to any $f \in L^2(0,\infty;dx)$ and any $z \in \mathbb{C}\setminus[0,+\infty)$.

We now study the boundary values $a(x+i0,y)$, $a(x-i0,y)$, $x,y \geq 0$. We could use the expression of the K Bessel functions in terms of the Hankel functions $H^{(1)}$ and $H^{(2)}$, go to the boundary, and then recover the Bessel functions J_0 and J_1. But we shall proceed in a more direct manner. Let us first examine

$$\begin{aligned} d(z,y) &= \frac{1}{2\pi i}\frac{1}{2}\int_0^\infty e^{zt-\frac{1}{2}y(t+\frac{1}{t})}\,\frac{1}{t}\,dt \\ &= \frac{1}{2\pi i}\frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\sqrt{y(y-2z)}(u+\frac{1}{u})}\frac{1}{u}\,du = \frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\frac{dt}{\sqrt{t^2-1}} \end{aligned} \tag{254a}$$

$$= \frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\,t^{-1}\,dt + \frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\Big(\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big)\,dt \tag{254b}$$

$$= \frac{1}{2\pi i}\frac{e^{-\sqrt{y(y-2z)}} - \int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\,t^{-2}\,dt}{\sqrt{y(y-2z)}} + \frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\Big(\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big)\,dt \tag{254c}$$

We now look at the (distributional) boundary values $z \to x$ with $z = x+i\epsilon$, $\epsilon \to 0^+$ or $z = x - i\epsilon$ and $\epsilon \to 0^+$. We shall take $x > 0$. Here the singularities at $y = 2x$ and at $y = 0$ are integrable and we need only take the limit in the naive sense. We distinguish $y > 2x$ from $0 < y < 2x$. In the former case, nothing happens:

$$d(x+i0,y) = d(x-i0,y) = \frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2x)}\,t}\frac{dt}{\sqrt{t^2-1}} \tag{255a}$$

In the latter case:

$$d(x+i0,y) = \frac{1}{2\pi}\frac{e^{+i\sqrt{y(2x-y)}} - \int_1^\infty e^{+i\sqrt{y(2x-y)}\,t}\,t^{-2}\,dt}{\sqrt{y(2x-y)}} + \frac{1}{2\pi i}\int_1^\infty e^{+i\sqrt{y(2x-y)}\,t}\Big(\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big)\,dt \tag{255b}$$

$$d(x-i0,y) = -\frac{1}{2\pi}\frac{e^{-i\sqrt{y(2x-y)}} - \int_1^\infty e^{-i\sqrt{y(2x-y)}\,t}\,t^{-2}\,dt}{\sqrt{y(2x-y)}} + \frac{1}{2\pi i}\int_1^\infty e^{-i\sqrt{y(2x-y)}\,t}\Big(\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big)\,dt \tag{255c}$$

So $d(x+i0,y) - d(x-i0,y)$ is supported in $(0,2x)$ and has values there

$$\frac{1}{\pi}\frac{\cos\sqrt{y(2x-y)} - \int_1^\infty \cos(\sqrt{y(2x-y)}\,t)\,t^{-2}\,dt}{\sqrt{y(2x-y)}} + \frac{1}{\pi}\int_1^\infty \sin(\sqrt{y(2x-y)}\,t)\Big(\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big)\,dt \tag{255d}$$

We used this method to have a clear control not only of the pointwise behavior but also of the limit as a distribution. There is no necessity now to keep working with absolutely convergent integrals and we have the simple result, using the very classical Mehler formula:[31]

$$d(x+i0,y)-d(x-i0,y)=\mathbf{1}_{0<y<2x}(y)\frac{1}{\pi}\int_1^\infty\frac{\sin(\sqrt{y(2x-y)}\,t)}{\sqrt{t^2-1}}\,dt=\frac{1}{2}\,\mathbf{1}_{0<y<2x}(y)J_0(\sqrt{y(2x-y)}) \tag{256}$$

Let us now consider the behavior of

$$e(z,y)=\frac{1}{2\pi i}\frac{1}{2}\int_0^\infty e^{zt-\frac{1}{2}y(t+\frac{1}{t})}\,dt=\frac{1}{2\pi i}\sqrt{\frac{y}{y-2z}}\,\frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\sqrt{y(y-2z)}(u+\frac{1}{u})}\,du \tag{257}$$

We make the simple observation that $e(z,y)=\frac{\partial}{\partial z}d(z,y)$. So we shall have (as is confirmed by a more detailed examination):

$$\begin{aligned}e(x+i0,y)-e(x-i0,y)&=\frac{\partial}{\partial x}\frac{1}{2}\,\mathbf{1}_{0<y<2x}(y)J_0(\sqrt{y(2x-y)})\\&=\delta_{2x}(y)-\frac{1}{2}\,\mathbf{1}_{0<y<2x}(y)\sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})\end{aligned} \tag{258}$$

Combining all those elements we obtain that the function $k(x)=\psi(f)(x)$ is given as:

$$k(x)=f(2x)+\frac{1}{2}\int_0^{2x}J_0(\sqrt{y(2x-y)})f(y)\,dy-\frac{1}{2}\int_0^{2x}\sqrt{\frac{y}{2x-y}}J_1(\sqrt{y(2x-y)})f(y)\,dy \tag{259}$$

Some pointwise regularity of f at x is necessary to fully justify the formula; in order to check if continuity of f at $2x$ is enough we can not avoid examining $e(z,y)$ more closely as $z\to x$.

$$\begin{aligned}e(z,y)&=\sqrt{\frac{y}{y-2z}}\frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\frac{t\,dt}{\sqrt{t^2-1}}\\&=\frac{1}{2\pi i}\frac{e^{-\sqrt{y(y-2z)}}}{y-2z}+\sqrt{\frac{y}{y-2z}}\frac{1}{2\pi i}\int_1^\infty e^{-\sqrt{y(y-2z)}\,t}\,\frac{t-\sqrt{t^2-1}}{\sqrt{t^2-1}}\,dt\end{aligned} \tag{260}$$

The integral term on the right causes no problem at all. And writing $\frac{e^{-\sqrt{y(y-2z)}}}{y-2z}=\frac{1}{y-2z}+\frac{e^{-\sqrt{y(y-2z)}}-1}{y-2z}$, again the term on the right has no problem, so there only remains $\frac{1}{y-2z}$, and of course, this is very well-known, the difference between $+i0$ and $-i0$ gives the Poisson kernel, so for non-tangential convergence, continuity of f at $2x$ is enough. Of course this discussion was quite superfluous if we wanted to understand k as an L^2 function, here we have the information that non tangential boundary value of $G(x+i0)-G(x-i0)$ does give pointwise the formula (259) if f is continuous at $y=2x$. We can also rewrite (259) as:

$$k(x)=(1+\frac{d}{dx})\frac{1}{2}\int_0^{2x}J_0(\sqrt{y(2x-y)})f(y)\,dy \tag{261}$$

This is exactly one half of equation (20c), where k was obtained from (f,g) as $\psi(f)+w\cdot\psi(g)$. Let us observe that $w=\frac{\lambda-i}{\lambda+i}$ verifies, as an operator, $(\frac{d}{dx}+1)\cdot w=w\cdot(\frac{d}{dx}+1)=\frac{d}{dx}-1$. So the isometry corresponding to $g(w)\mapsto w\,G(w^2)$, which is the composite $w\cdot\psi$, sends g to $(-1+\frac{d}{dx})\frac{1}{2}\int_0^{2x}J_0(\sqrt{y(2x-y)})f(y)\,dy$. This is indeed the second half of equation (20c).

The formulas (20a) and (20b) may be established in an exactly analogous manner (taking k with compact support to simplify the discussion). But this would be a repetition of the arguments

[31] we are mainly interested in the boundary value as a distribution and we skip the discussion of the pointwise behavior at the borders $y=0$ and $y=2x$.

we just went through, so rather I will conclude the paper with a method allowing to go directly from $\mathcal{A}_a(s)$, $-i\mathcal{B}_a(s)$, $\mathcal{E}_a(s)$ to the distributions $A_a(x)$, $-iB_a(x)$, $E_a(x)$, and this will show that they are in a natural manner (differences of) boundary values of an analytic function.

From the expression $\mathcal{E}_a(s) = \Gamma(s)\widehat{E_a}(s) = 2\sqrt{a}K_s(2a) = \sqrt{a}\int_0^\infty e^{-a(t+\frac{1}{t})}t^{s-1}\,dt$, we shall recover $\widehat{E_a}(s)$ as a right Mellin transform with the help of the Hankel formula $\Gamma(s)^{-1} = \int_\mathcal{C} e^v v^{-s}\,dv$, where $\mathcal{C}$ is a contour coming from $-\infty$ along the lower edge of the cut along $(-\infty, 0]$ turning counterclockwise around the origin and going back to $-\infty$ along or slightly above the upper edge of the cut. Let us write the Hankel formula as

$$\frac{t^{s-1}}{\Gamma(s)} = \frac{1}{2\pi i}\int_\mathcal{C} e^{tv}v^{-s}\,dv \qquad (t>0) \tag{262}$$

So we have:

$$\widehat{E_a}(s) = \sqrt{a}\frac{1}{2\pi i}\int_0^\infty \left(\int_\mathcal{C} e^{tv}v^{-s}\,dv\right)e^{-a(t+\frac{1}{t})}\,dt \tag{263}$$

Let us suppose $\Re(s) > 1$. Then the contour $\mathcal{C}$ can be deformed into the contour $\mathcal{C}_\epsilon$ coming from $-i\infty$ to $-i\epsilon$, then turning counterclockwise from $e^{-i\frac{\pi}{2}}\epsilon$ to $e^{i\frac{\pi}{2}}\epsilon$, then going to $+i\infty$. Also we impose $0 < \epsilon < a$. The integrals may then be permuted:

$$\widehat{E_a}(s) = \sqrt{a}\frac{1}{2\pi i}\int_{\mathcal{C}_\epsilon}\left(\int_0^\infty e^{tv}e^{-a(t+\frac{1}{t})}\,dt\right)v^{-s}\,dv \tag{264}$$

and using $e(z,y)$ from (257) this gives:

$$\Re(s) > 1 \implies \quad \widehat{E_a}(s) = \sqrt{a}\int_{\mathcal{C}_\epsilon} 2e(v, 2a)v^{-s}\,dv \tag{265}$$

We have previously studied $e(z,y)$, which is also expressed as in (260). We see on this basis and simple estimates that we may deform $\mathcal{C}_\epsilon$ into a contour $\mathcal{C}_{a,\eta}$ going from $+\infty$ to $a+\eta$ along the lower border, turning clockwise around a from $a+\eta-i0$ to $a+\eta+i0$, then going from $a+\eta$ to $+\infty$ on the upper border ($\eta \ll 1$). We will have in particular from (260) a term $\frac{2}{2\pi i}\int_{a+\eta-i0}^{a+\eta+i0}\frac{v^{-s}}{2a-2v}\,dv$ which is a^{-s}. The final result is obtained:

$$\Re(s) > 1 \implies \quad \widehat{E_a}(s) = \sqrt{a}\left(a^{-s} - \int_a^\infty \sqrt{\frac{a}{x-a}}J_1(2\sqrt{a(x-a)})x^{-s}\,dx\right) \tag{266}$$

This identifies $\widehat{E_a}(s)$ as the right Mellin transform of the distribution

$$E_a(x) = \sqrt{a}\left(\delta_a(x) + \mathbf{1}_{x>a}(x)\frac{\partial}{\partial x}J_0(2\sqrt{a(x-a)})\right) = \sqrt{a}\frac{\partial}{\partial x}\left(\mathbf{1}_{x>a}(x)J_0(2\sqrt{a(x-a)})\right) \tag{267}$$

This proof reveals that the distribution $E_a(x)$ is expressed in a natural manner as the difference of boundary values $\sqrt{a}(2e(x+i0,2a) - 2e(x-i0,2a))$, with

$$\sqrt{a}\,2e(z,2a) = \sqrt{a}\,\frac{1}{2\pi i}\int_0^\infty e^{zt-a(t+\frac{1}{t})}\,dt = \sqrt{a}\,\frac{1}{2\pi i}2\sqrt{\frac{a}{a-z}}\,K_1(2\sqrt{a(a-z)}) \tag{268}$$

The formulas (176e) and (176f) are recovered in the same manner.

Theorem 31. *The distribution* $A_a(x) = \frac{\sqrt{a}}{2}(1+\mathcal{H})(\phi_a^+ \mathbf{1}_{0<x<\infty})$, $\phi_a^+(x) + \int_0^a J_0(2\sqrt{xy})\phi_a^+(y)\,dy = J_0(2\sqrt{ax})$, *is the difference of boundary values* $\sqrt{a}(a(x+i0,2a) - a(x-i0,2a))$, *with:*

$$\begin{aligned}\sqrt{a}\,a(z,2a) &= \sqrt{a}\,\frac{1}{2\pi i}\int_0^\infty \frac{1}{2}(1+\frac{1}{t})e^{zt-a(t+\frac{1}{t})}\,dt \\ &= \sqrt{a}\,\frac{1}{2\pi i}\left(\sqrt{\frac{a}{a-z}}\,K_1(2\sqrt{a(a-z)}) + K_0(2\sqrt{a(a-z)})\right)\end{aligned} \tag{269}$$

The distribution $-iB_a(x) = \frac{\sqrt{a}}{2}(-1+\mathcal{H})(\phi_a^- \mathbf{1}_{0<x<\infty})$, $\phi_a^-(x) - \int_0^a J_0(2\sqrt{xy})\phi_a^-(y)\,dy = J_0(2\sqrt{ax})$, *is the difference of boundary values* $\sqrt{a}(-ib(x+i0,2a) - (-ib(x-i0,2a)))$, *with:*

$$\begin{aligned}\sqrt{a}(-ib(z,2a)) &= \sqrt{a}\,\frac{1}{2\pi i}\int_0^\infty \frac{1}{2}(1-\frac{1}{t})e^{zt-a(t+\frac{1}{t})}\,dt \\ &= \sqrt{a}\,\frac{1}{2\pi i}\left(\sqrt{\frac{a}{a-z}}\,K_1(2\sqrt{a(a-z)}) - K_0(2\sqrt{a(a-z)})\right)\end{aligned} \tag{270}$$

10 Appendix: a remark on the resolvent of the Dirichlet kernel

In this paper we have studied a special transform on the positive half-line with a kernel of a multiplicative type $k(xy)$, following the method summarized in [5, 6]. We have associated to the kernel the investigation of its Fredholm determinants on finite intervals $(0,a)$, and have related them with first and second order differential equations leading to problems of spectral and scattering theory. There is a vast literature on kernels of the additive type $k(x-y)$, and on the related Fredholm determinants on finite intervals. The Dirichlet kernel on $L^2(-s,s;dx)$:

$$K_s(x,y) = \frac{\sin(x-y)}{\pi(x-y)} \tag{271}$$

has been the subject of many works (only a few references will be mentioned here.) The Fredholm determinant $\det(1-K_s)$, as a function of s (or more generally as a function of the endpoints of finitely many intervals), has many properties, and is related to the study of random matrices [22]. The Fredholm determinants of the even and odd parts

$$K_s^{\pm}(x,y) = \frac{\sin(x-y)}{\pi(x-y)} \pm \frac{\sin(x+y)}{\pi(x+y)} \tag{272}$$

on $L^2(0,s;dx)$ have been studied by Dyson [16]. He used the second derivatives of their logarithms to construct potentials for Schrödinger equations on the half-line, and studied their asymptotics with the tools of scattering theory. Jimbo, Miwa, Môri, and Sato [17] related $\det(1-K_s)$ to a Painlevé equation. Widom [34] obtained the leading asymptotics using the Krein continuous analog of orthogonal polynomials. Deift, Its, and Zhou [11] justified the Dyson asymptotic expansions using tools developed for Riemann-Hilbert problems. Tracy and Widom [32] established partial differential equations for the Fredholm determinants of integral operators arising in the study of the scaling limit of the distribution functions of eigenvalues of random matrices. We refer the reader to the cited references and to [12] for recent results and we apologize for not providing any more detailed information here.

We have, in the present paper, been talking a lot of scattering and determinants and one might wonder if this is not a re-wording of known things. In fact, our work is with the multiplicative kernels $k(xy)$, and (direct) reduction to additive kernels would lead to (somewhat strange) $g(t+u)$

kernels on semi-infinite intervals $(-\infty, \log(a)]$. So we are indeed doing something different; one may also point out that the entire functions arising in the present study are not of finite exponential type; and the scattering matrices do not at all tend to 1 as the frequency goes to infinity. In the case of the cosine and sine kernels the flow of information will presumably go from the additive to the multiplicative, as the additive situation is more flexible, and has stimulated the development of powerful tools, with relation to Painlevé equations, Riemann-Hilbert problems, Integrable systems [11].

Nevertheless, one may ask if the framework of reproducing kernels in Hilbert spaces of entire functions also may be used in the additive situation. This is the case indeed and it is very much connected to the method of Krein in inverse scattering theory, and his continuous analog of orthogonal polynomials (used by Widom in the context of the Dirichlet kernel in [34].) The Gaudin identities for convolution kernels ([22, App. A16]) play a rôle very analogous to the identities in the present paper (132a), (132b) involved in the study of multiplicative kernels. Widom in his proof [34] of the main term of the asymptotics as $s \to +\infty$ studied the Krein functions associated with the complement of the interval $(-1, +1)$ and he mentioned the interest of extremal properties. In this appendix, I shall point out that the resolvent of the Dirichlet kernel indeed does have an extremal property: it coincides exactly (up to complex conjugation in one variable) with the reproducing kernel of a certain (interesting) Hilbert space of entire functions. This could be a new observation, obviously closely related to the method of Widom [34].

The space mPW_s we shall use is, as a set, the Paley-Wiener space PW_s, but the norm is different:

$$\begin{aligned} mPW_s &= \{f(z) \text{ entire of exponential type at most } s \text{ with } \|f\| < \infty\} \\ \|f\|^2 &= \int_{\mathbb{R}\setminus(-1,1)} |f(t)|^2\, dt \end{aligned} \tag{273}$$

Let $X_s(z,w)$ be the element of mPW_s which is the evaluator at z: $\forall f \in mPW_s\ (f, X_s(z,\cdot)) = f(z)$. We shall compare $X_s(z,w)$ with the resolvent of the kernel

$$D_s(x,y) = \frac{\sin(s(x-y))}{\pi(x-y)} \tag{274}$$

on $L^2(-1,1;dx)$.

Let $f \in mPW_s$. It belongs to PW_s so

$$f(z) = \int_{\mathbb{R}} f(t)\frac{e^{is(t-z)} - e^{-is(t-z)}}{2\pi i(t-z)}\, dt = \int_{\mathbb{R}} f(t)\frac{\sin(s(t-z))}{\pi(t-z)}\, dt \tag{275}$$

On the other hand:

$$f(z) = \left(\int_{-\infty}^{-1} + \int_1^{\infty}\right) f(t)\overline{X_s(z,t)}\, dt \tag{276}$$

As $\overline{f(\overline{z})} = \left(\int_{-\infty}^{-1} + \int_1^{\infty}\right) \overline{f(t)}\ \overline{\overline{X_s(z,t)}}\, dt = \left(\int_{-\infty}^{-1} + \int_1^{\infty}\right) \overline{f(t)} X_s(\overline{z},t)\, dt$ one has $\overline{X_s(z,t)} = X_s(\overline{z},t)$ for $t \in \mathbb{R}$. We have for y_1 and y_2 real

$$X_s(y_1,y_2) = \int_{\mathbb{R}\setminus(-1,1)} X_s(y_1,t)\overline{X_s(y_2,t)}\, dt = \int_{\mathbb{R}\setminus(-1,1)} X_s(y_1,t)X_s(y_2,t)\, dt = X_s(y_2,y_1) \tag{277}$$

so more generally $X_s(\overline{z_1}, z_2) = X_s(\overline{z_2}, z_1)$.

We apply (276) to $f(z) = \frac{\sin(s(z-y))}{\pi(z-y)}$ for some $y \in \mathbb{C}$:

$$\frac{\sin(s(z-y))}{\pi(z-y)} = \int_{\mathbb{R}\setminus(-1,1)} \frac{\sin(s(t-y))}{\pi(t-y)} X_s(\overline{z},t)\,dt \tag{278}$$

We apply (275) to $f(y) = X_s(\overline{z},y)$ for some $z \in \mathbb{C}$:

$$X_s(\overline{z},y) = \int_{\mathbb{R}} X_s(\overline{z},t)\frac{\sin(s(t-y))}{\pi(t-y)}\,dt \tag{279}$$

Combining we obtain:

$$X_s(\overline{z},y) - \frac{\sin(s(z-y))}{\pi(z-y)} = \int_{-1}^{1} X_s(\overline{z},t)\,\frac{\sin(s(t-y))}{\pi(t-y)}\,dt \tag{280}$$

Restricting to $y \in (-1,1)$, $z = x \in (-1,1)$, this says exactly:

$$X_s(x,y) = R_s(x,y) \tag{281}$$

where $R_s(x,y)$ is the kernel of the resolvent: $1 + R_s = (1-D_s)^{-1}$, $R_s - D_s = R_s D_s$. The resolvent $R_s(x,y)$ is entire in (x,y) and the general formula is thus:

$$\forall z,w \in \mathbb{C} \qquad R_s(z,w) = X_s(\overline{z},w)\ . \tag{282}$$

References

[1] L. de Branges, Self-reciprocal functions, J. Math. Anal. Appl. **9** (1964) 433–457.

[2] L. de Branges, *Hilbert spaces of entire functions*, Prentice Hall Inc., Englewood Cliffs, 1968.

[3] J.-F. Burnol, Sur certains espaces de Hilbert de fonctions entières, liés à la transformation de Fourier et aux fonctions L de Dirichlet et de Riemann, C. R. Acad. Sci. Paris, Ser. I **333** (2001), 201–206.

[4] J.-F. Burnol, On Fourier and Zeta ('s), talk at University of Nice (18 Dec. 2001), Forum Mathematicum **16** (2004), 789–840.

[5] J.-F. Burnol, Sur les "espaces de Sonine" associés par de Branges à la transformation de Fourier, C. R. Acad. Sci. Paris, Ser. I **335** (2002), 689–692.

[6] J.-F. Burnol, Des équations de Dirac et de Schrödinger pour la transformation de Fourier, C. R. Acad. Sci. Paris, Ser. I **336** (2003), 919–924.

[7] J.-F. Burnol, Two complete and minimal systems associated with the zeros of the Riemann zeta function, Jour. Th. Nb. Bord. **16** (2004).

[8] J.-F. Burnol, *Entrelacement de co-Poisson*, Ann. Inst. Fourier, **57** no. 2 (2007), 525–602.

[9] J.-F. Burnol, Spacetime causality in the study of the Hankel transform, Ann. Henri Poincaré **7** (2006), 1013–1034.

[10] E. Coddington, N. Levinson, *Theory of ordinary differential equations*, McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955.

[11] P. Deift, A. Its, X. Zhou, A Riemann-Hilbert approach to asymptotic problems arising in the theory of random matrix models, and also in the theory of integrable statistical mechanics, Ann. Math. **146** (1997), 149–235.

[12] P. Deift, A. Its, I. Krasovsky, X. Zhou, The Widom-Dyson constant for the gap probability in random matrix theory, arXiv:math.FA/0601535

[13] R. J. Duffin, H. F. Weinberger, Dualizing the Poisson summation formula, Proc. Natl. Acad. Sci. USA **88** (1991), 7348–7350.

[14] H. Dym, An introduction to de Branges spaces of entire functions with applications to differential equations of the Sturm-Liouville type, Advances in Math. **5** (1971), 395–471.

[15] H. Dym, H.P. McKean, *Gaussian processes, function theory, and the inverse spectral problem*, Probability and Mathematical Statistics, Vol. 31. Academic Press, New York-London, 1976.

[16] F. Dyson, Fredholm determinants and inverse scattering problems, Comm. Math. Phys. **47** (1976), 171–183

[17] M. Jimbo, T. Miwa, Y. Môri, M. Sato, Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent, Physica **1D** (1980), 80-158.

[18] J. C. Lagarias, Hilbert spaces of entire functions and Dirichlet L-functions, in: *Frontiers in Number Theory, Physics and Geometry: On Random Matrices, Zeta Functions and Dynamical Systems* (P. E. Cartier, B. Julia, P. Moussa and P. van Hove, Eds.), Springer-Verlag: Berlin 2006, to appear.

[19] J. C. Lagarias, Zero spacing distributions for differenced L-functions, Acta Arithmetica 120 (2005), No. 2, 159–184.

[20] P. Lax, *Functional Analysis*, Wiley, 2002.

[21] B. M. Levitan, I. S. Sargsjan, *Introduction to Spectral Theory*, Transl. of Math. Monographs **39**, AMS 1975.

[22] M. L. Mehta, *Random Matrices*, Academic Press, 2nd ed., 1991.

[23] M. Morimoto, *An introduction to Sato's Hyperfunctions*, Transl. of Math. Monographs **129**, AMS 1993.

[24] G. Pólya, Bemerkung über die Integraldarstellung der Riemannschen ξ-Funktion, Acta Math. **48** (1926), 305-317.

[25] G. Pólya, Über trigonometrische Integralen mit nur reellen Nullstellen, J. Reine u. Angew. Math. **158** (1927), 6-18.

[26] M. Reed, B. Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press, New York-London, 1975.

[27] C. Remling, Schrödinger operators and de Branges spaces, J. Funct. Anal. **196** (2002), 323–394.

[28] V. Rovnyak, Self-reciprocal functions, Duke Math. J. **33** (1966) 363–378.

[29] J. Rovnyak, V. Rovnyak, Self-reciprocal functions for the Hankel transformation of integer order, Duke Math. J. **34** (1967) 771–785.

[30] J. Rovnyak, V. Rovnyak, Sonine spaces of entire functions, J. Math. Anal. Appl. **27** (1969) 68–100.

[31] G. Szegö, *Orthogonal Polynomials*, AMS Colloquium Publications **23**, 1939.

[32] C. Tracy, H. Widom, Fredholm determinants, differential equations, and matrix models, Commun. Math. Phys. **163** (1994), 33–72.

[33] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed., Cambridge University Press, 1944.

[34] H. Widom, The asymptotics of a continuous analogue of orthogonal polynomials, J. Approx. Theory **77** (1994), 51–64