Reconstructing Stieltjes functions from their approximate values: a search for a needle in a haystack

Yury Grabovsky

SIAM J. Appl. Math. Vol. 82, No. 4, pp.1135-1166.

Abstract

Material response of real, passive, linear, time-invariant media to external influences is described by complex analytic functions of frequency that can always be written in terms of Stieltjes functions—a special class of analytic functions mapping complex upper half-plane into itself. Reconstructing such functions from their experimentally measured values at specific frequencies is one of the central problems that we address in this paper. A definitive reconstruction algorithm that produces a certificate of optimality as well as a graphical representation of the uncertainty of reconstruction is proposed. Its effectiveness is demonstrated in the context of the electrochemical impedance spectroscopy.

Contents

1 Introduction

The three fundamental physical principles—linearity, time-invariance, and passivity—are responsible for the ubiquity of Stieltjes functions in physics and engineering. *Stieltjes class* refers to a special class of complex analytic functions that describe the response of linear media or devices to external influences. If $E(t)$ denotes such an influence, and $J(t)$ the response, then the linear, time-invariant dependence of $J(t)$ on $E(t)$ could be formally written (without regard to the function spaces to which $E(t)$ and $J(t)$ may belong) as

$$J(t) = \gamma_0 E(t) + \int_{-\infty}^{t} a(t-\tau)E(\tau)d\tau, \tag{1.1}$$

where the causality principle, limiting the dependence of $J(t)$ only on the present and past values of $E(\tau)$, has been applied. For a mathematically rigorous discussion of convolution-type formulas, like (1.1) we refer the reader to many treatises on linear systems theory, e.g., [66, 67].

Due to the resemblance of the integral in (1.1) to a convolution, it is convenient to extend the memory kernel $a(s)$ to negative values of s by zero

$$a_0(s) = \begin{cases} a(s), & s \geq 0, \\ 0, & s < 0, \end{cases}$$

and rewrite (1.1) as a convolution

$$J(t) = \gamma_0 E(t) + \int_{-\infty}^{\infty} a_0(t-\tau)E(\tau)d\tau. \tag{1.2}$$

Assuming now that $a_0 \in L^1(\mathbb{R})$ and $\{E, J\} \subset L^2(\mathbb{R})$ we can take the Fourier transform of (1.2):

$$\widehat{J}(\omega) = (\gamma_0 + \widehat{a}_0(\omega))\widehat{E}(\omega). \tag{1.3}$$

Two different definitions of the Fourier transform are common in physics, depending on the representation of the input $E(t)$ as a superposition of "elementary harmonics". In signal processing and electrical circuit theory the elementary harmonics are functions $e^{i\omega t}$, leading to the representation

$$E(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{E}(\omega)e^{i\omega t}d\omega, \qquad \widehat{E}(\omega) = \int_{-\infty}^{\infty} E(t)e^{-i\omega t}dt.$$

In electromagnetics the elementary harmonics are the plane waves $e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} = E_0(\boldsymbol{x})e^{-i\omega t}$. In this case one uses

$$E(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{E}(\omega)e^{-i\omega t}d\omega, \qquad \widehat{E}(\omega) = \int_{-\infty}^{\infty} E(t)e^{i\omega t}dt.$$

In the former case (e.g., impedance of electrical circuits) causality, $a_0(s) = 0$, when $s < 0$, implies that $b(\omega) = \gamma_0 + \widehat{a}_0(\omega)$ is analytic in the lower half-plane of the complex ω-plane. In

the latter (e.g., complex dielectric permittivity), $b(\omega)$ is analytic in the upper half-plane. In each case the fact that the memory kernel $a(s)$ is a real-valued function implies that $b(\omega)$ has the symmetry

$$\overline{b(\omega)} = b(-\overline{\omega}). \tag{1.4}$$

The passivity principle, which says that the medium can only absorb or dissipate energy is a much more delicate condition leading to the nonnegativity of the real or imaginary parts of functions related to $b(\omega)$. In one way or another in each and every application the description of the linear, time-invariant, passive media response can be formulated in terms of functions from the Stieltjes class[1] $\mathfrak{S}$.

Definition 1.1. *We say that a complex function f analytic in $\mathbb{C} \setminus \mathbb{R}_+$ belongs to the Stieltjes class $\mathfrak{S}$ if it is either a nonnegative real constant or has the following three properties.*

(i) $\mathfrak{Im}(f(z)) > 0$ *for all* $z \in \mathbb{C}$ *with* $\mathfrak{Im}(z) > 0$*;*

(ii) $f(x) > 0$ *for all* $x < 0$*;*

(iii) $\overline{f(z)} = f(\overline{z})$*.*

For example, the complex electromagnetic permittivity $\varepsilon(\omega)$ of dielectrics can be written as $\varepsilon(\omega) = f(\omega^2)$, where $f \in \mathfrak{S}$ and $\mathfrak{Im}(\omega) > 0$ [44, 28]. Both the complex impedance and admittance functions $Z(\omega)$ and $Y(\omega)$, respectively, of electrical circuits made of resistors, capacitors, and inductive coils can be written as $Z(\omega) = i\omega f(\omega^2)$, where $f \in \mathfrak{S}$ and $\mathfrak{Im}(\omega) < 0$ [15]. In high energy physics it is the energy (or momentum) that plays the role of the complex variable and the scattering amplitude is the Stieltjes function [41, 48, 37, 57, 16]. In the theory of binary conducting composites the dependence of the effective conductivity σ^* of the composite on the ratio $h = \sigma_1/\sigma_2$ of the conductivities of two constituents is also expressible in terms of Stieltjes functions [9, 52, 32, 46] $\sigma^*/\sigma_1 = 1 + (1-h)f(-h)$, where $f \in \mathfrak{S}$. There are many other applications (see, e.g., [54]), where the models are linear and causality, time-invariance, and passivity (together with real values of the memory kernel) lead to system descriptions in terms of functions from the Stieltjes class $\mathfrak{S}$.

In this paper we consider the central *discrete* problem of the theory of Stieltjes functions that arises in all applications: the identification of $f \in \mathfrak{S}$ from n measurements at the n distinct points $\{z_1, \ldots, z_n\} \subset \mathbb{H}_+$, where $\mathbb{H}_+$ denotes the complex upper half-plane. The analyticity of $f \in \mathfrak{S}$ places constraints on the values $f(z_j)$. It turns out that the constraints are so delicate that even if one truncates the infinite decimal representations of the values $w_j = f(z_j)$ in order to store them as floating point numbers in a computer, one violates these constraints when $n \geq 15$. In most applications the values $w_j = f(z_j)$ are obtained through experimental measurements where the noise level is much larger than the round-off errors in floating point arithmetic. In view of these considerations the central problem is not the recovery of $f \in \mathfrak{S}$ from its exact values $w_j = f(z_j)$ but rather the minimization of the sum of squares

$$\Sigma(\boldsymbol{w}, \boldsymbol{z}) = \inf_{f \in \mathfrak{S}} \sum_{j=1}^{n} |f(z_j) - w_j|^2 \tag{1.5}$$

[1] There is no universal agreement on the names attached to various related classes of analytic functions. That is why we give a full formal definition here.

for a given set of noisy measurements $\boldsymbol{w} \in \mathbb{C}^n$. The problem of solving (1.5) bears only superficial resemblance to the classical linear least squares problem. The main difficulty is that the Stieltjes class $\mathfrak{S}$ is not a vector space but a convex cone.

In various guises this problem has been studied continuously for almost a century; see, e.g., [37, 24, 1, 56, 26, 59, 68, 58, 60, 13, 63]. Yet, so far, no definitive algorithm for solving (1.5) has emerged, and new algorithms and new papers on the subject continue to appear with unerring regularity (e.g., [11, 47, 51, 69], to give a taste). In this paper we propose such a definitive algorithm, described in Section 4, that is aimed to settle the question once and for all. The algorithm comes with a "certificate of optimality" based on the work of Caprini [16, 17, 18, 19]. The FORTRAN implementation of the algorithm is available from Github [34]. The method is easily extendable to weighted sums of squares as in Caprini's papers.

The main issue lies in intricacies of the geometry of the *interpolation body*

$$V(\boldsymbol{z}) = \{(f(z_1), \ldots, f(z_n)) \in \mathbb{C}^n : f \in \mathfrak{S}\}, \quad \boldsymbol{z} = (z_1, \ldots, z_n), \tag{1.6}$$

which is known to be a closed convex cone in $\mathbb{C}^n$ with nonempty interior. In practice, however, $V(\boldsymbol{z})$ is massively dimensionally degenerate, shaped very much like a needle or a sword. Even for modest values of n the smallest thickness of $V(\boldsymbol{z})$ is well below double precision floating point arithmetic. The proposed algorithm harnesses this dimensional degeneracy and turns it from a curse into a blessing. The algorithm produces not only the solution $f \in \mathfrak{S}$ of (1.5) but also shows the uncertainty associated with the given data (see Figure 5). Typical for analytic continuation problems the uncertainty balloons and explodes once one goes outside of the frequency range containing the measurements [25, 64, 7, 35, 36] (see Figure 6).

The algorithm described in Section 4 is an outcome of the understanding of the geometry of the interpolation body $V(\boldsymbol{z})$ discussed in Sections 2 and 3 as well as the optimality conditions described in Theorem 2.6. The key ingredient in the algorithm is the use of the local minima of the Caprini function to augment the ad-hoc basis of the space of Stieltjes functions. The final step is based on the realization that the near-optimal solution for a given noisy data is an optimal solution for "nearby data" representing a slightly different realization of the noise. The FORTRAN implementation of the algorithm is publicly available [34].

The fact that points z_j lie in the upper half-plane, and not on the real line is essential for our analysis. When some or all of the points z_j lie on the negative semi-axis a modification of our analysis given in [43, Chap. V.3] and [42] is necessary. Complementary to the setting of this paper is the situation where the imaginary part of $f(z)$ is known on a finite subinterval of the positive real axis, while the real part is known only at finitely many points in that same interval. Another complementary situation is when measurements are done in the time domain. The former is studied in [56], and the latter is addressed in [54, Chapter 6] and [50], where the collapse onto a needle is reflected in the fact that the time dependent bounds for an appropriate input and at a particular time almost coincide: one is viewing from a direction along the line of the needle [55].

This paper is structured as follows. We begin our discussion with the recollection of known results about Stieltjes functions in Section 2. In Section 3 we show that the interpolation body $V(\boldsymbol{z})$ is shaped like a needle or maybe like a sword. (Our language has an inadequate vocabulary limited to two- and three-dimensional shapes.) In Sections 4 and 5

we describe the algorithm. The performance of the algorithm is demonstrated in Section 6 in the context of electrochemistry, where the processes of corrosion and electrolysis that occur in batteries and in many other natural and man-made systems can be modeled by Voigt circuits—electrical circuits made only of resistors and capacitors [61, 5, 6]. The electrochemical impedance spectrum (EIS) function $Z(\omega)$ can then be written as $f(-i\omega)$ for some $f \in \mathfrak{S}$. Thus, the values $Z_j = f(-i\omega_j)$, $j = 1, \ldots, n$, can be measured experimentally at particular frequencies $\omega_1, \ldots, \omega_n$. Our algorithm takes noisy measurements of $Z_1, \ldots, Z_n$ as the input and generates physically admissible EIS functions $Z(\omega)$, representing them both numerically and as explicit complex impedance functions of small Voigt circuits. It also displays the certificate of optimality as well as the uncertainty of reconstruction of the EIS function for the specific data. Figure 5 shows the typical graphical output of the algorithm.

2 Preliminaries and background

2.1 The Nevanlinna-Pick theorem for Stieltjes functions

We recall two equivalent characterizations of the Stieltjes class $\mathfrak{S}$. One exhibits the centrality of property (i) in Definition 1.1, which is an expression of passivity in frequency domain. The other gives an explicit representation of all Stieltjes functions. Let $\mathbb{H}_+ = \{z \in \mathbb{C} : \mathfrak{Im}(z) > 0\}$ denote the complex upper half-plane.

Definition 2.1. *We say that $f(z)$ analytic in $\mathbb{H}_+$ is a Nevanlinna function if it is either a real constant function or $\mathfrak{Im}(f(z)) > 0$ for all $z \in \mathbb{H}_+$.*

Other names for this class, such as Herglotz functions, Pick functions, and R-functions are also used by various communities.

THEOREM **2.2.** *$f \in \mathfrak{S}$ if and only if both f and $z \mapsto zf(z)$ are Nevanlinna functions.*

As a corollary we see that the Stieltjes class has an involutive symmetry

$$f(z) \mapsto -\frac{1}{zf(z)}. \tag{2.1}$$

The second characterization of $\mathfrak{S}$ is more explicit.

THEOREM **2.3** (Stieltjes)**.** *$f \in \mathfrak{S}$ if and only if there exist $\gamma \geq 0$ and a positive Radon measure σ on $[0, +\infty)$, such that*

$$f(z) = \gamma + \int_0^\infty \frac{d\sigma(t)}{t - z}, \qquad \int_0^\infty \frac{d\sigma(t)}{1 + t} < +\infty. \tag{2.2}$$

The proof of both theorems can be found in [2, Chapter III, Addendum] or in [43, Addendum, Section 2]. We remark that given $f \in \mathfrak{S}$ we have

$$\gamma = \lim_{z \to \infty} f(z), \qquad \sigma(x) = \frac{1}{\pi} \lim_{y \to 0^+} \mathfrak{Im} f(x + iy), \tag{2.3}$$

where the second limit above is understood in the sense of distributions.

Our goal is the recovery of a Stieltjes function f from its approximately known values $f(z_1), \dots, f(z_n)$ at distinct points $\{z_1, \dots, z_n\} \subset \mathbb{H}_+$. In this regard we recall a well-known Nevanlinna-Pick theorem that, combined with Theorem 2.2, gives a criterion for $\boldsymbol{w} \in \mathbb{C}^n$ to lie in the interpolation body $V(\boldsymbol{z})$, given by (1.6).

THEOREM **2.4** (Nevanlinna-Pick)**.** *Let $\{z_1, \dots, z_n\} \subset \mathbb{H}_+$ be all distinct, and let $\boldsymbol{w} \in \mathbb{C}^n$. Then $\boldsymbol{w} \in V(\boldsymbol{z})$ if and only if the Nevanlinna-Pick matrices $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})$ and $\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})$ are nonnegative definite, where*

$$N_{jk}(\boldsymbol{z}, \boldsymbol{w}) = \frac{w_j - \overline{w_k}}{z_j - \overline{z_k}}, \qquad P_{jk}(\boldsymbol{z}, \boldsymbol{w}) = \frac{z_j w_j - \overline{z_k} w_k}{z_j - \overline{z_k}}. \tag{2.4}$$

Moreover, if $\boldsymbol{w} \in \partial V(\boldsymbol{z})$, so that either $\operatorname{rank}(\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})) < n$ or $\operatorname{rank}(\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})) < n$, then there is a unique rational function $f \in \mathfrak{S}$, such that $w_j = f(z_j)$, $j = 1, \dots, n$.

For the proof see, e.g., [62, Chaps. 16–18] (see also [42]).

2.2 Bounds on Stieltjes function values

The question we want to address now is about the freedom one has for the value $w = f(z)$, provided $f \in \mathfrak{S}$ and satisfies $f(z_j) = w_j$, $j = 1, \dots, n$. This freedom is represented by the admissible set of values

$$\mathcal{A}(z; \boldsymbol{z}, \boldsymbol{w}) = \{f(z) : f \in \mathfrak{S},\ f(z_j) = w_j,\ j = 1, \dots, n\}. \tag{2.5}$$

Such admissible sets are well understood and widely used in the context of effective properties of composite materials [32, 33, 31, 49, 23]. Our analysis is inspired by the one in [53] and reaches somewhat similar conclusions. However, it is based on Theorem 2.4 rather than the explicit representation of Stieltjes functions from Theorem 2.3, used in prior work. The question of bounds on values of Stieltjes functions in the case when the spectral measure σ is known in an interval of frequencies is addressed in [56]. The bounds in the case when the phase of the analytic function is known on a part of the boundary, and on the modulus on the remaining part have been derived in [3] by means of a modified Nevanlinna-Pick problem.

Let us assume that the data $\boldsymbol{w}$ lies in the interior of $V(\boldsymbol{z})$. By Theorem 2.4 the matrices $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})$ and $\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})$, given by (2.4), are positive definite. Then, by Sylvester's criterion (see, e.g., [38]) we obtain that the $\boldsymbol{N}([\boldsymbol{z}, z], [\boldsymbol{w}, w])$ and $\boldsymbol{P}([\boldsymbol{z}, z], [\boldsymbol{w}, w])$ matrices corresponding to the extended data $([\boldsymbol{z}, z], [\boldsymbol{w}, w])$ are positive definite if and only if

$$\det \boldsymbol{N}([\boldsymbol{z}, z], [\boldsymbol{w}, w]) > 0, \qquad \det \boldsymbol{P}([\boldsymbol{z}, z], [\boldsymbol{w}, w]) > 0. \tag{2.6}$$

We can make inequalities (2.6) explicit, since the determinants above are quadratic functions of w. Expanding the determinants with respect to the last column and the last row, so that w enters explicitly, we obtain

$$\det \boldsymbol{N}([\boldsymbol{z}, z], [\boldsymbol{w}, w]) = \frac{\mathfrak{Im}(w)}{\mathfrak{Im}(z)} \det \boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w}) - \alpha |w|^2 + 2\mathfrak{Re}(aw) - \beta,$$

where

$$\alpha = (\mathrm{cof}(\boldsymbol{N})\boldsymbol{\xi}(z), \boldsymbol{\xi}(z)), \qquad a = (\mathrm{cof}(\boldsymbol{N})\boldsymbol{\xi}(z), \boldsymbol{\eta}(z)), \qquad \beta = (\mathrm{cof}(\boldsymbol{N})\boldsymbol{\eta}(z), \boldsymbol{\eta}(z)),$$

and $\boldsymbol{N}$ stands for $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})$, and

$$\xi_k(z) = \frac{1}{z - \overline{z_k}}, \qquad \eta_k(z) = \frac{\overline{w_k}}{z - \overline{z_k}}, \qquad k = 1, \dots, n.$$

We conclude that $\det \boldsymbol{N}([\boldsymbol{z}, z], [\boldsymbol{w}, w]) > 0$ if and only if $|w - w^{(N)}(z)| < r_N(z)$, where

$$w^{(N)}(z) = \frac{\overline{a}}{\alpha} + i\frac{\det \boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})}{2\alpha \mathfrak{Im}(z)}, \qquad r_N(z)^2 = |w^{(N)}(z)|^2 - \frac{\beta}{\alpha}. \tag{2.7}$$

A similar analysis for the $\boldsymbol{P}$-matrix gives $|w - w^{(P)}(z)| < r_P(z)$, where

$$w^{(P)}(z) = \frac{\overline{a'}}{\alpha'} + i\frac{\overline{z}\det \boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})}{2\alpha' \mathfrak{Im}(z)}, \qquad r_P(z)^2 = |w^{(P)}(z)|^2 - \frac{\beta'}{\alpha'}, \tag{2.8}$$

and

$$\alpha' = (\mathrm{cof}(\boldsymbol{P})\boldsymbol{\xi}'(z), \boldsymbol{\xi}'(z)), \qquad a' = (\mathrm{cof}(\boldsymbol{P})\boldsymbol{\xi}'(z), \boldsymbol{\eta}'(z)), \qquad \beta' = (\mathrm{cof}(\boldsymbol{P})\boldsymbol{\eta}'(z), \boldsymbol{\eta}'(z)),$$

$$\boldsymbol{\xi}'(z) = z\boldsymbol{\xi}(z), \qquad \boldsymbol{\eta}'(z) = z\boldsymbol{\eta}(z) - \overline{\boldsymbol{w}}, \qquad \boldsymbol{P} = \boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})$$

Let us now estimate $r_N(z)$. (The estimate for $r_P(z)$ would be fully analogous.) The key observation is the inequality between α, β and a: $|a|^2 \le \alpha\beta$. Then

$$r_N(z)^2 = \frac{|a|^2 - \alpha\beta}{\alpha^2} + \rho^2 - \frac{2\mathfrak{Im}(a)\rho}{\alpha} \le 2\rho\left(\rho - \frac{\mathfrak{Im}(a)}{\alpha}\right) = 2\rho\mathfrak{Im}(w^{(N)}(z)), \quad \rho = \frac{\det \boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})}{2\alpha\mathfrak{Im}(z)}.$$

Thus, we have obtained the estimate

$$r_N(z)^2 \le \frac{\mathfrak{Im}(w^{(N)}(z))}{\mathfrak{Im}(z)}(\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})^{-T}\boldsymbol{\xi}(z), \boldsymbol{\xi}(z))^{-1}. \tag{2.9}$$

A similar calculation for the $\boldsymbol{P}$ matrix gives the estimate

$$r_P(z)^2 \le \frac{\mathfrak{Im}(zw^{(P)}(z))}{\mathfrak{Im}(z)}(\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})^{-T}\boldsymbol{\xi}'(z), \boldsymbol{\xi}'(z))^{-1}. \tag{2.10}$$

The main feature of matrices $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})$ and $\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})$ is the exponential decay of their eigenvalues due to their rank two displacement structure [8]:

$$\boldsymbol{D}(\boldsymbol{z})\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w}) - \boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})\boldsymbol{D}(\boldsymbol{z})^* = \boldsymbol{w} \otimes \mathbf{1} - \mathbf{1} \otimes \overline{\boldsymbol{w}}, \tag{2.11}$$

$$\boldsymbol{D}(\boldsymbol{z})\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w}) - \boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})\boldsymbol{D}(\boldsymbol{z})^* = \boldsymbol{D}(\boldsymbol{z})\boldsymbol{w} \otimes \mathbf{1} - \mathbf{1} \otimes \boldsymbol{D}(\overline{\boldsymbol{z}})\overline{\boldsymbol{w}}, \tag{2.12}$$

where $\boldsymbol{D}(\boldsymbol{z})$ is a diagonal matrix with numbers z_j on the main diagonal and $\mathbf{1}$ is a vector of ones.

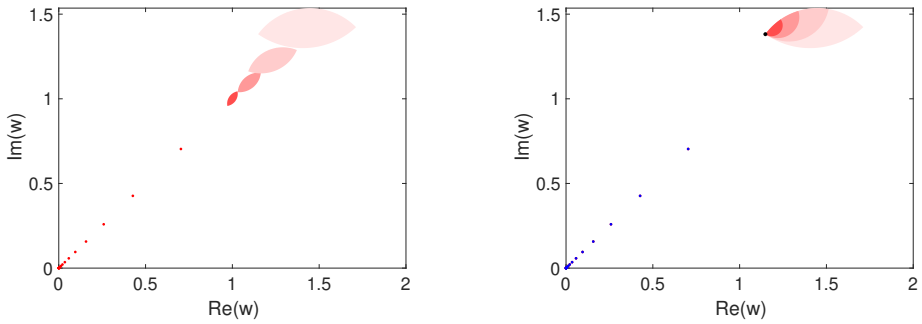

Figure 1: Dependence of the admissible set $\mathcal{A}(z;\boldsymbol{z},\boldsymbol{w})$ on the location of z, relative to the data z_j (left) and on the location of $\boldsymbol{w}$, relative to $\partial V(\boldsymbol{z})$ (right).

If the vector $\boldsymbol{\xi}(z)$ has a substantial projection onto the space spanned by the eigenvectors of $\boldsymbol{N}(\boldsymbol{z},\boldsymbol{w})$ and $\boldsymbol{P}(\boldsymbol{z},\boldsymbol{w})$ with exponentially small eigenvalues, then $(\boldsymbol{N}(\boldsymbol{z},\boldsymbol{w})^{-T}\boldsymbol{\xi}(z),\boldsymbol{\xi}(z))$ and $(\boldsymbol{P}(\boldsymbol{z},\boldsymbol{w})^{-T}\boldsymbol{\xi}'(z),\boldsymbol{\xi}'(z))$ will be exponentially large (as functions of n). This shows that $r_N(z)$ and $r_P(z)$ can easily become exponentially small even for relatively small values of n. In fact, $r_N(z)=0$ or $r_P(z)=0$ (or both) whenever $\boldsymbol{w}\in\partial V(\boldsymbol{z})$. This may lead one to think that fixing more than 15–20 values of a Stieltjes function determines it for all practical intents and purposes. The truth is more nuanced. It depends very strongly on the relative location of z and z_j and on the exact location of $\boldsymbol{w}\in V(\boldsymbol{z})$ relative to $\partial V(\boldsymbol{z})$. Formally, $V(\boldsymbol{z})$ is a closed convex cone in $\mathbb{C}^n$ with nonempty interior. In practice, its geometry resembles that of a thin knife blade, rather than a party hat, so that very small random perturbations of points in the interior of $V(\boldsymbol{z})$ will throw them outside. In other words, no matter where the point $\boldsymbol{w}$ is in $V(\boldsymbol{z})$, it is never far from $\partial V(\boldsymbol{z})$, where, as we have just observed, the region of admissible values $\mathcal{A}(z;\boldsymbol{z},\boldsymbol{w})$ degenerates to a point. What is somewhat counterintuitive is that for points $\boldsymbol{w}$ in the interior of $V(\boldsymbol{z})$ the set $\mathcal{A}(z;\boldsymbol{z},\boldsymbol{w})$ can be rather large, depending on the location of z relative to points z_j. The left panel of Figure 1 illustrates this effect in the simple example

$$z_j = ie^{0.01+j}, \quad w_j = f(z_j),\ j=0,1,\ldots,19, \quad f(z)=\frac{1}{\sqrt{-z}}. \tag{2.13}$$

We see how the shaded lens-shaped regions grow in size as the point z, taking values $i/2$, $i/2.4$, $i/3$, and $i/4$ moves "away" from the data z_j, given in (2.13). Our discussion also shows that if we move $\boldsymbol{w}$ from the interior of $V(\boldsymbol{z})$ to its boundary the admissible set will shrink to a point. The right panel of Figure 1 illustrates this effect when we move from $\boldsymbol{w}$, given in (2.13), which lies in the interior of $V(\boldsymbol{z})$, to $\partial V(\boldsymbol{z})$ along any random direction $\boldsymbol{u}$, which we have chosen (arbitrarily) to have all components equal to -1. The corresponding point $\widetilde{\boldsymbol{w}}\in\partial V(\boldsymbol{z})$ satisfies $|\boldsymbol{w}-\widetilde{\boldsymbol{w}}|/|\boldsymbol{w}|<10^{-4}$, as we have verified numerically. In the right panel of Figure 1 we plotted the original points w_j in red and the perturbed points $\widetilde{w}_j$ in blue, except one cannot see a difference between them in the figure. The set $\mathcal{A}(z;\boldsymbol{z},\widetilde{\boldsymbol{w}})$ degenerates to a point shown in black, while the sets $\mathcal{A}_t=\mathcal{A}(z;\boldsymbol{z},t\widetilde{\boldsymbol{w}}+(1-t)\boldsymbol{w})$ for three intermediate values of t are shown by progressively darker shading. The values we have chosen are $t_1=1-2\cdot10^{-5}$, $t_2=1-7\cdot10^{-6}$, and $t_3=1-3\cdot10^{-6}$. This indicates that if

we move uniformly from $\boldsymbol{w} \in V(\boldsymbol{z})$ to $\widetilde{\boldsymbol{w}} \in \partial V(\boldsymbol{z})$, the admissible sets $\mathcal{A}_t$ remain virtually unchanged until we get very close to $\partial V(\boldsymbol{z})$. The admissible set then collapses rather abruptly to a point corresponding to $\widetilde{\boldsymbol{w}} \in \partial V(\boldsymbol{z})$. This complicated, almost discontinuous behavior occurs as we move from $\boldsymbol{w}$ to $\widetilde{\boldsymbol{w}}$, which can barely be distinguished in right panel of Figure 1.

The computations needed to make Figure 1 have been done with the Advanpix Multiprecision Computing Toolbox for MATLAB (https://www.advanpix.com) using 100 digits of precision.

2.3 Interpolation

Let us assume now that the data $(\boldsymbol{z}, \boldsymbol{w}) \in \mathbb{C}^{2n}$ satisfies the conditions of Theorem 2.4, i.e., $\boldsymbol{w} \in V(\boldsymbol{z})$. Our goal is to construct an interpolant $f \in \mathfrak{S}$, such that $f(z_j) = w_j$ for all $j = 1, \dots, n$. We begin with the case $n = 1$. According to Theorem 2.4, the necessary and sufficient condition for the existence of such a function is $\mathfrak{Im}(w_1) \geq 0$ and $\mathfrak{Im}(z_1 w_1) \geq 0$. Of course, if $\mathfrak{Im}(w_1) = 0$, then $w_1 \geq 0$, according to the second inequality, and $f(z) = w_1$ for all z. If $\mathfrak{Im}(z_1 w_1) = 0$, then $zf(z)$ must be a real constant, and hence, according to the first inequality, $f = -\sigma/z$, where $\sigma = -z_1 w_1 \geq 0$. Let us now assume that

$$\mathfrak{Im}(w_1) > 0, \qquad \mathfrak{Im}(z_1 w_1) > 0, \tag{2.14}$$

and characterize the set

$$\mathfrak{S}(z_1, w_1) = \{f \in \mathfrak{S} : f(z_1) = w_1\}.$$

We look for the answer in the same form as in the case of polynomials $\mathcal{P}$, where the set $\mathcal{P}(z_1, w_1)$ of all polynomials $p \in \mathcal{P}$ satisfying $p(z_1) = w_1$ can be described as

$$\mathcal{P}(z_1, w_1) = \{p(z) = (z - z_1)q(z) + w_1 : q \in \mathcal{P}\}.$$

Moreover, distinct polynomials $q \in \mathcal{P}$ correspond to distinct polynomials $p \in \mathcal{P}(z_1, w_1)$. By analogy with polynomials, we want to parametrize the set $\mathfrak{S}(z_1, w_1)$ by elements of $\mathfrak{S}$ in the same fashion as $\mathcal{P}(z_1, w_1)$ is parametrized by elements of $\mathcal{P}$. Of course, we expect that the parametrization will be more complicated than in the case of polynomials. The desired parametrization has already been found in [42], but the derivation here is not a routine calculation, differing from the one in [42].

According to Theorem 2.4 the set of all admissible values $f(z)$ for $f \in \mathfrak{S}(z_1, w_1)$ is described by the inequalities

$$\det \boldsymbol{N}([z_1, z], [w_1, f(z)]) = \frac{\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)} \frac{\mathfrak{Im} f(z)}{\mathfrak{Im}(z)} - \left| \frac{f(z) - \overline{w_1}}{z - \overline{z_1}} \right|^2 \geq 0, \tag{2.15}$$

$$\det \boldsymbol{P}([z_1, z], [w_1, f(z)]) = \frac{\mathfrak{Im}(z_1 w_1)}{\mathfrak{Im}(z_1)} \frac{\mathfrak{Im}(zf(z))}{\mathfrak{Im}(z)} - \left| \frac{zf(z) - \overline{z_1 w_1}}{z - \overline{z_1}} \right|^2 \geq 0. \tag{2.16}$$

Inequalities (2.15), (2.16) place $f(z)$ inside closed disks $D_N(z_1, w_1, z)$ and $D_P(z_1, w_1, z)$, respectively. At the same time, Theorem 2.2 says that $f \in \mathfrak{S}$ if and only if $f(z)$ lies in the intersection of two closed half-planes $\overline{\mathbb{H}}_+ = \{w \in \mathbb{C} : \mathfrak{Im}(w) \geq 0\}$ and $\mathbb{H}_z = \{w \in$

$\mathbb{C} : \mathfrak{Im}(zw) \geq 0\}$, for every $z \in \mathbb{H}_+$. This gives the idea of the desired parametrization of $\mathfrak{S}(z_1, w_1)$ by elements of $\mathfrak{S}$. This idea is at the core of the so-called Potapov method of "fundamental matrix inequalities" [10]. It has been implemented for interpolation of matrix-valued Stieltjes functions in [27]. We present the argument and calculation both for the sake of completeness and because the formulas here are used in our algorithm.

For every $z \in \mathbb{H}_+$ there exists[2] a fractional-linear transformation

$$T_{z_1,w_1,z}(w) = \frac{L_{11}(z)w + L_{12}(z)}{L_{21}(z)w + L_{22}(z)}$$

that maps $\mathcal{A}(z; z_1, w_1) = D_N(z_1, w_1, z) \cap D_P(z_1, w_1, z)$ bijectively onto $\overline{\mathbb{H}}_+ \cap \mathbb{H}_z$. In order to derive the formula for $T_{z_1,w_1,z}(w)$ we exploit the simplicity of Stieltjes functions corresponding to the points on the boundary of the admissible regions $\mathcal{A}(z; z_1, w_1)$ and $\overline{\mathbb{H}}_+ \cap \mathbb{H}_z$. The idea is that while the set of functions in $\mathfrak{S}(z_1, w_1)$ is very large, if (2.14) is satisfied, it degenerates to a single point if any of the inequalities in (2.14) become equalities, as we have already discussed. The same holds for inequalities in (2.15), (2.16). If we have equality in (2.15), then there exists a nonzero vector $\boldsymbol{\xi} = (\xi_1, \xi_2) \in \ker \boldsymbol{N}([z_1, z_2], [f(z_1), f(z_2)])$, where for convenience of notation we replaced z with z_2. Using representation (2.2), we compute

$$\frac{f(z_j) - \overline{f(z_k)}}{z_j - \overline{z_k}} = \int_0^\infty \frac{d\sigma(t)}{(t - z_j)(t - \overline{z_k})}, \quad j, k = 1, 2.$$

Thus,

$$0 = (\boldsymbol{N}([z_1, z_2], [f(z_1), f(z_2)])\boldsymbol{\xi}, \boldsymbol{\xi})_{\mathbb{C}^2} = \int_0^\infty \left| \frac{\xi_1}{t - z_1} + \frac{\xi_2}{t - z_2} \right|^2 d\sigma(t).$$

This means that there is a nonzero vector $(\xi_1, \xi_2) \in \mathbb{C}^2$, such that the function

$$\phi(t) = \frac{\xi_1}{t - z_1} + \frac{\xi_2}{t - z_2}$$

is identically zero on the support of σ. Since $z_1 \neq z_2$ we conclude that the support of σ must be a single point, and the corresponding Stieltjes function must have the form

$$f(z) = \gamma + \frac{\sigma}{t - z}. \tag{2.17}$$

Conversely, if the spectral measure of $f \in \mathfrak{S}(w_1, z_1)$ is supported on a single point, then we have an equality in (2.15) for any $z \in \mathbb{H}_+$.

A similar analysis can be done for the case of an equality in (2.16):

$$0 = (\boldsymbol{P}([z_1, z_2], [f(z_1), f(z_2)])\boldsymbol{\xi}, \boldsymbol{\xi})_{\mathbb{C}^2} = \gamma|\xi_1 + \xi_2|^2 + \int_0^\infty \left| \frac{\xi_1}{t - z_1} + \frac{\xi_2}{t - z_k} \right|^2 t d\sigma(t).$$

This equality implies that $f(z)$ must have either of two forms

$$f(z) = \gamma - \frac{\sigma_0}{z} \text{ or } f(z) = -\frac{\sigma_0}{z} + \frac{\sigma_1}{t_1 - z}. \tag{2.18}$$

[2]Unique modulo $w \mapsto \alpha w$, $\alpha > 0$ and $w \mapsto -1/(zw)$.

We can regard the first form of $f(z)$ as a limit of the second one when $\sigma_1 = \gamma t_1$, as $t_1 \to +\infty$.

Now, since the fractional-linear transformation $T_{z_1,w_1,z}(w)$ maps the boundary of $\mathcal{A}(z; z_1, w_1)$ onto the boundary of $\mathbb{H}_+ \cap \mathbb{H}_z$, the set

$$S_N(z_1, w_1) = \{f \in \mathfrak{S}(z_1, w_1) : \det \boldsymbol{N}([z_1, z_2], [f(z_1), f(z_2)]) = 0\},$$

consisting of functions (2.17) must be mapped by $T_{z_1,w_1,z}$ onto the set $\{g \in \mathfrak{S} : \mathfrak{Im}(g(z)) = 0\}$, while the set

$$S_P(z_1, w_1) = \{f \in \mathfrak{S}(z_1, w_1) : \det \boldsymbol{P}([z_1, z_2], [f(z_1), f(z_2)]) = 0\},$$

consisting of functions (2.18) must be mapped by $T_{z_1,w_1,z}$ onto the set $\{g \in \mathfrak{S} : \mathfrak{Im}(zg(z)) = 0\}$. This gives us the desired equations. If we write $g(z) = T_{z_1,w_1,z}(f(z))$, then

$$f(z) = \frac{L_{22}(z)g(z) - L_{12}(z)}{L_{11}(z) - g(z)L_{21}(z)}. \tag{2.19}$$

Hence, the coefficients $L_{ij}(z)$ must satisfy the following properties: for any $\mu \geq 0$ the function $g(z) = \mu$ must be mapped into an element of $S_N(z_1, w_1)$, i.e., function of the form (2.17), while for any $\nu \geq 0$ the function $g(z) = -\nu/z$ must be mapped to an element of $S_P(z_1, w_1)$, i.e., function of the form (2.18). We therefore obtain the following system of equations for the unknown coefficients $L_{ij}(z)$:

$$\begin{cases} \dfrac{L_{22}(z)\mu - L_{12}(z)}{L_{11}(z) - \mu L_{21}(z)} = \gamma(\mu) + \frac{\sigma(\mu)}{t(\mu) - z}, \\ -\dfrac{L_{22}(z)\nu + zL_{12}(z)}{zL_{11}(z) + \nu L_{21}(z)} = -\dfrac{\sigma_0(\nu)}{z} + \dfrac{\sigma_1(\nu)}{t_1(\nu) - z}, \\ \dfrac{L_{22}(z_1)\mu - L_{12}(z_1)}{L_{11}(z_1) - \mu L_{21}(z_1)} = w_1, \\ -\dfrac{L_{22}(z_1)\nu + z_1 L_{12}(z_1)}{z_1 L_{11}(z_1) + \nu L_{21}(z_1)} = w_1. \end{cases} \tag{2.20}$$

The last two equations are easy to solve, since the coefficients $L_{ij}(z)$ depend neither on μ nor on ν. Thus, we must require that

$$\begin{cases} L_{11}(z_1)w_1 + L_{12}(z_1) = 0, \\ L_{21}(z_1)w_1 + L_{22}(z_1) = 0. \end{cases} \tag{2.21}$$

In order to solve the other two equations we first observe that equations

$$\begin{cases} \phi_N(z_1) = \gamma + \dfrac{\sigma}{t - z_1} = w_1, \\ \phi_P(z_1) = -\dfrac{\sigma_0}{z_1} + \dfrac{\sigma_1}{t_1 - z_1} = w_1 \end{cases} \tag{2.22}$$

determine two 1-parameter families of solutions $\phi_N(z;t)$ and $\phi_P(z;t_1)$, tracing the boundaries of D_N and D_P, respectively. Explicitly, we find

$$\begin{cases} \sigma = \dfrac{\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)}|t-z_1|^2, \\ \gamma = \dfrac{\mathfrak{Im}(z_1w_1)}{\mathfrak{Im}(z_1)} - t\dfrac{\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)}, \\ \sigma_0 = \left(\dfrac{\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)} - \dfrac{1}{t_1}\dfrac{\mathfrak{Im}(z_1w_1)}{\mathfrak{Im}(z_1)}\right)|z_1|^2, \\ \sigma_1 = \dfrac{|t_1-z_1|^2}{t_1}\dfrac{\mathfrak{Im}(z_1w_1)}{\mathfrak{Im}(z_1)}. \end{cases} \tag{2.23}$$

This shows that for functions $\phi_N(z)$ and $\phi_P(z)$ to be in $\mathfrak{S}$ it is necessary and sufficient that $t \in [0, t_*]$ and $t_1 \in [t_*, \infty]$, where

$$t_* = \frac{\mathfrak{Im}(z_1w_1)}{\mathfrak{Im}(w_1)}.$$

We then see that when $t = t_1 = t_*$ we have

$$\phi_N(z) = \phi_P(z) = \frac{\sigma_*}{t_*-z}, \quad \sigma_* = \frac{\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)}|t_*-z_1|^2 = \frac{|w_1|^2\mathfrak{Im}(z_1)}{\mathfrak{Im}(w_1)}, \tag{2.24}$$

while when $t = 0$ and $t_1 = \infty$ we have

$$\phi_N(z) = \phi_P(z) = \gamma_* - \frac{\sigma^*}{z}, \quad \gamma_* = \frac{\mathfrak{Im}(z_1w_1)}{\mathfrak{Im}(z_1)}, \quad \sigma^* = \frac{|z_1|^2\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)}. \tag{2.25}$$

The correspondence between the two points of intersection of ∂D_N and ∂D_P and the two points of intersection of $\partial\mathbb{H}_+$ and $\partial\mathbb{H}_z$, characterized by $\mu = \nu = 0$ and $\mu = \nu = \infty$, respectively, is determined unambiguously by the orientation-preserving property of fractional-linear transformations. We conclude that the point $t = t_1 = t_*$ corresponds to $\mu = \nu = 0$, while the point $\mu = \nu = \infty$ corresponds to $t = 0$, $t_1 = \infty$. Hence, we have the equations

$$-\frac{L_{12}(z)}{L_{11}(z)} = \frac{\sigma_*}{t_*-z}, \qquad -\frac{L_{22}(z)}{L_{21}(z)} = \gamma_* - \frac{\sigma^*}{z},$$

which permit us to eliminate L_{11} and L_{22}. Denoting $\Psi(z) = L_{21}(z)/L_{12}(z)$, we obtain from the first equation in (2.20)

$$\frac{\left(\frac{\sigma^*}{z} - \gamma_*\right)\mu\Psi(z) - 1}{\frac{(z-t_*)}{\sigma_*} - \mu\Psi(z)} = \gamma(t(\mu)) + \frac{\sigma(t(\mu))}{t(\mu)-z}, \quad \gamma(t) = \frac{\gamma_*}{t_*}(t_*-t), \quad \sigma(t) = \frac{\gamma_*}{t_*}|t-z_1|^2$$

Solving this equation for Ψ (on Maple) we obtain that $\Psi(z)/z$ is a ratio of two quadratic polynomials in z with

$$\lim_{z\to\infty}\frac{\Psi(z)}{z} = \frac{t(\mu)-t_*}{\sigma_*\mu t(\mu)}.$$

Since $\Psi(z)$ does not depend on μ we conclude that there exists $\alpha \in \mathbb{R}$, such that

$$t(\mu) = \frac{t_*}{1 - \alpha\sigma_*\mu}. \tag{2.26}$$

Then, substituting (2.26) together with the parameter values

$$t_* = \frac{\mathfrak{Im}(z_1 w_1)}{\mathfrak{Im}(w_1)}, \quad \sigma_* = \frac{|w_1|^2\mathfrak{Im}(z_1)}{\mathfrak{Im}(w_1)}, \quad \gamma_* = \frac{\mathfrak{Im}(z_1 w_1)}{\mathfrak{Im}(z_1)}, \quad \sigma^* = \frac{|z_1|^2\mathfrak{Im}(w_1)}{\mathfrak{Im}(z_1)} \tag{2.27}$$

into the formula for $\Psi(z)$ in Maple we obtain that $\Psi(z) = \alpha z$. We can now go back and recover the formulas for all of the coefficients $L_{ij}(z)$:

$$\frac{L_{11}(z)}{L_{12}(z)} = \frac{z - t_*}{\sigma_*}, \qquad \frac{L_{22}(z)}{L_{12}(z)} = \alpha(\sigma^* - \gamma_* z).$$

In this case it is easy to see that equations (2.21) will be satisfied. Thus, the desired fractional-linear transformation is given by

$$T_{z_1,w_1,z}(f(z)) = g(z) = \frac{1}{\alpha} \cdot \frac{(z - t_*)f(z) + \sigma_*}{zf(z) + \sigma^* - \gamma_* z}, \tag{2.28}$$

where the sign of α needs to be determined. It is easy to do when we examine the behavior of functions $f(z)$ and $g(z)$ at infinity. If we define

$$\gamma_g = \lim_{z\to\infty} g(z), \qquad \gamma_f = \lim_{z\to\infty} f(z),$$

then, according to (2.28)

$$\gamma_g = \frac{\gamma_f}{\alpha(\gamma_f - \gamma_*)}, \qquad \gamma_f = \frac{\alpha\gamma_*\gamma_g}{\alpha\gamma_g - 1}.$$

Since for any $g \in \mathfrak{S}$ we must get $f \in \mathfrak{S}(z_1, w_1)$ we conclude that we must have $\alpha < 0$. Since multiplication by $-\alpha > 0$ maps the intersection of the two half-planes $\mathbb{H}_+$ and $\mathbb{H}_z$ onto itself, any choice of $\alpha < 0$ will produce a valid parametrization of $f \in \mathfrak{S}(z_1, w_1)$ by $g \in \mathfrak{S}$. For simplicity we set $\alpha = -1$ and obtain the desired parametrization of $\mathfrak{S}(z_1, w_1)$:

$$\mathfrak{S}(z_1, w_1) = \left\{ f(z) = \frac{g(z)(\gamma_* z - \sigma^*) - \sigma_*}{zg(z) + z - t_*} : g \in \mathfrak{S} \right\}, \tag{2.29}$$

where the parameters γ_*, σ^*, σ_*, and t_* are given in (2.27), provided inequalities (2.14) hold. The exact same formula (but with different normalization α for $g(z)$) has been obtained[3] in [42].

The parametrization (2.29) has several useful properties. At infinity we obtain

$$\gamma_f = \frac{\gamma_*\gamma_g}{\gamma_g + 1}. \tag{2.30}$$

[3]There is a typo in [42]: $a_{12}^{(1)}$ should be $|c_1|^2/w_1$.

This can be important in applications in the context of complex electromagnetic susceptibility functions, where the physically mandated assumption on the interpolant $f \in \mathfrak{S}$ is $\gamma_f = 0$. Formula (2.30) shows that $\gamma_f = 0$ if and only if $\gamma_g = 0$. This means that starting with $g(z) = 0$ and iterating formula (2.29) will always result in a decaying Stieltjes function $f(z)$.

Another nice feature of (2.29) is the degree-reduction property. To exhibit it let us solve (2.29) for $g(z)$:

$$g(z) = \frac{f(z)(t_* - z) - \sigma_*}{zf(z) + \sigma^* - \gamma_* z}. \tag{2.31}$$

THEOREM **2.5.** *Suppose $f \in \mathfrak{S}(z_1, w_1)$ is a rational function of degree $n \geq 1$ in the sense that $f(z) = P_n(z)/Q_n(z)$, where the degree of Q_n is n, while the degree of P_n is either n or $n-1$, while P_n and Q_n have no common roots. Then $g(z)$, given by (2.31) is a rational function in $\mathfrak{S}$ of degree $n-1$ in the same sense as above.*

Proof. The essential feature of (2.31) is that all of its coefficients $L_{ij}(z)$ are linear in z. If $f(z) = P_n(z)/Q_n(z)$, then

$$L_{11}(z)f(z) + L_{12}(z) = \frac{L_{11}(z)P_n(z) + L_{12}(z)Q_n(z)}{Q_n(z)}.$$

Formulas (2.21) imply that the polynomial $L_{11}(z)P_n(z) + L_{12}(z)Q_n(z)$ will have a pair of complex conjugate roots z_1 and $\overline{z_1}$. We can therefore write

$$L_{11}(z)P_n(z) + L_{12}(z)Q_n(z) = (z - z_1)(z - \overline{z_1})T^+(z), \qquad L_{11} = t_* - z, \quad L_{12} = -\sigma_*.$$

It follows that $\deg(T^+) = n - 1$ if $\deg(P_n) = n$ and $\deg(T^+) \leq n - 2$ if $\deg(P_n) = n - 1$. Similarly,

$$L_{21}(z)P_n(z) + L_{22}(z)Q_n(z) = (z - z_1)(z - \overline{z_1})T^-(z), \qquad L_{21} = z, \quad L_{22} = \sigma^* - \gamma_* z,$$

and $\deg(T^-) \leq n - 1$, if $\deg(P_n) = n$, and $\deg(T^-) = n - 1$, if $\deg(P_n) = n - 1$. Since $g(z) = T^+(z)/T^-(z)$ is in $\mathfrak{S}$ the degree of $T^-(z)$ can be at most one above the degree of $T^+(z)$. This shows that we can only have equalities in the degree inequalities above. Finally, if T^+ and T^- have common roots, then formula (2.29) would imply that $f(z)$ is a rational function of degree strictly less than n, contradicting our assumption. □

The parametrization (2.29) of $\mathfrak{S}(z_1, w_1)$ by elements of $\mathfrak{S}$ leads to the recursive interpolation algorithm. Given the data $\boldsymbol{w} \in V(\boldsymbol{z})$, $\boldsymbol{z} = (z_1, \ldots, z_n)$ for n distinct points $\{z_1, \ldots, z_n\} \subset \mathbb{H}_+$, we define the interpolant $f(z)$ by (2.29), where $g(z) \in \mathfrak{S}$ satisfies $n - 1$ constraints

$$g(z_j) = \frac{L_{11}(z_j)w_j + L_{12}(z_j)}{w_j L_{21}(z_j) + L_{22}(z_j)}, \quad j = 2, \ldots, n, \tag{2.32}$$

provided

$$w_j L_{21}(z_j) + L_{22}(z_j) \neq 0, \qquad j = 2, \ldots, n.$$

In that case $f(z_j) = w_j$, $j = 2, \ldots, n$, and

$$f(z_1) = \frac{L_{22}(z_1)g(z_1) - L_{12}(z_1)}{L_{11}(z_1) - g(z_1)L_{21}(z_1)}.$$

Using equations (2.21) we obtain

$$f(z_1) = \frac{-L_{21}(z_1)w_1 g(z_1) + L_{11}(z_1)w_1}{L_{11}(z_1) - g(z_1)L_{21}(z_1)} = w_1,$$

provided $L_{11}(z_1) - g(z_1)L_{21}(z_1) \neq 0$. This condition is always satisfied, since linear functions $L_{ij}(z)$ are such that $f \in \mathfrak{S}$ for any $g \in \mathfrak{S}$. This requires that the denominator in (2.19) never vanishes when $z \in \mathbb{H}_+$.

In order to finish the analysis we need to consider the special case when there exists $k \in \{2, \ldots, n\}$, such that

$$L_{22}(z_k) + w_k L_{21}(z_k) = 0. \tag{2.33}$$

In this case the corresponding relation (2.32) will be undefined. But in this case the four real equations

$$\begin{cases} L_{22}(z_k) + w_k L_{21}(z_k) = 0, \\ L_{22}(z_1) + w_1 L_{21}(z_1) = 0 \end{cases}$$

form a linear homogeneous system of equations with four real unknowns a_{21}, a_{22}, b_{21}, and b_{22}, where

$$L_{21}(z) = a_{21}z + b_{21}, \qquad L_{22}(z) = a_{22}z + b_{22}.$$

Thus, the determinant of this system must vanish. Maple calculations show that this implies that $\det \boldsymbol{N} = 0$, where

$$\boldsymbol{N} = \begin{bmatrix} \dfrac{w_1 - \overline{w_1}}{z_1 - \overline{z_1}} & \dfrac{w_1 - \overline{w_k}}{z_1 - \overline{z_k}} \\ \dfrac{w_k - \overline{w_1}}{z_k - \overline{z_1}} & \dfrac{w_k - \overline{w_k}}{z_k - \overline{z_k}} \end{bmatrix}.$$

We have already proved that in this case the support of σ must be a single point. Thus, when (2.33) is satisfied we just return the rational function $\phi_N(z)$, given by (2.25). Indeed, (2.33) implies

$$w_k = \gamma_* - \frac{\sigma^*}{z_k} = \phi_N(z_k).$$

At the same time $\phi_N(z)$ also satisfies $\phi_N(z_1) = w_1$. It follows that $f(z) = \phi_N(z)$.

2.4 The least squares problem

For $\boldsymbol{w} \in \mathbb{C}^n$ there are two mutually exclusive logical possibilities. Either $\boldsymbol{w} \in V(\boldsymbol{z})$ or $\boldsymbol{w} \notin V(\boldsymbol{z})$. The former case, called the *interpolation problem* has been considered in the previous section. In the latter case, when there is no Stieltjes function f satisfying $\boldsymbol{w} = f(\boldsymbol{z})$, we want to solve the *least squares problem* (1.5), which can be also reformulated as

$$\Sigma(\boldsymbol{w}, \boldsymbol{z}) = \min_{\boldsymbol{p} \in V(\boldsymbol{z})} |\boldsymbol{p} - \boldsymbol{w}|. \tag{2.34}$$

The minimizer $\boldsymbol{p}^*$ of (2.34) exists because $V(\boldsymbol{z})$ is a closed subset of $\mathbb{C}^n$. It is unique because $V(\boldsymbol{z})$ is convex. Moreover, since $\boldsymbol{w} \notin V(\boldsymbol{z})$, the minimizer $\boldsymbol{p}^*$ must lie on the boundary of $V(\boldsymbol{z})$. In this case, the Nevanlinna-Pick theorem (Theorem 2.4) for the Stieltjes class says that there exists a unique Stieltjes function $f_* \in \mathfrak{S}$ satisfying $f_*(\boldsymbol{z}) = \boldsymbol{p}^*$.

Let us analyze the properties of this unique minimizer. Here we follow the analysis of Caprini [18], who derived the necessary and sufficient conditions for a minimizer in (2.34). Caprini's method is based on our ability to compute the effect of variations of γ and spectral measure σ in representation (2.2) on the functional we want to minimize. Suppose that

$$f_*(z) = \gamma + \int_0^\infty \frac{d\sigma(t)}{t - z}$$

is the minimizer in (1.5). Then $p_j^* = f_*(z_j)$ minimizes (2.34). Let

$$\widetilde{f}(z) = \widetilde{\gamma} + \int_0^\infty \frac{d\widetilde{\sigma}(t)}{t - z} \tag{2.35}$$

be a competitor in (1.5), and let $\widetilde{p}_j = \widetilde{f}(z_j)$. The variation $\phi = \widetilde{f} - f_*$ can then be written as

$$\phi(z) = \Delta\gamma + \int_0^\infty \frac{d\nu(t)}{t - z}, \qquad \nu = \widetilde{\sigma} - \sigma, \quad \Delta\gamma = \widetilde{\gamma} - \gamma.$$

We then compute

$$|\widetilde{\boldsymbol{p}} - \boldsymbol{w}|^2 - |\boldsymbol{p}^* - \boldsymbol{w}|^2 = |\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*|^2 + 2\mathfrak{Re}(\boldsymbol{p}^* - \boldsymbol{w}, \widetilde{\boldsymbol{p}} - \boldsymbol{p}^*).$$

Observing that

$$\widetilde{p}_j - p_j^* = \Delta\gamma + \int_0^\infty \frac{d\nu(t)}{t - z_j},$$

we see that

$$\mathfrak{Re}(\boldsymbol{p}^* - \boldsymbol{w}, \widetilde{\boldsymbol{p}} - \boldsymbol{p}^*) = (\Delta\gamma)\mathfrak{Re}\sum_{j=1}^n (p_j^* - w_j) + \int_0^\infty \mathfrak{Re}\sum_{j=1}^n \frac{p_j^* - w_j}{t - \overline{z_j}} d\nu(t).$$

The real rational function

$$C(t) = \mathfrak{Re}\sum_{j=1}^n \frac{p_j^* - w_j}{t - \overline{z_j}}, \qquad t \geq 0, \tag{2.36}$$

which we will call *the Caprini function*, will play an essential role in our algorithm for solving the least squares problem (1.5).

In terms of the Caprini function we obtain

$$|\widetilde{\boldsymbol{p}} - \boldsymbol{w}|^2 - |\boldsymbol{p}^* - \boldsymbol{w}|^2 = 2(\Delta\gamma)\lim_{t\to\infty} tC(t) + 2\int_0^\infty C(t)d\nu(t) + |\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*|^2. \tag{2.37}$$

This formula permits us to formulate and prove Caprini's necessary and sufficient conditions for the minimizer in (2.34). This is a particular version of Caprini's result [18], where the real and imaginary parts of each individual measurement could have a different weight in the least squares functional.

THEOREM **2.6.** *Suppose that the minimum in (1.5) is nonzero, then the unique minimizer $f_* \in \mathfrak{S}$ is given by*

$$f_*(z) = \gamma + \sum_{j=1}^{N} \frac{\sigma_j}{t_j - z} \tag{2.38}$$

for some $\sigma_j > 0$, $t_j \geq 0$ and $\gamma \geq 0$. Moreover, f_*, given by (2.38) is the minimizer in (1.5) if and only if its Caprini function $C(t)$ is nonnegative and vanishes at $t = t_j$, $j = 1, \ldots, N$, and "at infinity", in the sense that

$$\Re\mathfrak{e} \sum_{j=1}^{n} (p_j^* - w_j) = \lim_{t\to\infty} tC(t) = 0, \tag{2.39}$$

provided $\gamma > 0$.

Proof. If $\gamma > 0$, then we can consider the competitor (2.35) with $\widetilde{\sigma} = \sigma$. Formula (2.37) then implies that

$$2(\Delta\gamma) \lim_{t\to\infty} tC(t) + (\Delta\gamma)^2 > 0,$$

where $\Delta\gamma$ can be either positive or negative and can be chosen as small in absolute value as we want. This implies (2.39).

Next, suppose $t_0 \in [0, +\infty)$ is in the support of σ. For every $\epsilon > 0$ we define $I_\epsilon(t_0) = \{t \geq 0 : |t - t_0| < \epsilon\}$. Saying that t_0 is in the support of σ is equivalent to $m(t_0, \epsilon) = \sigma(I_\epsilon(t_0)) > 0$ for all $\epsilon > 0$. Then, there are two possibilities. Either

(i) $\lim_{\epsilon\to 0} m(t_0, \epsilon) = 0$, or

(ii) $\lim_{\epsilon\to 0} m(t_0, \epsilon) = \sigma_0 > 0$

In case (i) we construct a competitor measure

$$\sigma_\epsilon = \sigma - \sigma|_{I_\epsilon(t_0)} + \theta m(t_0, \epsilon)\delta_{t_0},$$

where $\theta > 0$ is an arbitrary constant. We then define

$$f_\epsilon(z) = \gamma + \int_0^\infty \frac{d\sigma_\epsilon(t)}{t - z}, \qquad p_j^\epsilon = f_\epsilon(z_j). \tag{2.40}$$

Formula (2.37) then implies

$$\lim_{\epsilon\to 0} \frac{|\boldsymbol{p}^\epsilon - \boldsymbol{w}|^2 - |\boldsymbol{p}^* - \boldsymbol{w}|^2}{m(t_0, \epsilon)} = 2(\theta - 1)C(t_0),$$

since $|\boldsymbol{p}^\epsilon - \boldsymbol{p}^*| \leq Cm(t_0, \epsilon)$, where C is independent of ϵ. If f_* is the minimizer, then we must have $(\theta - 1)C(t_0) \geq 0$ for all $\theta > 0$, which implies that $C(t_0) = 0$.

In case (ii) we have $\sigma(\{t_0\}) = \sigma_0 > 0$. Then, for every $|\epsilon| < \sigma_0$ we construct a competitor measure

$$\sigma_\epsilon = \sigma + \epsilon\delta_{t_0}, \tag{2.41}$$

as well as the corresponding f_ϵ and $\boldsymbol{p}^\epsilon$, given by (2.40). We then compute

$$\lim_{\epsilon\to 0}\frac{|\boldsymbol{p}^\epsilon-\boldsymbol{w}|^2-|\boldsymbol{p}^*-\boldsymbol{w}|^2}{\epsilon}=2C(t_0). \tag{2.42}$$

Since in this case ϵ can be both positive and negative we conclude that $C(t_0)=0$.

Hence, we have shown that $C(t_0)=0$ whenever $t_0\in[0,+\infty)$ is in the support of the spectral measure σ of the minimizer f_*. It remains to observe that for any $t\in\mathbb{R}$

$$C(t)=\sum_{j=1}^n\left\{\frac{p_j^*-w_j}{t-\overline{z_j}}+\frac{\overline{p_j^*}-\overline{w_j}}{t-z_j}\right\}.$$

Thus, $C(t)$ is a restriction to the real line of a rational function on the neighborhood of the real line in the complex t-plane. By assumption, $\boldsymbol{w}\notin V(\boldsymbol{z})$, and therefore $C(t)$ is not identically zero. In particular, $C(t)$ cannot have more than $2n-1$ zeros. We conclude that the support of the spectral measure of the minimizer f_* must be finite, and the minimizer must be a rational function.

Now let us consider the competitor (2.40) defined by (2.41), where $\epsilon>0$ and t_0 is not in the support of σ. Formula (2.42) then implies that

$$\lim_{\epsilon\to 0^+}\frac{|\boldsymbol{p}^\epsilon-\boldsymbol{w}|^2-|\boldsymbol{p}^*-\boldsymbol{w}|^2}{\epsilon}=2C(t_0)\geq 0.$$

This proves that $C(t)\geq 0$ for all $t\geq 0$. The necessity of the stated properties of the Caprini function $C(t)$ is now established.

Sufficiency is a direct consequence of formula (2.37). For any competitor measure $\widetilde{\sigma}$ we can write

$$\nu=\widetilde{\sigma}-\sigma=\sum_{j=1}^N(\Delta\sigma_j)\delta_{t_j}+\widetilde{\nu},$$

where $\widetilde{\nu}$ is a positive Radon measure without any point masses at $t=t_j$, $j=1,\dots,N$. It is obtained by eliminating point masses of $\widetilde{\sigma}$ at t_j, $j=1,\dots,N$, if it has any:

$$\widetilde{\nu}=\widetilde{\sigma}-\sum_{j=1}^N\widetilde{\sigma}(\{t_j\})\delta_{t_j}.$$

We then compute, via formula (2.37), taking into account that $C(t_j)=0$

$$|\widetilde{\boldsymbol{p}}-\boldsymbol{w}|^2-|\boldsymbol{p}^*-\boldsymbol{w}|^2=2(\Delta\gamma)\lim_{t\to\infty}tC(t)+2\int_0^\infty C(t)d\widetilde{\nu}(t)+|\widetilde{\boldsymbol{p}}-\boldsymbol{p}^*|^2\geq 0,$$

since $C(t)\geq 0$. If $\Delta\gamma<0$, then $\gamma=\widetilde{\gamma}-\Delta\gamma>0$, and therefore the first term on the right-hand side vanishes due to (2.39). □

We observe that that if $t_j > 0$, then we must also have $C'(t_j) = 0$, since $t = t_j$ is a point of local minimum of $C(t)$. If we write formula (2.38) in the form

$$f_*(z) = \gamma - \frac{\sigma_0}{z} + \sum_{j=1}^{N} \frac{\sigma_j}{t_j - z}, \qquad \gamma \geq 0,\ \sigma_0 \geq 0,\ t_j > 0,\ \sigma_j > 0,\ j = 1, \ldots, N,$$

then we have exactly $2(N+1)$ equations for $2(N+1)$ unknowns γ, σ_0, t_j, σ_j, $j = 1, \ldots, N$:

$$\gamma \lim_{t\to\infty} tC(t) = 0, \quad \sigma_0 C(0) = 0, \quad C(t_j) = 0, \quad C'(t_j) = 0, \quad j = 1, \ldots, N. \tag{2.43}$$

Obviously, these equations do not enforce the nonnegativity of $C(t)$ and may very well be satisfied when some t_j are points of local maxima and $C(t)$ is not nonnegative. Hence, the equations should not really be regarded as equations for the minimizer. Instead the intended use of Theorem 2.6 is to provide the certificate of optimality of a purported solution of (2.34) by exhibiting the graph of $C(t)$ that shows that the necessary and sufficient conditions of optimality are satisfied. In fact, equations (2.43) are used in our algorithm to make the final adjustments when a near-optimal solution is obtained.

2.5 Analytic structure of the boundary of $V(\boldsymbol{z})$

The analytic structure of the interpolation body $V(\boldsymbol{z})$ defined in (1.6) is well understood. The set $V(\boldsymbol{z})$ is a closed convex cone in $\mathbb{C}^n$ with nonempty interior $V^\circ(\boldsymbol{z})$, characterized by the inequalities $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w}) > 0$, $\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w}) > 0$ in the sense of quadratic forms. The set

$$\mathfrak{S}(\boldsymbol{z}, \boldsymbol{w}) = \{f \in \mathfrak{S} : f(\boldsymbol{z}) = \boldsymbol{w}\}$$

is parametrized by elements of $\mathfrak{S}$ via the recursive interpolation procedure described in Section 2.3. The function $f \in \mathfrak{S}(\boldsymbol{z}, \boldsymbol{w})$ corresponding to $0 \in \mathfrak{S}$ in such a parametrization has the form

$$f(z) = \sum_{j=1}^{n} \frac{\sigma_j}{t_j - z}, \quad \sigma_j > 0,\ 0 < t_1 < \cdots < t_n,$$

with the list of parameters σ_j and t_j above, in one-to-one correspondence with points $\boldsymbol{w} = f(\boldsymbol{z})$ in $V^\circ(\boldsymbol{z})$ [53, 52].

By contrast with $V^\circ(\boldsymbol{z})$, each point on $\partial V(\boldsymbol{z})$ can be realized as a list of values of a unique Stieltjes function, which must necessarily be rational. In view of Theorem 2.4 the boundary of $V(\boldsymbol{z})$ can be naturally written as a union of two overlapping sets

$$\partial V^N(\boldsymbol{z}) = \{\boldsymbol{w} \in V(\boldsymbol{z}) : \det \boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w}) = 0\},\ \partial V^P(\boldsymbol{z}) = \{\boldsymbol{w} \in V(\boldsymbol{z}) : \det \boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w}) = 0\}.$$

We can think of them as two sides of a clam shell that meet along the "rim"

$$\partial V^{NP}(\boldsymbol{z}) = \{\boldsymbol{w} \in V(\boldsymbol{z}) : \det \boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w}) = 0,\ \det \boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w}) = 0\}.$$

Each point $\boldsymbol{w} \in \partial V^N(\boldsymbol{z})$ is attained by a unique rational function $f \in \mathfrak{S}_n^N$, where

$$\mathfrak{S}_n^N = \left\{ \gamma + \sum_{k=1}^{n-1} \frac{\sigma_k}{t_k - z} : \gamma \geq 0,\ t_k \geq 0,\ \sigma_k \geq 0 \right\}. \tag{2.44}$$

Similarly, each point $\boldsymbol{w} \in \partial V^P(\boldsymbol{z})$ is attained by a unique rational function $f \in \mathfrak{S}_n^P$. Unfortunately, a simple representation, like (2.44) of functions in $\mathfrak{S}_n^P$ is not possible. This is because the parameter space $(\gamma, \boldsymbol{\sigma}, \boldsymbol{t})$ in (2.44) is noncompact, and it is an accident that the set $\mathfrak{S}_n^N$ happens to be closed (in the space of holomorphic functions on $\mathbb{C} \setminus \mathbb{R}_+$). The most concise, but somewhat indirect description of $\mathfrak{S}_n^P$ can be formulated using the "reflection" symmetry $\mathcal{R} : f \mapsto -1/(zf)$ of class $\mathfrak{S}$: $\mathfrak{S}_n^P = \mathcal{R}(\mathfrak{S}_n^N)$. Another description of $\mathfrak{S}_n^P$ is the closure of the set

$$\widetilde{\mathfrak{S}}_n^P = \left\{ -\frac{\sigma_0}{z} + \sum_{k=1}^{n-1} \frac{\sigma_k}{t_k - z} : \sigma_0 \geq 0,\ t_k \geq 0,\ \sigma_k \geq 0 \right\} \tag{2.45}$$

with respect to the uniform convergence on compact subsets of $\mathbb{C} \setminus \mathbb{R}_+$. Explicitly, the set $\mathfrak{S}_n^P$ can be described as $\mathfrak{S}_n^P = \widetilde{\mathfrak{S}}_n^P \cup \mathfrak{S}_{n-1}^N$.

Similarly, each point $\boldsymbol{w} \in \partial V^{NP}(\boldsymbol{z})$ is attained by a unique rational function $f \in \mathfrak{S}_n^{NP}$, where $\mathfrak{S}_n^{NP}$ can be described implicitly as the closure of

$$\widetilde{\mathfrak{S}}_n^{NP} = \left\{ \sum_{k=1}^{n-1} \frac{\sigma_k}{t_k - z} : t_k \geq 0,\ \sigma_k \geq 0 \right\}, \tag{2.46}$$

or explicitly, as $\mathfrak{S}_n^{NP} = \widetilde{\mathfrak{S}}_n^{NP} \cup \mathfrak{S}_{n-1}^N$.

If we define the evaluation operator $E_{\boldsymbol{z}} : \mathfrak{S} \to \mathbb{C}^n$ by $E_{\boldsymbol{z}} f = f(\boldsymbol{z})$, then we have both

$$V(\boldsymbol{z}) = E_{\boldsymbol{z}}(\mathfrak{S}) \text{ and } V(\boldsymbol{z}) = E_{\boldsymbol{z}}(\mathfrak{S}_{n+1}^{NP}).$$

Moreover, $E_{\boldsymbol{z}} : \mathfrak{S}_{n+1}^{NP} \to V(\boldsymbol{z})$ is a bijection. The statements above are all consequences of the following classical theorem [39, 40].

THEOREM **2.7.** *Suppose that $f \in \mathfrak{S}$ is a rational function. Then it can be written uniquely in the form*

$$f(z) = \gamma + \sum_{j=1}^{n} \frac{\sigma_j}{t_j - z}, \qquad \gamma \geq 0,\ \sigma_j > 0,\ 0 \leq t_1 < t_2 < \cdots < t_n, \tag{2.47}$$

where $n \geq 0$ is an integer. If $\gamma > 0$, then $f(z)$ has exactly n distinct real zeros $z = x_j$, $j = 1, \ldots, n$, satisfying the interlacing property

$$0 \leq t_1 < x_1 < t_2 < x_2 < \cdots < t_n < x_n < +\infty,$$

so that $f(z)$ can also be written as a product

$$f(z) = \gamma \prod_{j=1}^{n} \frac{z - x_j}{z - t_j}. \tag{2.48}$$

If $\gamma = 0$ and $n \geq 1$, then there are exactly $n - 1$ distinct real zeros $z = x_j$ and

$$f(z) = \frac{A}{t_n - z} \prod_{j=1}^{n-1} \frac{z - x_j}{z - t_j}, \quad A > 0,\ 0 \leq t_1 < x_1 < t_2 < \cdots < x_{n-1} < t_n < +\infty. \tag{2.49}$$

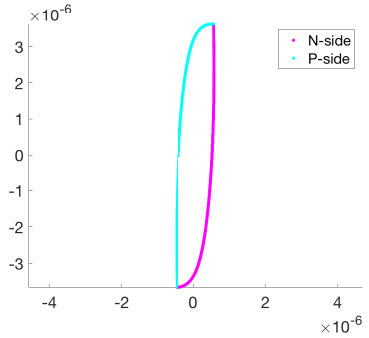

Figure 2: A random two-dimensional cross-section of $V(\boldsymbol{z})$. The origin corresponds to $\boldsymbol{w} \in V(\boldsymbol{z})$ and the cross-section is spanned by a random unit vector $\boldsymbol{d} \in \mathbb{C}^n$ and a normal $\boldsymbol{n}$ to $\partial V(\boldsymbol{z})$ at the point where $\boldsymbol{w} + t\boldsymbol{d}$ intersects $\partial V(\boldsymbol{z})$.

3 A needle in a haystack

In Section 2 we have summarized a substantial body of existing knowledge about the Stieltjes class $\mathfrak{S}$ and the closed convex cone $V(\boldsymbol{z})$. Can one harness this knowledge to devise an algorithm solving the least squares problem (2.34)? Surprisingly the answer is not apparent. What has been described so far is an interpolation algorithm for constructing functions $f(z)$, satisfying $f(z_j) = p_j^*$, once the solution $\boldsymbol{p}^*$ of (2.34) has been found. In this section we take a closer look at the geometry of $V(\boldsymbol{z})$. Here will show that in effect, the set $V(\boldsymbol{z}) \subset \mathbb{C}^n$ has a very small (real) dimension compared to $2n$. The proverbial needle analogy is apt here. Even though the needle is a three-dimensional body, we can approximate it well by an interval of a straight line. To illustrate our point we return to our simple example (2.13). Figure 2 shows a two-dimensional cross-section of $V(\boldsymbol{z})$, where the $\partial V^N(\boldsymbol{z})$ and $\partial V^P(\boldsymbol{z})$ parts of the boundary of $V(\boldsymbol{z})$ are shown in magenta and cyan and are on the left and the right sides of $V(\boldsymbol{z})$, respectively. The origin in the figure is placed at $\boldsymbol{w}$ in the interior of $V(\boldsymbol{z})$. When we added a 2% noise to $\boldsymbol{w}$, the noisy data $\widetilde{\boldsymbol{w}}$ would lie about 25,000 thicknesses of the cross-section away. If an ordinary sawing needle is the analogy for $V(\boldsymbol{z})$, the point $\widetilde{\boldsymbol{w}}$ would be about 25 meters away!

To see the dimensional degeneracy of $V(\boldsymbol{z})$ mathematically we recall that the rank-two displacement structure (2.11) and (2.12) of $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{w})$ and $\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{w})$, respectively, implies that their eigenvalues decay exponentially fast [8]. Hence, numerically, these matrices will always have eigenvalues which are indistinguishable from 0 up to the floating point precision, when $n > 15$. Thus, numerically, all points in $V(\boldsymbol{z})$ will appear to lie on its boundary.

The crucial point here is that the dimensional degeneracy of the geometry of $V(\boldsymbol{z})$ handily defeats typical minimization algorithms that start with some initial guess $\boldsymbol{p}_0 \in \partial V(\boldsymbol{z})$ and choose the direction in which we want to travel "along" $\partial V(\boldsymbol{z})$ in order to make the distance to $\boldsymbol{w} \notin V(\boldsymbol{z})$ smaller. Indeed, even if we are travelling along one of the "long dimensions" of the needle, a tiny generic perturbation of the direction of travel will cause us to exit $V(\boldsymbol{z})$ after an extremely short distance. For example, when $n \approx 20$ our numerical experiments showed that we needed to perform 10^{10} steps to make even a barely noticeable change in the distance of $|\boldsymbol{p} - \boldsymbol{w}|$.

Milton [55] suggested that since $V(\boldsymbol{z})$ is a convex cone which is dimensionally degenerate it must effectively lie in a low-dimensional subspace of $\mathbb{C}^n$, in the same way as the needle whose point is at the origin effectively lies in a one-dimensional subspace of $\mathbb{R}^3$. In order to capture this low-dimensional subspace (or rather its orthogonal complement) we look for vectors $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n) \in \mathbb{C}^n$, such that $|\boldsymbol{\xi}| = 1$ and $\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi})$ is negligibly small for all $\boldsymbol{w} \in V(\boldsymbol{z})$ with $|\boldsymbol{w}| = 1$. Let us explore this idea.

Suppose $\gamma \geq 0$ and σ is the Stieltjes spectral measure. For given nodes $z_j \in \mathbb{H}_+$ we define

$$w_j[\sigma, \gamma] = \gamma + \int_0^\infty \frac{d\sigma(t)}{t - z_j}.$$

We estimate

$$|w_j[\sigma, \gamma]| \leq \gamma + \|\sigma\| \max_{t \geq 0} \left| \frac{t+1}{t - z_j} \right| \leq M(z_j)(\gamma + \|\sigma\|), \tag{3.1}$$

where

$$\|\sigma\| = \int_0^\infty \frac{d\sigma(t)}{t+1}, \qquad M(z_j) = \max_{t \geq 0} \left| \frac{t+1}{t - z_j} \right|.$$

It is not hard to compute the constant $M(z)$ explicitly, when $\mathfrak{Im}(z) > 0$, using the theory of fractional-linear maps. We can also derive the reverse estimate from the formulas

$$\mathfrak{Im}(w_j) = \mathfrak{Im}(z_j) \int_0^\infty \frac{d\sigma(t)}{|t - z_j|^2},$$

and

$$\mathfrak{Re}(w_j) = \gamma + \int_0^\infty \frac{t - \mathfrak{Re}(z_j)}{|t - z_j|^2} d\sigma(t) = \gamma + \int_0^\infty \left| \frac{t+1}{t - z_j} \right|^2 \frac{d\sigma(t)}{t+1} - \frac{1 + \mathfrak{Re}(z_j)}{\mathfrak{Im}(z_j)} \mathfrak{Im}(w_j).$$

Denoting

$$m(z_j) = \min_{t \geq 0} \left| \frac{t+1}{t - z_j} \right| = \min \left\{ \frac{1}{|z_j|}, 1 \right\},$$

we obtain

$$m(z_j)(\gamma + \|\sigma\|) \leq \frac{\mathfrak{Im}(w_j z_j) + \mathfrak{Im}(w_j)}{\mathfrak{Im}(z_j)} \leq \frac{|z_j + 1|}{\mathfrak{Im}(z_j)} |w_j|. \tag{3.2}$$

Inequalities (3.1) and (3.2) imply that there exist constants $c(\boldsymbol{z})$ and $C(\boldsymbol{z})$, such that

$$c(\boldsymbol{z}) \|\boldsymbol{w}[\sigma, \gamma]\|_\infty \leq \gamma + \|\sigma\| \leq C(\boldsymbol{z}) \|\boldsymbol{w}[\sigma, \gamma]\|_\infty, \tag{3.3}$$

where

$$\|\boldsymbol{w}\|_\infty = \max_{1 \leq j \leq n} |w_j|, \quad c(\boldsymbol{z}) = \min_{1 \leq j \leq n} \frac{1}{M(z_j)}, \quad C(\boldsymbol{z}) = \min_{1 \leq j \leq n} \left\{ \frac{|z_j + 1|}{m(z_j) \mathfrak{Im}(z_j)} \right\}. \tag{3.4}$$

This means that $\gamma + \|\sigma\|$ and $\|\boldsymbol{w}\|_\infty$ are equivalent norms of $f \in \mathfrak{S}$, given by (2.2), provided $\boldsymbol{w} = f(\boldsymbol{z})$.

We recall that our goal is to understand how the convex set $V_1(\boldsymbol{z}) = V(\boldsymbol{z}) \cap B(\mathbf{0}, 1)$ would look geometrically as a subset of the $2n$-dimensional Euclidean space $\mathbb{C}^n$. We claim that this set, which is technically of full real dimension $2n$, is "very flat". To quantify just how flat it is we look for unit vectors $\boldsymbol{\xi} \in \mathbb{C}^n$, such that $\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi})$ is very small for all $\boldsymbol{w} \in V_1(\boldsymbol{z})$. We compute

$$\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi}) = \mathfrak{Re}\left(\gamma S + \int_0^\infty \sum_{k=1}^n \frac{\xi_k}{t - \overline{z_k}} d\sigma(t)\right), \quad S = \sum_{k=1}^n \xi_k.$$

Since it is the measure $d\sigma(t)/(1+t)$ that is finite it will be convenient to rewrite the above formula as follows:

$$\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi}) = \mathfrak{Re}\left(\gamma S + \int_0^\infty (\psi[\boldsymbol{\xi}](t) + S) \frac{d\sigma(t)}{t+1}\right),$$

where

$$\psi[\boldsymbol{\xi}](t) = \sum_{k=1}^n \frac{\xi_k(\overline{z_k} + 1)}{t - \overline{z_k}}.$$

Thus,

$$|\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi})| \le (\gamma + \|\sigma\|)|\mathfrak{Re}(S)| + \|\sigma\| \max_{t \ge 0} |\mathfrak{Re}(\psi[\boldsymbol{\xi}](t))|.$$

Since $\theta[\boldsymbol{\xi}](t) = \mathfrak{Re}(\psi[\boldsymbol{\xi}](t))$ is a complicated function of t whose maximum is impossible to compute directly we observe that both $\theta[\boldsymbol{\xi}]$ and $\theta'[\boldsymbol{\xi}]$ are in $L^2(0, +\infty)$ and use the inequality

$$\max_{t \ge 0} |\theta(t)|^2 \le \|\theta\|_{1,2}^2 = \|\theta\|_{L^2(0,+\infty)}^2 + \|\theta'\|_{L^2(0,+\infty)}^2,$$

valid for all $\theta \in W^{1,2}(0, +\infty)$. The inequality is sharp. It becomes equality when $\theta(t) = e^{-t}$. Hence,

$$|\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi})|^2 \le 2(\gamma + \|\sigma\|)^2 \left(\mathfrak{Re}(S)^2 + \|\theta[\boldsymbol{\xi}]\|_{1,2}^2\right) \le 2C(\boldsymbol{z})^2 \|\boldsymbol{w}\|_\infty^2 \boldsymbol{Q}(\boldsymbol{z})[\boldsymbol{\xi}], \tag{3.5}$$

where

$$\boldsymbol{Q}(\boldsymbol{z})[\boldsymbol{\xi}] = \mathfrak{Re}(S)^2 + \|\theta[\boldsymbol{\xi}]\|_{1,2}^2$$

is a positive definite real quadratic form in $\boldsymbol{\xi}$ and $C(\boldsymbol{z})$ is given in (3.4), in accordance with (3.2). Let $\lambda_1 > \lambda_2 > \cdots > \lambda_{2n} > 0$ be the eigenvalues of $\boldsymbol{Q}(\boldsymbol{z})$. For each $\delta_m = C(\boldsymbol{z})\sqrt{2\lambda_{m+1}}$ taken as the "negligibility threshold", we can regard m as the effective dimension of $V(\boldsymbol{z})$, since the $2n-m$-dimensional span $\mathcal{W}_m$ of all eigenvectors of $\boldsymbol{Q}(\boldsymbol{z})$ corresponding to eigenvalues λ_k, $k > m$ is effectively orthogonal to $V(\boldsymbol{z})$. Indeed, for any $\boldsymbol{\xi} \in \mathcal{W}_m$ and any $\boldsymbol{w} \in V_1(\boldsymbol{z})$ we have the inequality[4] $|\mathfrak{Re}(\boldsymbol{w}, \boldsymbol{\xi})| \le \delta_m$. For the example (2.13) the quadratic form is of full rank, its 40 eigenvalues decreasing from $\lambda_1 \approx 3.37 \cdot 10^8$ to $\lambda_{40} = 4.73 \cdot 10^{-5}$. If the number of data points increases to 40, $z_j = ie^{0.01+0.5j}$, $j = 0, 1, \ldots, 39$, then numerical rank of the 80×80 matrix $\boldsymbol{Q}(z)$ is 56. It also remains 56 for the 100×100 matrix $\boldsymbol{Q}(z)$, corresponding

[4]Obviously, the estimate holds in a larger convex subset $V(\boldsymbol{z}) \cap B_\infty(\mathbf{0}, 1)$ of $V(\boldsymbol{z})$, where B_∞ denotes a ball in $\|\cdot\|_\infty$ norm of $\mathbb{C}^n$.

to $z_j = ie^{0.01+0.4j},\ j = 0, 1, \ldots, 49$. These results show that the theoretical bound (3.5) is fairly conservative and overestimates the perceived dimension of $V(\boldsymbol{z})$ quite a bit.

The quadratic form $\boldsymbol{Q}(\boldsymbol{z})$ is not hard to compute explicitly using the residue formula

$$\int_0^\infty R(x)dx = -\sum_{r=1}^{N} \mathrm{Res}[R(z)\ln(-z), z = p_r], \tag{3.6}$$

where $R(z)$ is a rational function with at least $1/|z|^2$ decay at infinity and poles p_r none of which lie on $[0, +\infty)$. Even with the exact formula for $\boldsymbol{Q}(\boldsymbol{z})$, the accurate computation of its eigenvalues requires many more digits of precision than the floating point allows even for $n = 20$. In our examples we have used the Advanpix Multiprecision Computing Toolbox for MATLAB (https://www.advanpix.com) using 200 digits of precision.

4 The least squares algorithm

In this section we describe the algorithm that solves the least squares problem (2.34), displays the graph of the Caprini function $C(t)$ certifying that the minimum in (2.34) has indeed been reached (see Theorem 2.6), and exhibits the "uncertainty band" where the least squares minimizer might belong for different realizations of the random noise in the data.

The first step in the algorithm is to replace $V(\boldsymbol{z})$ by a much simpler object: the positive span of an *ad-hoc basis* of $V(\boldsymbol{z})$.

Definition 4.1. *An ad-hoc basis of $V(\boldsymbol{z})$ is a finite set of positive spectral measures $\mathfrak{B} = \{\sigma_1, \ldots, \sigma_N\}$, whereby $V(\boldsymbol{z})$ is replaced by*

$$V_{\mathfrak{B}}(\boldsymbol{z}) = \left\{ \boldsymbol{w} \in \mathbb{C}^n : w_j = x_0 + \sum_{\alpha=1}^{N} x_\alpha \phi_\alpha(z_j),\ j = 1, \ldots, n,\ x_\alpha \geq 0,\ \alpha = 0, \ldots, N \right\}, \tag{4.1}$$

where

$$\phi_\alpha(z) = \int_0^\infty \frac{d\sigma_\alpha(t)}{t - z},\ \alpha = 1, \ldots, N.$$

The adjective "ad-hoc" indicates that our choice of the basis $\mathfrak{B}$ is nothing more than an educated guess, and other choices could be at least as effective as our choice. The choice that appears to work well consists of

- measures $\delta_\tau(t)$—unit point mass at $t = \tau$—where τ is either the real or the imaginary part of one of z_j for some j,
- measures $\chi_{[s_1,s_2]}(t)dt$, where both s_1 and s_2 are either one of the τs or a midpoint between adjacent τs.

We will denote this construction of an ad-hoc basis by $\mathfrak{B}(\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ stands for a list of τ's used in the above construction.

Imagining $V(\boldsymbol{z})$ as a needle explains why the choice of an ad-hoc basis can be fairly arbitrary. Indeed, selecting a point $\boldsymbol{w}_0$ at random inside a needle and replacing the needle

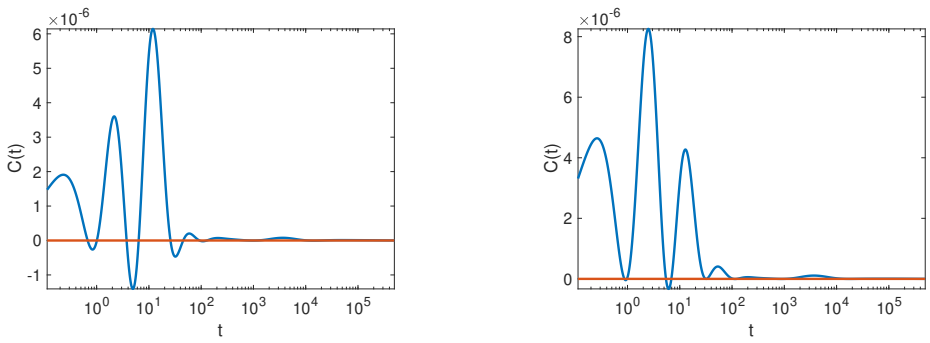

Figure 3: The Caprini function for the ad-hoc (left) and for the Caprini-augmented (right) bases projections.

with the ray $\{s\boldsymbol{w}_0 : s \geq 0\}$ gives a fairly accurate representation of the needle. The more accurately we want to approximate $V(\boldsymbol{z})$ the more important the choice of an ad-hoc basis becomes. Our choice above is just an attempt to tie the ad-hoc basis to the data in a somewhat natural and algorithmic fashion. Many existing algorithms (e.g., [12, 65]) make an effort to choose a better basis, but in the absence of any rigorous approximation error analysis they also remain largely ad-hoc. In the new algorithm the ad-hoc basis is only needed as a stepping stone for the construction of a much better basis tailor-made for the specific experimental data.

Once the above ad-hoc basis has been chosen, we compute

$$p_j(\boldsymbol{x}) = x_0 + \sum_{\alpha=1}^{N} x_\alpha \phi_\alpha(z_j), \quad j = 1, \ldots, n,$$

by solving the nonnegative least squares problem

$$\min_{\boldsymbol{x} \geq 0} |\boldsymbol{p}(\boldsymbol{x}) - \boldsymbol{w}|^2. \tag{4.2}$$

The above least squares problem is solved by a well established and widely implemented nonnegative least squares algorithm [45].

Naturally, we would like to know how good our ad-hoc approximation is. For illustration we once again turn to our simple example (2.13). We use the same noisy version $\widetilde{\boldsymbol{w}}$ of $\boldsymbol{w}$ as in the example of Figure 2. The optimality conditions described in Theorem 2.6 require the Caprini function $C(t)$ to be nonnegative and equal to zero on the support of the spectral measure. The graph of $C(t)$ shown in the left panel of Figure 3 suggests that we are not too far away from the true minimum but not there yet. Had we hit the minimum exactly, the local minima of $C(t)$ would also be both the global minima and the zeros of $C(t)$ and would comprise the support of the optimal spectral measure $\sigma(t)$. This observation leads to the next step in our algorithm: we add the points of local minima of $C(t)$ to the list of τ's in our ad-hoc basis $\mathfrak{B}(\boldsymbol{\tau})$ and recompute $\boldsymbol{p}(\boldsymbol{x})$, solving (4.2) using the augmented ad-hoc basis $\mathfrak{B}(\boldsymbol{\tau}_{\text{aug}})$ for $V(\boldsymbol{z})$. The Caprini function for the new approximation is shown in the right panel of Figure 3. We see both the substantial improvement and the fact that the

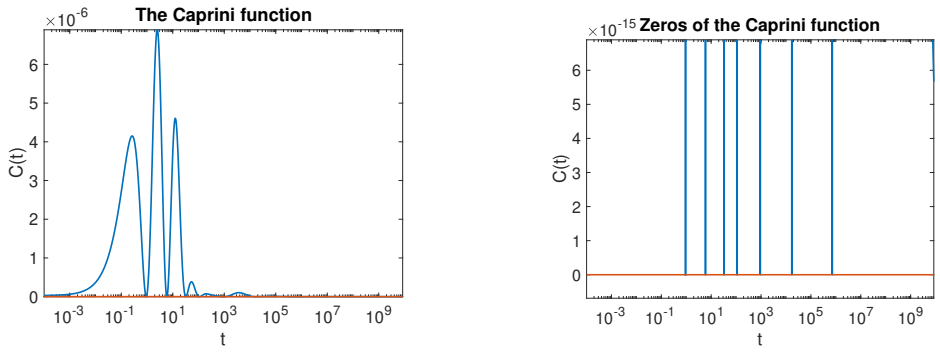

Figure 4: Achieving optimality for the "alternative data".

new approximation $\boldsymbol{p}^*$ is still not the true minimum in (2.34). We can repeat this step by adjoining the local minima of the improved Caprini function in the right panel of Figure 3 to the list of τ's. The improvement after the second application of the augmentation of the ad hoc basis is significantly smaller, and more repetitions no longer lead to discernible improvements.

To achieve certifiable optimality we cheat by "moving the goalposts". In the author's experience the Caprini function is very sensitive to even the tiniest deviations from the true optimum. The idea is to exploit this sensitivity and achieve optimality by means of making negligible changes, but not in $\boldsymbol{p}^*$, which is required to be in $V(\boldsymbol{z})$. Changing $\boldsymbol{w}$ instead of $\boldsymbol{p}^*$ leads to a *linear problem*! We therefore look for the *alternative data* $\widetilde{\boldsymbol{w}}$ near $\boldsymbol{w}$, so that the same $\boldsymbol{p}^*$ is a true minimizer in (2.34), where $\boldsymbol{w}$ is replaced by $\widetilde{\boldsymbol{w}}$ and where $\widetilde{\boldsymbol{w}}$ is computed by requiring that the local minima t_j of the original $C(t)$ satisfy equations (2.43). In other words we are looking for the vector $d\boldsymbol{w} \in \mathbb{C}^n$ of smallest norm, satisfying the following equations:

$$\begin{cases} \mathfrak{Re} \sum_{j=1}^{n} \dfrac{dw_j}{(t_k - \overline{z_j})^2} = 0, \\ \mathfrak{Re} \sum_{j=1}^{n} \dfrac{dw_j}{t_k - \overline{z_j}} = C(t_k), \end{cases} \qquad k = 1, \ldots, N. \tag{4.3}$$

If we want to enforce the $\gamma > 0$ condition, we need to add the equation

$$\mathfrak{Re} \sum_{j=1}^{n} dw_j = \mathfrak{Re} \sum_{j=1}^{n} (p_j - w_j). \tag{4.4}$$

If $C(0) < 0$ for the original data, we add $t = 0$ to the support of the spectral measure σ and require

$$\mathfrak{Re} \sum_{j=1}^{n} \frac{dw_j}{\overline{z_j}} = -C(0). \tag{4.5}$$

Vector $d\boldsymbol{w}$ can then be computed using the standard least norm least squares solver.

Our simulations show that the "alternative data" $\widetilde{\boldsymbol{w}} = \boldsymbol{w} + d\boldsymbol{w}$ is indeed sufficiently close to the actual data to justify replacing one with the other. In other words, if we regard $\boldsymbol{w}$ to

be equal to $\boldsymbol{p}^*$ plus random measurement errors, then $\widetilde{\boldsymbol{w}}$ is also equal to $\boldsymbol{p}^*$ plus a different realization of random measurement errors. At the same time the Caprini function for the alternative data $\widetilde{\boldsymbol{w}}$ in Figure 4 shows that our formerly imperfect solution $\boldsymbol{p}^*$ of (2.34) is now optimal to within computer precision[5], while $|\boldsymbol{w} - \widetilde{\boldsymbol{w}}|/|\boldsymbol{w}| \approx 6.5 \cdot 10^{-4}$, where $\boldsymbol{w}$ is given by (2.13) plus 2% noise.

On rare occasions during the algorithm testing the change from $\boldsymbol{w}$ to $\widetilde{\boldsymbol{w}}$ caused a point of local minimum $t = t_j$ of the original $C(t)$ to become a point of local maximum of the modified $C(t)$, while creating two new points of local minima to the right and to the left of t_j. If the new local minima are nonnegligibly negative, then we update the list of local minima of $C(t)$ and apply the same "alternative data" procedure to $\widetilde{\boldsymbol{w}}$, solving (4.3)–(4.5) again. In our numerical tests no more than two iterations of "data-fixing" was ever necessary to bring the graph of $C(t)$ into the desired shape.

In order to capture all local minima of $C(t)$ on $[0, +\infty)$ we observe that $C(t)$ will be a monotone function on $[T, +\infty)$ for sufficiently large T. Let us estimate the value of T. We will assume that $\gamma > 0$ and therefore

$$\Re\mathfrak{e} \sum_{j=1}^{n} \delta_j = 0, \qquad \delta_j = p_j - w_j.$$

In this case we can write $C'(t) = D_\infty(t) + O(t^{-4})$, as $t \to \infty$, where

$$D_\infty(t) = -\frac{2}{t^3} \Re\mathfrak{e} \sum_{j=1}^{n} \delta_j \overline{z_j}.$$

Estimating $|C'(t) - D_\infty(t)|$, it is not hard to show that

$$|C'(t) - D_\infty(t)| < |D_\infty(t)|, \quad \forall t > T = (M_0 + 1) \max_{1 \le j \le n} |z_j|, \tag{4.6}$$

where

$$M_0 = \frac{2 \sum_{j=1}^{n} |\delta_j||z_j|}{|\Re\mathfrak{e} \sum_{j=1}^{n} \delta_j \overline{z_j}|}.$$

Inequality (4.6) shows that $C'(t)$ cannot be 0 when $t > T$. Hence, if we want to make sure that we missed no local minima of $C(t)$, we need to examine it only on the finite interval $[0, T]$.

In order to construct the function $f \in \mathfrak{S}$ satisfying $f(\boldsymbol{z}) = \boldsymbol{p}^*$ we run the recursive interpolation algorithm described in Section 2.3. In practice, even though matrices $\boldsymbol{N}(\boldsymbol{z}, \boldsymbol{p}^*)$ and $\boldsymbol{P}(\boldsymbol{z}, \boldsymbol{p}^*)$ have no numerically significant negative eigenvalues, feasibility gets lost after a number of iterations due to the amplification of round-off errors. This may happen even when n is as small as 10. When this occurs, we replace the currently infeasible data $\boldsymbol{w}$ by its "projection" $\boldsymbol{p}^*$ as described above and continue the recursion using the projected feasible data.

Finally, our algorithm tries to estimate the degree of uncertainty of the output. If we regard the discrepancies $w_j - f_*(z_j)$ as a random noise, then the fact that the measured

[5]The right graph's vertical scale in Figure 4 is 10^{-9} times the left graph's vertical scale.

values w_j are exactly what they are is in part an outcome of a random event. Simulating normal random noise with variance

$$\rho^2 = \frac{1}{2n-1}\sum_{j=1}^{n} |w_j - f_*(z_j)|^2$$

we produce other "realizations" of the error of measurement, each of which leads to its own least squares solution $f_*(z)$. Plotting these functions for 500 different realizations of the random noise gives one an idea of the degree to which we can trust the output of the algorithm. These potential realizations are shown in grey in Figures 5 and 6. While in [35, 36] we estimated the *worst case* error of extrapolation, these Monte Carlo simulations are a simple and direct way to estimate the uncertainty for *specific data*. The use of Monte Carlo simulations to exhibit the uncertainty in the analytic continuation due to the statistical errors in the data has also been used in particle physics [4].

5 Direct computation of spectral measure

While the interpolation algorithm computes values $f(\zeta)$ for any specified list of points ζ in the upper half-plane, one would also want to have an explicit formula for $f(z)$. The goal of this section is to describe an algorithm for computing the spectral representation (2.47) of the function $f \in \mathfrak{S}$ satisfying $f(\boldsymbol{z}) = \boldsymbol{p}$. The algorithm computes this representation recursively following the algorithm described in Section 2.3. It is based on the following theorem

THEOREM **5.1.** *Suppose*

$$g(z) = \gamma_g - \frac{\sigma_0}{z} + \sum_{j=1}^{n} \frac{\sigma_j}{t_j - z}, \qquad \gamma_g \geq 0,\ \sigma_0 \geq 0,\ \sigma_j > 0,\ 0 < t_1 < t_2 < \cdots < t_n,$$

Suppose $f(z)$ is given by (2.29). Then

$$f(z) = \gamma_f - \frac{\nu_0}{z} + \sum_{j=1}^{n+1} \frac{\nu_j}{\tau_j - z},$$

where

$$\gamma_f = \frac{\gamma_* \gamma_g}{\gamma_g + 1}, \qquad \nu_0 = \frac{\sigma_0 \sigma^*}{\sigma_0 + t_*}, \tag{5.1}$$

and

$$0 < \tau_1 < t_1 < \tau_2 < t_2 < \cdots < t_n < \tau_{n+1}.$$

Proof. Formulas (5.1) are obtained by taking limits of $f(z)$ as $z \to \infty$ and $zf(z)$ as $z \to 0$ using formula (2.29). We have also proved in Theorem 2.5 that the degree of $f(z)$ is exactly 1 higher than $g(z)$. Thus, proving that the intervals $(0, t_1), (t_1, t_2), \ldots, (t_n, +\infty)$ contain at least one pole of $f(z)$ would imply that these intervals must contain exactly one pole.

Formula (2.29) shows that the poles of $f(z)$ can only come either from the poles of $g(z)$ or from the zeros of the denominator

$$\phi(z) = zg(z) + z - t_*.$$

It is easy to compute that

$$\lim_{z \to t_j} f(z) = \frac{\gamma_* t_j - \sigma^*}{t_j} \neq \infty.$$

Hence, only the zeros of $\phi(z)$ can be the positive poles of $f(z)$. The existence of zeros τ_j in the indicated intervals follows from the following observations

$$\lim_{x \to 0^+} \phi(x) = -\sigma_0 - t_* < 0, \quad \lim_{x \to t_j^{\pm}} \phi(x) = \mp\infty, \quad \lim_{x \to +\infty} \phi(x) = +\infty.$$

□

Once the intervals containing single zeros of $\phi(x)$ are isolated, the zeros can be computed using the standard zero finding algorithm [14, 29]. We only need to derive the upper bound for the last pole τ_{n+1}. We observe that

$$\phi(x) = (\gamma_g + 1)x - t_* - \sigma_0 + \sum_{j=1}^{n} R_j(x)$$

is monotone increasing on $(t_n, +\infty)$, since all functions $R_j(x) = x\sigma_j/(t_j - x)$, $j = 1, \ldots, n$, are monotone increasing on $(t_n, +\infty)$. When $x > t_n$ we have

$$R_j(x) \geq -\frac{\sigma_j x}{x - t_n}, \quad j = 1, \ldots, n.$$

Therefore, $\phi(x) > 0$ when $x > T_{\max}$, where $T_{\max} > t_n$ is the larger root of the quadratic equation

$$(\gamma_g + 1)T^2 - \left(\sum_{j=0}^{n} \sigma_j + t_* + (\gamma_g + 1)t_n \right) T + t_n(\sigma_0 + t_*) = 0.$$

The spectral representation of $f(z)$ is then computed recursively, using (2.29), with the explicit formula in the case when $g(z) = \gamma_g - \sigma_0/z$:

$$f(z) = \frac{\gamma_* \gamma_g}{\gamma_g + 1} - \frac{\sigma_0 \sigma^*}{(\sigma_0 + t_*)z} + \frac{\nu_1}{\tau_1 - z}, \tag{5.2}$$

where

$$\tau_1 = \frac{\sigma_0 + t_*}{\gamma_g + 1}, \quad \nu_1 = \frac{\sigma^* \gamma_g + \sigma_* + \gamma_* \sigma_0}{\gamma_g + 1} - \frac{\gamma_* \gamma_g (\sigma_0 + t_*)}{(\gamma_g + 1)^2} - \frac{\sigma_0 \sigma^*}{\sigma_0 + t_*}.$$

In our numerical simulations the values of $f_*(z)$ at specified points computed from the spectral representation of $f_*(z)$ are indistinguishable (graphically) from the values computed using the recursion algorithm from Section 2.3.

6 Case study: Electrochemical impedance spectroscopy

Electrochemistry studies the electrical behavior of systems where the motion of charges occurs not only due to the applied electric field but also due to chemical reactions that occur on sometimes vastly different time scales. One of the key characteristics of such systems is the electrochemical impedance spectrum $Z(\omega)$ that has the meaning of resistance to an applied sinusoidal current. Combining the sine and cosine functions into a complex exponential the steady response of such a system to the current $I(t) = e^{i\omega t}$ is the voltage $U(t) = R(\omega)e^{i(\omega t+\phi(\omega))}$. The resistance $R(\omega)$ and the phase shift $\phi(\omega)$ are combined into a single complex valued function $Z(\omega) = R(\omega)e^{i\phi(\omega)}$—the EIS. The theory of electrochemical cells, including batteries, electrodes and electrolytes [6, Sec. 2.1.2.3] says that $Z(\omega)$ has the spectral representation

$$Z(\omega) = \frac{1}{iC_0\omega} + \int_0^\infty \frac{d\sigma(\tau)}{1+i\omega\tau}, \quad \int_0^\infty \frac{d\sigma(\tau)}{1+\tau} < +\infty, \quad 0 < C_0 \le \infty, \tag{6.1}$$

where σ is a positive Borel-regular measure on $[0, +\infty)$, called the distribution of relaxation times (DRT). This formula shows that if $Z(\omega)$ is the EIS, then $Z(\omega) = f(-i\omega)$ for some $f \in \mathfrak{S}$. It is also a continuum version of the complex impedance of an electrical circuit made of a series of Voigt elements, each being a resistor and a capacitor connected in parallel.

Definition 6.1. *A Voigt circuit is an electrical circuit made of finitely many resistors and capacitors.*

The following theorem has long been known [30, 20, 21] (see also [22, Statement 2, p. 196, Vol. 1]).

THEOREM **6.2.** *The complex impedance functions $Z(\omega)$ of Voigt circuits are in one-to-one correspondence with rational Stieltjes functions $f \in \mathfrak{S}_\mathcal{R}$ via $Z(\omega) = f(-i\omega)$.*

In electrochemistry there are several explicit EIS functions representing important electrochemical cells, each serving as a building block of more complex devices. The ideal capacitor's EIS $Z(\omega) = 1/(iC\omega)$ is often replaced by the more realistic constant phase element (CPE) with $Z_{\mathrm{CPE}}(\omega) = R/(i\tau\omega)^\phi$, $\phi \in [0, 1]$. Connecting it in parallel with a resistor gives the ZARC or Cole-Cole element

$$Z_{\mathrm{ZARC}}(\omega) = \frac{R}{1 + (i\tau\omega)^\phi}, \quad \phi \in [0, 1].$$

A generalization of the ZARC element is the Havriliak-Negami element

$$Z_{\mathrm{HN}}(\omega) = \frac{R}{(1 + (i\tau\omega)^\phi)^\psi}, \quad \phi \in [0, 1],\ \psi \in [0, 1].$$

Following examples in [65], we test our algorithm on a double Havriliak-Negami element

$$Z_{\mathrm{DHN}}(\omega) = R_\infty + \frac{R_0}{(1 + (i\tau_1\omega)^\phi)^\psi} + \frac{R_0}{(1 + (i\tau_2\omega)^\phi)^\psi}, \tag{6.2}$$

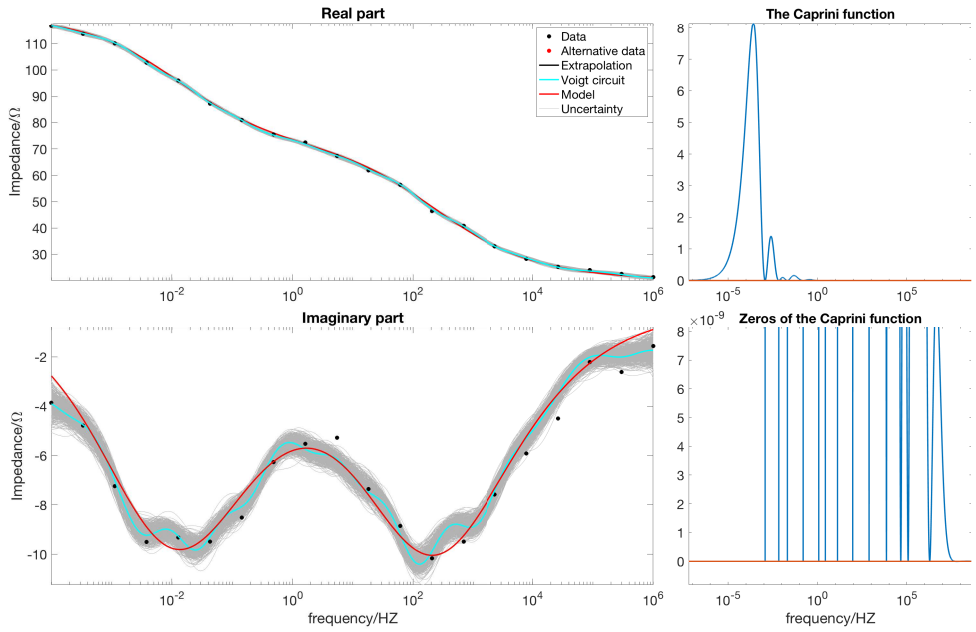

Figure 5: The output of the algorithm for a Voigt circuit.

where we chose $R_\infty = 20,\ R_0 = 50,\ \phi = 0.5,\ \psi = 0.8,\ \tau_1 = 20,\ \tau_2 = 0.001$. This element operates on two very different times scales (20 seconds and 1 millisecond) differing by four orders of magnitude.

The "experimental data" was produced by computing $Z_{\mathrm{DHN}}(2\pi f)$ at 20 frequencies f_j equispaced on the logarithmic scale from $f_{\min} = 10^{-4}$Hz to $f_{\max} = 10^{6}$Hz and then polluting the exact values with 1% random noise on the relative scale. Figure 5 shows the result of the implementation of the algorithm. The real and imaginary parts of the exact EIS function (6.2) are shown in red. The imaginary part has exactly two local minima at $1/(2\pi\tau_1)$ and $1/(2\pi\tau_2)$. Since the random noise is complex-valued and $\mathfrak{Im}(Z_{\mathrm{DHN}})$ is 10 times smaller than $\mathfrak{Re}(Z_{\mathrm{DHN}})$, the relative size of the noise for the imaginary part is actually 10%. This is why the algorithm's performance seems to be better for the real part than for the imaginary part. While absolute errors of reconstruction for both the real and the imaginary parts are the same, the relative errors differ by a factor of 10.

There is no discernible difference between the actual and the "alternative data" for which the plots of the Caprini function at the global and local scales show certified optimality. The grey band indicates the uncertainty of the extrapolation shown by the cyan curve. The cyan curve is a plot of a rational function whose spectral measure is supported on 20 points. It coincides to computer precision with values computed by the recursion algorithm of Section 2.3. It is important to keep in mind that the results in Figure 5 look nice because we are "filling the gaps" between measurements. The situation changes if we try to extrapolate beyond the largest or the smallest frequency at which the impedance function has been measured. Figure 6 illustrates what happens with exactly the same "experimental data" when we ask the algorithm to reconstruct the EIS function on a larger frequency band. The uncertainty of reconstruction "explodes", but our two methods of extrapolation: the

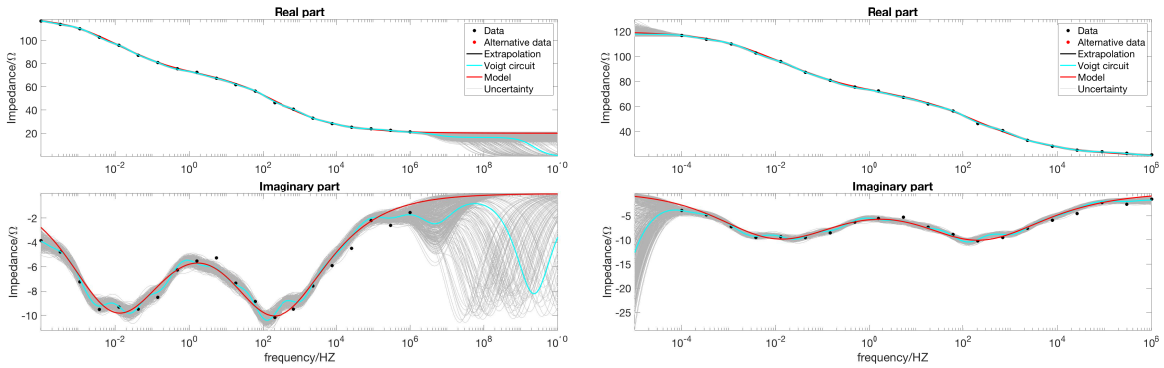

Figure 6: Extrapolation beyond the experimentally accessible frequency band.

recursive and spectral representations continue to agree. Both panels in Figure 6 show a pronounced disagreement between the theoretical and the extrapolated curves away from the experimentally accessible frequency band, confirming that it is in general impossible to extrapolate to the entire frequency spectrum reliably.

Acknowledgments. The author is grateful to Graeme Milton, Mihai Putinar, and Vladimir Bolotnikov for their comments and suggestions. A special thanks goes to the referee whose detailed report improved the paper significantly. This material is based upon work supported by the National Science Foundation under Grant No. DMS-2005538.

References

[1] P. Agarwal, M. E. Orazem, and L. H. Garcia-Rubio, *Application of measurement models to impedance spectroscopy: III. Evaluation of consistency with the Kramers-Kronig relations*, Journal of the Electrochemical Society, 142 (1995), p. 4159.

[2] N. I. Akhiezer, *Classical moment problem and some related questions in analysis*, SIAM, Philadelphia, 2020, `https://doi.org/10.1137/1.9781611976397`.

[3] B. Ananthanarayan, I. Caprini, and D. Das, *Test of analyticity and unitarity for the pion form-factor data around the ρ resonance*, Physical Review D, 102 (2020), p. 096003.

[4] B. Ananthanarayan, I. Caprini, D. Das, and I. S. Imsong, *Precise determination of the low-energy hadronic contribution to the muon $g-2$ from analyticity and unitarity: An improved analysis*, Physical Review D, 93 (2016), p. 116007.

[5] A. Bard and L. Faulkner, *Electrochemical Methods; Fundamentals and Applications*, Wiley Interscience, Hoboken, NJ, 2nd ed., 2000.

[6] E. Barsoukov and J. R. Macdonald, eds., *Impedance spectroscopy: theory, experiment, and applications*, John Wiley & Sons Inc., 2nd ed., 2005.

[7] D. Batenkov, L. Demanet, and H. N. Mhaskar, *Stable soft extrapolation of entire functions*, Inverse Problems, 35 (2019), p. 015011.

[8] B. Beckermann and A. Townsend, *On the Singular Values of Matrices with Displacement Structure*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 1227–1248.

[9] D. J. Bergman, *The dielectric constant of a composite material — A problem in classical physics*, Phys. Rep., 43 (1978), pp. 377–407.

[10] V. Bolotnikov and L. Sakhnovich, *On an operator approach to interpolation problems for Stieltjes functions*, Integral Equations and Operator Theory, 35 (1999), pp. 423–470.

[11] B. A. Boukamp, *A linear Kronig-Kramers transform test for immittance data validation*, Journal of the electrochemical society, 142 (1995), p. 1885.

[12] B. A. Boukamp, *Fourier transform distribution function of relaxation times; application and limitations*, Electrochimica acta, 154 (2015), pp. 35–46.

[13] B. A. Boukamp, *Distribution (function) of relaxation times, successor to complex nonlinear least squares analysis of electrochemical impedance spectroscopy?*, Journal of Physics: Energy, 2 (2020), p. 042001.

[14] R. P. Brent, *Algorithms for Minimization Without Derivatives*, Prentice-Hall, 1973.

[15] O. Brune, *Synthesis of a finite two-terminal network whose driving-point impedance is a prescribed function of frequency*, Journal of Mathematics and Physics, 10 (1931), pp. 191–236.

[16] I. Caprini, *On the best representation of scattering data by analytic functions in L_2-norm with positivity constraints*, Nuovo Cimento A (11), 21 (1974), pp. 236–248.

[17] I. Caprini, *Integral equations for the analytic extrapolation of scattering amplitudes with positivity constraints*, Nuovo Cimento A (11), 49 (1979), pp. 307–325.

[18] I. Caprini, *General method of using positivity in analytic continuations*, Rev. Roumaine Phys., 25 (1980), pp. 731–740.

[19] I. Caprini, *Constraints on physical amplitudes derived from a modified analytic interpolation problem*, J. Phys. A, 14 (1981), pp. 1271–1279.

[20] W. Cauer, *Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit*, Archiv für Elektrotechnik, 17 (1926), pp. 355–388.

[21] W. Cauer, *Über eine Klasse von Funktionen, die die Stieltjesschen Kettenbrüche als Sonderfall enthält.*, Jahresbericht der Deutschen Mathematiker-Vereinigung, 38 (1929), pp. 63–72.

[22] W. Cauer, *Synthesis of Linear Communication Networks*, vol. I and II, 2nd Ed., McGraw-Hill, 1958.

[23] E. Cherkaeva and K. M. Golden, *Inverse bounds for microstructural parameters of composite media derived from complex permittivity measurements*, Waves Random Media, 8 (1998), pp. 437–450.

[24] S. Ciulli, *A stable and convergent extrapolation procedure for the scattering amplitude.—I*, Il Nuovo Cimento A (1965-1970), 61 (1969), pp. 787–816.

[25] L. Demanet and A. Townsend, *Stable extrapolation of analytic functions*, Foundations of Computational Mathematics, 19 (2018), pp. 297–331.

[26] A. Dienstfrey and L. Greengard, *Analytic continuation, singular-value expansions, and Kramers-Kronig analysis*, Inverse Problems, 17 (2001), p. 1307.

[27] Y. M. Dyukarev and V. Katsnelson, *Multiplicative and additive classes of Stieltjes analytic matrix-valued functions and interpolation problems associated with them.*, Transactions of the American Mathematical Society, 131 (1986), pp. 55–70.

[28] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics. Vol. 2: Mainly electromagnetism and matter*, Addison-Wesley Publishing Co., Inc., Reading, Mass.-London, 1964.

[29] G. E. Forsythe, M. A. Malcolm, and C. B. Moler, *Computer methods for mathematical computations.*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1977.

[30] R. M. Foster, *Theorems regarding the driving-point impedance of two-mesh circuits*, The Bell System Technical Journal, 3 (1924), pp. 651–685.

[31] K. M. Golden, *Bounds on the complex permittivity of sea ice*, J. Geophys. Res. (Oceans), 100 (1995), pp. 699–711.

[32] K. M. Golden and G. Papanicolaou, *Bounds for effective parameters of heterogeneous media by analytic continuation*, Comm. Math. Phys., 90 (1983), pp. 473–491.

[33] K. M. Golden and G. Papanicolaou, *Bounds for effective parameters of multicomponent media by analytic continuation*, J. Statist. Phys., 40 (1985), pp. 655–667.

[34] Y. Grabovsky, *Fortran implementation of the Stieltjes function reconstruction algorithm.* `https://github.com/YuryGrabovsky/Stieltjes`, February 2021.

[35] Y. Grabovsky and N. Hovsepyan, *Explicit power laws in analytic continuation problems via reproducing kernel Hilbert spaces*, Inverse Problems, 36 (2020), p. 035001.

[36] Y. Grabovsky and N. Hovsepyan, *Optimal error estimates for analytic continuation in the upper half-plane*, Comm Pure Appl Math, (2021). to appear.

[37] J. Hamilton, P. Menotti, T. Spearman, and W. Woolcock, *Evidence for pion-pion interactions froms-wave pion-nucleon scattering*, Il Nuovo Cimento (1955-1965), 20 (1961), pp. 519–528.

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

[39] I. S. Kac and M. G. Krein, *R-functions–analytic functions mapping the upper half-plane into itself*, Amer. Math. Soc. Transl.(2), 103 (1974), p. 18.

[40] I. V. Kac and M. G. Krein, *On the spectral functions of the string*, vol. 103 of Translations, Amer Mathematical Society, 1974.

[41] N. N. Khuri, *Analyticity of the Schrödinger scattering amplitude and nonrelativistic dispersion relations*, Physical Review, 107 (1957), p. 1148.

[42] M. Krein and A. Nudelman, *An interpolation problem in the class of Stieltjes functions and its connection with other problems*, Integral Equations and Operator Theory, 30 (1998), pp. 251–278.

[43] M. G. Krein and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*, Translation of Mathematical Monographs, 50, American Mathematical Society, Providence, RI, 1977.

[44] L. D. Landau and E. M. Lifshitz, *Electrodynamics of continuous media*, vol. 8, Pergamon, New York, 1960. Translated from the Russian by J. B. Sykes and J. S. Bell.

[45] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, vol. 15 of Classics in Applied Mathematics, SIAM, 1995.

[46] R. Lipton, *Optimal inequalities for gradients of solutions of elliptic equations occurring in two-phase heat conductors*, SIAM Journal on Mathematical Analysis, 32 (2001), pp. 1081–1093.

[47] V. Lucarini, J. J. Saarinen, K.-E. Peiponen, and E. M. Vartiainen, *Kramers-Kronig relations in optical materials research*, vol. 110, Springer Science & Business Media, 2005.

[48] S. W. MacDowell, *Analytic properties of partial amplitudes in meson-nucleon scattering*, Phys. Rev., 116 (1959), pp. 774–778.

[49] J. V. Mantese, A. L. Micheli, D. F. Dungan, R. G. Geyer, J. Baker-Jarvis, and J. Grosvenor, *Applicability of effective medium theory to ferroelectric/ferromagnetic composites with composition and frequency-dependent complex permittivities and permeabilities*, J. Appl. Phys., 79 (1996), pp. 1655–1660.

[50] O. Mattei, G. W. Milton, and M. Putinar, *An extremal problem arising in the dynamics of two-phase materials that directly reveals information about the internal geometry*, Comm Pure Appl Math, (2021).

[51] A. Mecozzi, C. Antonelli, and M. Shtaif, *Kramers-Kronig coherent receiver*, Optica, 3 (2016), pp. 1220–1227.

[52] G. W. Milton, *Bounds on the complex permittivity of a two-component composite material*, J. Appl. Phys., 52 (1981), pp. 5286–5293.

[53] G. W. Milton, *Bounds on the transport and optical properties of a two-component composite material*, Journal of Applied Physics, 52 (1981), pp. 5294–5304.

[54] G. W. Milton, *Extending the Theory of Composites to Other Areas of Science*, Milton-Patton publishers, Salt Lake City, UT, USA, 2016.

[55] G. W. Milton, *Private communication*, 2020.

[56] G. W. Milton, D. J. Eyre, and J. V. Mantese, *Finite frequency range Kramers Kronig relations: bounds on the dispersion*, Phys. Rev. Lett., 79 (1997), pp. 3062–3065.

[57] H. M. Nussenzveig, *Causality and Dispersion Relations*, Academic Press, New York, 1972.

[58] C. Orum, E. Cherkaev, and K. M. Golden, *Recovery of inclusion separations in strongly heterogeneous composites from effective property measurements*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 468 (2012), pp. 784–809.

[59] M.-J. Ou and E. Cherkaev, *On the integral representation formula for a two-component elastic composite*, Math. Methods Appl. Sci., 29 (2006), pp. 655–664.

[60] M.-J. Y. Ou, *On reconstruction of dynamic permeability and tortuosity from data at distinct frequencies*, Inverse Problems, 30 (2014), p. 095002.

[61] J. Scully, D. Silverman, and M. Kendig, eds., *Electrochemical Impedance: Analysis and Interpretation*, ASTM, 1993.

[62] B. Simon, *Loewner's Theorem on Monotone Matrix Functions*, Springer, 2019.

[63] A. Srivastava, *Causality and passivity: From electromagnetism and network theory to metamaterials*, Mechanics of Materials, 154 (2021), p. 103710.

[64] L. N. Trefethen, *Quantifying the ill-conditioning of analytic continuation*, BIT Numerical Mathematics, (2020).

[65] T. H. Wan, M. Saccoccio, C. Chen, and F. Ciucci, *Influence of the discretization methods on the distribution of relaxation times deconvolution: implementing radial basis functions with drttools*, Electrochimica Acta, 184 (2015), pp. 483–499.

[66] M. Wohlers and E. Beltrami, *Distribution theory as the basis of generalized passive-network analysis*, IEEE Transactions on Circuit Theory, 12 (1965), pp. 164–170.

[67] A. H. Zemanian, *Realizability Theory for Continuous Linear Systems*, Academic Press, New York, NY, 1972.

[68] D. Zhang and E. Cherkaev, *Reconstruction of spectral function from effective permittivity of a composite material using rational function approximations*, J. Comput. Phys., 228 (2009), pp. 5390–5409.

[69] M. Žic, S. Pereverzyev, V. Subotić, and S. Pereverzyev, *Adaptive multiparameter regularization approach to construct the distribution function of relaxation times*, GEM-International Journal on Geomathematics, 11 (2020), p. 2.