Geometry of rational functions in l_2 and Walsh approximation

A. P. Mullhaupt

Abstract. The geometry of rational functions in l_2 naturally relates to the unit disc. More to come in this note...

1. Interpretation of geometry in l_2

We start with the sequence space l_2, with inner product

$$(f,g) = \sum_{k\geq 0} f_k g_k^*. \tag{1}$$

which is naturally included in the space of functions analytic in the disc

$$f(z) = \sum_{k\geq 0} f_k z^k \tag{2}$$

where the inner product can be calculated as the contour integral

$$(f,g) = \frac{1}{2\pi i}\int_{|z|=1} f(z)\, g(z)^* \,\frac{dz}{z}. \tag{3}$$

Remark 1. *We have taken for granted the fact that if $\sum_{k\geq 0} |f_k|^2$ converges, then the corresponding function $f(z)$ is analytic in the disc. There are clearly power series about the origin that converge in the unit disc, and thus define analytic functions there, which however do not have coefficients in l_2; for example $f_k = 1$ is not in l_2 but the series converges to $(1-z)^{-1}$ which is analytic in the disc.*

We now consider the functions of the form

$$\begin{aligned} f(z) &= 1 + az + a^2 z^2 + \cdots && (4)\\ &= \frac{1}{1-az} && (5)\end{aligned}$$

where $|a| < 1$ is necessary for convergence in the unit disc, and sufficient for convergence in l_2. We compute the inner product of two such functions

$$\begin{aligned} \frac{1}{2\pi i}\int_{|z|=1} f(z)\, g(z)^* \,\frac{dz}{z} &= \frac{1}{2\pi i}\int_{|z|=1} \left(\frac{1}{1-az}\right)\left(\frac{1}{1-bz}\right)^* \frac{dz}{z} && (6)\\ &= \frac{1}{2\pi i}\int_{|z|=1} \left(\frac{1}{1-az}\right)\left(\frac{1}{z-b^*}\right) dz && (7)\\ &= \frac{1}{1-ab^*}. && (8)\end{aligned}$$

Remark 2. *The Schwartz inequality*

$$|(f,g)| \leq |(f,f)|^{1/2}\, |(g,g)|^{1/2} \tag{9}$$

provides

$$\left|\frac{1}{1-ab^*}\right| \leq \left|\frac{1}{1-|a|^2}\right|^{1/2} \left|\frac{1}{1-|b|^2}\right|^{1/2} \tag{10}$$

and

$$0 \leq \frac{\left(1-|a|^2\right)\left(1-|b|^2\right)}{|1-ab^*|^2} \leq 1. \tag{11}$$

Using the fact

$$\frac{\left(1-|a|^2\right)\left(1-|b|^2\right)}{|1-ab^*|^2} = 1 - \left|\frac{a-b}{1-ab^*}\right|^2 \tag{12}$$

we have

$$0 \leq \left|\frac{a-b}{1-ab^*}\right| \leq 1 \tag{13}$$

for any a, b in the unit disc.

Additionally, we have that the length of the function $(1-az)^{-1}$ is $\left(1-|a|^2\right)^{-1/2}$, so the unit vector in the direction $(1-az)^{-1}$ is

$$\frac{\left(1-|a|^2\right)^{1/2}}{1-az}. \tag{14}$$

The inner product of two unit vectors is thus

$$\left(\frac{\left(1-|a|^2\right)^{1/2}}{1-az}, \frac{\left(1-|b|^2\right)^{1/2}}{1-bz}\right) = \frac{\left(1-|a|^2\right)^{1/2}\left(1-|b|^2\right)^{1/2}}{1-ab^*} \tag{15}$$

which defines the angle between the directions. There are several choices for how to associate an angle with this (complex) number, but whichever one is chosen, the angle is then determined by the value of the inner product. And whichever definition is chosen, there is no choice of a and b inside the unit disc such that the functions $(1-az)^{-1}$ and $(1-bz)^{-1}$ are orthogonal in l_2.

The pole of $(1-az)^{-1}$ is at $z = a^{-1}$, which is out of the unit disc, since $|a| < 1$.

Definition 3. *The space of functions analytic in the unit disc which have power series coefficients in l_2 is a Hilbert space called the Hardy space on the disc, a space which we will denote H^2.*

2. Walsh Approximation

We consider here an approximation due to Walsh. We choose a set of functions

$$S_n = \left\{1, \frac{1}{1-\lambda_1 z}, ..., \frac{1}{1-\lambda_n z}\right\} \tag{16}$$

and for any function $f \in H^2$ we want to know what is the nearest function to f in span S_n, and how close is it? Suppose that r_n is the function, then the condition that r_n is the nearest function to f in span S_n is

$$(f - r_n) \perp \operatorname{span} S_n \tag{17}$$

and the distance is

$$(f - r_n, f - r_n)^{1/2}. \tag{18}$$

We introduce the representation of r_n:

$$r_n(z) = w_0 + \sum_{j=1}^{n} \frac{w_j}{1 - \lambda_j z}. \tag{19}$$

The perpendicularity condition calculation is simplified by noting the following inner products:

$$\frac{1}{2\pi i} \int_{|z|=1} (f(z) - r_n(z)) \frac{dz}{z} = f(0) - r_n(0) \tag{20}$$

$$\begin{aligned} \frac{1}{2\pi i} \int_{|z|=1} (f(z) - r_n(z)) \left(\frac{1}{1-\lambda_k z}\right)^* \frac{dz}{z} &= \frac{1}{2\pi i} \int_{|z|=1} (f(z) - r_n(z)) \frac{dz}{z - \lambda_k^*} && (21) \\ &= f(\lambda_k^*) - r_n(\lambda_k^*). && (22) \end{aligned}$$

Then the perpendicularity condition becomes the interpolation conditions

$$r_n(0) = f(0) \tag{23}$$

$$r_n(\lambda_k^*) = f(\lambda_k^*) \ \text{ for } 1 \le k \le n \tag{24}$$

which in view of the representation of r_n amount to the linear system

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \frac{1}{1-|\lambda_1|^2} & \cdots & \frac{1}{1-\lambda_1 \lambda_n^*} \\ \vdots & \vdots & & \vdots \\ 1 & \frac{1}{1-\lambda_n \lambda_1^*} & \cdots & \frac{1}{1-|\lambda_n|^2} \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} f(0) \\ f(\lambda_1^*) \\ \vdots \\ f(\lambda_n^*) \end{pmatrix}. \tag{25}$$

This introduces the (Hermitian) matrix of this system P, which we will call the Pick matrix a bit further on, as well as some other names. It is a Gram matrix, that is it is a matrix of inner products

$$P_{jk} = \left(\frac{1}{1-\lambda_j z}, \frac{1}{1-\lambda_k z}\right) \tag{26}$$

(where we have included $\lambda_0 = 0$ to take into account the constant term), as long as the λ_k are distinct, this matrix is positive definite since for any $v \neq 0$,

$$v^* P v = \left(v_0 + \sum_{k=1}^{n} \frac{v_k}{1-\lambda_k z}, v_0 + \sum_{k=1}^{n} \frac{v_k}{1-\lambda_k z}\right) > 0. \tag{27}$$

The determinant of P is known:

$$\det P = \left(\prod_{1 \le j \le n} \frac{|\lambda_j|^2}{1 - |\lambda_j|^2} \right) \left(\prod_{1 \le j < k \le n} \left| \frac{\lambda_j - \lambda_k}{1 - \lambda_j \lambda_k^*} \right|^2 \right) \tag{28}$$

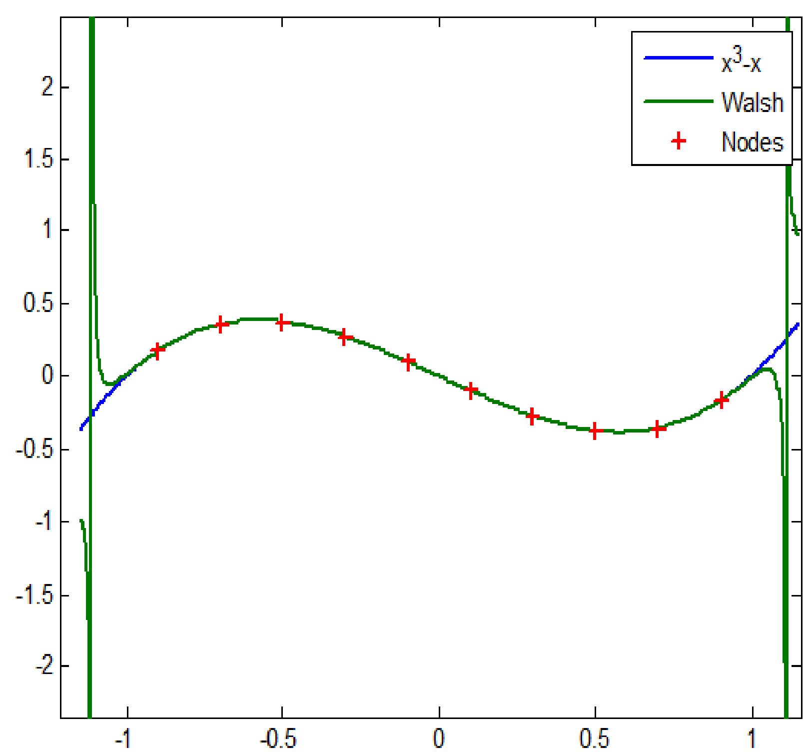
Since the largest eigenvalue of P is at least $n+1$, if the points λ_k are not well separated in the pseudohyperbolic metric, then the determinant will be very small, and the system will be ill-conditioned. We have more to say about this elsewhere.

But the solution of the system does determine r_n, the nearest point to f in span S_n. But not how close it comes to f. This distance is given by

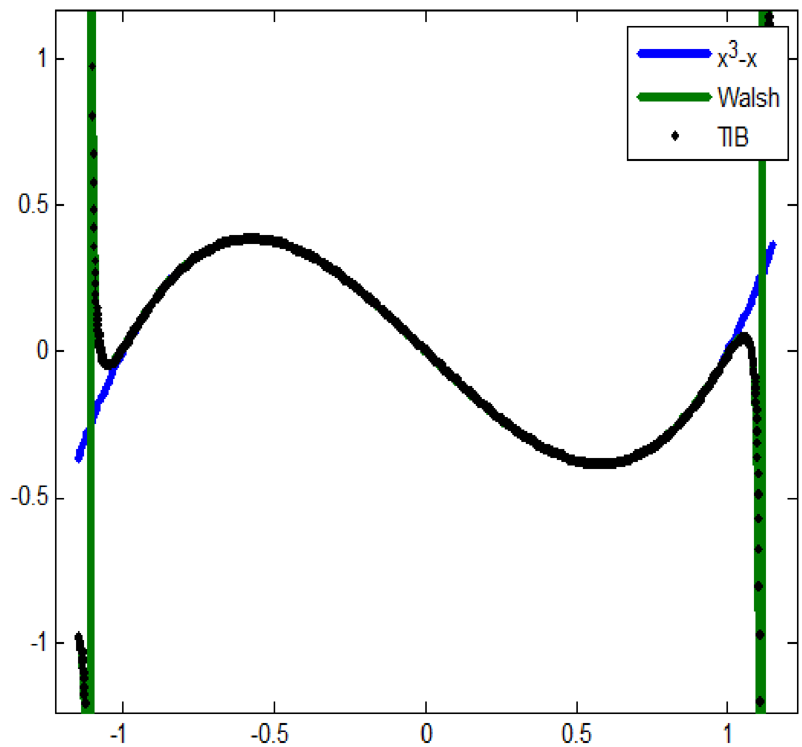
$$(f - r_n, f - r_n) = (f, f) - (r_n, r_n) \tag{29}$$

which in view of the positive definiteness of P, we have $0 < (f - r_n, f - r_n)$ unless $f = r_n$. And unless $r_n = 0$, which fan only happen if f vanishes at each of $\{0, \lambda_1^*, ..., \lambda_n^*\}$,we have $(f - r_n, f - r_n) < (f, f)$.

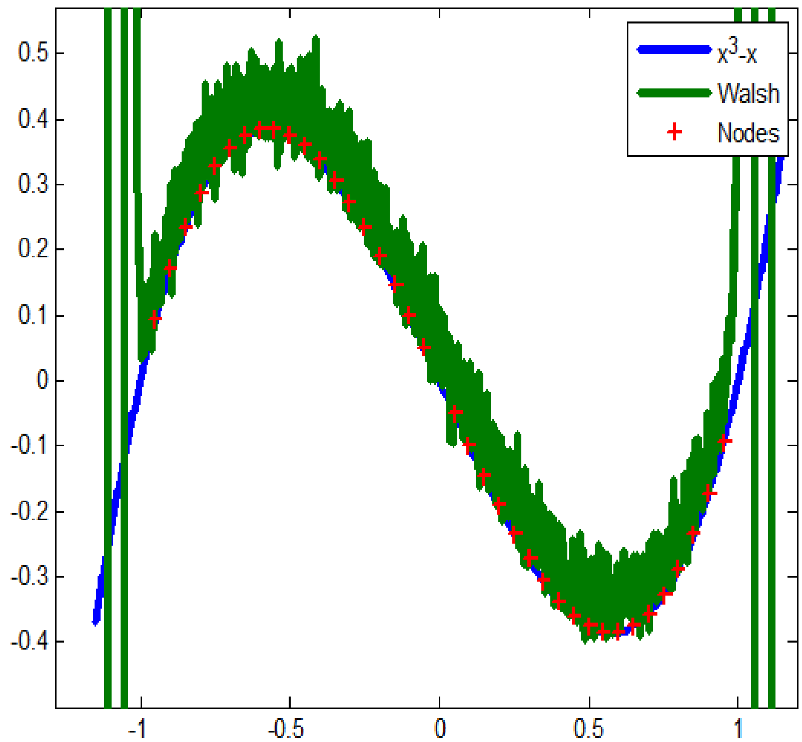
As an example, we have computed this approximation to the function $f(x) = x^3 - x$ using $\{\lambda_1, ..., \lambda_n\} = \left\{-\frac{9}{10}, -\frac{7}{10}, -\frac{5}{10} - \frac{3}{10}, -\frac{1}{10}, \frac{1}{10}, \frac{3}{10}, \frac{5}{10}, \frac{7}{10}, \frac{9}{10}\right\}$, as pictured here:



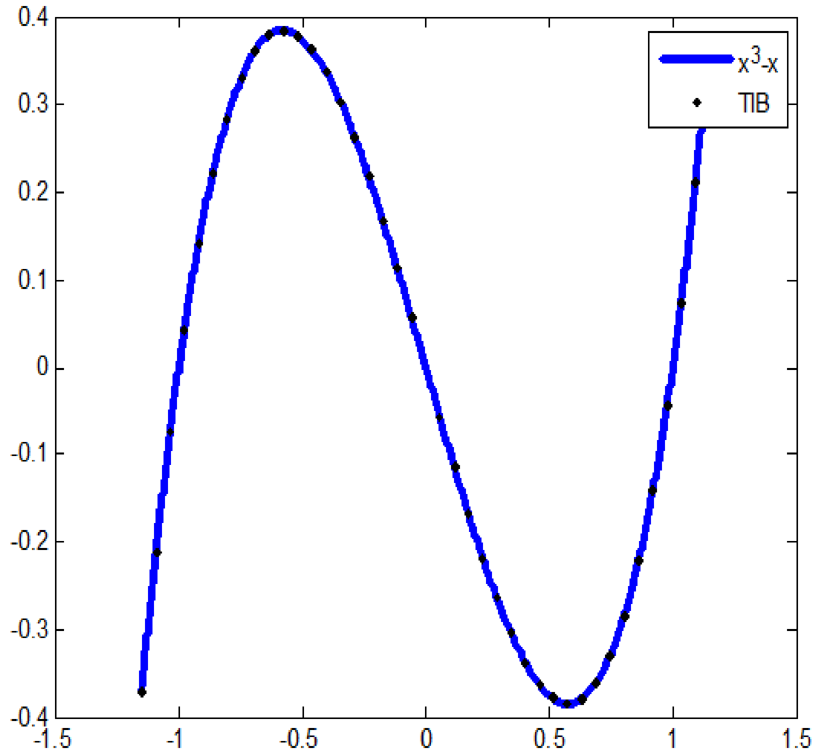
There is reasonable agreement over the interval $[-1, 1]$ which is expected due to the minimization of the Hardy norm, and it is not surprising that the rational approximation has poles near the unit disc. The (Euclidean) condition number of the Pick matrix in this case is about 2.86×10^9. We have another way of calculating this estimate which we call 'TIB', and will discuss later, which we picture here as black dots overlaying the Walsh estimate:



Problems due to the ill-conditioning of the Pick matrix are inevitable, as can be seen by the choice of $\lambda \in \left\{-\frac{9}{20}, \ldots, -\frac{1}{20}, \frac{1}{20}, \ldots, \frac{19}{20}\right\}$ resulting in this picture:



where it is clearly seen that the calculation of the Walsh estimate has failed; the condition number of the Pick matrix is about 7.5×10^{32}, which is enough to exhaust the available floating point arithmetic. In actual fact we have used a somewhat special method to solve the system which is better behaved numerically but not enough to arrive at a decent estimate. But if we resort to the 'TIB' calculation of the estimate, we see that there is no numerical difficulty:



These numerical examples serve as a reminder that although we have shown that the form of the Walsh estimator is a fact, it is not to be used numerically.

A hint of what is happening is available if we look at the coefficients w that were computed in this Walsh estimator:

```
-1.1706e+014       0.80375       -119.08        6534.1 -1.9384e+005
 3.6664e+006 -4.8505e+007  4.7615e+008 -3.6122e+009  2.1808e+010
-1.0712e+011    4.354e+011  -1.484e+012  4.2864e+012 -1.0579e+013
 2.2462e+013 -4.1238e+013  6.5735e+013 -9.1255e+013  1.1056e+014
 1.0837e+014 -8.7709e+013  6.1995e+013 -3.8205e+013  2.0476e+013
-9.5109e+012  3.8115e+012 -1.3101e+012   3.834e+011 -9.4594e+010
 1.9428e+010 -3.2659e+009  4.3909e+008 -4.5732e+007  3.5269e+006
-1.8886e+005        6417.3       -124.09        1.2538
```

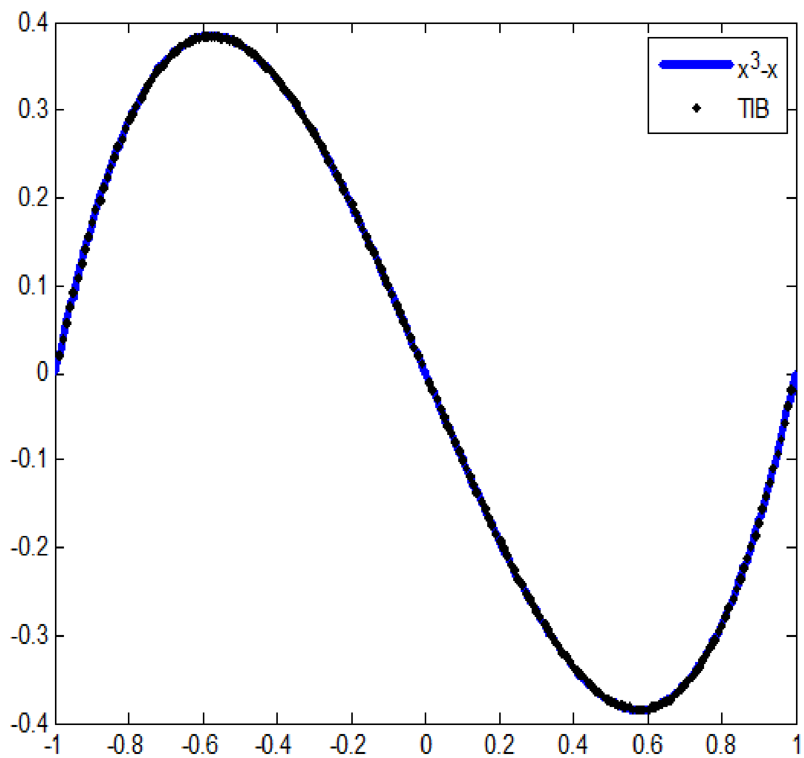
which are some very large magnitudes, but their sum (which is the value $r_n(0)$) is necessarily small (in this case 0.032555). So the values of $r_n(z)$ for z in the disc are the result of adding together very large magnitude numbers to get a small result - which is one of the more usual ways that finite precision arithemetic can be made to fail.

With this in mind, we realize that only if the vector of values

$$\begin{pmatrix} f(0) \\ f(\lambda_1^*) \\ \vdots \\ f(\lambda_n^*) \end{pmatrix} \tag{30}$$

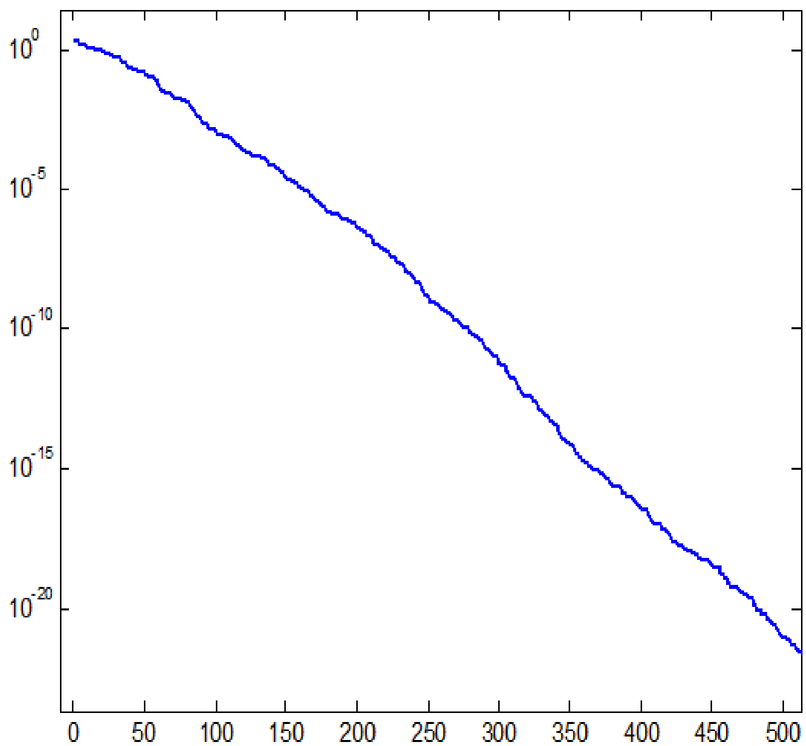
is in the span of the eigenvalues of P which are not too small will the coefficient vector w have a comparable size, leading to a well behaved approximation. The news on this front is drastically bad. For many choices of λ there are typically $O(1)$ such eigenvalues, an observation which we will make precise later. In practice this means that only low degree rational functions can be used in this way.

Having introduced the 'TIB' calculation of the rational approximant to f, we can go to much higher degree. For example, we choose λ from a set of 2000 points distributed (Euclidean) uniformly in the annulus $0.9 < |z| < 1$ and their complex conjugates. This is actually a relatively bad choice of λ, but we will make a point from it. First, here is the picture:



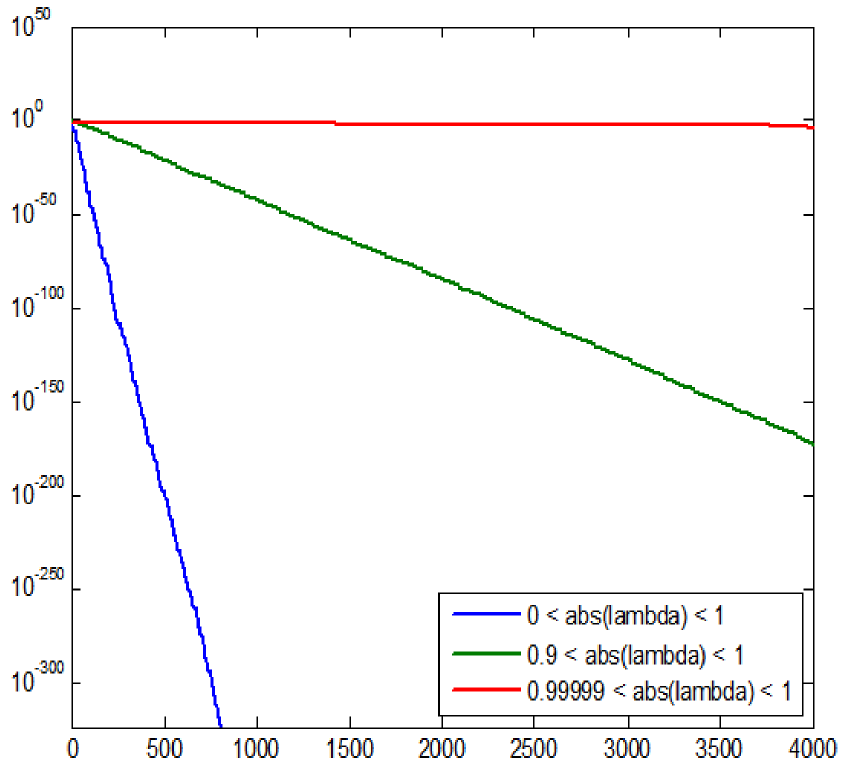
There is no evidence of any numerical pathology in this 4000 degree rational approximation to the cubic $x^3 - x$. In fact, despite having a very fast method for computing this approximation, we would lose patience before we would compute an approximation of degree high enough to be inaccurate. One rough estimate of the condition number of the Pick matrix in this case is above 6.98×10^{290}.

We can plot the H^2 norm error of the rational approximation as a function of the degree of the approximation:



which strongly suggests that this function is in the closure of the span of the rational functions in H^2 (which it is), as well as suggesting a rate of approximation if the λ are chosen in the annulus.

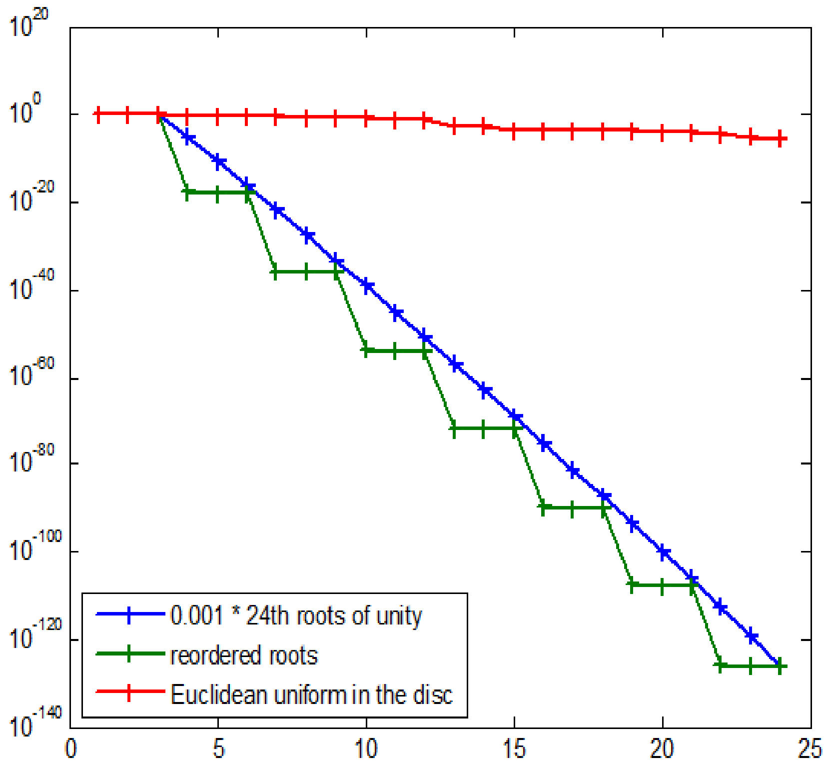
However we can choose a different sets for λ, drawn from the disc, the annulus $0.9 < |z| < 1$ and the annulus $0.99999 < |z| < 1$ resulting in this approximation picture:



which has the interpretation that the rate of approximation depends on the set of λ.It appears that some sets might not even be dense in H^2, (which is true).

The suggestion from these calculations is that to approximate a function in H^2 by rational functions with prescribed poles, then the locations of the poles matters.

We can go a little further with our current example. The function z^3-z is a polynomial, so the only singularity is a pole at ∞, with the degree of the pole being the degree of the polynomial, in this case, three. One expects that by placing three poles near ∞, one would get a close approximation to the function. This picture:



compares the approximation corresponding to λ given as the 24^{th} roots of unity multiplied by 0.001 in the usual order; the approximation is not particularly good until the third pole, after which the accuracy is very high, and increases steadily. If we reorder these roots, so that the scaled cube roots of unity appear first, then those three multiplied by a 24^{th} root of unity, etc., then the approximation is extremely accurate after the third pole, and increases every successive third pole, giving the stairstep pattern. 12 points chosen (Euclidean uniformly) randomly in the unit disc and their conjugates provide the red curve, which shows that approximation is much slower, although the Euclidean uniform disc was the best of the three random annuli for this function. It should not come as a surprise to the reader that if we chose to compute the best approximation by a triple pole at ∞ that it is exactly $z^3 - z$.

Remark 4. *Multiple poles are in the closure of the span of the simple poles, as can be easily seen from the formulae for derivatives*

$$\frac{z}{(1-az)^2} = \frac{d}{da}\left(\frac{1}{1-az}\right) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\left(\frac{1}{1-(a+\varepsilon)z} - \frac{1}{1-(a-\varepsilon)z}\right) \tag{31}$$

$$\frac{2z^2}{(1-az)^3} = \frac{d^2}{da^2}\left(\frac{1}{1-az}\right) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\left(\frac{1}{1-(a+\varepsilon)z} - \frac{2}{1-az} + \frac{1}{1-(a-\varepsilon)z}\right) \tag{32}$$

$$\vdots \tag{33}$$

which express the derivatives of $(1-az)^{-1}$ as approximate linear combinations of nearby simple poles. Since the poles are outside the disc, the convergence is uniform on compact subsets of the disc, etc. An immediate consequence is that for any $\rho > 0$, the polynomials are in the closure of the span of the rational functions $(1-az)^{-1}$ with $|a| < \rho$. The polynomials are dense in H^2 (the convergence of power series of analytic functions), so the rational functions of the form $(1-az)^{-1}$ with $|a| < 1$ are dense in H^2. Note to the sophisticated reader: This is neither new, nor sharp. The sharp answer is that a sequence of poles will generate rational functions with dense span if and only if $\sum_k (1-|\lambda_k|)$ diverges - it is about whether the corresponding Blaschke product converges (see below). There is some other stuff in the background that I have not brought in yet, like Beurling's characterization of closed subspaces of Hardy space, etc.

3. Walsh Approximation by Other Means

Given a function $f \in H^2$ we see that we can find the nearest rational function r_n with specified poles by Walsh interpolation, resulting in a partial fraction expansion. This meant solving the linear system

$$P \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} f(0) \\ f(\lambda_1^*) \\ \vdots \\ f(\lambda_n^*) \end{pmatrix} \tag{34}$$

where P was the Gram matrix of the component functions in the partial fractions

$$P_{jk} = \left(\frac{1}{1-\lambda_j z}, \frac{1}{1-\lambda_k z} \right). \tag{35}$$

We can just as well write this as the Gram matrix of the power series coefficients of those functions (considering them as elements in l_2):

$$P = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 1 & \lambda_1^* & \lambda_1^{*2} & \cdots \\ \vdots & \vdots & & \\ 1 & \lambda_n^* & \lambda_n^{*2} & \cdots \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & \lambda_1 & \cdots & \lambda_n \\ 0 & \lambda_1^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} \tag{36}$$

and view the Walsh interpolation system as

$$\begin{pmatrix} 1 & 0 & 0 & \cdots \\ 1 & \lambda_1^* & \lambda_1^{*2} & \cdots \\ \vdots & \vdots & & \\ 1 & \lambda_n^* & \lambda_n^{*2} & \cdots \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & \lambda_1 & \cdots & \lambda_n \\ 0 & \lambda_1^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} f(0) \\ f(\lambda_1^*) \\ \vdots \\ f(\lambda_n^*) \end{pmatrix}. \tag{37}$$

We recognize that

$$\begin{pmatrix} 1 & 0 & 0 & \cdots \\ 1 & \lambda_1^* & \lambda_1^{*2} & \cdots \\ \vdots & \vdots & & \\ 1 & \lambda_n^* & \lambda_n^{*2} & \cdots \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} f(0) \\ f(\lambda_1^*) \\ \vdots \\ f(\lambda_n^*) \end{pmatrix} \tag{38}$$

where f_k are the coefficients in the power series

$$f(z) = f_0 + f_1 z + f_2 z^2 + \cdots . \tag{39}$$

So the Walsh interpolation system is just the normal equation for the least squares (l_2 norm) approximation

$$\min_w \left\| \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & \lambda_1 & \cdots & \lambda_n \\ 0 & \lambda_1^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{pmatrix} - \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix} \right\| . \tag{40}$$

To determine how close f is to rational functions with specified poles, we do not need to restrict ourselves to the partial fraction representation, instead we can change the coordinates. We can trivialize the least squares problem by orthogonalizing the design matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & \lambda_1 & \cdots & \lambda_n \\ 0 & \lambda_1^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} = QR \tag{41}$$

and then solving for $C = Rw$ instead of w; because the solution then becomes

$$C = Q^* \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix} . \tag{42}$$

Up to some (nontrivial) computational efficiencies, this is what the 'TIB' method of calculation does. The reason this is numerically so robust amounts to the fact that Q is a partial isometry, so that the Euclidean norm of C is bounded by the l_2 norm of f; and small perturbations in the representation of f have likewise small effect on C. So instead of the Walsh system, we can regard the l_2 minimization

$$\min_C \left\| QC - \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix} \right\| \tag{43}$$

as determining the nearest rational approximation to f.

For any unitary transformation U on l_2, we have

$$\left\| UQC - U \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix} \right\| = \left\| QC - \begin{pmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \end{pmatrix} \right\| \tag{44}$$

so a unitary transformation preserves nearest approximation, but unless it also preserves the power series for poles, it does not preserve rationality. The largest subgroup of the unitary group on l_2 which preserves rational approximation turns out to be easy to find, using TIB systems, and even easier to describe them even without using TIB systems - these unitary transformation amounts to a conformal automorphisms of the disc applied to the (inverses) of the poles of the disc. To actually construct such a U we exploit the fact that

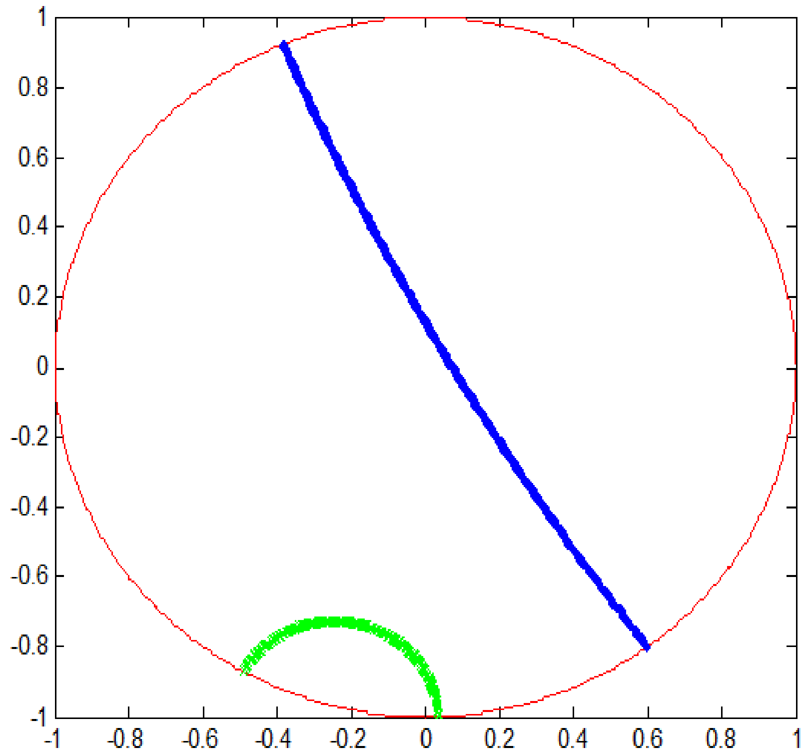
$$U = \begin{pmatrix} b^* \\ b^*A^* \\ b^*A^{*2} \\ \vdots \end{pmatrix} \tag{45}$$

where A is lower triangular with only one eigenvector, and (A, b) is input balanced, that is $AA^* + bb^* = I$. At first, one might wonder why A has to have only one eigenvector, but it turns out that it is because $U^* = U^{-1}$ is likewise a unitary transformation on l_2 which preserves the power series coefficients of poles, hence the components of b are a geometric progression. Although we do not have a fancy reason why this forces A to have constant diagonal, a boring calculation is availabe to verify it. The suspicion is that this should be a consequence of Schur's Lemma if the problem is formulated nicely enough.

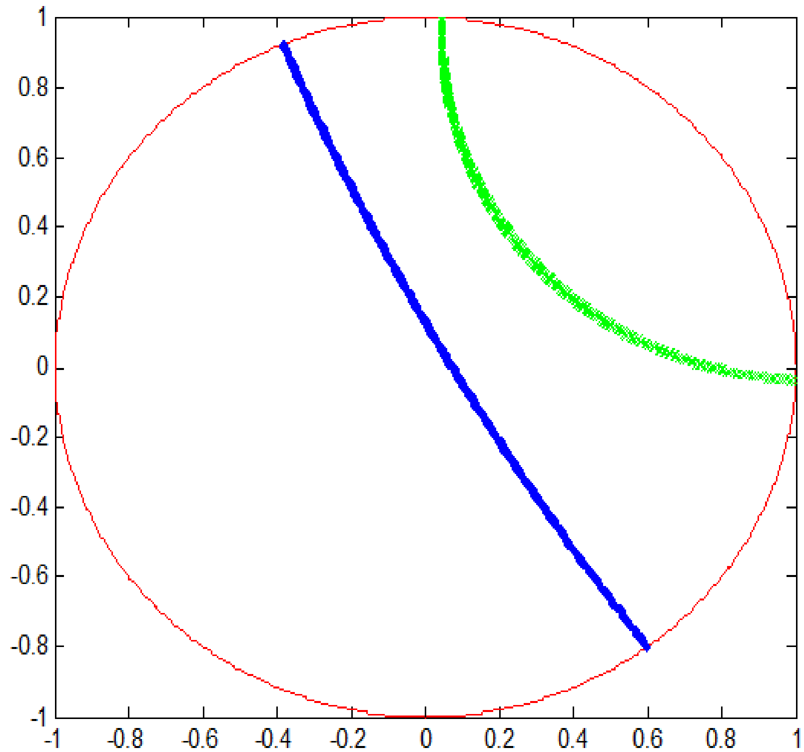
It is a bit glib to claim without proof that these unitary transformations correspond to conformal automorphisms, so we can check by actually seeing what they do to the power series coefficients of simple poles. We chose one point in the unit disc ζ and generated the unitary U described above corresponding to the conformal automorphism:

$$z \mapsto \frac{\zeta - z}{1 - \zeta^* z}. \tag{46}$$

We chose some points λ_k along a circular arc inside and orthogonal to the unit circle, that is, along a hyperbolic geodesic. To each point λ_k we generated the power series coefficients of $(1 - \lambda_k z)^{-1}$, which are the geometric sequence $1, \lambda_k, \lambda_k^2, ...$ Then we transformed the coefficient sequences by multiplying each by U, resulting in sequences which are themselves geometric; the common ratios being denoted by μ_k. Here we plot the unit circle in red, the sequence λ_k in blue, and the sequence μ_k in green:



That the green points lie along a circular arc inside and orthogonal to the unit circle is suggestive that our unitary transformation does in fact correspond to the conformal automorphism. A second ζ chosen (Euclidean uniform) randomly in the disc provides:



which shows that the green points lie along a hyperbolic geodesic.

It is perhaps instructive to look at a small part of the upper left hand corner of this unitary matrix:

```
>> real(U(1:6, 1:6))

ans =

    0.9006    0.1353   -0.1294   -0.0645    0.0051    0.0137
   -0.1353    0.5604    0.1940   -0.2905   -0.1970    0.0196
   -0.1294   -0.1940    0.0727    0.1475   -0.3569   -0.3277
    0.0645   -0.2905   -0.1475   -0.2982    0.0322   -0.2635
    0.0051    0.1970   -0.3569   -0.0322   -0.3792   -0.0768
   -0.0137    0.0196    0.3277   -0.2635    0.0768   -0.1862

>> imag(U(1:6, 1:6))

ans =

         0    0.3673    0.1104   -0.0362   -0.0317   -0.0027
    0.3673         0    0.5264    0.2478   -0.1106   -0.1227
```

```
-0.1104    0.5264    0.0000    0.4004    0.3044   -0.1840
-0.0362   -0.2478    0.4004    0.0000    0.0875    0.2247
 0.0317   -0.1106   -0.3044    0.0875    0.0000   -0.2084
-0.0027    0.1227   -0.1840   -0.2247   -0.2084    0.0000
```

Which suggests that U is up to some sign pattern, a complex symmetric matrix, although I haven't proved that yet.

Remark 5. *It would seem that the fact that the unitary transformations on l_2 which preserve the geometric sequences correspond to the conformal automorphisms of the disc would not be a new fact. However I haven't found it in the literature I have seen.*

Remark 6. *That the unitary transformations we have preserve l_2 distances and rational approximations implies that our density result from can be extended from a small disc around the origin to any small disc in the unit disc.*

4. Orthogonal Rational Functions

For any choice of pole a^{-1} outside the unit disc we can expand the rational function in a power series

$$\frac{1}{1-az} = z + az + a^2z^2 + a^3z^3 + a^4z^4 + O\left(z^5\right) \tag{47}$$

and in the case of a repeated pole we can add additional power series as required:

$$\frac{d}{da}\left(\frac{1}{1-az}\right) = \frac{z}{(1-az)^2} = z + 2az^2 + 3a^2z^3 + 4a^3z^4 + 5a^4z^5 + O\left(z^6\right) \tag{48}$$

$$\frac{1}{2}\frac{d^2}{da^2}\left(\frac{1}{1-az}\right) = \frac{z^2}{(1-az)^3} = z^2 + 3az^3 + 6a^2z^4 + 10a^3z^5 + 15a^4z^6 + O\left(z^7\right) \tag{49}$$

$$\frac{1}{6}\frac{d^3}{da^3}\left(\frac{1}{1-az}\right) = \frac{z^3}{(1-az)^4} = z^3 + 4az^4 + 10a^2z^5 + 20a^3z^6 + 35a^4z^7 + O\left(z^8\right) \tag{50}$$

$$\vdots \tag{51}$$

We represent the coefficients of n such functions with simple poles $\lambda_1^{-1}, ..., \lambda_n^{-1}$ as the matrix

$$\begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots \\ \vdots & \vdots & \vdots & \\ 1 & \lambda_n & \lambda_n^2 & \cdots \end{pmatrix} = \begin{pmatrix} e & \Lambda e & \Lambda^2 e & \cdots \end{pmatrix} \tag{52}$$

where $\Lambda = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ and, as usual, e is the vector with all components equal to one. And we obtain the rational functions

$$\begin{pmatrix} \frac{1}{1-\lambda_1 z} \\ \vdots \\ \frac{1}{1-\lambda_n z} \end{pmatrix} = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots \\ \vdots & \vdots & \vdots & \\ 1 & \lambda_n & \lambda_n^2 & \cdots \end{pmatrix} \begin{pmatrix} 1 \\ z \\ z^2 \\ \vdots \end{pmatrix}. \tag{53}$$

Not quite as obvious, but necessary for us to note is

$$\begin{pmatrix} 1 & a & a^2 & a^3 & a^4 \\ 0 & 1 & 2a & 3a^2 & 4a^3 \\ 0 & 0 & 1 & 3a & 6a^2 \\ \vdots & & \ddots & \ddots & \ddots \end{pmatrix} = \begin{pmatrix} e_1 & Je_1 & J^2e_1 & \cdots \end{pmatrix} \tag{54}$$

where e_1 is the first standard basis vector, that is $e_1 = \begin{pmatrix} 1 & 0 & 0 & \cdots \end{pmatrix}^T$, and

$$J = \begin{pmatrix} a & 0 & 0 & 0 \\ 1 & a & 0 & 0 \\ 0 & 1 & a & 0 \\ 0 & 0 & 1 & a \end{pmatrix}. \tag{55}$$

Taken together, we can construct power series coefficients for a basis for the span of any finite given set of rational functions in H^2 by specifying the Jordan form of a matrix A and a choosing a vector b which is the obvious sum of the cyclic vectors:

$$\begin{pmatrix} b & Ab & A^2b & \cdots \end{pmatrix} \tag{56}$$

which we refer to as a (semi-infinite) Krylov matrix. We can associate rational functions with any such Krylov matrix in the obvious way

$$f(z) = \begin{pmatrix} b & Ab & A^2b & \cdots \end{pmatrix} \begin{pmatrix} 1 \\ z \\ z^2 \\ \vdots \end{pmatrix} = \sum_{k \geq 0} z^k A^k b \tag{57}$$

which in view of the l_2 convergence of the coefficient series defines function analytic on compact subsets of the disc $|z| < 1$, we are thus entitled to write

$$f(z) = (I - zA)^{-1} b. \tag{58}$$

This is a rational function with denominator given by the characteristic polynomial of A.

We see that the Pick matrix of Walsh interpolation is of the form

$$P = \begin{pmatrix} b & Ab & A^2b & \cdots \end{pmatrix} \begin{pmatrix} b^* \\ b^*A^* \\ b^*A^{*2} \\ \vdots \end{pmatrix} \tag{59}$$

where the convergence is guaranteed by the fact that the power series coefficients are in l_2. It is immediate that P is positive definite if and only if the coefficients are independent in l_2, equivalently, the corresponding rational functions in H^2. The Pick matrix satisfies the Stein equation

$$P = bb^* + APA^* \tag{60}$$

which is sometimes called the 'discrete Lyapunov equation'.

In principle one can then compute a Cholesky factorization $P = RR^*$ and obtain an orthogonal basis of rational functions:

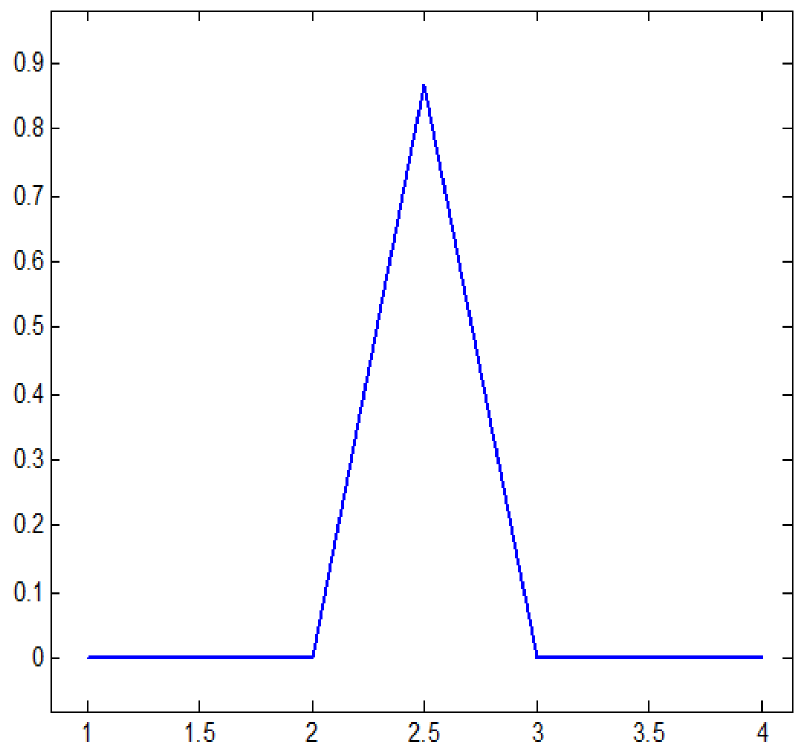
$$R^{-1}\begin{pmatrix} b & Ab & A^2b & \cdots \end{pmatrix} = \begin{pmatrix} R^{-1}b & \left(R^{-1}AR\right)\left(R^{-1}b\right) & \left(R^{-1}AR\right)^2\left(R^{-1}b\right) & \cdots \end{pmatrix} \tag{61}$$

which is clearly orthogonal, since

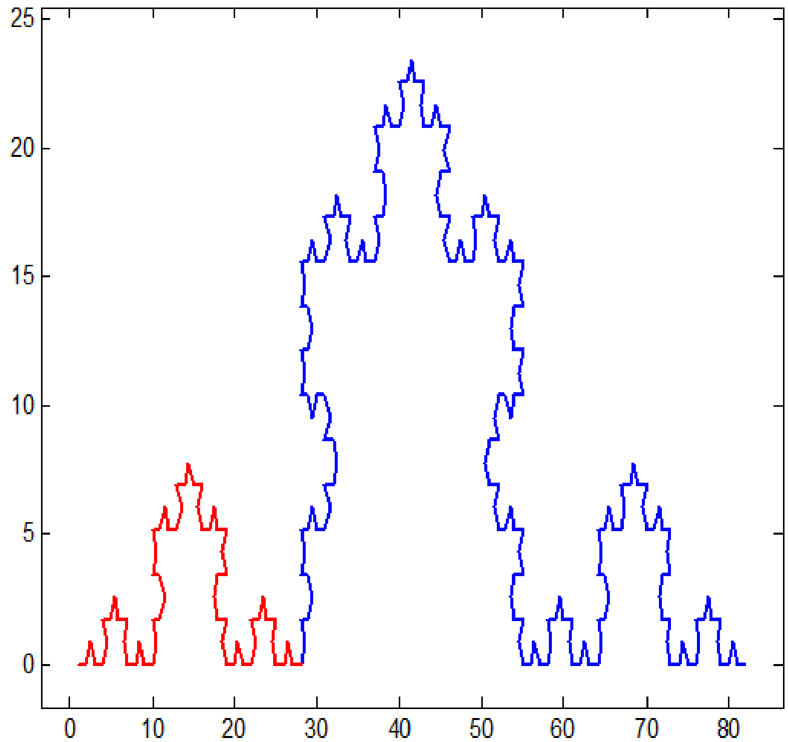
$$\begin{pmatrix} R^{-1}b & \left(R^{-1}AR\right)\left(R^{-1}b\right) & \left(R^{-1}AR\right)^2\left(R^{-1}b\right) & \cdots \end{pmatrix} \begin{pmatrix} \left(b^*R^{-*}\right) \\ \left(b^*R^{-*}\right)\left(R^{-1}AR\right)^* \\ \left(b^*R^{-*}\right)\left(R^{-1}AR\right)^{*2} \\ \vdots \end{pmatrix} = R^{-1}PR^{-*} = I. \tag{62}$$

The problem with this simple idea is that the Pick matrices in question are essentially always ill-conditioned - more to the point the eigenvalues of the Pick matrices decay geometrically (Cf. A. P. Mullhaupt, K.S. Riedel, "Exponential condition number of solutions of the discrete Lyapunov equation", IEEE Transactions on Signal Processing, 52 (2004) pp. 1257-1265.). This is one reason that an arbitrary choice of the representation a lot of rational functions analytic in the unit disc typically does not lead to useful computations.

4.1. An illustrative computation. So far, we have computed rational approximations to a function which turned out to be a low degree rational function itself. Here we will approximate a function which is clearly not rational, to exhibit good approximation and the dependence of that approximation on the choice of λ. The Koch snowflake curve is a fractal; recursively generated by replacing a line segment (in this case starting from the segment from (0,1) to (0,2)) by four segments in this pattern:



and then at each stage, replacing each segment in the same way, which after several stages results in this curve:



where we have shown the third stage in red, and the new part of the fourth stage in blue. We can represent this curve as the headings one follows in traversing the curve from left to right: $\left(0, \frac{\pi}{3}, -\frac{\pi}{3}, 0, \frac{\pi}{3}, \frac{2\pi}{3} 0, \frac{\pi}{3}, -\frac{\pi}{3}, ...\right)$. This sequence is recursively defined in stages

$$\psi_0 = (0) \tag{63}$$
$$\psi_{k+1} = \left(\psi_k, \psi_k + \frac{\pi}{3}, \psi_k - \frac{\pi}{3}, \psi_k\right). \tag{64}$$

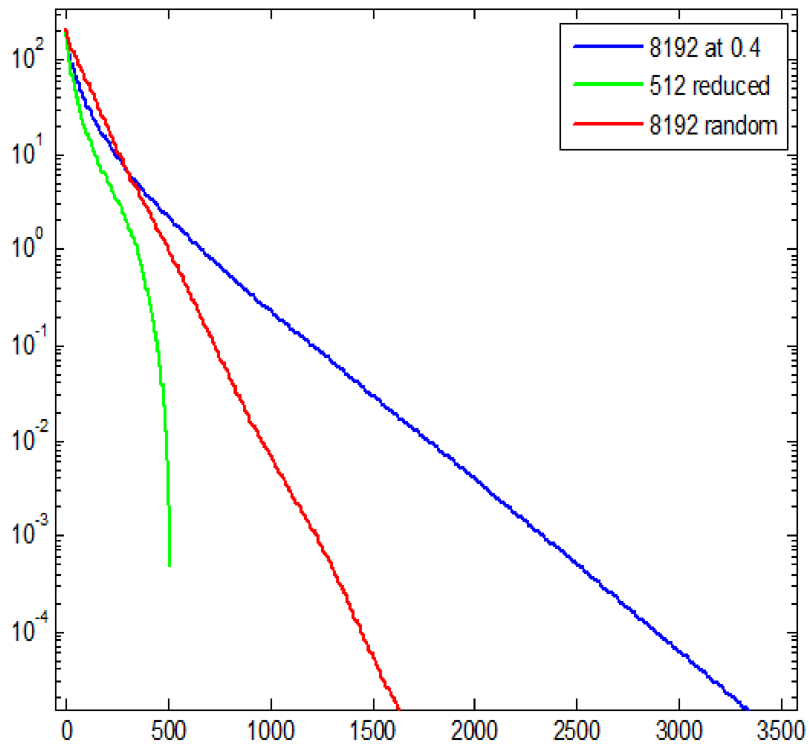
We define the power series with coefficients given by the sequence ψ_k as the (polynomial) $\kappa_k(z)$, and define the limit of these polynomials as $\kappa(z)$. Since the maximum absolute value of the coefficients of the $\left(4^k\right)^{th}$ degree polynomial ψ_k is clearly less than or equal to $k\frac{\pi}{3}$, the radius of convergence of $\kappa(z)$ is unity, and $\kappa(z)$ is analytic in the unit disc. The recursion provides the functional equation

$$\begin{aligned}
\kappa_{k+1}(z) &= \kappa_k\left(z^{4^k}\right)\kappa_k(z) && (65)\\
&= \kappa_1\left(z^{4^k}\right)\kappa_1\left(z^{4^{k-1}}\right)\kappa_{k-1}(z) && (66)\\
&= \kappa_1(z)\,\kappa_1\left(z^4\right)\cdots\kappa_1\left(z^{4^k}\right) && (67)
\end{aligned}$$

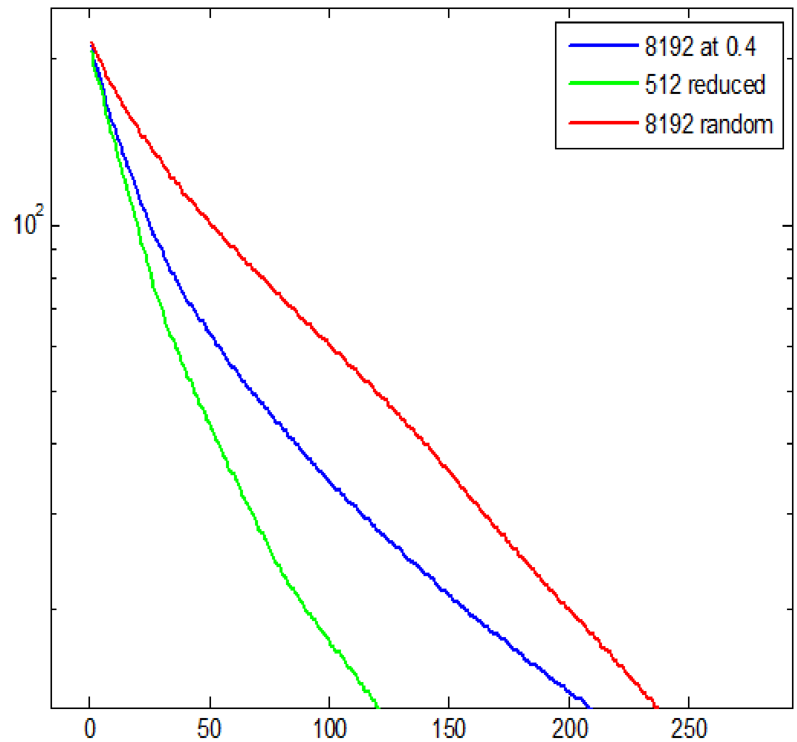
where $\kappa_1(z) = \frac{\pi}{3}\left(z - z^2\right)$. It is clear that $\kappa(z)$ is not rational, since for each k it vanishes at all the $\left(4^k\right)^{th}$ roots of unity; therefore $\kappa(z)$ has the unit disc as a natural boundary.

The power series coefficients of $\kappa(z)$ are not in l_2, since their magnitude does not tend to zero with increasing degree. However for any $\rho \in (0,1)$ the function $\kappa(\rho z)$ is not rational (the natural boundary is now at $|z| = \rho^{-1}$), but is in H^2.

Having established the function $\kappa(\rho z)$ as irrational but in H^2, we now compute some approximations to it, in this case using $\rho = 0.995$:

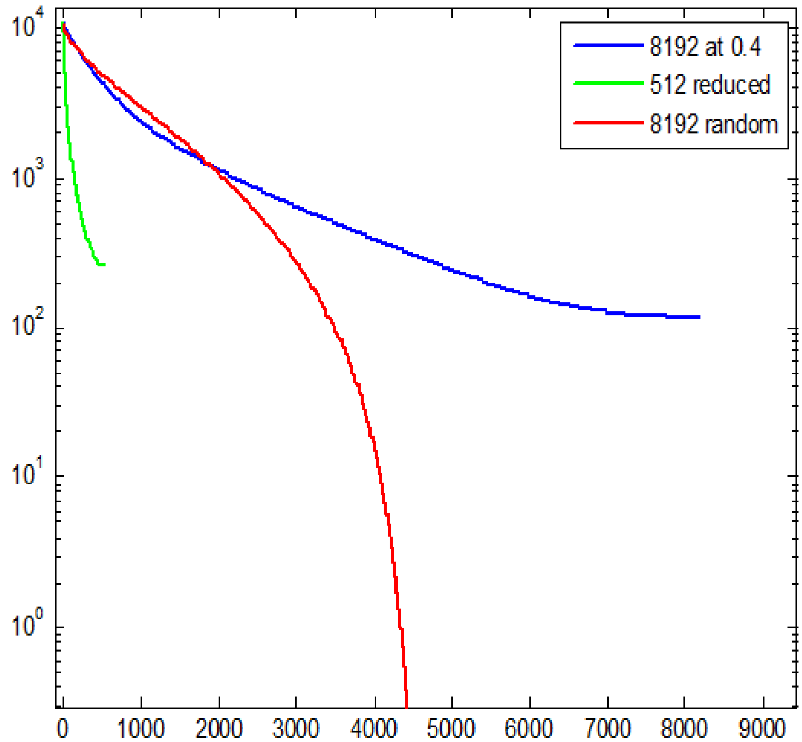


The blue line corresponds to a single pole at $(0.4)^{-1}$ of order 8192, the red line corresponds to 4096 Euclidean uniform points in the unit disc and their conjugates as the location of the pole inverses, and the green line represents poles which have been computed to optimize the approximation in some sense. We can see that the single high order pole is initially better than the random poles, but the random poles provide a better asymptotic approximation. If we examine the initial approximation more closely:



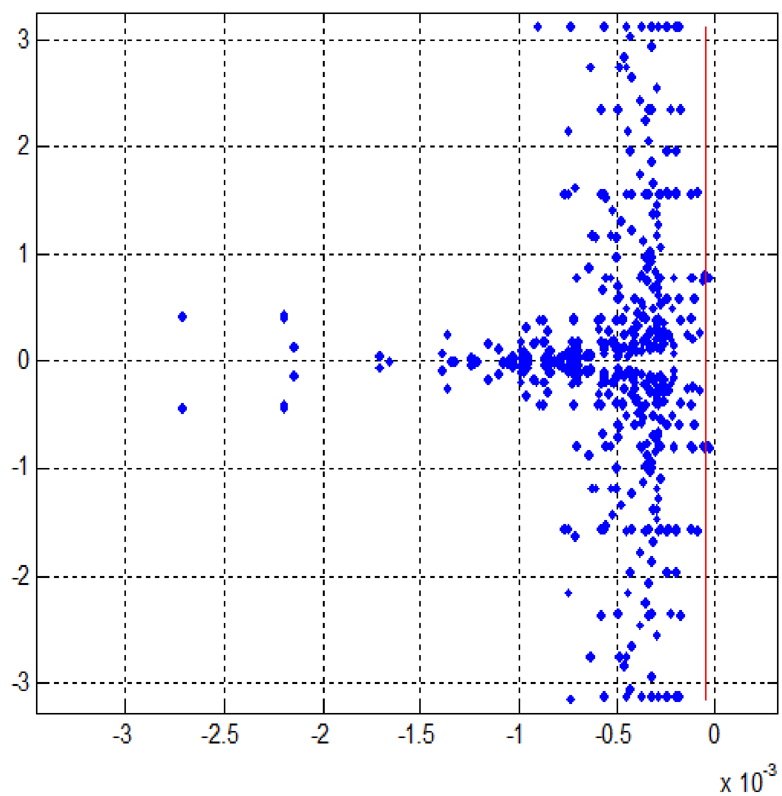
we see that the initial rate of approximation is about as good for the single high order pole as it is for the 'optimized' choice of poles.

If we increase ρ to 0.99995, then the rational approximation will be more difficult. The picture of error against degree then becomes:



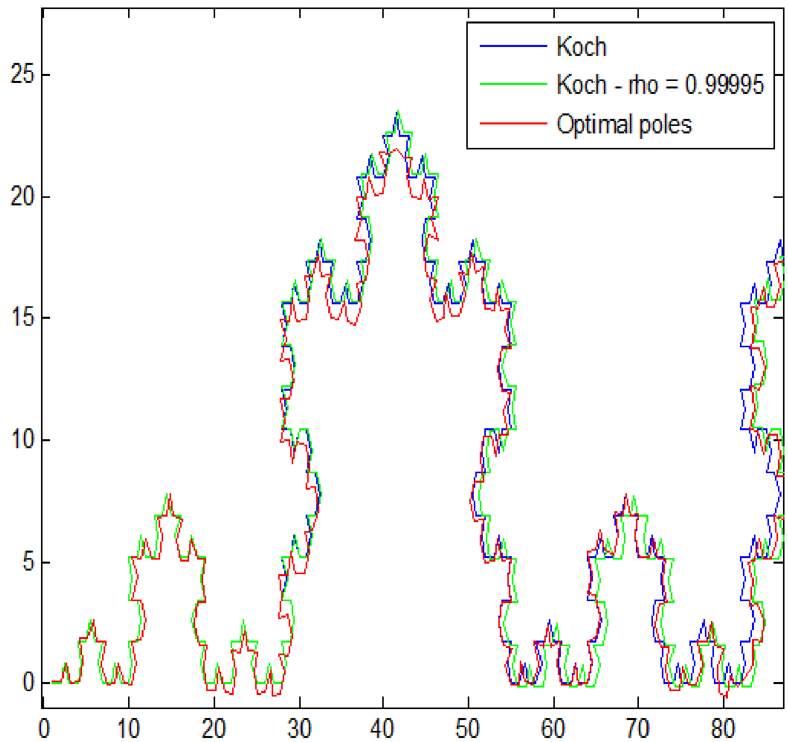
which shows again that the high order pole is initially a bit better than the random distribution of poles, and that the 'optimized' poles are significantly better for the same degree of approximation.

Since we know that $\kappa(\rho z)$ has a natural boundary at $|z| = \rho^{-1}$ we might expect that the poles in the 'optimized' rational approximation lie close to that boundary; we plot the logarithms of the poles:

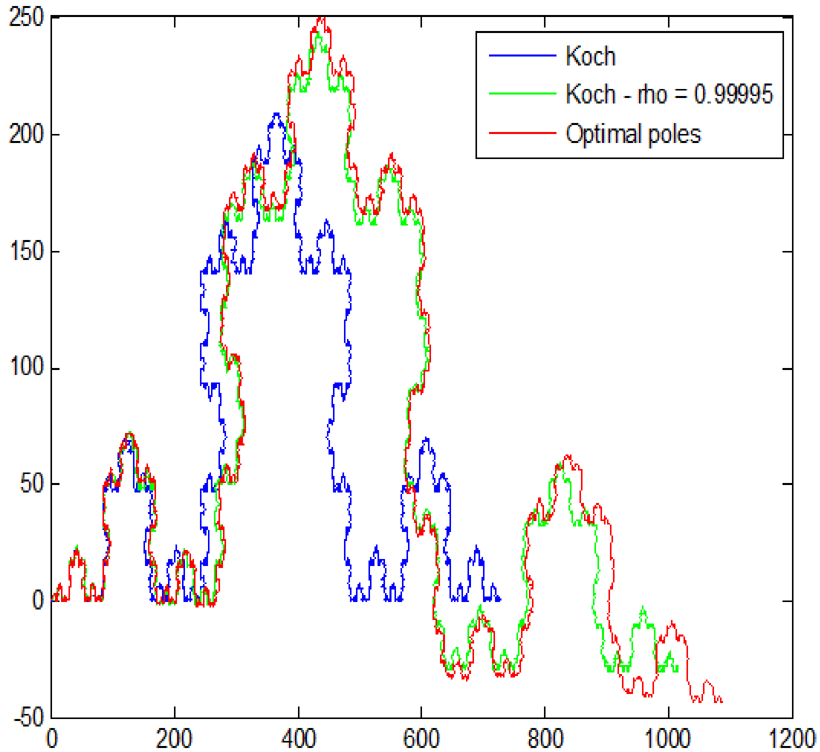


where the red line indicates the location of the natural boundary. In the optimization to produce these poles, we did not force the poles to be complex conjugate pairs, however this is essentially the case. It also appears that the arguments of the optimum poles are clustered at the $\left(4^k\right)^{th}$ roots of unity, which would be a reasonable expectation.

Finally we can display these results in a way that gives some idea how close the approximations are with respect to the Koch fractal:



We can see that the first few stages in some sort of agreement, if we look at the entire curve:



we can see that the approximation (in red) tracks $\kappa(\rho z)$ more closely than $\kappa(\rho z)$ tracks $\kappa(z)$.

Remark 7. *The point here is not that rational functions are close to the Koch snowflake; it is that even though the Koch snowflake corresponds to a fairly un-rational function in ways which are easy to understand, we can nevertheless compute high degree rational approximations to it, which should be compared to the naive application of Walsh's interpolation system to the function* $z^3 - z$.

5. SCHWARZ, SCHUR, AND PICK-NEVANLINNA

Consider the rational functions with simple poles at a_j^{-1} for $1 \leq j \leq n$:

$$\left\{(1-a_1 z)^{-1}, (1-a_2 z)^{-1}, ..., (1-a_n z)^{-1}\right\}. \tag{68}$$

These functions span a linear space, which is the same as the column space:

$$\operatorname{col}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix}. \tag{69}$$

We can obtain an orthogonal basis for this space by, for example, using Gram-Schmidt, or what is almost the same thing, the QR factorization. Recall that the QR factorization of a matrix M is a factorization $M = QR$, where in this case Q is column-unitary (that is $Q^*Q = I$, and R is upper triangular. This factorization can be computed recursively by using Householder reflections, or other elementary unitary transformations (e.g. Givens rotations, etc.). Here we have to use other unitary transformations.

The QR factorization is easily defined recursively, we determine a transformation K_1 such that

$$K_1 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \left(1-|a_1|^2\right)^{1/2} & * & \cdots & * \\ 0 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \end{pmatrix}, \tag{70}$$

then we determine another unitary transformation of the form $1 \oplus K_2$ so that

$$(1 \oplus K_2) \begin{pmatrix} \left(1-|a_1|^2\right)^{1/2} & * & \cdots & * \\ 0 & * & \cdots & * \\ 0 & 0 & \cdots & * \\ 0 & 0 & & * \\ \vdots & \vdots & & \vdots \end{pmatrix} \tag{71}$$

and so on. After n such transformations, we have

$$(I_{n-1} \oplus K_n) \cdots (1 \oplus K_2)\, K_1 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} R \\ 0 \\ \vdots \end{pmatrix} \tag{72}$$

from which it follows

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} = Q \begin{pmatrix} R \\ 0 \\ \vdots \end{pmatrix} \tag{73}$$

where Q^* is the first n rows of the product of the transformations

$$Q^* = \begin{pmatrix} I_n & 0 \end{pmatrix} (I_{n-1} \oplus K_n) \cdots (1 \oplus K_2)\, K_1. \tag{74}$$

It is a bit perplexing how to proceed without exploiting the rational structure. A general unitary transformation acting on a general sequence in l_2 is an interesting idea, but not one that springs instantly into computation.

Remark 8 [Gram-Schmidt and QR]**.** *The Gram-Schmidt orthogonalization of a sequence in a Hilbert space is a special case of the QR factorization. As a result, QR factorizations of matrices X with infinitely many columns exist, and can be had by the algorithm described.*

However we do have the rational structure - that the columns of X are geometric sequences, and we have unitary transformations which preserve the geometric sequences as described above. So the choice of K_1 is obvious, it corresponds to the conformal automorphism of the disc

$$z \mapsto \frac{a_1 - z}{1 - a_1^* z} \tag{75}$$

which maps a_1 to 0, since

$$K_1 \begin{pmatrix} 1 \\ a_1 \\ a_1^2 \\ \vdots \end{pmatrix} = e_1 = \begin{pmatrix} 1 \\ 0 \\ 0^2 \\ \vdots \end{pmatrix}, \tag{76}$$

and by way of preserving geometric series and their norms in l_2:

$$K_1 \begin{pmatrix} 1 \\ a_k \\ a_k^2 \\ \vdots \end{pmatrix} = \left(\frac{1 - \left| \frac{a_1 - a_k}{1 - a_1^* a_k} \right|^2}{1 - |a_k|^2} \right)^{1/2} \begin{pmatrix} 1 \\ \left(\frac{a_1 - a_k}{1 - a_1^* a_k} \right) \\ \left(\frac{a_1 - a_k}{1 - a_1^* a_k} \right)^2 \\ \vdots \end{pmatrix}. \tag{77}$$

Thus the first stage of the QR factorization corresponds to applying the conformal automorphism

$$(a_1, ..., a_n) \mapsto \left(0, \left(\frac{a_1 - a_2}{1 - a_1^* a_2} \right), ..., \left(\frac{a_1 - a_n}{1 - a_1^* a_n} \right) \right) \tag{78}$$

which we will denote

$$\alpha_{1k} = \left(\frac{a_1 - a_k}{1 - a_1^* a_k} \right) \tag{79}$$

and

$$R_{1k} = \left(\frac{1 - \left| \frac{a_1 - a_k}{1 - a_1^* a_k} \right|^2}{1 - |a_k|^2} \right)^{1/2} \tag{80}$$

for simplicity. So the first stage can be written

$$K_1 \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \left(1 - |a_1|^2\right)^{1/2} & R_{12} & \cdots & R_{1n} \\ 0 & R_{12}\alpha_{12} & \cdots & R_{1n}\alpha_{1n} \\ 0 & R_{12}\alpha_{12}^2 & \cdots & R_{1n}\alpha_{1n}^2 \\ \vdots & \vdots & & \vdots \end{pmatrix}. \tag{81}$$

The second stage is the same algorithm applied to the matrix

$$\begin{pmatrix} R_{12}\alpha_{12} & \cdots & R_{1n}\alpha_{1n} \\ R_{12}^2\alpha_{12}^2 & \cdots & R_{1n}\alpha_{1n}^2 \\ \vdots & & \vdots \end{pmatrix} \tag{82}$$

which can clearly be done. It is a straightforward task to keep track of the values of R_{jk} as the algorithm progresses. The second stage will apply the conformal automorphism

$$z \mapsto \frac{\alpha_{12} - z}{1 - \alpha_{12}^* z} = \frac{\left(\frac{a_1 - a_2}{1 - a_1^* a_2}\right) - z}{1 - \left(\frac{a_1 - a_2}{1 - a_1^* a_2}\right)^* z}. \tag{83}$$

At the end of n stages, all of the initial points $a_1, ..., a_n$ have been mapped to zero. This has the interpretation that the unitary factor

$$(I_{n-1} \oplus K_n) \cdots (1 \oplus K_2)\, K_1 \tag{84}$$

corresponds to the Blaschke product

$$e^{i\theta} \prod_{k=1}^{n} \left(\frac{z - a_k}{1 - a_k^* z}\right) \tag{85}$$

for some real θ.

One of the most common uses of the QR factorization is to find the orthogonal projection of a vector y on to the span of a set of vectors. In particular, if $X = QR$ then the projection of y onto span X is $Q\,(Q^* y)$ and the distance from y to span X is

$$\left(y^*\,(I - QQ^*)\, y\right)^{1/2}. \tag{86}$$

When solving the orthogonal projection problem, the overdetermined system

$$Xa = y \tag{87}$$

is transformed to the system

$$Ra = Q^* y \tag{88}$$

which, if the columns of X are independent, is nonsingular with solution

$$a = R^{-1} Q^* y. \tag{89}$$

For comparison, the Walsh approximation system was of the normal equations form

$$(X^* X)\, a = X^* y \tag{90}$$

which are (up to numerical issues) equivalent.

Remark 9 [Blaschke Convergence and Density]**.** *It seems then that the orthogonal basis for the rational functions $(1 - a_k z)^{-1}$ corresponds to composition of the functions with a specific sequence of finite Blaschke products so that the sequence of composed functions*

$$\left(1 - a_k e^{i\theta} \prod_{j=1}^{k} \left(\frac{z - a_j}{1 - a_j^* z}\right)\right)^{-1} \tag{91}$$

are polynomials of (exact) degree k. The obvious extension to infinite sequences $a_1, a_2, ...$ would seem to correspond to the convergence of the Blaschke product. This makes sense

- if the span of the sequence of functions $(1 - a_j z)^{-1}$ *is dense in* H^2 *then the infinite Blaschke product must diverge - since the corresponding unitary transformation*

$$\cdots (I_{n-1} \oplus K_n) \cdots (1 \oplus K_2) K_1 \tag{92}$$

maps every sequence in l_2 *to zero. It is well known that an infinite Blaschke product converges if and only if* $\sum_k (1 - |a_k|)$ *converges.*

Remark 10 [Frostman's Theorem]**.** *If* $\phi(z)$ *is an inner function, then aside from a set of logarithmic capacity zero, for any* $|a| < 1$ *the function*

$$\frac{a - \phi(z)}{1 - a^* \phi(z)} \tag{93}$$

is a convergent Blaschke product. This has an interpretation from the point of view of the geometric sequence preserving QR *algorithm but I haven't worked out what it is. Perhaps it is a simple consequence. One way or another, the situation of Frostman's theorem ought to be examined from the* QR *point of view.*

Remark 11 [Equivalence of QR and Nevanlinna Algorithms]**.** *This is essentially the* QR *factorization algorithm for Householder reflections except that we have to substitute for Householder reflections, other elementary unitary transformations which preserve geometric series in* l_2 *(i.e. preserve rational functions in* H^2*). Nevanlinna's algorithm for Pick's theorem, which we will get to in a bit, determines the existence of a rational function* ϕ *that satisfies interpolation conditions*

$$\phi(a_j) = b_j \tag{94}$$

and the bound $|\phi(z)| \le 1$ *for* $|z| \le 1$*, and in the* QR *algorithm, we have described the steps are literally the same as Nevanlinna's algorithm for the case* $b_j = 0$*, an observation which I believe is new. Nevanlinna's algorithm is essentially a repeated application of Pick's theorem (also known as the invariant form of Schwartz's lemma).*

Remark 12. *We didn't handle the case of multiple poles to avoid extra notation, but it should be a straightforward generalization.*

6. The Tyranny of the Stein Equation

Much of the foregoing has connection to, and is in some ways, illuminated by the Stein equation. This equation is usually given in the symmetric (or more properly, Hermitian,) form:

$$P - APA^* = BB^* \tag{95}$$

where A and B are given, and P is to be solved. There is also the nonsymmetric form:

$$P - APF = BG. \tag{96}$$

It is common for some literature to refer to the Stein equation as the 'Discrete Lyapunov' equation. This is because the discrete analog of the Lyapunov equation for stability is a Stein equation.

In the case that A is stable, that is $A^k \to 0$ as $k \to \infty$, then the solution by iteration

$$\begin{aligned} P &= BB^* + APA^* && (97)\\ &= BB^* + A\left(BB^* + APA^*\right)A^* && (98)\\ &= BB^* + ABB^*A^* + A^2BB^*A^{*2} + \cdots && (99)\end{aligned}$$

exists, is unique, Hermitian, and nonnegative definite. The representation

$$P = \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \begin{pmatrix} B^* \\ B^*A^* \\ B^*A^{*2} \\ \vdots \end{pmatrix} \tag{100}$$

shows that P is positive definite if and only if $\begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}$ has full rank.

In general, the positivity of P places a constraint on the dimension of B for a given A; in order for P to be positive, then B must have at least as many columns as the maximum number of distinct Jordan blocks in A that have a common eigenvalue. (Otherwise you can choose a left eigenvector $u^*A = \lambda u^*$ which is orthogonal to B, hence $u^* \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \cdots \end{pmatrix}$ and $u^*Pu = 0$.) However in the case of a set of rational functions in l_2, then as long as the poles are distinct, then a B exists with a single column such that P is positive definite. This is a *topological* constraint, but *geometrically* we would like to know how close P is to singularity, or some measure of how 'positive definite' is P. One way or another, this amounts to getting information about the spectrum of P.

Now we provide a (new) result which includes the central geometric fact of the Stein equation.

Theorem 13. *Let $P - APA^* = BB^*$, and $U^*A = \Lambda U^*$ be a left eigenvalue-eigenvector decomposition of A with nonsingular U. Then*

$$U^*PU - \Lambda\left(U^*PU\right)\Lambda^* = \left(U^*B\right)\left(U^*B\right)^* \tag{101}$$

and

$$\begin{aligned} \frac{\left|\left(U^*PU\right)_{jk}\right|^2}{\left(U^*PU\right)_{jj}\left(U^*PU\right)_{kk}} &= \frac{\left|e_j^*\left(U^*B\right)\left(U^*B\right)^*e_k\right|^2}{\left|1-\lambda_j\lambda_k^*\right|^2} \frac{\left(1-\left|\lambda_j\right|^2\right)\left(1-\left|\lambda_k\right|^2\right)}{\left(e_j^*\left(U^*B\right)\left(U^*B\right)^*e_j\right)\left(e_k^*\left(U^*B\right)\left(U^*B\right)^*e_k\right)} && (102)\\ &= \left(\frac{\left(1-\left|\lambda_j\right|^2\right)\left(1-\left|\lambda_k\right|^2\right)}{\left|1-\lambda_j\lambda_k^*\right|^2}\right)\left(\frac{\left|\left[\left(U^*B\right)^*e_j\right]^*\left[\left(U^*B\right)^*e_k\right]\right|^2}{\left\|\left(U^*B\right)^*e_j\right\|^2\left\|\left(U^*B\right)^*e_k\right\|^2}\right). && (103)\end{aligned}$$

In the case $d = 1$, where B is a vector, then

$$\frac{\left|\left[\left(U^*B\right)^*e_j\right]^*\left[\left(U^*B\right)^*e_k\right]\right|^2}{\left\|\left(U^*B\right)^*e_j\right\|^2\left\|\left(U^*B\right)^*e_k\right\|^2} = 1 \tag{104}$$

and

$$\frac{\left|\left(U^*PU\right)_{jk}\right|^2}{\left(U^*PU\right)_{jj}\left(U^*PU\right)_{kk}} = \frac{\left(1-\left|\lambda_j\right|^2\right)\left(1-\left|\lambda_k\right|^2\right)}{\left|1-\lambda_j\lambda_k^*\right|^2} = 1 - \left|\frac{\lambda_j - \lambda_k}{1-\lambda_j^*\lambda_k}\right|^2. \tag{105}$$

These are similarity invariants of the system (they are similarity invariants of A). In particular if we change coordinates by a nonsingular transformation T, $(A,B) \mapsto \left(TAT^{-1}, TB\right)$ and $P \mapsto TPT^*$ then $U \mapsto T^{-*}UD$ for some diagonal matrix D (which reflects the possibility of rescaling the eigenvectors), and

$$U^*PU \mapsto D^*U^*T^{-1}TPT^*T^{-*}UD = D^*\left(U^*PU\right)D. \tag{106}$$

Thus

$$\frac{\left|\left(U^*PU\right)_{jk}\right|^2}{\left(U^*PU\right)_{jj}\left(U^*PU\right)_{kk}} \mapsto \frac{\left|D^*_{jj}\left(U^*PU\right)_{jk}D_{kk}\right|^2}{\left(D^*_{jj}U^*PUD_{jj}\right)_{jj}\left(D^*_{kk}U^*PUD_{kk}\right)_{kk}} = \frac{\left|\left(U^*PU\right)_{jk}\right|^2}{\left(U^*PU\right)_{jj}\left(U^*PU\right)_{kk}} \tag{107}$$

is preserved under coordinate change. It follows that the angle of P-conjugacy of the eigenvectors of A is invariant under coordinate change. So if we have orthogonal eigenvectors of A (A is diagonal) then we have highly nonorthogonal eigenvectors of P, and vice versa.

Remark 14. *The impact of the situation described by the result appears in (A. P. Mullhaupt, K.S. Riedel, "Exponential condition number of solutions of the discrete Lyapunov equation", IEEE Transactions on Signal Processing, 52 (2004) pp. 1257-1265.) and gives rise to the widespread belief of the ill-conditioning of Pick matrices, and Pick-like matrices. However this result in principle allows for the most precise picture, which at some point we ought to elaborate.*

Remark 15 [Amplification]**.** *In the theorem, we can choose T to diagonalize A, and then again to diagonalize P. What results is the (previously known) characterization of the eigenvectors of an input balanced (i.e. $P = I$) pair (A,B). What we get from this new result is that other choices of T still preserve the 'correlation' matrix. In some sense, this is obvious. Suppose you have two left eigenvectors $u_j^*A = \lambda u_j^*$ and $u_k^*A = \lambda u_k^*$, then*

$$u_j^*\left(\begin{array}{cccc} B & AB & A^2B & \cdots \end{array}\right) = \left(u_j^*B\right)\left(\begin{array}{cccc} 1 & \lambda_j & \lambda_j^2 & \cdots \end{array}\right) \tag{108}$$

and

$$u_j^*Pu_k = \frac{u_j^*\left(BB^*\right)u_k}{1-\lambda_j\lambda_k^*} \tag{109}$$

and

$$\frac{\left|u_j^*Pu_k\right|^2}{\left(u_j^*Pu_j\right)\left(u_k^*Pu_k\right)} = \frac{\left(1-\left|\lambda_j\right|^2\right)\left(1-\left|\lambda_k\right|^2\right)}{\left|1-\lambda_j\lambda_k^*\right|^2}. \tag{110}$$

For a nontrivial Jordan block in A, the result is slightly different.

7. About those Unitary Transformations