Annales Univ. Sci. Budapest., Sect. Comp. 49 (2019) 35–51

ON THE POLE STABILITY OF RATIONAL APPROXIMATION

Gergő Bognár[1] **and Sándor Fridli**[2]
(Budapest, Hungary)

Communicated by László Szili
(Received March 27, 2019; accepted June 25, 2019)

Abstract. In this paper we investigated the stability of the system parameters of the adaptive rational transformation, i.e. how specific perturbation of the inverse poles affects the approximation error. We provided some estimations and sufficient confidence intervals on the perturbation of the absolute values of the inverse poles with respect to the approximation error. The research was motivated by biomedical (primarily ECG) signal processing problems, where the results may be utilized to improve the approximation and representation techniques.

1. Introduction

Rational systems, like Malmquist–Takenaka, Laguerre, and Kautz systems are well known and frequently used in system and control theory, and signal processing. For details, we refer the reader to [4, 8]. In this paper, we focus on the signal processing applications, namely the modelling with rational transformation, as introduced in [6]. Here, the signal is modelled as the linear combination with specific rational functions, i.e. it is approximated by the orthogonal projection onto a subspace spanned by rational systems. The previous results in biomedical signal processing prove that the adaptive rational

Key words and phrases: Rational functions, Malmquist–Takenaka systems, Laguerre systems, ECG signal processing.
2010 Mathematics Subject Classification: 41A20.
1998 CR Categories and Descriptors: J.3.
[1]EFOP-3.6.3-VEKOP-16-2017-00001: Talent Management in Autonomous Vehicle Control Technologies - The Project is supported by the Hungarian Government and co-financed by the European Social Fund.
[2]This research was supported by the Hungarian Scientific Research Funds (OTKA) No K115804.

transformation is an effective tool for ECG and EEG signal processing, including modelling [5, 7], compression [10], and parameter extraction [1, 2, 3]. The most important advantage of the rational systems, compared to other transformation methods, is its flexibility and adaptivity. Namely, we have arbitrary number of free system parameters at hand. These parameters, that directly define the system, allow the adaptation to the specific signals. This, e.g. in case of ECG and EEG signals, leads to an efficient and advantageous representation: low-dimension approximations of the signals that contain the relevant details, and reflect directly to the morphological behavior of them.

The key issue of the rational transformation is the identification of the system parameters, the so-called inverse poles. In general, pole identification can be addressed as a non-linear optimization problem that can be solved using well-known optimization methods, like Monte-Carlo, gradient methods [11], Hyperbolic Nelder–Mead [7, 12], Hyperbolic Particle Swarm Optimization [9, 10], etc. Although these methods are usually efficient in practice, they can only provide an estimation of the actual optimum. In this paper we discuss some questions about the stability of the inverse poles around the theoretical optimum, i.e. how the perturbation of the inverse poles affects the rational approximation. To this order, we give estimations of the approximation error based on the perturbation of the inverse poles. Our motivations came from the ECG signal processing practice, namely the signal compression [10], heartbeat classification [1], and waveform segmentation problems [2, 3]. In these problems, approximation error plays a direct or indirect role: the utilization of the non-optimal inverse pole combination may affect the rational model curves, the quality of the compression, the accuracy of morphological feature extraction (heartbeat classification) and fiducial point detection (waveform segmentation). However, that practical experiences show that the rational system is only partially sensitive to the perturbation of the absolute values of the inverse poles. In this paper we investigate the acceptable perturbation of these absolute values corresponding to a given level of approximation error. Towards this, we introduce two general approximation problems, and we provide some corresponding formulae and sufficient confidence intervals for the perturbation. The results may have direct impacts in the applications of rational transformation for ECG processing problems. Namely, the quantization and pole identification techniques may be fine-tuned and time-optimized according to the expected level of approximation error.

2. Rational approximation

Let us consider a least squares approximation of complex analytic, square integrable functions on the unit circle (torus) $\mathbb{T}$, i.e. in the $H^2(\mathbb{T})$ Hardy-space along with the usual scalar product

$$\langle F, G\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} F(e^{it})\overline{G(e^{it})}\,dt \qquad \bigl(F, G \in H^2(\mathbb{T})\bigr).$$

We discuss the approximation by means of an orthogonal projection onto specific subspaces of rational functions. To this order, let us define the *basic rational functions* on the closed unit disk $\overline{\mathbb{D}}$ as

$$r_a(z) := \frac{1}{1-\overline{a}z}, \qquad r_{a,k}(z) := r_a^k(z) = \frac{1}{(1-\overline{a}z)^k} \qquad (z \in \overline{\mathbb{D}}),$$

where the free parameter $a \in \mathbb{D}$ is the so-called *inverse pole*, and $k \in \mathbb{N}^+$ is the order of the functions. For a signal processing purpose, it is enough to take the restriction of these functions to the unit circle: $r_{a,k}(e^{it})$ $(t \in [-\pi, \pi))$. The *elementary rational functions* are the linear combinations of basic rational functions having the same inverse pole $a \in \mathbb{D}$, up to a given order $m \in \mathbb{N}^+$ (multiplicity):

$$\sum_{k=1}^{m} c_k \cdot r_a^k \qquad (c_k \in \mathbb{C};\ k = 1, 2, \ldots, m).$$

We remark that every proper rational function can be expressed as the linear combination of basic rational functions, i.e. the set $\mathcal{R}_0$ of proper rational functions is

$$\mathcal{R}_0 = \operatorname{span}\{r_a^k : a \in \mathbb{D}\setminus\{0\};\ k \in \mathbb{N}^+\}.$$

Moreover, $\mathcal{R}_0$ is dense in $H^2(\mathbb{T})$. We consider the orthogonal projection onto the subspace

$$S_{\mathfrak{m}}^{\mathfrak{a}} := \operatorname{span}\{r_{a_j}^{k_j} : j = 1, 2\ldots, n;\ k_j = 1, 2, \ldots, m_j\} \subset H^2(\mathbb{T}),$$

where $n \in \mathbb{N}^+$ is the number of the given (distinct) inverse poles $\mathfrak{a} = (a_1, \ldots, a_n) \in \mathbb{D}^n$, and $\mathfrak{m} = (m_1, \ldots, m_n) \in (\mathbb{N}^+)^n$ is the multiplicities associated with the inverse poles. Since $S_{\mathfrak{m}}^{\mathfrak{a}}$ is a finite-dimensional subspace of the Hilbert-space $H^2(\mathbb{T})$, thus the orthogonal projection always exists. It is natural to take an orthonormal basis in $S_{\mathfrak{m}}^{\mathfrak{a}}$, that can be generated using the Gram–Schmidt orthogonalization method applied to the linearly independent generator set of basic rational functions. The orthonormal basis can be expressed in an explicit way by the Malmquist–Takenaka (MT) functions. To this order, define the (not necessarily distinct) sequence of inverse poles $\mathfrak{b} = (b_0, \ldots, b_{N-1}) \in \mathbb{D}^N$, so that each inverse pole a_j appears m_j times in the sequence, i.e.

$$N := \sum_{j=1}^{n} m_j, \qquad \sum_{\substack{k=0\\ b_k = a_j}}^{N-1} 1 = m_j \qquad (j = 1, 2\ldots, n).$$

Then the MT system generated by $\mathfrak{b}$ is of the form

$$\Phi_k^{\mathfrak{b}}(z) := \frac{\sqrt{1-|b_k|^2}}{1-\overline{b_k}z}\prod_{j=0}^{k-1} B_{b_j}(z) \qquad (z\in\overline{\mathbb{D}};\ k=0,1,\ldots,N-1),$$

where B_a is the Blaschke function of inverse pole $a\in\mathbb{D}$:

$$B_a(z) := \frac{z-a}{1-\overline{a}z} \qquad (z\in\overline{\mathbb{D}}).$$

The MT system above is orthonormal with respect to the scalar product of $H^2(\mathbb{T})$, and spans the subspace $S_{\mathfrak{m}}^{\mathfrak{a}}$. Thus the orthogonal projection of a function $f\in H^2(\mathbb{T})$ can be expressed as an MT-Fourier partial sum:

$$f \approx \sum_{k=0}^{N-1}\langle f,\Phi_k^{\mathfrak{b}}\rangle\Phi_k^{\mathfrak{b}}.$$

3. Approximation problems

In the following, we investigate two approximation problems involving rational systems. In both cases, we consider two subspaces of basic or elementary rational functions: $S_1, S_2 \subset \mathcal{R}_0$, and we investigate the least squares approximation of functions $f\in S_1$ with respect to S_2. This configuration simulates the signal processing approach, where we model the signal (or its parts) by basic or elementary rational functions, but only an estimation of the optimal pole combination is given by an optimization method. Here, S_1 and S_2 represents the subspaces generated by the optimal and the estimated inverse poles, respectively. Later, we discuss the connection between the approximation error and the difference between the optimal and estimated inverse poles, in some special cases. Let us express the approximation error in a relative way, in terms of *percent root difference (PRD)*:

$$PRD(f,g) := \frac{\|f-g\|_2}{\|f\|_2} \qquad \left(f,g\in H^2(\mathbb{T})\right).$$

Here, if $g\in S_2$ is the least squares approximation of $f\in S_1$ with respect to S_2, then $(f-g)$ and g are orthogonal, and

$$PRD^2(f,g) = \frac{\|f-g\|_2^2}{\|f\|_2^2} = \frac{\|f\|_2^2-\|g\|_2^2}{\|f\|_2^2} = 1-\frac{\|g\|_2^2}{\|f\|_2^2}.$$

We remark that we focus on the case when S_1 and S_2 are generated by basic rational functions of the same order or elementary functions with the same multiplicity, but with different inverse poles. These fit the ECG signal processing problems that served as a motivation, where the waveforms of the ECG heartbeats were modelled by elementary rational functions. Another interesting questions could be the combination of multiple inverse poles, and the variation of the multiplicities.

Let us introduce the two approximation problems investigated.

Approximation problem A. *Approximate a signal modelled by a single basic rational function of a given inverse pole and order, by another basic rational function of the same order but different inverse pole.*

Let $a, b \in \mathbb{D} \setminus \{0\}$ be distinct inverse poles, $n \in \mathbb{N}^+$, $c \in \mathbb{C} \setminus \{0\}$, and consider the approximation

$$f(z) = c \cdot r_a^n(z) \approx g(z) = d \cdot r_b^n(z) \qquad (z \in \mathbb{T}),$$

where the coefficient $d \in \mathbb{C}$ is determined by means of least squares:

$$\langle f, r_b^n \rangle = \langle g, r_b^n \rangle \implies c \cdot \langle r_a^n, r_b^n \rangle = d \cdot \langle r_b^n, r_b^n \rangle \implies d = c \cdot \frac{\langle r_a^n, r_b^n \rangle}{\langle r_b^n, r_b^n \rangle}.$$

Then the relative error of the approximation:

$$PRD_{a,b,n}^2 = 1 - \frac{|d|^2 \cdot \langle r_b^n, r_b^n \rangle}{|c|^2 \cdot \langle r_a^n, r_a^n \rangle} = 1 - \frac{|\langle r_a^n, r_b^n \rangle|^2}{\langle r_a^n, r_a^n \rangle \cdot \langle r_b^n, r_b^n \rangle}.$$

Approximation problem B. *Approximate a signal modelled by an elementary rational function of a given inverse pole and multiplicity, by another elementary rational function of the same multiplicity but different inverse pole.*

Let $a, b \in \mathbb{D} \setminus \{0\}$ be distinct inverse poles having the same multiplicity $m \in \mathbb{N}^+$, $c \in \mathbb{C}^m \setminus \{0\}$ and consider the approximation

$$f(z) = \sum_{k=1}^{m} c_k \cdot r_a^k(z) \approx g(z) = \sum_{k=1}^{m} d_k \cdot r_b^k(z) \qquad (z \in \mathbb{T}),$$

where the coefficients $d \in \mathbb{C}^m$ are determined by means of least squares:

$$\langle f, r_b^\ell \rangle = \langle g, r_b^\ell \rangle \implies \sum_{k=1}^{m} c_k \cdot \langle r_a^k, r_b^\ell \rangle = \sum_{k=1}^{m} d_k \cdot \langle r_b^k, r_b^\ell \rangle \ (\ell = 1, 2, \ldots, m) \implies$$
$$\implies G(a, b)c = G(b, b)d \implies d = G^{-1}(b, b)G(a, b)c,$$

where $G : (\mathbb{D} \setminus \{0\})^2 \to \mathbb{C}^{m\times m}$ is the twovariable Gram matrix of the basic rational functions:

$$[G(a,b)]_{\ell,k} := \langle r_a^k, r_b^\ell \rangle \qquad (k,\ell = 1,2,\ldots,m).$$

Here we mention that $G(a,b)$ is a Hermitian matrix. Then the relative error of the approximation:

$$PRD^2_{a,b,m}(c) = 1 - \frac{\langle G(b,b)d,d\rangle}{\langle G(a,a)c,c\rangle} = 1 - \frac{\langle G(a,b)G^{-1}(b,b)G(a,b)c,c\rangle}{\langle G(a,a)c,c\rangle}.$$

4. Scalar products, Gram matrices

The relative approximation error in (A) and (B) is expressed using scalar products of the basic rational functions, and the Gram matrices involving the scalar products. Here we provide formulae for the scalar products with Gaussian hypergeometric, and rational functions, and then we give the Cholesky-decomposition of the Gram matrix, in a special case.

Theorem 1. *Let $a,b \in \mathbb{D}\setminus\{0\}$, $n,m \in \mathbb{N}^+$, then*

$$\langle r_a^n, r_b^m\rangle = {}_2F_1(n,m,1;\overline{a}b),$$

where ${}_2F_1$ is the Gaussian hypergeometric function.

Proof. The series expansion of r_a^n on the unit disk is given as

$$r_a^n(z) = \left(\frac{1}{1-\overline{a}z}\right)^n = \sum_{k=0}^{\infty}\binom{n+k-1}{k}\overline{a}^k z^k \qquad (z \in \overline{\mathbb{D}}).$$

If $a \in \mathbb{D}\setminus\{0\}$, then the series converges absolutely for every $z \in \overline{\mathbb{D}}$, thus the scalar product takes the form

$$\langle r_a^n, r_b^m\rangle = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\binom{n+k-1}{k}\binom{m+\ell-1}{\ell}\overline{a}^k b^\ell \langle z^k, z^\ell\rangle =$$

[based on the ortogonality of the trigonometric system: $\langle z^k, z^\ell\rangle = \delta_{k\ell}$]

$$= \sum_{k=0}^{\infty}\binom{n+k-1}{k}\binom{m+k-1}{k}(\overline{a}b)^k = \sum_{k=0}^{\infty}\frac{(n)_k(m)_k}{(1)_k}\frac{(\overline{a}b)^k}{k!} = {}_2F_1(n,m,1;\overline{a}b),$$

where $(x)_k$ is the (increasing) Pochhammer symbol:

$$(x)_k = x(x+1)\ldots(x+k-1) = k!\cdot\binom{x+k-1}{k} \qquad (x \in \mathbb{R}). \qquad \blacksquare$$

Corollary 1.1. *The Gram matrix function G defined in (B) can actually be considered as a function of one variable, also denoted as G:*

$$[G(a,b)]_{kl} = {}_2F_1(k,l,1;\overline{a}b) \implies G(a,b) =: G(\overline{a}b) \qquad (a,b \in \mathbb{D} \setminus \{0\}).$$

Theorem 2. *Let $|z| < 1$, $n, m \in \mathbb{N}^+$, then*

$${}_2F_1(n,m,1;z) = \frac{P(z)}{(1-z)^{n+m-1}},$$

where P is a polynomial of order $\min\{n-1, m-1\}$. *Specially,*

$${}_2F_1(n,n,1;z) = \frac{\sum_{k=0}^{n-1} \binom{n-1}{k}^2 z^k}{(1-z)^{2n-1}}.$$

Proof. The proof is straitforward based on the Euler-transformation of hypergeometric functions:

$${}_2F_1(n,m,1;z) = (1-z)^{1-n-m} {}_2F_1(1-n,1-m,1,z) \quad (|z| < 1).$$

Here the arguments $(1-n)$ and $(1-m)$ are both nonpositive integers, so the expression ${}_2F_1(1-n,1-m,1,z)$ reduces to a polynomial of order $\min\{n-1, m-1\}$ in the variable z. ■

We note that another way to explore the scalar products is based on the discretization property of the rational system [5]. Namely, we can construct a sampling of the interval $[-\pi, \pi)$ and a proper discrete scalar product that is equivalent to the scalar product of $H^2(\mathbb{T})$ on the subspaces generated by the rational systems. Moreover, depending on $a, b \in \mathbb{D} \setminus \{0\}$, an explicit expression of the discretization points, and the scalar product can be given with the Blaschke functions. Although this concept results the same formulae as above, it is important to express the Gram matrix. Namely, in the sense of the scalar products, we can handle the basic rational functions and the MT functions as finite vectors, sampled on the discretization points above.

Theorem 3. *Let $x \in (0,1)$ and $m \in \mathbb{N}^+$. Then the Cholesky decomposition of Gram matrix $G(x) \in \mathbb{R}^{m \times m}$ is:*

$$G(x) = L(x)L^T(x), \qquad L^T(x) = \frac{1}{\sqrt{1-x}} D(\sqrt{x}) P D\left(\frac{1}{1-x}\right),$$

where the parametric diagonal matrix $D : \mathbb{R} \to \mathbb{R}^{m \times m}$ is:

$$D(x) = diag(1, x, x^2, \ldots, x^{m-1}) \qquad (x \in \mathbb{R}),$$

and $P \in \mathbb{R}^{m\times m}$ is the upper Pascal matrix:

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & \cdots & 1 \\ & 1 & 2 & 3 & 4 & \cdots & m-1 \\ & & 1 & 3 & 6 & \cdots & \binom{m-1}{2} \\ & & & 1 & 4 & \cdots & \binom{m-1}{3} \\ & & & & 1 & \cdots & \binom{m-1}{4} \\ & & & & & \ddots & \vdots \\ & & & & & & 1 \end{bmatrix}.$$

Proof. Let $a \in \mathbb{D} \setminus \{0\}$, and consider the rational system $\{r_a^k : k = 1, \ldots, m\}$, i.e. the MT (or Laguerre) system generated by the pole sequence $\mathfrak{a} = (a, \ldots, a) \in \in \mathbb{D}^m$:

$$\Phi_k^{\mathfrak{a}}(z) = \frac{\sqrt{1-|a|^2}}{1-\overline{a}z} B_a^k(z) \qquad (z \in \mathbb{T};\ k = 0, 1 \ldots, m-1).$$

Based on the connection between the basic rational and Blaschke functions:

$$B_a = \frac{(1-|a|^2)r_a - 1}{\overline{a}},$$

the funtions $\Phi_k^{\mathfrak{a}}$ can be expressed with r_a:

$$\Phi_k^{\mathfrak{a}} = \sqrt{1-|a|^2} r_a B_a^k = \frac{(-1)^k\sqrt{1-|a|^2}}{\overline{a}^k} \sum_{\ell=0}^{k} \binom{k}{\ell} (1-|a|^2)^\ell (-1)^\ell r_a^{\ell+1}.$$

Equivalently, written in a matrix form:

$$[\Phi_0^{\mathfrak{a}}\ \Phi_1^{\mathfrak{a}}\ \cdots\ \Phi_{m-1}^{\mathfrak{a}}] = [r_a\ r_a^2\ \cdots\ r_a^m] \cdot U,$$

where

$$U = \sqrt{1-|a|^2} D(1-|a|^2) EPED\left(\frac{1}{\overline{a}}\right), \quad E = D(-1).$$

Based on the orthogonality of the MT system:

$$I = U^* G(a,a) U \implies G(a,a) = (U^*)^{-1} U^{-1}.$$

Using the $P^{-1} = EPE$ inverse formula of the Pascal matrix:

$$U^{-1} = \frac{1}{\sqrt{1-|a|^2}} D(\overline{a}) PD\left(\frac{1}{1-|a|^2}\right).$$

Finally, combine the term $D(\overline{a})$ in U^{-1}, and the term $(D(\overline{a}))^* = D(a)$ in $(U^*)^{-1} = \left(U^{-1}\right)^*$: $D(a)D(\overline{a}) = D^2(|a|)$. Thus, the Cholesky decomposition of $G(a,a) = G(|a|^2)$ is given by

$$L^T(|a|^2) = \frac{1}{\sqrt{1-|a|^2}} D(|a|) P D\left(\frac{1}{1-|a|^2}\right),$$

where it is easy to verify that $L^T \in \mathbb{R}^{m\times m}$ is an upper triangular matrix with positive diagonal elements. The actual statement is given by the substitution $x = |a|^2 \in (0,1)$. ∎

Corollary 3.1. *If the complex argument of the inverse poles $a, b \in \mathbb{D}\setminus\{0\}$ are the same (especially if $a = b$), then the Cholesky decomposition of the Gram matrix $G(a,b) = G(\overline{a}b) = G(|ab|)$:*

$$G(|ab|) = L(|ab|)L^T(|ab|), \quad L^T(|ab|) = \frac{1}{\sqrt{1-|ab|}} D(\sqrt{|ab|}) P D\left(\frac{1}{1-|ab|}\right).$$

5. Pole stability analysis

In the following, we provide estimations on the approximation error based on the perturbation of the inverse poles as introduced in approximation problem (A) and (B). These estimations lead to sufficient conditions on the perturbation.

In (A) and (B), we introduced the approximation problems in a general way. Now, we discuss restricted cases motivated by signal processing (primarily ECG processing) applications. These applications show that the rational systems seem to be more sensitive to the perturbation of the complex argument than the absolute value of the inverse poles. Theoretical reasons also explain this behavior. Namely, the peak locations of the basic rational functions are defined by the argument of their inverse pole, while the absolute values are accountable for the shapes. Based on the time-localization property of the basic and elementary rational functions, we can conclude that small changes of the arguments can lead to the rapid increase of the approximation error. Meanwhile, as the practical experiences show, the small changes of the absolute values may be compensated by the coefficients of the approximation. Therefore, our interest is to give a confidence interval of the perturbation of the absolute values of the inverse poles, based on the acceptable change of the approximation error. To this order, we restrict the approximation problems to the variations of the absolute values only, i.e. we consider the case when the complex arguments of $a, b \in \mathbb{D} \setminus \{0\}$ are the same. We note that based on Theorem 1, only the product $\overline{a}b$ matters, thus we can assume that the common complex argument is 0, i.e. a and b are real numbers in the interval $(0, 1)$. Considering the absolute values only, it is reasonable to introduce further restrictions on them. ECG signal processing experiences [1, 3] show that neither too small neither too big absolute values are feasible in practice. Namely, if the absolute value is close to 0 or 1, than the corresponding rational function will become too wide or too narrow (i.e. its effective support widens or narrows, respectively) that do not fit the shape of the ECG waveforms. Moreover, as the absolute values tend to 0, the system tends to the trigonometric system, and due to the discretization, numerical uncertainty arise close to 1. Therefore, we restrict the absolute values of the inverse poles to the $[0.5, 1)$ interval that usually meets the practical goals.

Let us state the two estimations and sufficient conditions provided.

Theorem 4. *Let $a \in [0.5, 1)$, $h \in (0, a)$, $b = a - h \in (0, 1)$, $n \in \mathbb{N}^+$, and consider the approximation problem (A). Then*

$$h < \frac{1 - a^2}{\sqrt{4(n-1)^2(1-a^2)^2 + (2n-1)}}\varepsilon$$

is sufficient for $PRD_{a,b,n} < \varepsilon < 1$.

Proof. Based on Theorem 1, since a, b are real numbers:

$$PRD^2_{a,b,n} = 1 - \frac{\langle r_a^n, r_b^n \rangle^2}{\langle r_a^n, r_a^n \rangle \cdot \langle r_b^n, r_b^n \rangle} = 1 - \frac{{}_2F_1(n, n, 1; ab)^2}{{}_2F_1(n, n, 1; a^2) \cdot {}_2F_1(n, n, 1; b^2)} =$$

[using Theorem 2]

$$= 1 - \frac{\left(\sum_{k=0}^{n-1} \binom{n-1}{k}^2 (ab)^k \right)^2}{\left(\sum_{k=0}^{n-1} \binom{n-1}{k}^2 a^{2k} \right) \cdot \left(\sum_{k=0}^{n-1} \binom{n-1}{k}^2 b^{2k} \right)} \cdot \left(\frac{(1-a^2)(1-b^2)}{(1-ab)^2} \right)^{2n-1}.$$

Let Z and W be the numerator and the denominator of the first term, respectively:

$$Z = \left(\sum_{k=0}^{n-1} \underbrace{\binom{n-1}{k}^2}_{=:\alpha_k} (ab)^k \right)^2 = \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \alpha_k \alpha_j a^{k+j} b^{k+j} =$$

$$= \sum_{k=0}^{n-1} \alpha_k^2 a^{2k} b^{2k} + 2 \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j a^{k+j} b^{k+j},$$

$$W = \left(\sum_{k=0}^{n-1} \alpha_k a^{2k} \right) \cdot \left(\sum_{k=0}^{n-1} \alpha_k b^{2k} \right) = \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \alpha_k \alpha_j a^{2k} b^{2j} =$$

$$= \sum_{k=0}^{n-1} \alpha_k^2 a^{2k} b^{2k} + \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j (a^{2k} b^{2j} + a^{2j} b^{2k}).$$

Then

$$W - Z = \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j (a^{2k} b^{2j} - 2a^{k+j} b^{k+j} + a^{2j} b^{2k}) =$$

$$= \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j (a^j b^k - a^k b^j)^2 = \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j a^{2k} b^{2k} (a^{j-k} - b^{j-k})^2 \leq$$

[using $a^m - b^m = (a-b)(a^{m-1} + a^{m-2} b + \cdots + ab^{m-2} + b^{m-1}) \leq$
$\leq h \cdot m \cdot a^{m-1} \leq 2 \cdot h \cdot m \cdot a^m$, with the substitution $m := j - k$]

$$\leq 4h^2 \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j a^{2k} b^{2k} (j-k)^2 a^{2(j-k)} \leq$$

$$\leq 4h^2 (n-1)^2 \sum_{k=0}^{n-2} \sum_{j=k+1}^{n-1} \alpha_k \alpha_j a^{2j} b^{2k} \leq 4h^2 (n-1)^2 W,$$

thus

$$\frac{Z}{W} = \frac{W - (W - Z)}{W} \geq 1 - 4(n-1)^2 h^2.$$

Estimate the second term:

$$\left(\frac{(1-a^2)(1-b^2)}{(1-ab)^2} \right)^{2n-1} = \left(1 - \frac{(a-b)^2}{(1-ab)^2} \right)^{2n-1} \geq$$

[using the Bernoulli inequality with $-(a-b)^2/(1-ab)^2 \geq -1$]

$$\geq 1 - (2n-1) \frac{(a-b)^2}{(1-ab)^2} \geq 1 - \frac{(2n-1)}{(1-a^2)^2} h^2.$$

Then, if h is small enough:

$$PRD_{a,b,n}^2 \leq 1 - \left(1 - 4(n-1)^2 h^2\right) \left(1 - \frac{(2n-1)}{(1-a^2)^2} h^2 \right) =$$

$$= \left(4(n-1)^2 + \frac{(2n-1)}{(1-a^2)^2}\right) h^2 - 4(n-1)^2 \frac{(2n-1)}{(1-a^2)^2} h^4 \leq$$
$$= \left(4(n-1)^2 + \frac{(2n-1)}{(1-a^2)^2}\right) h^2,$$

thus, a sufficient condition for $PRD_{a,b,n} < \varepsilon$:

$$h < \frac{\varepsilon}{\sqrt{4(n-1)^2 + \frac{(2n-1)}{(1-a^2)^2}}} = \frac{1-a^2}{\sqrt{4(n-1)^2(1-a^2)^2 + (2n-1)}}\varepsilon. \qquad \blacksquare$$

Fig. 1. demonstrates the tolerable error rate, i.e. the upper limit to the perturbation compared to the expected level of approximation error:

$$e_n(a) := \frac{1-a^2}{\sqrt{4(n-1)^2(1-a^2)^2 + (2n-1)}}, \qquad h < e_n(a)\varepsilon.$$

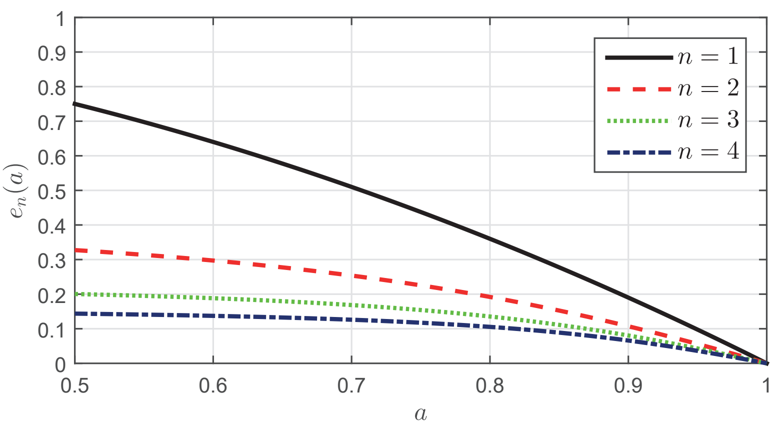

Figure 1. The tolerated error rate depending on the absolute value of the inverse pole, for $n = 1, 2, 3, 4$.

Theorem 5. *Let $a \in [0.5, 1)$, $h \in (0, a)$, $b = a - h \in (0, 1)$, $m \in \mathbb{N}^+$, and consider the approximation problem (B). Then*

$$h < \frac{1-a^2}{\sqrt{3m^2 - 5m + 3}}\varepsilon$$

is suffucient for $PRD_{a,b,m} < \varepsilon < 1$.

Proof. Similarly to Theorem 4, look for an estimation of PRD of the form:

$$PRD^2_{a,b,m} < g(a)h^2 \leq \varepsilon^2$$

with a proper function $g : (0,1) \to \mathbb{R}^+$. The estimation holds for a coefficient $c \in \mathbb{C}^m \setminus \{0\}$, if:

$$PRD^2_{a,b,m}(c) < g(a)h^2 \iff 1 - PRD^2_{a,b,m}(c) > 1 - g(a)h^2 \iff$$

$$\iff \frac{\langle G(ab)G^{-1}(b^2)G(ab)c, c\rangle}{\langle G(a^2)c, c\rangle} > 1 - g(a)h^2 \iff$$

[using that $G(a^2)$ is positive definite, and assuming that $1 - g(a)h^2 > 0$]

$$\iff \left\langle \left(G(ab)G^{-1}(b^2)G(ab) - (1 - g(a)h^2)G(a^2)\right) c, c\right\rangle > 0.$$

Let $c = L^{-T}(a^2)\gamma$ with a proper $\gamma \in \mathbb{C}^m\{\backslash 0\}$, where $L^T(a^2)$ is the Cholesky decomposition of $G(a^2)$, as in Theorem 3. Then the statement, expressed with γ:

$$\left\langle \left(L^{-1}(a^2)G(ab)G^{-1}(b^2)G(ab)L^{-T}(a^2) - (1 - g(a)h^2)I\right) \gamma, \gamma\right\rangle > 0$$

The quadratic form is positive for every $c \in \mathbb{C}^m\{\backslash 0\}$, or equivalently for every $\gamma \in \mathbb{C}^m\{\backslash 0\}$, if and only if the symmetric generator matrix is positive definit, i.e.

$$\lambda_{min}\left(L^{-1}(a^2)G(ab)G^{-1}(b^2)G(ab)L^{-T}(a^2) - (1 - g(a)h^2)I\right) > 0 \iff$$

$$\lambda_{min}\left(L^{-1}(a^2)G(ab)G^{-1}(b^2)G(ab)L^{-T}(a^2)\right) > 1 - g(a)h^2.$$

where λ_{min} denotes the smallest eigenvalue. The matrix on the left can be expressed as AA^T, with

$$A = L^{-1}(a^2)G(ab)L^{-T}(b^2).$$

Here we again note the symmetric positive definite property of $G(a^2)$ and $G(ab)$. Then an equivalent condition can be given using the singular values of A:

$$\lambda_{min}\left(AA^T\right) = \sigma^2_{min}(A) = \frac{1}{\sigma^2_{max}\left(A^{-1}\right)} > 1 - g(a)h^2.$$

Express A^{-1} using the Cholesky decompositions, as in the proof of Theorem 3:

$$A^{-1} = L^T(b^2)G^{-1}(ab)L(a^2) = L^T(b^2)L^{-T}(ab)L^{-1}(ab)L(a^2) =$$

$$= \frac{1-ab}{\sqrt{(1-a^2)(1-b^2)}} D(b)PD\left(\frac{1-ab}{1-b^2}\right)EPD\left(\frac{1}{ab}\right)P^TED\left(\frac{1-ab}{1-a^2}\right)P^TD(a).$$

In order to simplify the expression, notice that the elements of $EPD(x)EP$ ($x \in (0,1)$) can be expressed as:

$$[EPD(x)EP]_{ij} = \binom{j}{i} x^i (1-x)^{j-i} \qquad (i \le j).$$

Specially, with the substitution $x = (1-ab)/(1-b^2)$:

$$\left[EPD\left(\frac{1-ab}{1-b^2}\right)EP\right]_{ij} = \binom{j}{i} \frac{(1-ab)^i (a-b)^{(j-i)} b^{j-i}}{(1-b^2)^j} \qquad (i \le j),$$

and

$$\left[ED(b)PD\left(\frac{1-ab}{1-b^2}\right)EPD\left(\frac{1}{b}\right)\right]_{ij} = \binom{j}{i} \frac{(1-ab)^i (a-b)^{(j-i)}}{(1-b^2)^j} \qquad (i \le j),$$

thus

$$ED(b)PD\left(\frac{1-ab}{1-b^2}\right)EPD\left(\frac{1}{b}\right) = D\left(\frac{1-ab}{a-b}\right)PD\left(\frac{a-b}{1-b^2}\right).$$

Similarly

$$ED\left(\frac{1}{a}\right)P^TD\left(\frac{1-ab}{1-a^2}\right)EP^TD(a) = D\left(\frac{a-b}{1-a^2}\right)P^TD\left(\frac{1-ab}{a-b}\right).$$

Then A^{-1} takes the form:

$$A^{-1} = \frac{1-ab}{\sqrt{(1-a^2)(1-b^2)}} ED\left(\frac{1-ab}{a-b}\right)PD\left(\frac{(a-b)^2}{(1-a^2)(1-b^2)}\right)EP^TD\left(\frac{1-ab}{a-b}\right).$$

Let

$$t = \frac{a-b}{\sqrt{(1-a^2)(1-b^2)}}, \qquad u = \frac{1-ab}{\sqrt{(1-a^2)(1-b^2)}},$$

then $1+t^2 = u^2$, and

$$A^{-1} = uED(u)\underbrace{D\left(\frac{1}{t}\right)PD(t)}_{=:B}E\underbrace{D(t)P^TD\left(\frac{1}{t}\right)}_{=B^T}D(u) = uED(u)BEB^TD(u).$$

The greatest singular value of A^{-1} can be estimated as:

$$\sigma_{max}\left(A^{-1}\right) = \left\|A^{-1}\right\|_2 \le u\|D(u)\|_2^2\|BEB^T\|_2 = u^{2m-1}\|BEB^T\|_2 \le$$

[using $2\sigma_{max}(XY^T) \le \sigma_{max}(X^TX + Y^TY)$ with $X = BE$, $Y = B$, and $\|M\| \le \sqrt{\|M\|_1 \cdot \|M\|_\infty} = \|M\|_\infty$ ($M \in \mathbb{R}^{m\times m}$ symmetric)]

$$\le u^{2m-1} \left\| \frac{EB^TBE + B^TB}{2} \right\|_2 \le u^{2m-1} \left\| \frac{EB^TBE + B^TB}{2} \right\|_\infty .$$

Here we skip the side calculations, and provide only the idea and a sketch. The transformation $M \mapsto EME$ modifies the sign of the elements of the matrix $M \in \mathbb{R}^{m\times m}$ according to a checkerboard pattern. Thus, the transformation $M \mapsto (EME + M)/2$ clears the odd indiced elements of M, according to the checkerboard pattern. Let us discuss the structure of matrix B, B^TB and $EB^TBE + B^TB$. B is a positive upper triangular matrix, and similarly to the Pascal matrix P, the its diagonals consist of the same exponents of t. More precisely, B can be expressed in the matrix exponential form

$$B = e^{tQ}, \qquad Q = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 2 & & \\ & & \ddots & \ddots & \\ & & & 0 & m-1 \\ & & & & 0 \end{bmatrix}.$$

B^TB is positive, and the remaining even indiced elements of $EB^TBE + B^TB$ (according to the checkerboard pattern) contain only even exponents of t. Moreover, its infinity norm is simply the maximum of the row sums, because of its nonnegativity. According to its structure it is easy to see that maximal row sum correspond to the last row, where we can prove the following estimation by induction:

$$\left\| \frac{EB^TBE + B^TB}{2} \right\|_\infty < 1 + \left(\frac{3m^2 - 7m + 4}{2} \right) t^2 + \left(\frac{3m^2 - 7m + 4}{2} \right)^2 t^4 + \cdots \le$$

$$\le \frac{1}{1 - \frac{3m^2-7m+4}{2} t^2},$$

if t is small enough. Then

$$\sigma^2_{max}\left(A^{-1}\right) < \frac{u^{2(2m-1)}}{\left(1 - \frac{3m^2-7m+4}{2} t^2\right)^2} = \frac{(1+t^2)^{2m-1}}{\left(1 - \frac{3m^2-7m+4}{2} t^2\right)^2} <$$

$$< \frac{1}{(1-t^2)^{2m-1}\left(1 - \frac{3m^2-7m+4}{2} t^2\right)^2},$$

and

$$\sigma^2_{min}(A) = \frac{1}{\sigma^2_{max}\left(A^{-1}\right)} > (1-t^2)^{2m-1} \left(1 - \frac{3m^2 - 7m + 4}{2} t^2 \right)^2 \ge$$

[using the Bernoulli inequality]

$$\geq (1-(2m-1)t^2)(1-(3m^2-7m+4)t^2) > 1-(3m^2-5m+3)t^2 \geq$$

$$\geq 1 - \frac{3m^2-5m+3}{(1-a^2)^2}h^2.$$

Thus, a sufficient condition for $PRD_{a,b,m} < \varepsilon$:

$$h < \frac{1-a^2}{\sqrt{3m^2-5m+4}}\varepsilon. \qquad \blacksquare$$

6. Conclusion, future work

We investigated some stability problems related to the rational approximation: the perturbation of the absolute values of the inverse poles, assuming that the other dimensions of the systems are fixed. A linear connection is found between the relative approximation error and the perturbation: we provided a sufficient confident interval for the perturbation depending on the expected approximation level. The results may directly affect the applications of the adaptive rational transformation.

Further research goals related to the results may include the perturbation of the complex argument of the inverse poles, the variation of the order and multiplicities, the perturbation of complex systems with multiple inverse poles, and to provide necessary conditions, as well.

References

[1] **Bognár, G. and S. Fridli,** Heartbeat classification of ECG signals using rational function systems, in: R. Moreno Díaz et al. (Eds.) *Proc. 16th EUROCAST 2017, Part II*, LNCS **10672**, Springer, 2018, 187–195.

[2] **Bognár, G. and F. Schipp,** Geometric interpretation of QRS complexes in ECG signals by rational functions, *Ann. Univ. Sci. Budapest., Sect. Comp.*, **47** (2018), 155–166.

[3] **Bognár, G. and S. Fridli,** ECG segmentation by adaptive rational transform, in: R. Moreno Díaz et al. (Eds.) *Proc. 17th EUROCAST 2019*, LNCS, Springer, (to appear).

[4] **Bokor, J. and F. Schipp,** Approximate identification in Laguerre and Kautz bases, *IFAC Automatica J.*, **34** (1998), 463–468.

[5] **Fridli, S. and F. Schipp,** Biorthogonal systems to rational functions, *Ann. Univ. Sci. Budapest., Sect. Comp.*, **35** (2011), 95–105.

[6] **Fridli, S., L. Lócsi and F. Schipp,** Rational function system in ECG processing, in: R. Moreno Díaz et al. (Eds.) *Proc. 13th EUROCAST 2011, Part I*, LNCS **6927**, Springer, 2012, 88–95.

[7] **Fridli, S., P. Kovács, L. Lócsi and F. Schipp,** Rational Modeling of multi-lead QRS complexes in ECG signals, *Ann. Univ. Sci. Budapest., Sect. Comp.*, **37** (2012), 145–155.

[8] **Heuberger, P.S.C., P.M.J. van den Hof, and B. Wahlberg,** (Eds.) Modelling and identification with rational orthogonal basis functions, Springer-Verlag London Limited, 2005.

[9] **Kovács, P., S. Kiranyaz and M. Gabbouj,** Hyperbolic particle swarm optimization with application in rational identification, in: *Proc. 21st Euro. Sign. Proc. Conf.*, 2013, 1–5.

[10] **Kovács, P.,** *Transformation Methods in Signal Processing*, Ph.D. thesis, ELTE Eötvös Loránd University, Budapest, Hungary, 2016.

[11] **Kovács, P.,** Rational variable projection methods in ECG signal processing, in: R. Moreno Díaz et al. (Eds.) *Proc. 16th EUROCAST 2017, Part II*, LNCS **10672**, Springer, 2018, 196–203.

[12] **Lócsi L.,** A hyperbolic variant of the Nelder–Mead simplex method in low dimensions, *Acta Univ Sapientiae, Mathematica*, **5** (2013), 169–183.

G. Bognár and S. Fridli
Department of Numerical Analysis
Faculty of Informatics
ELTE Eötvös Loránd University
H-1117 Budapest
Pázmány P. sétány 1/C
Hungary
bognargergo@caesar.elte.hu
fridli@inf.elte.hu