

# Cyclic AFD Algorithm for Best Rational Approximation

Tao Qian \*

## Abstract

We propose a practical algorithm of best rational approximation of a given order to a function in the Hardy  $H^2$  space on the unit circle or on the real line. The type approximation is proved to be equivalent with Blaschke form approximation. The algorithm is called Cyclic AFD as it adaptively selects one parameter during each cycle based on the Maximal Selection Principle used in adaptive Fourier decomposition (AFD).

**Key Words** Blaschke Form, Best Rational Approximation, Hardy Space, Rational Orthonormal System, Maximal Selection Principle

## 1 Introduction

For a function  $f$  in the Hardy space  $H^2(\mathbf{D})$ , where  $\mathbf{D}$  stands for the open unit disc, the problem of finding a rational function  $p/q$  of order less than or equal to  $n$  with the least distance to the function  $f$  has a long history. The same question is asked for the Hardy space on a half of the complex plane. The existence of such optimal approximation has been proved by using various methods ([17], [1]). There are also recent proofs based on Blaschke form approximation ([15], [12]). A practical algorithm for finding a solution, however, has been an open problem. Some proofs themselves, such as those based on weak convergence or by contradiction, do not directly imply an algorithm. The existing ones, including [2], [10], are based on the second derivative test, parameterized by the coefficients of the polynomial  $q$  in the denominator. Each of the existing algorithms does converge to critical points. So far there have not been algorithms that convergence to the global minimum point. Starting from different initial status does lead to different local minima. To find a global minimal point is then reduced to find appropriate initial status to start with ([2]). Among all possible initial status finding the optimal ones to start with itself is an NP hard problem ([9]). Yet, one has to show that for each global minimum there is at least one

---

\*Department of Mathematics, University of Macau, Macao (Via Hong Kong). Email: fsttq@umac.mo. The work was partially supported by Research Grant of University of Macau UL017/08-Y3/MAT/QT01/FST, Macao Science and Technology Fund FDCT/014/2008/A1, FDCT/056/2010/A3

initial status corresponding to it under the performed algorithm. This, in the Cyclic AFD Algorithm case, however, is straightforward, based on existence of a solution, because it directly uses the poles as parameters.

There could be two ways to get around from the embarrassing of converging to a local critical point. One is to find a method to effectively examine whether a local critical point is a global minimum point, and, if not, subsequently find the other critical points. Great computational complexity is encountered with initial status selection and calculation of second order partial derivatives. In particular, one has to have a practical algorithm to obtain all the possible critical points for comparison. The other way to get around is to find easy sufficient conditions for Hardy space functions to possess only one critical point. It is regarded as *uniqueness condition* of best rational approximation. If there exists only one critical point, then the critical point has to be the unique global minimum point. Note that critical points include all local minima and maxima and saddle points as well as those at which there are no derivatives. The sufficient condition for uniqueness of critical point would be too strong in practice. Among others Baratchart et al have been devoting to the uniqueness problem. Below we will give an account on their studies in this direction.

The first approach is made in [7] in which they deal with the Hardy space functions of Markov type, that is the Cauchy transform of a positive measure on a segment supported within some absolute bounds. The work [4] develops certain techniques connecting a local minimum with classical interpolation theory that is used to prove asymptotic uniqueness of a certain type on  $L^2$  rational approximation. The criterion developed in [4] is further refined to show that the asymptotic uniqueness holds for Markov functions whose defining measure satisfies precisely the Szegő condition ([6], [5]). Further development along this line is given in [8] in which Markov functions defined by the Cauchy transformation are generalized to Cauchy transforms of complex measures that are absolutely continuous with respect to the equilibrium distributions on a real segment  $[a, b] \subset (-1, 1)$  with the density function being Dini-smooth and non-vanishing. Another type of the uniqueness conditions is given in [1]. It requires the logarithm derivative of the approximated function being bounded inside the unit disc, and non-vanishing. Unfortunately, it is only available for  $n = 1$ . The result of [3] concerns uniqueness of the critical point when approximating a function at its own order.

The present work proposes a new algorithm for best rational approximation via Blaschke forms. On one hand, it converges to a *coordinate-minimum point* (see below) which is local. A global minimum point can be selected from all those based on comparison. On this aspect it does no better than the existing algorithms. It, however, has practical advantages. First, the parameters in the objective function are directly related to the poles of the approximating rational functions. Once an optimal collection of the poles is known, the corresponding best rational approximating function is obtained as orthogonal

projection onto the related orthonormal rational system (or, in an alternative terminology, the Takenaka-Malmquist system). Having poles as parameters of the objective function simplifies the theory and algorithm. The second advantage is that the objective function is explicit, being a sum of  $n$  terms of which each is a product of the polynomial  $(1 - x^2 - y^2)$  and square of the module of an analytic function, being recursively obtained (12). The objective function is symmetric as well as smooth in the  $n$  parameters. The third advantage is that there is no computation of gradients involved: the algorithm is based on recursive formulas and cyclic one-parameter improvement. The improvement is based on Maximal Selection Principle ([13], [14]). The Maximal Selection at the moment is based on comparison of function values that is elementary ([12], [16]). The whole computation does not require a high speed computer. We note that uniqueness conditions on critical point can be applied with our algorithm although our formulation does not involve differentiation.

**Acknowledgement** The author wishes to sincerely thank L. Baratchart, M. Olivi, E. Wegert, W. Mi and W-X. Mai. Baratchart and Olivi generously shared with the author their knowledge and results on this topic during our reciprocal visits. The collaboration with E. Wegert and Mi Wen on existence of best  $n$ -Blaschke form are helpful for the author to develop the cyclic algorithm. Mr Mai generously helped the author to produce all the numerical experiment results. The discussions with him were also helpful to the understanding.

## 2 Preparations

The Hardy space  $H^2(\mathbf{D})$  consists of the holomorphic functions

$$f(z) = \sum_{k=0}^{\infty} c_k z^k, \quad |z| < 1, \quad \sum_{k=0}^{\infty} |c_k|^2 < \infty. \quad (1)$$

Below we will sometimes abbreviate  $H^2(\mathbf{D})$  as  $H^2$ . There are two other equivalent definitions, namely

$$H^2 = \{f : \mathbf{D} \rightarrow \mathbf{C} \mid \text{analytic and } \|f\|^2 = \sup_{0 \leq r < 1} \int_0^{2\pi} |f(re^{it})|^2 dt < \infty\} \quad (2)$$

$$= \{f : \mathbf{D} \rightarrow \mathbf{C} \mid \text{analytic and } M_\alpha f \in L^2(\partial\mathbf{D})\}, \quad (3)$$

where  $M_\alpha f(e^{it})$  is the non-tangential maximal function of  $f$  with respect to the cone at  $e^{it}$  being orthogonal to the circle with any but fixed opening angle  $\alpha < \pi$ . Facilitated with the norm given in (2) the Hardy space is a Hilbert space. Functions in  $H^2(\mathbf{D})$  have non-tangential boundary limits that constitute the space  $H^{2,+}(\partial\mathbf{D})$ , being a proper and closed subspace of  $L^2(\partial\mathbf{D})$ , where the latter is a Hilbert space under the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(e^{it}) \bar{g}(e^{it}) dt. \quad (4)$$

The mapping sending  $f \in H^2(\mathbf{D})$  to its boundary limit in  $H^{2,+}(\partial\mathbf{D})$  is an isometric isomorphism. One similarly defines the Hardy space outside the closed unit disc,  $H^2(\overline{\mathbf{D}}^c)$ , corresponding to the functions

$$f(z) = \sum_{k=-\infty}^{-1} c_k z^k, \quad |z| < 1, \quad \sum_{k=-\infty}^{-1} |c_k|^2 < \infty. \quad (5)$$

The space  $H^2(\overline{\mathbf{D}}^c)$  has alternative definitions analogous with (2) and (3). Functions in  $H^2(\overline{\mathbf{D}}^c)$  also have non-tangential boundary limits that constitute the space  $H^{2,-}(\partial\mathbf{D})$ , being a proper and closed subspace of  $L^2(\partial\mathbf{D})$ . The mapping from  $f \in H^2(\overline{\mathbf{D}}^c)$  to its boundary limit in  $H^{2,-}(\partial\mathbf{D})$  is again an isometric isomorphism. Functions in  $H^2(\overline{\mathbf{D}}^c)$  satisfy the condition  $\lim_{z \rightarrow \infty} f(z) = 0$ , that, in terms of the boundary limit, is equivalent with the condition

$$\int_0^{2\pi} f(e^{it}) dt = 0.$$

The two spaces  $H^{2,+}(\partial\mathbf{D})$  and  $H^{2,-}(\partial\mathbf{D})$  are orthogonal complements to each other in the boundary Hilbert space  $L^2(\partial\mathbf{D})$ .

Now we recall the classical definition of *best  $n$ -rational approximation*. Let  $p$  and  $q$  be polynomials and the zeros of  $q$  are all outside the closed unit disc. We say that  $p/q$  is *non-degenerate* if  $p$  and  $q$  are coprime to each other, meaning that  $p$  and  $q$  have no common divisors, or zeros. The *order of a non-degenerate rational function  $p/q$*  is defined to be  $\max\{\deg(p), \deg(q)\}$ . An  *$n$ -rational approximation to  $f \in H^2(\mathbf{D})$*  is an approximation by a non-degenerate rational function of order **less or equal to  $n$** . We will also use the expressions “exact or precise order  $n$ ” for the cases where the order of the approximating rational function is exactly equal to  $n$ . A *best  $n$ -rational approximation* is an  $n$ -rational approximation  $p_1/q_1$  that satisfies

$$\|f - p_1/q_1\| \leq \|f - p/q\| \quad (6)$$

for all non-degenerate rational functions  $p/q$  of order less than or equal to  $n$ .

Before defining  $n$ -Blaschke form we first introduce  $n$ -dimensional linear space spanned by *Szegő kernels* and their derivatives. Let  $a_1, \dots, a_n$  be a set of  $n$  points in the open unit disc where repeating is allowed. There are two separate cases. One is that none of the  $a_k$ 's are zero. In the case a rational function  $p/q$ , where  $p$  and  $q$  are polynomials, is said to be in  $L(a_1, \dots, a_n)$  if  $p/q$  is a linear combination of the functions in the  $n$ -linearly independent set

$$\frac{1}{1 - \bar{b}_1 z}, \dots, \frac{1}{(1 - \bar{b}_1 z)^{l_1}}, \dots, \frac{1}{1 - \bar{b}_m z}, \dots, \frac{1}{(1 - \bar{a}_m z)^{l_m}}, \quad (7)$$

where  $b_1, \dots, b_m, m \leq n$ , are all the distinguished ones among  $a_1, \dots, a_n$ ,  $l_k$  is the total repeating time of  $b_k$  in  $a_1, \dots, a_n$ , or *the multiple of  $b_k$  in  $(a_1, \dots, a_n)$* ,  $1 \leq k \leq m$ ,  $l_1 + \dots + l_m = n$ . The second case is that one of the  $a_k$ 's is zero. In

the case we say that there are poles at infinity. Without loss of generality, we assume that  $b_1 = 0$ , repeating altogether  $l_1$  times (with the multiple  $l_1$  in the  $n$ -sequence). In the case  $p/q$  is said to be in  $L(a_1, \dots, a_n)$  if it is a linear combination of some functions in

$$1, \dots, z^{l_1-1}, \frac{1}{1 - \bar{b}_2 z}, \dots, \frac{1}{(1 - \bar{b}_2 z)^{l_2}}, \dots, \frac{1}{1 - \bar{b}_m z}, \dots, \frac{1}{(1 - \bar{a}_m z)^{l_m}}, \quad (8)$$

where  $l_1 + \dots + l_m = n$ . In both cases the definitions  $L(a_1, \dots, a_n)$  is irrelevant with the ordering of  $a_1, \dots, a_n$ . We note that both of the two sequences (7) and (8) have poles outside the closed unit disc  $\bar{\mathbf{D}}$ . We say that  $p/q$  is *precisely in*  $L(a_1, \dots, a_n)$  if its linear expansion in  $L(a_1, \dots, a_n)$  has non-zero coefficients with all the highest order terms in the respective list (7) or (8).

Given an  $n$ -tuple  $(a_1, \dots, a_n)$ , the *position multiple* of  $a_k$ , denoted by  $l(a_k)$ , is defined to be the repeating time of  $a_k$  in the  $k$ -tuple  $(a_1, \dots, a_k)$ . In other words, it is the number of all the occurrences of  $a_k$  in the procession from  $a_1$  to  $a_k$ . Below when we say that *an  $n$ -tuple  $(a_1, \dots, a_n)$  is in  $\mathbf{D}$*  we mean that all the  $a_k, k = 1, \dots, n$ , are in  $\mathbf{D}$ .

An  $n$ -tuple  $(a_1, \dots, a_n)$  in  $\mathbf{D}$  corresponds to the following four objects.

1. One of the two sets of the partial fractions given in (7) and (8), depending on whether there is an  $a_k$  being zero.

2. The sequence of the functions  $e_{a_k}(z)$  in  $H^2(\mathbf{D}), k = 1, \dots, n$ , where

$$e_{a_k}(z) = \frac{\sqrt{1 - |a_k|^2}}{1 - \bar{a}_k z}.$$

For  $a \in \mathbf{D}$ ,  $e_a$  is an  $L^2$ -normalized Szegő kernel at  $a$ .

3. The sequence of the functions  $E_{a_k}(z)$  in  $H^2(\mathbf{D}), k = 1, \dots, n$ , where

$$E_{a_k}(z) = \frac{1}{(1 - \bar{a}_k z)^{l(a_k)}},$$

if  $a_k \neq 0$ , and  $l(a_k)$  is the position multiple of  $a_k$ ; or

$$E_{a_k}(z) = z^{l(a_k)-1},$$

if  $a_k = 0$ . Note that for  $l(a_k) > 1$ ,  $E_{a_k}$  is essentially the  $l(a_k)$ -th derivative of the Szegő kernel  $e_{a_k}$ . When all the entries  $a_k$  have multiple 1, the system  $\{E_{a_k}\}_{k=1}^n$  coincides with the system  $\{e_{a_k}\}$ .

4. The  $n$ -tuple of the corresponding  *$n$ -orthonormal rational function system*, or the  *$n$ -Takenaka-Malmquist system* (or the  *$n$ -TM system* in brief),  $B_k, k =$

$1, \dots, n$ , where

$$B_k(z) = B_{a_1, \dots, a_k}(z) = \frac{\sqrt{1 - |a_k|^2}}{1 - \bar{a}_k z} \prod_{j=1}^{k-1} \frac{z - a_j}{1 - \bar{a}_j z}.$$

Each  $B_k$  is called a *modified Blaschke product*. We note that the system  $\{B_l\}_{l=1}^k$  is the result of the Gram-Schmidt orthogonalization process used to the system  $\{E_{a_l}\}_{l=1}^k$  ([14], [15]).

Note that the normalized Szegő kernel  $e_a$  plays an important role in the theory due to the reproducing relation

$$\langle f, e_a \rangle = \sqrt{1 - |a|^2} f(a),$$

proven through a direct application of Cauchy's formula.

For any  $n$ -tuple of complex numbers  $(c_1, \dots, c_n)$  with  $c_n \neq 0$  the function

$$\sum_{k=1}^n c_k B_k$$

is called a *non-degenerate Blaschke form of order  $n$* . Without the assumption  $c \neq 0$  it is called an  $n$ -Blaschke form.

For any function  $f$  in  $H^2$ , due to the orthonormal property of  $\{B_k\}_{k=0}^n$ , the Blaschke form

$$\sum_{k=1}^n \langle f, B_{a_1, \dots, a_k} \rangle B_{a_1, \dots, a_k}$$

is the orthogonal projection of  $f$  onto  $L(a_1, \dots, a_n)$ . The corresponding square-distance from  $f$  to the linear span by the partial fractions is given by

$$A(f; a_1, \dots, a_n) = \|f\|^2 - \sum_{k=1}^n |\langle f, B_k \rangle|^2. \quad (9)$$

$A(f; \dots)$  is the *objective function* in question that is smooth in  $n$  variables  $a_1, \dots, a_n$ . The objective function is, in fact, symmetric in the  $n$  variables: For any permutation  $P$  of the  $n$  variables the value of  $A(f; \dots)$  is invariant, ie.

$$A(f; a_1, \dots, a_n) = A(f; Pa_1, \dots, Pa_n).$$

This is due to the fact that the  $n$ -dimensional linear spaces  $L(a_1, \dots, a_n)$  and  $L(Pa_1, \dots, Pa_n)$  are identical. As the orthogonal projections of  $f$  onto the two spaces,  $\sum_{k=1}^n \langle f, B_{a_1, \dots, a_k} \rangle B_{a_1, \dots, a_k}$  and  $\sum_{k=1}^n \langle f, B_{Pa_1, \dots, Pa_k} \rangle B_{Pa_1, \dots, Pa_k}$  have to be identical, too.

An  $n$ -Blaschke approximation is an approximation by a Blaschke form of order  $n$  or a non-degenerate approximation of order  $m, m \leq n$ . We say that  $(b_1, \dots, b_m)$  gives rise to a *best  $n$ -Blaschke approximation to  $f \in H^2(\mathbf{D})$*  if

$$\|f - \sum_{k=1}^m \langle f, B_{b_1, \dots, b_k} \rangle B_{b_1, \dots, b_k}\| \leq \|f - \sum_{k=1}^l \langle f, B_{a_1, \dots, a_k} \rangle B_{a_1, \dots, a_k}\| \quad (10)$$

for all possible choices of  $l \leq n$  and  $l$ -tuples  $(a_1, \dots, a_l)$  in  $\mathbf{D}$ .

Note that on the both sides of the above inequality the Blaschke forms can be degenerate. The definition amounts to say that a best  $n$ -Blaschke approximation is an approximation by an  $n$ -Blaschke form (degenerate or non-degenerate), that is one of the best approximations among all  $n$  Blaschke approximations (degenerate or non-degenerate). It is proved in [15] and [12] that such minimizer exists and is attainable at an  $m$ -tuple  $(b_1, \dots, b_m)$  in  $\mathbf{D}, m \leq n$ . The analogous result for a half plane, is proved in [12] by a Poisson integral argument.

**Proposition 1** Let  $(a_1, \dots, a_n)$  be an  $n$ -tuple in  $\mathbf{D}$ . Then i) if none of the  $a_k$ 's are zero, then a non-degenerate rational function  $f = p/q$  is precisely in  $L(a_1, \dots, a_n)$  if and only if  $f = \sum_{k=1}^n \langle f, B_k \rangle B_k$  is non-degenerate  $n$ , and if and only if  $p/q$  is a non-degenerate rational function with order  $n$ , where  $q$  is of the degree  $n$  and  $p$  is of a degree less than or equal to  $n - 1$ ; and, ii) if one of the  $a_k$ 's is zero, say,  $a_1 = 0$  with the multiple  $l_1$  in the  $n$ -tuple, then a non-degenerate rational function  $f = p/q$  is precisely in  $L(a_1, \dots, a_n)$  if and only if  $f = \sum_{k=1}^n \langle f, B_k \rangle B_k$  is non-degenerate, and if and only if  $p/q$  is a non-degenerate rational function with order  $n - 1$ , where  $q$  is of the degree  $n - l_1$  and  $p$  is of the degree  $n - 1$ .

**Proof** The assertions can be proved through the partial fraction expansions of  $p/q$  in, respectively, (7) and (8) for the two separate cases.

So, if there are no poles at infinity then the orders of a non-degenerate rational function  $p/q$  as Blaschke form and as rational function are consistent. In the opposite case the order as rational function is one degree less. In the case to obtain an  $n$ -best rational approximation one may need to consider both the  $n$ - and  $n + 1$ -best Blaschke approximations. Given the above mentioned partial fractions formulation in terms of poles the  $n$ -Blaschke approximation seems to be more natural.

Notwithstanding the order difference between best  $n$ -Blaschke and best  $n$ -rational functions, the "order lemma" still holds in the best  $n$ -Blaschke form approximation as proved in ([15], [12]):

An  $n$ -best Blaschke form approximation to a function that is not itself a Blaschke form of degree less than  $n$  is, in fact, of degree precisely  $n$ .

One can similarly define best  $n$ -rational and best  $n$ -Blaschke approximations to functions in the Hardy space outside the closed unit disc,  $H^2(\overline{\mathbf{D}}^c)$ . Note that a non-degenerate Blaschke form of order  $n$ , or a non-degenerate  $n$ -Blaschke, outside the unit disc is a linear combination

$$\sum_{k=1}^n c_k \tilde{B}_k,$$

where  $c_k, k = 1, \dots, n$ , are complex numbers,  $c_n \neq 0$ ,

$$\tilde{B}_k(z) = \tilde{B}_{a_1, \dots, a_k}(z) = \frac{\sqrt{|a_k|^2 - 1}}{\bar{a}_k z - 1} \prod_{j=1}^{k-1} \frac{z - a_j}{\bar{a}_j z - 1},$$

where  $a_j, j = 1, \dots, k$ , are complex numbers outside the closed unit disc. Without the assumption  $c_n \neq 0$  the above linear combination of  $\tilde{B}_k$  is called an  $n$ -Blaschke product outside the unit disc.

It may be easily shown that all the Blaschke forms of order  $n$  outside the disc are identical with all the rational functions of order  $n$  with poles inside the disc. This can also be seen through the roles of the Kelvin inversion defined by

$$Kf(z) = \frac{1}{z} f\left(\frac{1}{z}\right).$$

Simple computations show that the Kelvin inversion maps the functions in  $H^2(\mathbf{D})$  to those in  $H^2(\overline{\mathbf{D}}^c)$ , and vice versa. It maps Blaschke forms inside the disc to Blaschke forms outside the disc of the same degrees, and vice versa. On rational functions, we have the following

**Proposition 2** (i) The Kelvin inversion maps the rational functions  $p/q$  in  $H^2(\mathbf{D})$  of order  $n - 1$  with  $p$  being of degree  $n - 1$ , corresponding to the non-degenerate Blaschke forms of order  $n$  with zero parameters  $a_k$ , to rational functions of order  $n$  in  $H^2(\overline{\mathbf{D}}^c)$ ; and maps the rational functions  $p/q$  in  $H^2(\mathbf{D})$  of order  $n$  with  $q$  being of degree  $n$  and  $p$  being of degree less than  $n$  with poles outside the closed unit disc, corresponding to the non-degenerate Blaschke forms of order  $n$  without zero parameters  $a_k$ , to rational functions of order  $n$  in  $H^2(\overline{\mathbf{D}}^c)$ . (ii) Conversely, the Kelvin inversion maps rational functions of order  $n$  in  $H^2(\overline{\mathbf{D}}^c)$  to one of the two types of rational functions in  $H^2(\mathbf{D})$  of, respectively, the order  $n - 1$  or  $n$ .

The uniform order-correspondence between Blaschke forms and rational functions is based on the fact that a rational function of degree  $n$  with poles inside the disc must satisfy degree of  $p_1 <$  degree of  $q_1 = n$ . The Blaschke form format outside the unit disc does not have two cases regarding some  $a_k$  being zero or not. In spite of the advantage of the outside unit disc setting, we choose to work inside the unit disc for the consistency with our previous work. Some



researchers, however, including Baratchart et al, do work outside the unit disc.

The proposed algorithm works for functions in the Hardy spaces. It also gives rise to rational approximation of fixed order to functions in  $L^2$ . For  $f \in L^2(\partial\mathbf{D})$ , we have the Hardy space decomposition  $f = f^+ + f^-$ , where  $f^+ \in H^2(\mathbf{D})$ , and  $f^- \in H^2(\overline{\mathbf{D}}^c)$ .  $f^\pm$  may be obtained by the corresponding Cauchy integrals, or may be obtained by the Fourier series expansion of  $f$  restricted to the summation non-negative indices and the negative indices, respectively. If  $f$  is real-valued, we have

$$f = 2\operatorname{Re}f^+ - c_0.$$

Then the best  $n$ -rational approximation gives a corresponding approximation result for  $f$ .

We note that the approach given below for  $H^2(\mathbf{D})$  may be adapted to treat the same problem, with necessary changes in accordance with the context, in the Hardy spaces of other domains, including  $\overline{\mathbf{D}}^c$ ,  $\mathbf{C}^+$ ,  $\mathbf{C}^-$ , where the last two notations are for the upper- and lower-half complex planes.

**Definition 1** Assume that  $f \in H^2$  and  $f$  is not an  $m$ -Blaschke form for any  $m < n$ . Then  $a_1, \dots, a_n$  is said to be being a coordinate-minimum point of  $A(f; z_1, \dots, z_n)$  if for any permutation  $P$  whenever fixing  $n-1$  points  $(z_1, \dots, z_{n-1}) = (Pa_1, \dots, Pa_{n-1})$  and performing the Maximal Selection Principle to  $|\langle f_n, e_{z_n} \rangle|$  for the remaining complex variable  $z_n$ , then the missing point  $Pa_n$  is one of the optimal choices for  $z_n$ .

As already mentioned, the proposed algorithm converges to a coordinate-minimum, which is usually a local minimum, but not necessarily to be a global minimum of the objective function. Practically, starting from a large collection of  $n$ -tuples with sufficient density, the algorithm results in a collection of coordinate-minimum points, and a global minimum can be obtained through the comparison between the coordinate-minimum points.

Write the objective function (9) as  $A(f; z_1, \dots, z_n)$ , where  $z_k = x_k + iy_k$ ,  $k = 1, \dots, n$ , are complex variables.  $A(f; z_1, \dots, z_n)$  is a smooth function of  $2n$  real variables. Then  $\{a_1, \dots, a_n\}$  is a critical, or equivalently, a stationary point because of smoothness of the objective function  $A(f; \dots)$ , if and only if

$$\partial_k A(f; a_1, \dots, a_n) = 0, \quad \partial_k = \frac{\partial}{\partial z_k} = \frac{1}{2} \left( \frac{\partial}{\partial x_k} + i \frac{\partial}{\partial y_k} \right), \quad k = 1, \dots, n.$$

Because  $A(f; a_1, a_2, \dots, a_n)$  is real-valued, the above relations are equivalent with

$$\frac{\partial_k}{\partial x_k} A(f; a_1, a_2, \dots, a_n) = 0, \quad \frac{\partial_k}{\partial y_k} A(f; a_1, a_2, \dots, a_n) = 0.$$

In [15] and [12] we show that for any  $f \in H^2$  and any fixed  $n$  one can find  $m$  numbers,  $a_1, \dots, a_m$ , in  $\mathbf{D}$ ,  $m \leq n$ , repeating is allowed, such that at this  $m$ -tuple

(9) reaches a global minimum out of all possible choices of  $l$ -tuples in  $\mathbf{D}$ ,  $l \leq n$ , where repeating is allowed. When  $m < n$ ,  $f$  itself is an  $m$ -Blaschke form; and otherwise  $m = n$ .

In [15] the inner products  $\langle f, B_k \rangle$  are written in the alternative way  $\langle f_k, e_{a_k} \rangle$ , where the Hardy space function, the  $k$ -th *reduced remainder*  $f_k$ , which is different from the *standard remainder* (see (15)), is given by the recursive formula

$$f_k(z) = \frac{f_{k-1}(z) - \langle f_{k-1}, e_{a_{k-1}} \rangle e_{a_{k-1}}(z)}{\frac{z - a_{k-1}}{1 - \bar{a}_{k-1}z}}, \quad f_1 = f, \quad (11)$$

which we call *the generalized backward shift of  $f_{k-1}$  through  $a_{k-1}$* . When  $a_{k-1} = 0$ , this reduces to the classical backward shift. With the replacement of  $f$  by  $f_k$ , the objective function has an alternative form

$$A(f; a_1, \dots, a_n) = \|f\|^2 - \sum_{k=1}^n (1 - |a_k|^2) |f_k(a_k)|^2. \quad (12)$$

**Remark** By using mathematical induction we can show that in  $H^2$  the functions  $f_k$ 's are continuously dependent on the parameters  $a_1, \dots, a_{k-1}$ .

The expression (12) was first used in [14] and [13], in which *adaptive Fourier decomposition (AFD)* and *unwinding AFD* are proposed. The AFD algorithm depends on a *Maximal Selection Principle*: For any  $f \in H^2$ ,  $b = \arg \max\{|\langle f, e_a \rangle| : a \in \mathbf{D}\}$  is attainable inside  $\mathbf{D}$  (see [14] or [13]). To proceed AFD starting from a given signal  $f \in H^2$ , each time when having selected  $a_1, \dots, a_{k-1}$  an adaptive selection of  $a_k$  is an application of Maximal Selection Principle to  $f_k$ . It is a consecutive selection process of the parameters  $a_1, \dots, a_n$ . To solve the best  $n$ -Blaschke approximation problem, or, equivalently, the best  $n$ -rational approximation problem, one is reduced to select all the parameters  $a_1, \dots, a_n$  inside  $\mathbf{D}$  at one time to give rise to the global minimum value of (12). As mentioned, the work [15] and [12] show that such simultaneous selections exist.

### 3 The Algorithm

For  $n = 1, 2, 3$ , optimal selections of  $n$ -tuples of parameters to minimize the objective function given by (9) or (12) can be done by comparing the values of the objective function on a collection of  $n$ -tuples with sufficient density ([12]). When  $n$  gets large, this comparison function value method cannot be applied because of the exponential increasing of the cardinal number of a dense enough set of  $n$ -tuples. The proposed Cyclic AFD Algorithm is based on the AFD algorithm that cyclically improves elements of an  $n$ -tuple. The process is based on the formulation and techniques developed in [14], [13], [12] and [15].

In the AFD algorithm we repeat the following procedure: Along with choosing  $a_1, \dots, a_{k-1}$  in  $\mathbf{D}$ , we produce the reduced remainders  $f_2, \dots, f_k$ . Then to  $f_k$  we apply the Maximal Selection Principle to find an  $a_k$  that giving rise to  $\max\{|\langle f_k, e_a \rangle| : a \in \mathbf{D}\}$ . The Cyclic AFD Algorithm repeats such procedure for  $k = n$ : Whenever  $a_1, \dots, a_{n-1}$  are fixed from previous steps we inductively obtain the reduced remainders  $f_2, \dots, f_n$ , and then use the Maximal Selection Principle to select an optimal  $a_n$ .

Denote by *LMP* a local minimum points, by *CMP* a coordinate-minimum point, and *CP* a critical point of the objective function. Denote by  $\mathcal{LM}, \mathcal{CM}$  and  $\mathcal{C}$  the sets, of, respectively, all LMPs, CMPs and Cs. Then we have the following inclusion relations.

**Proposition 3**

$$\mathcal{LM} \subset \mathcal{CM} \subset \mathcal{C}. \quad (13)$$

**Proof** The first inclusion is obvious. Now show the second. Since the objective function is symmetric with respect to permutations, the assertion of coordinate-minimum implies

$$\partial_z A(f; z, a_2, \dots, a_n)|_{z=a_1} = 0, \quad \partial_z = \frac{1}{2} \left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right).$$

We therefore have

$$\frac{\partial A(f; \cdot, a_2, \dots, a_n)}{\partial x_1} \Big|_{z=a_1} = 0, \quad \frac{\partial A(f; \cdot, a_2, \dots, a_n)}{\partial y_1} \Big|_{z=a_1} = 0.$$

Thus  $(a_1, \dots, a_n)$  is a stationary point of  $A(f; z_1, \dots, z_n)$ . The proof is complete.

**Corollary 4** If the objective function has only one CP, then there is only one CMP that gives rise to the global minimum of the objective function.

**Theorem 5** Suppose that  $f$  is not an  $m$ -Blaschke form for any  $m < n$ . Let  $s_0 = \{b_1^{(0)}, \dots, b_n^{(0)}\}$  be any  $n$ -tuple of parameters inside  $\mathbf{D}$ . Fix some  $n - 1$  parameters of  $s_0$  and make an optimal selection of the single remaining parameter according to the Maximum Selection Principle. Denote the obtained new  $n$ -tuple of parameters by  $s_1$ . We repeat this process and make cyclic optimal selections over the  $n$  parameters. We thus obtain a sequence of  $n$ -tuples  $s_0, s_1, \dots, s_l, \dots$ , with decreasing objective function values  $d_l$  that tend to a limit  $d \geq 0$ , where, in the notation and formulation of (12),

$$d_l = A(f; b_1^{(l)}, \dots, b_n^{(l)}) = \|f\|^2 - \sum_{k=1}^n (1 - |b_k^{(l)}|^2) |f_k^{(l)}(b_k^{(l)})|^2. \quad (14)$$

Then, (i) If  $\bar{s}$ , as an  $n$ -tuple, is a limit of a subsequence of  $\{s_l\}_{l=0}^\infty$ , then  $\bar{s}$  is in  $\mathbf{D}$ ; (ii)  $\bar{s}$  is a CMP of  $A(f; \dots)$ ; (iii) If the correspondence between a CMP and

the corresponding value of  $A(f; \dots)$  is one to one, then the sequence  $\{s_l\}_{l=0}^\infty$  itself converges to the CMP, being dependent of the initial  $n$ -tuple  $s_0$ ; (iv) If  $A(f; \dots)$  has only one CMP, then  $\{s_l\}_{l=0}^\infty$  converges to a limit  $\bar{s}$  in  $\mathbf{D}$  at which  $A(f; \dots)$  attains its global minimum value.

**Proof** (i) The proof uses a Poisson kernel argument. We have noted that for any  $n$ -tuple  $\{b_1, \dots, b_n\}$  in  $\mathbf{D}$ , although the system  $\{B_{b_1, \dots, b_k}\}_{k=1}^n$  is relevant to the ordering of the  $b_k$ 's in the  $n$ -tuple, the projection

$$\sum_{k=1}^n \langle f, B_{b_1, \dots, b_k} \rangle B_{b_1, \dots, b_k}$$

is irrelevant to it. Now we show that if an  $n$ -tuple  $\bar{s}$  is the limit of some subsequence of  $s_0, s_1, \dots, s_l, \dots$ , then  $\bar{s}$  is in  $\mathbf{D}$ . We show this by introducing a contradiction. If  $d = 0$ , then the existence result of [15] or [12] shows that the function  $f$  itself is the solution of the best  $n$ -rational approximation. In particular,  $f$  is a rational function with  $n$  poles of which all are in  $\mathbf{D}$ . Next we assume  $d > 0$ . Suppose that there exists a sequence of  $n$ -tuples  $\{b_1^{(l)}, \dots, b_n^{(l)}\}$  such that the corresponding objective function values decreasingly tend to  $d > 0$  as  $l \rightarrow \infty$ , while for at least one index  $k_0$  there exists a subsequence  $\{b_{k_0}^{(l_j)}\}_{j=1}^\infty$  converging to a boundary point. We divide the indices  $k = 1, \dots, n$  into two groups, denoted by the letters  $\mathcal{B}$  and  $\mathcal{I}$ , respectively, where if  $\{b_k^{(l)}\}_{l=1}^\infty$  contains a subsequence converging to a boundary point of  $\mathbf{D}$ , then  $k \in \mathcal{B}$ , and otherwise  $k \in \mathcal{I}$ . Note that  $k_0 \in \mathcal{B} \neq \emptyset$ . Due to the observation made at the beginning of the proof, we may alter the ordering of  $\{b_1^{(l)}, \dots, b_n^{(l)}\}$ , if necessary, and may assume that the indices in  $\mathcal{I}$  are all smaller than those in  $\mathcal{B}$ . By a diagonal process we can choose a subsequence  $l_j \rightarrow \infty$  such that for  $k \in \mathcal{B}$  the sequences  $\{b_k^{(l_j)}\}$  converge to boundary points, and for  $k \in \mathcal{I}$ , converge to interior points. Without loss of generality we may assume that the original sequence  $\{l\}_{l=1}^\infty$  has such property. That is, as  $l \rightarrow \infty$ , the sequence  $\{b_k^{(l)}\}_{l=1}^\infty$  converges to  $b_k \in \partial\mathbf{D}$  if  $k \in \mathcal{B}$ , and converges to  $b_k \in \mathbf{D}$  if  $k \in \mathcal{I}$ . For each  $l$  adopt the notation for the non-zero *standard remainders*

$$\tilde{f}_m^{(l)} = f - \sum_{k=1}^{m-1} \langle f_j^{(l)}, e_{b_j^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_j^{(l)}}, \quad m = 1, \dots, n. \quad (15)$$

There holds

$$\tilde{f}_m^{(l)}(z) = f_m^{(l)}(z) \prod_{j=1}^{m-1} \frac{z - b_j^{(l)}}{1 - \bar{b}_j^{(l)} z}, \quad (16)$$

where  $f_m^{(l)}$  is associated with  $(b_1^{(l)}, \dots, b_{m-1}^{(l)})$ , recursively defined through the generalized backward shift operator (11). We also use the notation

$$R_{\mathcal{I}}^{(l)} = f - \sum_{j \in \mathcal{I}} \langle f_j^{(l)}, e_{b_j^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_j^{(l)}}.$$

There follows

$$\|R_{\mathcal{I}}^{(l)}\|^2 = \|f\|^2 - \sum_{j \in \mathcal{I}} |\langle f_j^{(l)}, e_{b_j^{(l)}} \rangle|^2 \geq \|\tilde{f}_k^{(l)}\|^2 = \|f_k^{(l)}\|^2, \quad k \in \mathcal{B}. \quad (17)$$

We have, owing the properties of the Poisson kernel  $P_r$  and the inequality (17), for any given  $\epsilon > 0$ ,

$$\begin{aligned} \|R_{\mathcal{I}}^{(l)}\| &\geq \|R_{\mathcal{I}}^{(l)} - \sum_{k \in \mathcal{B}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}}\| \\ &\geq \|P_r * (R_{\mathcal{I}}^{(l)} - \sum_{k \in \mathcal{B}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}})\| \\ &\geq \|P_r * R_{\mathcal{I}}^{(l)}\| - \sum_{k \in \mathcal{B}} \|f_k^{(l)}\| \|P_r * B_{b_1^{(l)}, \dots, b_k^{(l)}}\| \\ &\geq \|P_r * R_{\mathcal{I}}^{(l)}\| - \sum_{k \in \mathcal{B}} \|R_{\mathcal{I}}^{(l)}\| \|P_r * B_{b_1^{(l)}, \dots, b_k^{(l)}}\| \\ &\geq (1 - \frac{\epsilon}{2}) \|R_{\mathcal{I}}^{(l)}\| - \|R_{\mathcal{I}}^{(l)}\| \sum_{k \in \mathcal{B}} \|P_r * B_{b_1^{(l)}, \dots, b_k^{(l)}}\|, \end{aligned}$$

if  $r$  is sufficiently close to 1. Let  $r$  be such fixed.

Now, since  $B_{b_1^{(l)}, \dots, b_k^{(l)}} \in H^\infty(\mathbf{D})$ , we have (Corollary 3.2, p58, [11])

$$P_r * B_{b_1^{(l)}, \dots, b_k^{(l)}}(e^{it}) = B_{b_1^{(l)}, \dots, b_k^{(l)}}(re^{it}),$$

and thus

$$\begin{aligned} \|P_r * B_{b_1^{(l)}, \dots, b_k^{(l)}}\|^2 &= \|B_{b_1^{(l)}, \dots, b_k^{(l)}}(re^{i(\cdot)})\|^2 \\ &\leq \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |b_k^{(l)}|^2}{|1 - \bar{b}_k^{(l)} re^{it}|^2} dt \\ &= \frac{1 - |b_k^{(l)}|^2}{1 - r^2 |b_k^{(l)}|^2}. \end{aligned}$$

For the fixed  $r$ , since  $\lim_{l \rightarrow \infty} |b_k^{(l)}| = 1, k \in \mathcal{B}$ , we can choose  $l$  large enough so that

$$\|P_r * B_{b_1^{(l)}, \dots, b_k^{(l)}}\| \leq \frac{\epsilon}{2n}$$

hold for all  $k \in \mathcal{B}$ . Therefore, for such  $l$ ,

$$\|R_{\mathcal{I}}^{(l)}\| \geq \|R_{\mathcal{I}}^{(l)} - \sum_{k \in \mathcal{B}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}}\| \geq (1 - \epsilon) \|R_{\mathcal{I}}^{(l)}\|.$$

This shows that

$$\lim_{l \rightarrow \infty} \left\| \sum_{k \in \mathcal{B}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}} \right\| = 0. \quad (18)$$

Since the modified Blaschke products are orthogonal to each other, for any subset  $\mathcal{B}' \subset \mathcal{B}$  the above limit also holds. As consequence of (18),

$$\begin{aligned}
\lim_{l \rightarrow \infty} \|\tilde{f}_n^{(l)}\|^2 &= \lim_{l \rightarrow \infty} \left\| f - \sum_{k \in \mathcal{I}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}} \right\|^2 \\
&= \lim_{l \rightarrow \infty} \left( \|f\|^2 - \sum_{k \in \mathcal{I}} |\langle f_k^{(l)}, e_{b_k^{(l)}} \rangle|^2 \right) \\
&= \|f\|^2 - \sum_{k \in \mathcal{I}} |\langle f, B_{b_1, \dots, b_k} \rangle|^2 \\
&= d > 0.
\end{aligned} \tag{19}$$

This last relation (19), however, shows that the selections of  $b_k^{(l)}$ 's and their corresponding limits for  $k \in \mathcal{B}$  all have no contribution and the coordinate-minimum value  $d$  can be attained at an  $m$ -tuple inside the unit disc, where  $m < n$ .

Let  $m_0$  be the first index of  $\mathcal{B}$ , and  $\mathcal{B}' = \mathcal{B} \setminus \{m_0\}$ . For a large  $l$ , we have

$$\begin{aligned}
d_{l+1} &= \left\| f - \sum_{k=1}^n \langle f_k^{(l+1)}, e_{b_k^{(l+1)}} \rangle B_{b_1^{(l+1)}, \dots, b_k^{(l+1)}} \right\|^2 \\
&\leq \min_{b \in \mathbf{D}} \left\| f - \sum_{k \in \mathcal{I}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}} - \langle f_{m_0}^{(l)}, e_b \rangle B_{a_1^{(l)}, \dots, a_{m_0-1}^{(l)}, b} \right. \\
&\quad \left. - \sum_{k \in \mathcal{B}'} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}} \right\|^2 \\
&= \left\| f - \sum_{k \in \mathcal{I}} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}} \right\|^2 - \max_{b \in \mathbf{D}} |\langle f_{m_0}^{(l)}, e_b \rangle|^2 - \left\| \sum_{k \in \mathcal{B}'} \langle f_k^{(l)}, e_{b_k^{(l)}} \rangle B_{b_1^{(l)}, \dots, b_k^{(l)}} \right\|^2 \\
&\leq d + \epsilon - \alpha,
\end{aligned}$$

where  $f_{m_0}$  is recursively defined through the generalized backward shift operator depending on the parameters  $b_j, j \in \mathcal{I}$ , and we note that since  $f_{m_0}$  is a non-zero function in the Hardy space,

$$\alpha = \max_{b \in \mathbf{D}} |\langle f_{m_0}, e_b \rangle|^2 > 0,$$

where the error term  $\epsilon$  is collected from the limit (19), the limit (18) for  $\mathcal{B}'$ , and from the estimate

$$\begin{aligned}
\max_{b \in \mathbf{D}} |\langle f_{m_0}^{(l)}, e_b \rangle|^2 &= \max_{b \in \mathbf{D}} |\langle (f_{m_0}^{(l)} - f_{m_0}), e_b \rangle + \langle f_{m_0}, e_b \rangle|^2 \\
&\geq \max_{b \in \mathbf{D}} (|\langle f_{m_0}, e_b \rangle| - \|f_{m_0}^{(l)} - f_{m_0}\|)^2 \\
&\geq \alpha - \epsilon',
\end{aligned}$$

where  $\epsilon'$  is small depending on  $l$  the parameters  $b_k, b_k^{(l)}, k \in \mathcal{I}$ , that are all inside  $\mathbf{D}$  (See the Remark below (12)). For large  $l$  we have  $\epsilon < \alpha$  that implies  $d_{l+1} < d$ ,

being a contradiction. (ii) If  $\bar{s}$  is not a CMP, then fix some  $n - 1$  entries of it the exceptional one,  $b_{k_0}$ , has a positive distance from what is selected, say  $b'_{k_0}$ , according to Maximal Selection Principle. Now examine  $s_l$  in the subsequence tending to  $\bar{s}$  whose  $n - 1$  entries are respectively close to the fixed  $n - 1$  entries and the exceptional one  $b_{k_0}^{(l)}$  is to be replaced in the next  $n$ -tuple in the subsequence. On one side, the replacing entry  $b_{k_0}^{(l+1)}$  in the next  $n$ -tuple in the subsequence should be close to  $b'_{k_0}$  according to Maximal Selection principle. On the other hand, because of convergence of the subsequence,  $b_{k_0}^{(l+1)}$  should be close to  $b_{k_0}$ . This is a contradiction. (iii) is obvious as, in the case, a sequence without limit contains at least two subsequences tending to different CMPs, contradictory to the one to one assumption. (iv) Let now  $A(f; \dots)$  have only one CMP. The same reasoning as in (iii) concludes that starting from any initial  $n$ -tuple the sequence  $\{s_l\}_{l=0}^{\infty}$  converges to the same limit  $\bar{s}$  in  $\mathbf{D}$  at which  $A(f; \dots)$  attains its global minimum value. The proof is complete.

**Definition 2** The algorithm described in Theorem 1 is called *Cyclic AFD Algorithm*.

**Corollary 6** If the objective function has only one critical point, then Cyclic AFD Algorithm starting from any initial  $n$ -tuple generates a sequence of  $n$ -tuples converging to the unique CMP  $n$ -tuple giving rise to the global minimum of the objective function.

**Algorithm** Below we give a step by step description of the algorithm. It is a process to produce a sequence of  $n$ -tuples inside  $\mathbf{D}$  whose subsequence-limits are CMPs. If there is only one critical point or only one CMP, then the sequence itself converges to the unique CMP giving rise to the global minimum value of  $A(f; \dots)$ .

Step 1. We start from an initial  $n$ -tuple  $s_0 = (b_1^{(0)}, \dots, b_n^{(0)})$  that can be obtained, for instance, from one of the following three methods. (i) Use AFD to create  $b_1^{(0)}$ , then  $b_2^{(0)}, \dots$ , and  $b_n^{(0)}$ ; (ii) In an equally distributing pattern pick up  $n$  points in  $\mathbf{D}$ ; or (iii) Randomly pick up  $n$  points in  $\mathbf{D}$ .

Step 2. Fix  $b_2^{(0)}, \dots, b_n^{(0)}$ . Under any but convenient order of the  $n - 1$  points renamed as  $(a'_1, \dots, a'_n)$  compute

$$f(z) - \sum_{k=1}^{n-1} \langle f_k, e_{a'_k} \rangle B_k(z),$$

where  $f_1 = f$ , and  $f_k, k = 2, \dots, n - 1$ , is the generalized backward shift of  $f_{k-1}$  through  $a'_{k-1}$ , and  $B_k$  is the modified Blaschke product corresponding to

$(a'_1, \dots, a'_k)$ . From [15] or [13] we know

$$f(z) - \sum_{k=1}^{n-1} \langle f_k, e_{a'_k} \rangle B_k(z) = f_n(z) \prod_{k=1}^{n-1} \frac{z - a'_k}{1 - a'_k z}.$$

Step 3. Using Maximal Selection Principle to  $f_n$  we obtain  $b_1^{(1)}$ . We then form a new  $n$ -tuple  $s_1 = (b_1^{(1)}, b_2^{(1)}, \dots, b_n^{(1)})$ , where  $(b_2^{(1)}, \dots, b_n^{(1)}) = (b_2^{(0)}, \dots, b_n^{(0)})$ . With this new  $n$ -tuple we have the improvement  $A(f; b_1^{(1)}, b_2^{(1)}, \dots, b_n^{(1)}) \leq A(f; b_1^{(0)}, \dots, b_n^{(0)})$ .

Step 4. Keeping  $b_1^{(1)}$  and  $b_3^{(1)}, \dots, b_n^{(1)}$  unchanged in there positions, we proceed to replace  $b_2^{(1)}$  by  $b_2^{(2)}$  in virtue of Maximal Selection Principle to get the improvement  $A(f; b_1^{(2)}, b_2^{(2)}, \dots, b_n^{(2)}) \leq A(f; b_1^{(1)}, \dots, b_n^{(1)})$ . We set  $s_2 = (b_1^{(2)}, b_2^{(2)}, \dots, b_n^{(2)})$ , where  $b_1^{(2)} = b_1^{(1)}$  and  $(b_3^{(2)}, \dots, b_n^{(2)}) = (b_3^{(1)}, \dots, b_n^{(1)})$ .

We carry on this process to the time  $l$  when within a threshold none of the continuous  $n$  replacements of the  $b_k^{(l)}$  can improve the value of  $A(f; \dots)$ . Then  $(b_1^{(l)}, \dots, b_n^{(l)})$  is a CMP, and  $A(f; b_1^{(l)}, \dots, b_n^{(l)})$  is the value of a critical point that has to be the global minimum under the one-critical-point assumption.

In practice such cyclic process of entry optimization for an  $n$ -tuple keeps going and improves the approximation, and the sequence of the values  $A(f; \dots)$  has the global minimum value as its limit, and, in the mean time, the sequence of the arguments  $(b_1^{(j)}, \dots, b_n^{(j)})$  has a CMP limit  $(b_1, \dots, b_n)$  in  $\mathbf{D}$  giving rise to the global minimum. If there are more than one critical points, then the process leads to CMPs and therefore CPs as subsequence limits. The process can end up with different groups of critical points, depending on the initial  $n$ -tuple  $s_0$  to start with. If the approximated function  $f$  itself is an  $m$ -Blaschke form,  $m < n$ , then all subsequence limits  $\bar{s}$  have  $n - m$  components resting on the boundary no matter the process starting from which initial  $n$ -tuple  $s_0$ . The codes of Cyclic AFD Algorithm are available through the link <http://www.fst.umac.mo/en/staff/fsttq.html>.

## 4 The Experiments

We present three examples. All experiments are done on a PC with Intel(R) Core(TM) i7 CPU 860 @ 2.80GHz 2.79GHz and 4.00 GB memory under windows 7(64 bit) used Matlab 7.11.1(r2010b) service pack 1.

**Experiment 1** The approximation function to be approximated is a 4-Blaschke form

$$f(z) = \sum_{k=1}^4 c_k B_{b_1, b_k}(z)$$



with the poles  $(b_1, b_2, b_3, b_4) = (0.6800 + 0.5200i, 0.3900 + 0.8100i, -0.1300 - 0.8700i, 0.5500 - 0.1000i)$  and the coefficients  $(c_1, c_2, c_3, c_4) = (0.1440 + 0.5197i, -1.6387 - 0.0142i, -0.7601 - 1.1555i, -0.8188 - 0.0095i)$  For Cyclic AFD Algorithm the initial 4-tuple is randomly chosen as  $s_0 = (a_1^{(0)}, a_2^{(0)}, a_3^{(0)}, a_4^{(0)}) = (-0.6444 + 0.6790i, -0.1194 + 0.4501i, -0.6619 + 0.3502i, 0.4162 + 0.8329i)$ .

For Cyclic AFD Algorithm the running time is 4.966592 seconds with the result  $(0.3900 + 0.8100i, -1300 - 0.8700i, 0.6800 + 0.5200i, 0.5500 - 0.1000i)$

For RARL2 Algorithm (INRIA, France) the running time is 5.685793 seconds with the result  $(-0.1470 - 0.8330i, 0.6572 + 0.2798i, 0.4212 + 0.9055i, 0.4329 + 0.7721i)$ .

In the experiments the domain of the function, viz., the unit circle, is discretized by using 1024 points.

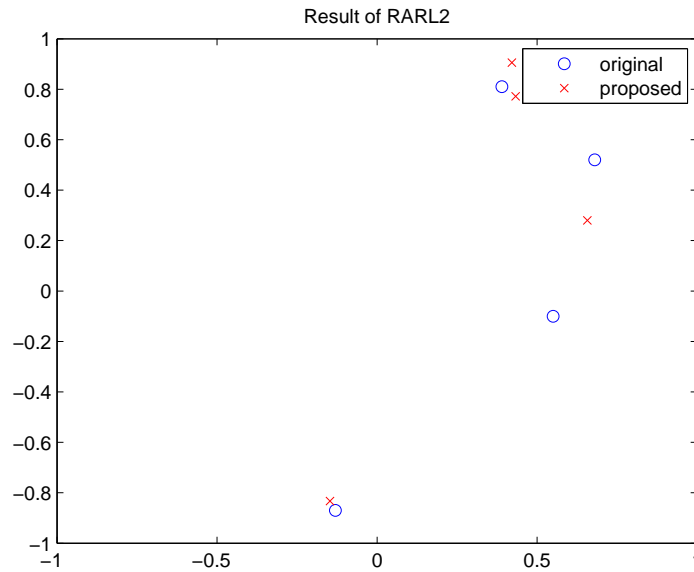


Figure 1: Experiment Result of The First Case of Example 1

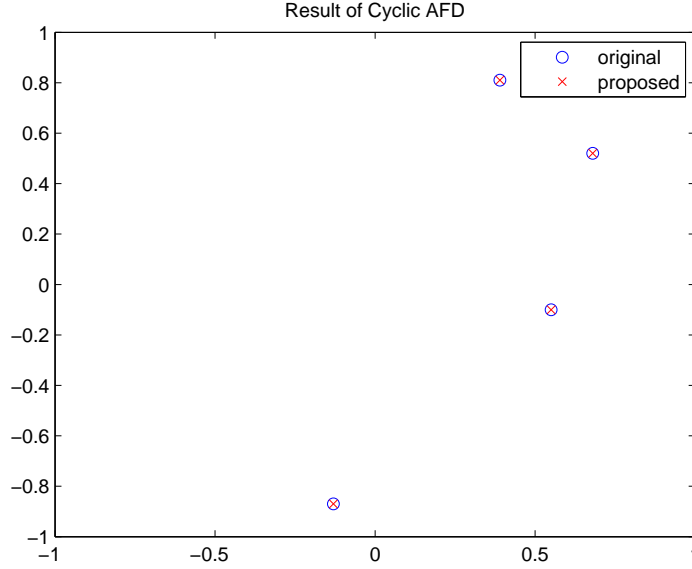


Figure 2: Experiment Result of The First Case of Example 1

**Experiment 2** The function to be approximated is a 4-Blaschke form

$$f(z) = \sum_{k=1}^4 c_k B_{b_1, b_k}(z)$$

with the poles  $(b_1, b_2, b_3, b_4) = (-0.4900 - 0.8000i, 0.3100 + 0.1400i, -0.9400 - 0.2900i, 0.2300 - 0.6900i)$  and the coefficients  $(c_1, c_2, c_3, c_4) = (1.0470 + 0.55587i, -0.2269 - 1.1203i, -0.1625 - 1.5327i, 0.6901 - 1.0979i)$ . To perform Cyclic AFD Algorithm the initial 4-tuple is randomly chosen to be  $s_0 = (a_1^{(0)}, a_2^{(0)}, a_3^{(0)}, a_4^{(0)}) = (0.0938 + 0.5303i, 0.4849 - 0.6714i, 0.5201 - 0.1500i, 0.2621 - 0.7549i)$ .

The running time of Cyclic AFD Algorithm is 8.711609 seconds with the result  $(-0.9400 - 0.2900i, -0.4900 - 0.8000i, 0.2300 - 0.6900i, 0.3100 + 0.1400i)$ .

The running time of RARL2 is 1.382164 seconds with the result  $(-0.9400 - 0.2900i, -0.4900 - 0.8000i, 0.2300 - 0.6900i, 0.3100 + 0.1400i)$ .

In Experiment 2 the function domain  $z = e^{it}$ ,  $t \in [0, 2\pi]$ , is discretized by using 2048 points.

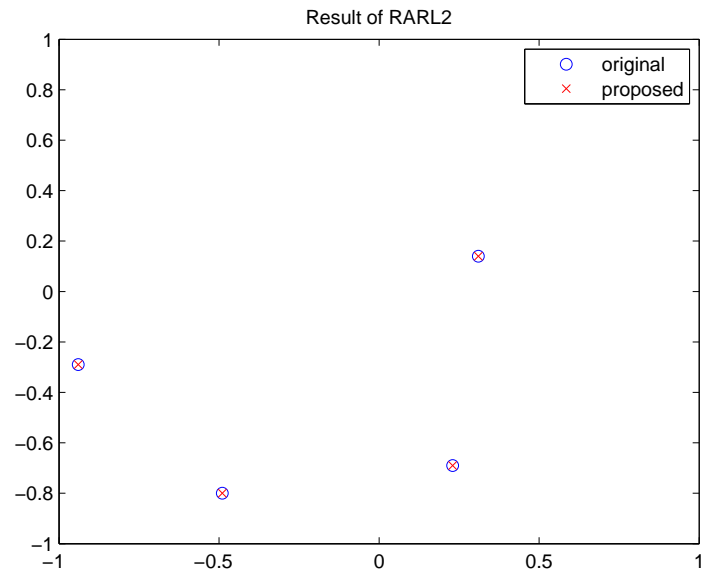


Figure 3: Experiment Result of The First Case of Example 2

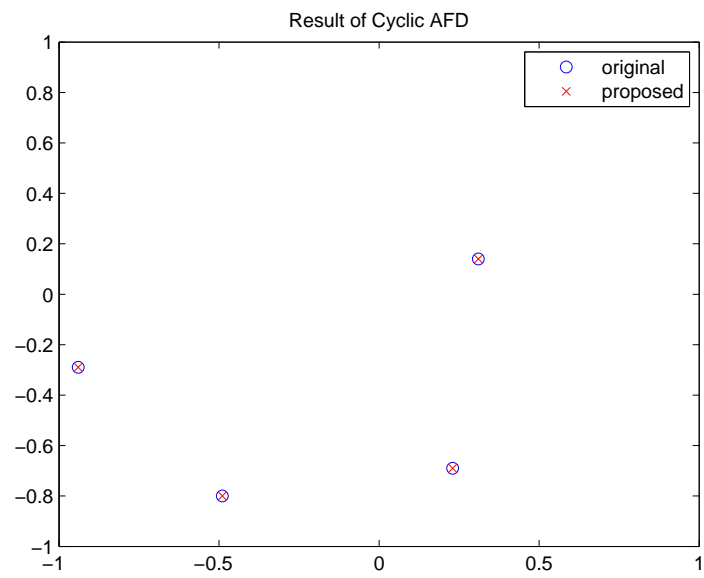


Figure 4: Experiment Result of The First Case of Example 2

**Experiment 3** The function to be approximated is a 4-Blaschke form

$$f(z) = \sum_{k=1}^4 c_k B_{b_1, b_k}(z)$$

with the poles  $(b_1, b_2, b_3, b_4) = (-0.1800 + 0.7700i, -0.0200 - 0.1800i, 0.1000 + 0.2400i, 0.1800 - 0.5300i)$  and the coefficients  $(c_1, c_2, c_3, c_4) = (0.1097 + 0.4754i, 1.1287 + 1.1741i, -0.2900 + 0.1269i, 1.2616 - 0.6568i)$ . For Cyclic AFD Algorithm the initial  $n$ -tuple is randomly chosen to be  $s_0 = (a_1^{(0)}, a_2^{(0)}, a_3^{(0)}, a_4^{(0)}) = (-0.4424 + 0.3592i, 0.3947 + 0.7040i, -0.0196 - 0.4243i, -0.1971 + 0.4299i)$ .

For Cyclic AFD Algorithm the running time is 13.689860 seconds with the result  $(-0.1100 + 0.7300i, 0.2200 - 0.4900i, 0.1100 + 0.3600i, -0.0300 - 0.2900i)$ .

For RARL2 the running time is 25.067327 seconds with the result  $(-0.2659 - 0.9629i, -0.1193 + 0.3195i, 0.1099 - 0.5024i, 0.1864 - 0.2356i)$ .

In Experiment 3 the domain the unit circle is discretized by using 1024 points.

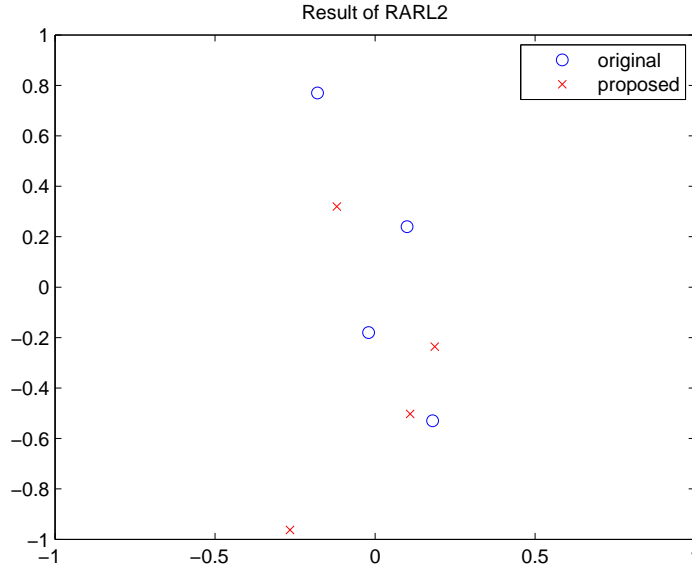


Figure 5: Experiment Result of The First Case of Example 3

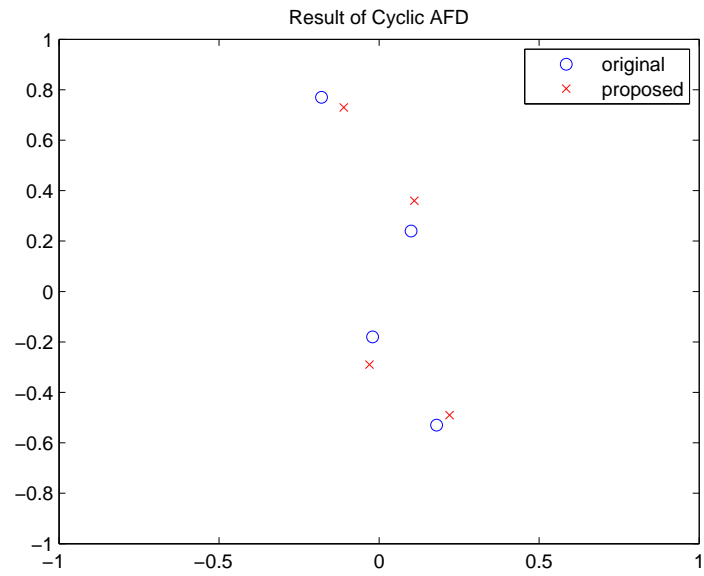


Figure 6: Experiment Result of The First Case of Example 3

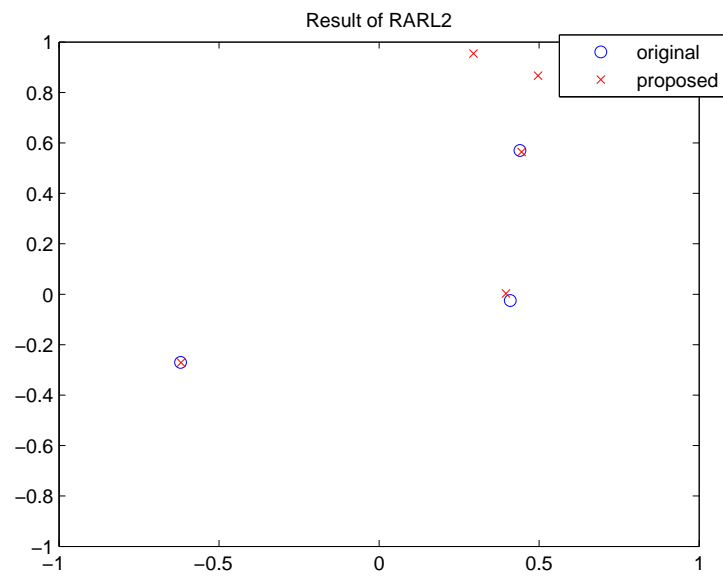


Figure 7: Experiment Result of The First Case of Example 4

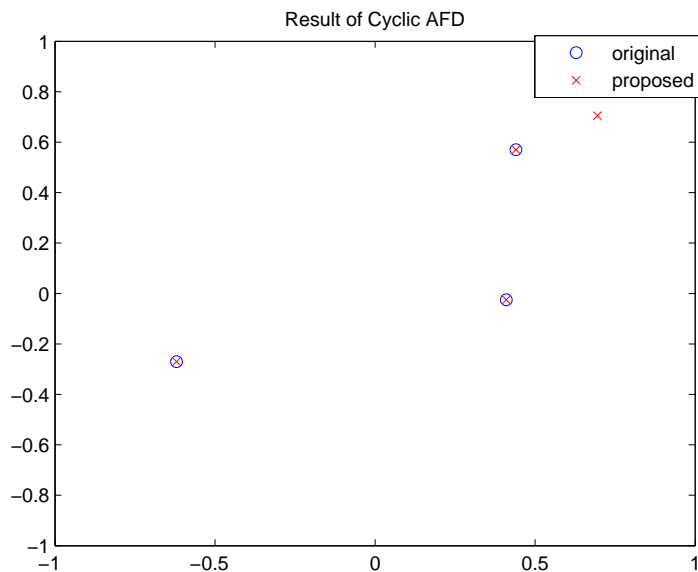


Figure 8: Experiment Result of The First Case of Example 4

The experiments show that when involving a large amount discretization points both algorithms can accurately locate the poles. In Experiment 1 and 3 Cyclic AFD Algorithm performs better than RARL2. When poles are close to zero neither of the two algorithms can accurately locate the poles. In most situations the results are case by case.

## 5 Conclusion

Cyclic AFD Algorithm is introduced to partially solve the  $H^2$ -best  $n$ -rational approximation problem, or rather  $H^2$ -best  $n$ -Blaschke form approximation problem. We rigourously proved that if the critical point of the objective function is unique, then the algorithm converges to the global minimum point. In practice the algorithm with suitable initial status leads to a global minimum point. Comparison based on three experiments shows that the proposed algorithm outperform to the existing RARL2 algorithm.

## References

- [1] L. Baratchart, A remark on uniqueness of best rational approximation of degree 1 in  $L^2$  of the circle, Electronic Transactions on Numerical Analysis, Volume 25, pp. 54-66, 2006.

- [2] L. Baratchart, M. Cardelli, M. Olivi, Identification and rational  $L^2$  approximation, a gradient algorithm, *Automatica*, 27(1991), pp. 413-418.
- [3] L. Baratchart, M. Olivi. Critical points and error rank in matrix  $H^2$  rational approximation and the strong differential consistency of output error identification from white noise inputs, *Constructive Approximation* 14, pp. 273-300 (1998).
- [4] L. Baratchart, E.B. Saff, F. Wielonsky, A criterion for uniqueness of a critical point in  $H^2$  rational approximation, *J. Analuse Mathématique*, 70:225-266, 1996.
- [5] L. Baratchart, E.B. Saff, F. Wielonsky, Non-uniqueness of rational approximations, *J. Comput. Appl. Math.*, 105:141-154, 1999.
- [6] L. Baratchart, E.B. Saff, F. Wielonsky, Asymptotic uniqueness of best rational approximations of given degree to Markov functions in  $L^2$  of the circle, *Constr. Approx.*, 17:103-138, 2001.
- [7] L. Baratchart, F. Wielonsky, Rational approximation in real Hardy space  $H^2$  and Stieltjes integrals: A uniqueness theorem, *Constr. Approx.*, 9:1-21, 1993.
- [8] L. Baratchart, M. Yattselev, Asymptotic uniqueness of best rational approximations to complex Cauchy transforms in  $L^2$  of the circle, *Recent Trends in Orthogonal Polynomials and Approximation Theory*, volume 507 of *Contemporary Mathematics*, page 87-111, Amer. Math. Soc., Providence, RI, 2010.
- [9] G. Davis, S. Mallet, M. Avellaneda, Adaptive Greedy Approximations, *Constr. Approx.* (1997) 13:57-98.
- [10] P. Fulcheri, M. Olivi, Matrix rational  $H^2$  approximation: a gradient algorithm based on schur analysis, *SIAM I. Control Optim.*, Vol 36, No. 6, pp. 2103-2127, November 1998.
- [11] J.B. Garnett, *Bounded Analytic Functions*, Academic Press, 1981.
- [12] W. Mi, T. Qian, F. Wan, A Fast Adaptive Model Reduction Method Based on Takenaka-Malmquist Systems, by W. Mi, T. Qian and F. Wan, *Systems and Control Letters*, Volume 61, Issue 1, January 2012, Page 223-230.
- [13] T. Qian, *Intrinsic mono-component decomposition of functions: An advance of Fourier theory*, *Mathematical Methods in Applied Sciences*, 2010, 33, 880-891, DOI: 10.1002/mma.1214.
- [14] T. Qian, Yanbo Wang, *Adaptive Fourier Series-A Variation of Greedy Algorithm*, *Advances in Computational Mathematics*, 34(2011), no.3, 279-293.

- [15] T. Qian and E. Wegert, *Optimal Approximation by Blaschke Forms*, Complex Variables and Elliptic Equations, Volume 58, Issue 1, 2013, page 123-133.
- [16] T. Qian, Liming Zhang and Zhi-Xiong Li, *Algorithm of Adaptive Fourier Decomposition*, IEEE Transaction on Signal Processing, Dec., 2011, Volume 59, Issue 12, page 5899-5902.
- [17] J.L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*,