

Chebyshev approximation by non-Chebyshev systems

Rinat Kamalov¹, Vladimir Protasov^{2,*}

1. Gran Sasso Science Institute, L'Aquila, Italy

2. University of L'Aquila, L'Aquila, Italy

Abstract: We address the problem of uniform approximation by linear combinations of a finite system of functions (possibly non-Chebyshev) under arbitrary linear constraints. The proposed solution is based on the modified concept of alternance and of the Remez iterative method. A criterion for the best approximation is proved and the full set of polynomials of best approximation (which may be non-singleton in the non-Chebyshev case) is characterized. We derive a regularization procedure that makes the generalized Remez algorithm avoid cycling. The efficiency is demonstrated in examples with systems of complex exponents, Gaussian functions, and lacunar polynomials. As an application, the Markov-Bernstein type inequality is obtained for systems of complex exponents. Applications to signal processing, linear ODEs, switching dynamical systems are considered.

Keywords: uniform approximation, polynomial of best approximation, non-Chebyshev system, linear constraints, Remez algorithm, regularization, exponential polynomials, recovery of signals

2010 Mathematics Subject Classification: 41A29, 41A50, 93D20

1. Introduction

The uniform finite-dimensional approximation of functions on compact domains is well-understood for approximations by Chebyshev systems on a segment. In this case, for every $f \in C[a, b]$, the best polynomial approximation always exists and is always unique, and is characterized by the alternance criterion (the Chebyshev - Vallée-Poussin theorem). Finding the polynomial of best approximation is realized by the Remez algorithm [44]. A system of n functions on a set K is *Chebyshev* (also called *Haar*) if every nontrivial polynomial by this system, i.e., a linear combination of its functions, has at most $n - 1$ zeros on K . They are, for example, algebraic polynomials on a segment, trigonometric polynomials on the period, shifts of Gaussian functions on a line, etc. See [13, 26, 29] for the general theory of Chebyshev approximations.

For non-Chebyshev systems, the situation is different. Neither the alternance criterion nor the Remez algorithm works. Many important problems in signal processing, numerical ODEs, etc. deal with non-Chebyshev systems such as quasipolynomials, band-limited functions, lacunary algebraic polynomials (polynomials spanned by some nonnegative integer powers t^{m_k} , $k = 1, \dots, n$), Müntz polynomials (spanned by arbitrary nonnegative powers), etc. All multivariate systems are, as a rule, non-Chebyshev. Another class of problems which is not covered by the Chebyshev analysis is the constrained uniform approximation. For example, an algebraic polynomial of best approximation with a prescribed value of the derivative at some point may not possess an alternance and may not be unique. Actually, imposing a linear constraint can destroy all properties of the Chebyshev approximation.

Approximation by non-Chebyshev systems has been studied in the literature [12, 45, 48, 50, 53]. The main difficulty is that a direct generalization of the Remez algorithm may not converge to the polynomial of best approximation. This is caused by a possible non-uniqueness of the optimal polynomial and by the degeneracy of alternance, in which case, the algorithm slows down or even cycles. Moreover, the degenerate alternance leads to difficulties in the numerical implementation (linear systems solved in each iteration become ill-conditioned, etc.) Our main contribution is the regularization method which overcomes all those problems. The performance of the new approach is shown through the numerical experiments.

*Corresponding author.

E-mail addresses: R. Kamalov: rinat.kamalov@gssi.it; V. Protasov: vladimir.protasov@univaq.it

<https://doi.org/10.68381/jca34025>

Received date: May 16, 2025; Accepted date: August 14, 2026

© 2027 Published by Banach Press LLC. (Dedicated to Alexander Plakhov on the occasion of his 65th birthday.)

In the vast majority of examples, including degenerate ones, the polynomial of best approximation is found within a few iterations. We still have not succeeded in proving the convergence of the algorithm for all initial data. This must be a challenging problem, although we believe that the answer is affirmative. In all numerical examples the algorithm converges, which is also demonstrated in applications to the signal processing and to the stability of dynamical systems (see Appendix).

2. The roadmap of the main results

We consider an arbitrary metric compact space K and a finite system of continuous linearly independent real-valued functions $\Phi = \{\varphi_1, \dots, \varphi_n\}$ on it. A linear combination $p = \sum_{k=1}^n p^k \varphi_k$, $p^k \in \mathbb{R}$, will be referred to as a *polynomial by the system* Φ . All polynomials form a linear space $\mathcal{P}$, which is naturally identified with $\mathbb{R}^n$. The vector of coefficients of the polynomial p will be denoted by the bold letter: $\mathbf{p} = (p^1, \dots, p^n) \in \mathbb{R}^n$. All the results are also valid for unbounded closed domains provided that all $\varphi_k(t)$ and f tend to zero as $t \rightarrow \infty$.

By $C(K)$ we denote the space of continuous functions on a metric compact space K with a standard uniform norm. For a function $f \in C(K)$, the problem of the *best uniform polynomial approximation* is

$$\begin{cases} \|p - f\|_{C(K)} \rightarrow \min, \\ p \in \mathcal{P}. \end{cases} \quad (1)$$

Note that the objective function is continuous and coercive since the optimal polynomial $\hat{p}$ (the solution of the problem (1)) always exists, although may not be unique. If Φ is a Chebyshev system on a segment $[a, b]$ (every nontrivial polynomial by Φ has at most $n - 1$ zeros on $[a, b]$), then $\hat{p}$ is unique and can be found by the Remez algorithm, which has a linear convergence and normally computes the solution fast and with a high accuracy. It is based on the classical Chebyshev criterion: $\hat{p}$ is a polynomial of best approximation precisely when there exists a sequence of $n + 1$ points $\tau_1 < \dots < \tau_{n+1}$ called *alternance* such that the difference $\hat{p}(t) - f(t)$ reaches the maximal absolute value at the points $t = \tau_j$ and its sign alternates in j , i.e., $\hat{p}(\tau_j) - f(\tau_j) = (-1)^{j+\delta} \|\hat{p} - f\|$, where δ is either zero or one. Let us briefly describe the Remez algorithm. By the k th iteration, it produces points $\{t_i^{(k)}\}_{i=1}^{n+1}$ and a polynomial p_k such that the values $p_k(t_i^{(k)}) - f(t_i^{(k)})$ are equal in modulus (not necessary equal to $\|p_k - f\|$) and their signs alternate in i . We take the point of maximum of $|p_k - f|$, denote it by $t_0^{(k)}$, and replace some point $t_s^{(k)}$ by $t_0^{(k)}$ to preserve the sign alternation property. The point $t_s^{(k)}$ is chosen to be the neighboring node to $t_0^{(k)}$ with the same sign of the difference $p - f$ or to be the farthest point from $t_0^{(k)}$ if $t_0^{(k)}$ has no neighboring node with the same sign of the difference $p - f$ (this may happen if $t_0^{(k)}$ lies to the left or to the right of all points $\{t_i^{(k)}\}_{i=1}^{n+1}$). Then we construct the next polynomial p_{k+1} which possesses the alternating property on the new system of points.

The Chebyshev property holds for several important functional systems, however, in general, Chebyshev systems are rather exceptional. For arbitrary systems, neither of the aforementioned facts take place. The optimal polynomial $\hat{p}$ is not necessarily unique and may not possess an alternance. The “Refinement theorem” (Theorem B in Section 3) implies the existence of a set of points $\{\tau_i\}_{i=1}^N$, $N \leq n + 1$, such that no polynomial from $\mathcal{P}$ has all values $|p(\tau_i) - f(\tau_i)|$ smaller than $\|\hat{p} - f\|$. This set could be considered as a weakened version of alternance, however, it may be degenerate (when $N < n + 1$) and the signs of $p(\tau_i) - f(\tau_i)$, $i = 1, \dots, n + 1$, may not alternate. For instance, in Example 7.1, the point $\tau_1 = 0$ satisfies the condition that no polynomial from $\mathcal{P}$ has value $|p(\tau_1) - f(\tau_1)|$ smaller than $\|\hat{p} - f\|$. However, this point does not characterize the set of optimal polynomials, because for any polynomial p from $\mathcal{P}$ the value $|p(\tau_1) - f(\tau_1)|$ is equal to 1. Therefore, this “weakened alternance” is not sufficient to characterize the best approximation. Moreover, the routine of Remez’ algorithm becomes inapplicable, because the choice of the point $t_s^{(k)}$ to be replaced by $t_0^{(k)}$ is not valid without the sign alternating.

Theorem 3.2 proved in Section 3 formulates the best approximation criterion in the case of a non-Chebyshev system. The signs alternating property is replaced by the following geometric condition: the simplex formed by the oriented moment vectors in $\mathbb{R}^n$ contains the origin. Here we need some further notation. For an arbitrary point $t \in K$, we consider the *moment vector* $\mathbf{u}(t) = (\varphi_1(t), \dots, \varphi_n(t))^T$. If, in addition, we are given a function $f \in C(K)$ and a polynomial $p \in \mathcal{P}$, then we define the *oriented moment vector* $\mathbf{a}(t) = \sigma(t) \mathbf{u}(t)$, where $\sigma(t) = \text{sign}(p(t) - f(t))$. For a given polynomial $\hat{p} \in \mathcal{P}$, the *generalized alternance* is a set $\{\tau_i\}_{i=1}^m \subset K$, $m \leq n + 1$, such that $\hat{p}(\tau_i) - f(\tau_i) = \sigma_i \|\hat{p} - f\|$, $\sigma_i \in \{-1, 1\}$, for all i , and the

$(m - 1)$ -dimensional simplex Δ with vertices $\{\sigma_i \mathbf{u}(\tau_i)\}_{i=1}^m$ contains the origin in $\mathbb{R}^n$. In other words, some convex combination of the vectors $\mathbf{a}(\tau_i)$, $i = 1, \dots, m$, is equal to zero.

Theorem 3.2 asserts that $\hat{p}$ is a best approximation polynomial if and only if it possesses the generalized alternance. The results similar to Theorem 3.2 appeared in the literature in various forms, see [9, p.12], [11, theorem 5.2], [51, theorem 4.2], [53]. Its proof follows in a rather straightforward manner by applying the Refinement theorem and the Farkas lemma (see Section 3). Other possible proofs involve the Kolmogorov criterion [28, 50] or subdifferentials (Remark 3.3). If the system is non-Chebyshev, then the best approximation may not be unique, see Example 7.1.

A complete classification of all polynomials of the best approximation is given in Theorem 3.4. The classification is surprisingly simple: all polynomials of the best approximation must have the same generalized alternance with the same values on it. Note that each polynomial may also have other alternances. Theorems 3.2 and 3.4 make it possible to generalize Remez algorithm for an arbitrary system of functions Φ on a compact set. In each iteration we take the point t_0 of maximum modulus $|p_k(t) - f(t)|$ and replace one of the vertices of the simplex Δ_k with $\sigma_0 \mathbf{u}(t_0)$, $\sigma_0 = \text{sign}(p_k(t_0) - f(t_0))$, so that the new simplex Δ_{k+1} still contains the origin. The algorithm halts when we localize the distance from f to $\mathcal{P}$ to a segment of a given length ε . We show that this algorithm has a linear rate of convergence in the nondegenerate case, the estimates are presented in Theorem 4.4. The latter condition is significant: in many situations the simplex Δ_k that appears at the k th iteration may be close to degenerate, after which the algorithm slows down or stacks. In this case, algorithm also faces technical problems because of the ill-conditioned linear systems. Moreover, if $m < n + 1$, then this simplex always degenerates for large k since the limit simplex is $(m - 1)$ -dimensional. To resolve these issues we present the method of regularization and modify the algorithm. The key novelty is the choice of the new point $t_0^{(k)}$, see Section 5. The efficiency of the new method is confirmed numerically (Section 6).

In Section 5 we address the constrained approximation. For given linear functionals $\ell_j : \mathcal{P} \rightarrow \mathbb{R}$ and numbers b_j , $j = 1, \dots, r$, we consider the problem

$$\begin{cases} \|p - f\|_{C(K)} \rightarrow \min, \\ \ell_j(p) = b_j, \quad j = 1, \dots, r, \\ p \in \mathcal{P}. \end{cases} \quad (2)$$

Even if the system $\{\varphi_i\}_{i=1}^n$ is Chebyshev, problem (2) cannot be solved by the standard approach: the optimal polynomial may not have an alternance (even generalized one) and the Remez algorithm may not be applicable (Example 7.2). Nevertheless, it turns out that the generalized alternance approach and the regularization technique can be adopted to constrained problems and can efficiently solve them. Theorems 5.2 and 5.3 characterize polynomial of the best approximation. Then we present the algorithm that can be used to compute them.

In Appendix we provide illustrative examples, a description of Algorithms 1 and 2 suitable for implementation, and possible real-life applications. Particular attention is paid to the signal processing, where signals often have the form of quasipolynomials (real parts of complex exponentials). If the signal is given by observation, with possible errors and noise, the problem is to recover it in an optimal (in some sense) way. Our method restores it with the minimal uniform distance from the observed function. If, in addition, the signal is given by means of a linear sampling, then the optimal recovery is done by solving the constrained problem (2). The same technique works for other types of signals, for instance, the ones spanned by shifts and contractions of the Gaussian function or of the Cauchy function. Other applications are on the stability of linear dynamical systems and analyzing the growth of their trajectories. Moreover, in this section we discuss why we choose Remez-like recursive methods over convex optimization methods, even though the problem is convex.

3. The generalized alternance

We begin with a criterion of the best uniform approximation. Let K be a metric compact space, $f \in C(K)$, $\Phi = \{\varphi_i\}_{i=1}^n$ be a system of linearly independent functions from $C(K)$ and $\mathcal{P}$ be the space of polynomials by the system Φ . We denote with $\text{co}(X)$ the convex hull of a set X .

Definition 3.1. For given $p \in \mathcal{P}$, a set of points $\{\tau_1, \dots, \tau_m\} \subset K$, $m \leq n + 1$, is called *generalized alternance* if

1. $p(\tau_i) - f(\tau_i) = \sigma_i \|p - f\|$, $\sigma_i \in \{-1, 1\}$, for all $i = 1, \dots, m$;
2. the convex hull of the vectors $\{\sigma_i \mathbf{u}(\tau_i)\}_{i=1}^m$ in $\mathbb{R}^n$, where $\mathbf{u} = (\varphi_1(t), \dots, \varphi_n(t))^T$, contains the origin.

In the case $\|p - f\| = 0$, in the Definition 3.1, for each $\tau \in K$, instead of taking a single vector $\sigma(\tau)\mathbf{u}(\tau)$, we take a pair of vectors $\mathbf{u}(\tau), -\mathbf{u}(\tau)$.

Using the notation $\mathbf{a}_i = \sigma_i \mathbf{u}(\tau_i)$, we have $\mathbf{O} \in \Delta = \text{co}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$. In what follows we deal with generalized alternance only and drop the word “generalized”.

Theorem 3.2. *A polynomial $p \in \mathcal{P}$ is of best approximation for a function $f \in C(K)$ if and only if it has an alternance.*

As we know, for a Chebyshev system on a segment the alternance possesses the sign alternating property. For general systems this property is replaced by that $\mathbf{O} \in \Delta$. In fact, already for two-element systems Φ on $[-1, 1]$, there are examples of polynomials of best approximation with non-alternating signs of $p(\tau_i) - f(\tau_i)$. Example 7.4 in Appendix demonstrates this situation for the system $\Phi = \{t, t^2\}$.

The proof of Theorem 3.2 is based on two results of convex analysis formulated below. The first one is the well-known Farkas’ lemma, see, for instance [46]. We enumerate them with letters as known results taken from the literature.

Lemma A (Farkas’ lemma) *For an arbitrary set of vectors $\mathbf{b}_1, \dots, \mathbf{b}_m$ in $\mathbb{R}^n$ the following are equivalent:*

- 1) *their convex hull contains the origin;*
- 2) *the system of linear inequalities $(\mathbf{b}_i, \mathbf{x}) < 0$, $i = 1, \dots, m$, does not have a solution.*

The second auxiliary result is the Refinement theorem; its proof can be found in [33].

Theorem B (Refinement theorem) *Let K be a compact subset of a metric space, a function $F : K \times \mathbb{R}^n \rightarrow \mathbb{R}$ be such that for each $t \in K$, the function $F(t, \cdot)$ is convex and for each $\mathbf{p} \in \mathbb{R}^n$, the function $F(\cdot, \mathbf{p})$ is continuous. Then there exist $N \leq n + 1$ points $\tau_1, \dots, \tau_N \in K$ such that*

$$\inf_{\mathbf{p} \in \mathbb{R}^n} \left(\max_{t \in K} F(t, \mathbf{p}) \right) = \inf_{\mathbf{p} \in \mathbb{R}^n} \left(\max_{i=1, \dots, N} F(\tau_i, \mathbf{p}) \right).$$

Applying Theorem B to the function $F(t, \mathbf{p}) = |p(t) - f(t)|$, where p is the polynomial from $\mathcal{P}$ with the vector of coefficients $\mathbf{p}$, we see that finding the polynomial of best approximation on K can always be reduced to that on some finite subset $\{\tau_i\}_{i=1}^m \subset K$. Therefore, some kind of “weakened alternance” always exists. The problem is that Theorem B asserts only the existence and does not give any recipe how to find such points. Moreover, these points do not necessarily characterize the set of optimal polynomials, as we discussed in Section 2. Nevertheless, it gives a bridge to Theorem 3.2 for characterizing polynomials of the best approximation.

Proof of Theorem 3.2. *Sufficiency.* If a polynomial p possesses an alternance $\{\tau_i\}_{i=1}^m$ but does not provide the best approximation, then there exists $h \in \mathcal{P}$ such that $\|p + h - f\| < \|p - f\|$. Therefore, at each point τ_i , we have

$$|p(\tau_i) - f(\tau_i) + h(\tau_i)| < \|p - f\| = |p(\tau_i) - f(\tau_i)|.$$

This implies that the value $h(\tau_i)$ has an opposite sign to $(p - f)(\tau_i) = \sigma_i \|p - f\|$. Hence, $\text{sign } h(\tau_i) = -\sigma_i$ and consequently, $\sigma_i h(\tau_i) < 0$. If $\mathbf{h}$ denotes the vector of coefficients of the polynomial h , then $h(\tau_i) = (\mathbf{h}, \mathbf{u}(\tau_i))$ and $\sigma_i h(\tau_i) = (\sigma_i \mathbf{h}, \mathbf{u}(\tau_i)) = (\mathbf{h}, \sigma_i \mathbf{u}(\tau_i)) = (\mathbf{h}, \mathbf{a}_i)$. We see that $(\mathbf{h}, \mathbf{a}_i) < 0$ for all $i = 1, \dots, m$, which contradicts to Farkas’ lemma.

Necessity. Let $\hat{p}$ be the best approximation polynomial. Applying the Refinement theorem to the function $F(t, \mathbf{p}) = |p(t) - f(t)|$, we obtain a set of points $\tau_i \in K$, $i = 1, \dots, N$, where $N \leq n + 1$, such that $\hat{p}$ gives the best approximation of the function f on this set. After renumbering it can be assumed that $|\hat{p}(\tau_i) - f(\tau_i)| = \|\hat{p} - f\|$ for $i \leq m$ and $|\hat{p}(\tau_i) - f(\tau_i)| < \|\hat{p} - f\|$ for $i > m$, where $0 \leq m \leq N$. If $m = 0$, then $\max_{i=1}^N |\hat{p}(\tau_i) - f(\tau_i)| < \|\hat{p} - f\|$, which is impossible, hence $m \geq 1$. If there exists a polynomial $h \in \mathcal{P}$ such that for each $i = 1, \dots, m$, the value $h(\tau_i)$ has the sign $-\sigma_i$, then for sufficiently small $\lambda > 0$, we have $|\hat{p}(\tau_i) - f(\tau_i) + \lambda h(\tau_i)| < \|\hat{p} - f\|$ for all $i = 1, \dots, N$, which contradicts to the best approximation

property of $\hat{p}$ on the set $\{\tau_i\}_{i=1}^N$. Thus, there is no polynomial $h \in \mathcal{P}$ for which $\sigma_i h(\tau_i) < 0$, $i = 1, \dots, m$. In terms of the corresponding vectors of coefficients $\mathbf{h}, \mathbf{p}$, we have $(\mathbf{h}, \mathbf{a}_i) = (\mathbf{h}, \sigma_i \mathbf{u}(\tau_i)) = \sigma_i h(\tau_i) < 0$. Hence, the system of linear inequalities $(\mathbf{h}, \mathbf{a}_i) < 0$, $i = 1, \dots, m$, has no solution. By the Farkas' lemma, $\mathbf{O} \in \text{co}\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$.

□

Remark 3.3. Another proof of Theorem 3.2 can be done by involving subdifferentials. Let us describe the idea. The point $\hat{p} \in \mathcal{P}$ gives an absolute minimum of the function $F(p) = \|p - f\|$ if and only if $\mathbf{O} \in \partial F(\hat{p})$, where ∂F denotes the subdifferential. Since $F(\hat{p}) = \max_{t \in K} |\hat{p}(t) - f(t)|$, it follows that the subdifferential $\partial F(p)$ can be computed by the Dubovitsky-Milyutin theorem on the subdifferential of maximum. Denote

$$\Gamma = \left\{ \sigma(t)\mathbf{u}(t) \in \mathbb{R}^n : t \in K, |\hat{p}(t) - f(t)| = \|\hat{p} - f\| \right\}.$$

We obtain $\partial F(\hat{p}) = \text{co}\Gamma$ (in the case $f \in \mathcal{P}$, the set Γ includes the vectors $\mathbf{u}(t), -\mathbf{u}(t)$ for each $t \in K$). Thus, $\mathbf{O} \in \text{co}\Gamma$. Now we invoke the Caratheodory theorem and conclude that there exist $m \leq n + 1$ points from Γ , whose convex hull contains $\mathbf{O}$.

Clearly, for every $f \in C(K)$, the set of polynomials of the best approximation is a convex compact subset of $\mathcal{P}$. Due to Theorem 3.2, each element p of this subset possesses an alternance, which is not necessarily unique and depends on p . It turns out, however, that all those alternances have a nonempty intersection, which is a common alternance for all polynomials of best approximation.

Theorem 3.4. Let $\mathcal{P}$ be the space of polynomials by a system Φ . For every $f \in C(K)$, all polynomials of best approximation have the same alternance with the same values on it.

This means that for every function f , there exists a set of points $\{\tau_i\}_{i=1}^m$ such that all polynomials of approximation have this set as an alternance and for every i , the value of all those polynomials at the point τ_i is the same.

Proof. Let $B = \{p_v \in \mathcal{P} : v \in \mathcal{V}\}$ be the set of all polynomials of best approximation, $\mathcal{V}$ is some index set. Without loss of generality it can be assumed that $\|p_v - f\| = 1$, $v \in \mathcal{V}$ (if $\|p_v - f\| = 0$, $v \in \mathcal{V}$, then any set of points $\{\tau_i\}_{i=1}^m \subset K$, $m \leq n + 1$, is a "minimal" alternance set in the sense of Theorem 2). Every point of absolute maximum of $|p_v - f|$ will be called *extreme point* of p_v . Thus, at every extreme point τ , we have $p_v(\tau) - f(\tau) = \sigma_v(\tau) \in \{-1, 1\}$. Denote by $\mathcal{T}_v$ the set of pairs $\{(\tau, \sigma_v(\tau)) : \tau \text{ is an extreme point of } p_v\}$. By Theorem 3.2, for every $v \in \mathcal{V}$, the set of extreme points of p_v contains an alternance of p_v . Therefore, the set $C_v = \text{co}\{\mathbf{a}(\tau) : \tau \text{ is an extreme point of } p_v\}$ contains the origin. By the convexity of $F(p) = \|f - p\|$, we have that for every $v_1, \dots, v_N \in \mathcal{V}$, $N \geq 1$, $\|f - \frac{1}{N} \sum_{j=1}^N p_{v_j}\| \leq \sum_{j=1}^N \frac{1}{N} \|f - p_{v_j}\| = 1$. Thus, the polynomial $\frac{1}{N} \sum_{j=1}^N p_{v_j}$ belongs to the set B . Denote it by p_a , $a \in \mathcal{V}$ and use the short notation $v_j = j$. Since $|p_j(\tau) - f(\tau)| \leq 1$, $j = 1, \dots, N$, for every extreme point τ of p_a (at least one such point τ exists), all the values $p_j(\tau) - f(\tau)$, $j = 1, \dots, N$ are equal (either to 1 or to -1). Hence, τ is an extreme point for all p_j with the same sign $\sigma_j(\tau) = \sigma_a(\tau)$. Therefore, $\mathcal{T}_v$ is non-empty and $\mathcal{T}_v \subset \bigcap_{j=1}^N \mathcal{T}_j$. We see that the family of compact sets $M = \{\mathcal{T}_v : v \in \mathcal{V}\}$ possesses the following property: the intersection of an arbitrary finite subfamily of M contains the set from M . Then, by the theorem on the finite intersection property [14], all $\mathcal{T}_v \in M$ have a nonempty intersection, which contains a set from M . Let it be $\mathcal{T}_b$, $b \in \mathcal{V}$. The set C_b contains the origin, therefore, by Caratheodory's theorem [33], there are extreme points $\tau_1, \dots, \tau_m$ of p_b , $m \leq n + 1$ such that $\mathbf{O} \in \text{co}\{\sigma(\tau_i)\mathbf{u}(\tau_i)\}_{i=1}^m$. Since they are extreme points for every p_v with the same signs $\sigma_v(\tau_i)$, $v \in \mathcal{V}$, we see that they form an alternance for p_v .

4. Algorithm for the best approximation. The regular case

We derive Algorithm 1 that follows the Remez scheme, but with another rule of choosing the next point of alternance at the k th iteration. The sign alternating property is replaced by the requirement that $\mathbf{O} \in \Delta_k$, where the simplex Δ_k has vertices at the oriented moment vectors $\mathbf{a}_i = \sigma(\tau_i)\mathbf{u}(\tau_i)$. We show that in the regular case, i.e., when all the simplices Δ_k are nondegenerate and the barycentric coordinates of $\mathbf{O}$ in Δ_k are bounded away from zero independently of k , Algorithm 1 converges with a linear rate. Otherwise, if Δ_k is close to degenerate in some iterations, we apply a regularization procedure given in the next section (Algorithm 2). We begin Algorithm 1 with several auxiliary results.

4.1. Preliminary facts

For an arbitrary vector $\mathbf{b} \in \mathbb{R}^n$, its *negative extension* is $\{\lambda \mathbf{b} : \lambda \leq 0\}$. Thus, the negative extension of a nonzero vector is a ray directed opposite to $\mathbf{b}$; for a zero vector, it is zero.

A system of N vectors $\mathbf{b}_1, \dots, \mathbf{b}_N \in \mathbb{R}^n$, $N \geq n$, is said to be *nondegenerate* if every n of its vectors are linearly independent. Clearly, an independent system cannot contain zero vectors or a subset of $m \leq n - 1$ dependent vectors.

Lemma 4.1. *Suppose the convex hull of vectors $\mathbf{b}_1, \dots, \mathbf{b}_{n+1} \in \mathbb{R}^n$ contains the origin; then for an arbitrary vector $\mathbf{b}_0 \in \mathbb{R}^n$, there exists $s \in \{1, \dots, n+1\}$, such that after the replacement of $\mathbf{b}_s$ by $\mathbf{b}_0$, the convex hull still contains the origin. Moreover, if the enlarged system $\{\mathbf{b}_i\}_{i=0}^{n+1}$ is nondegenerate, then such an s is unique.*

Proof. First, consider the case when $\{\mathbf{b}_i\}_{i=0}^{n+1}$ is nondegenerate. In this case, the system $\{\mathbf{b}_i\}_{i=1}^{n+1}$ is also nondegenerate, and consequently the simplex $\Delta = \text{co}\{\mathbf{b}_i\}_{i=1}^{n+1}$ contains $\mathbf{O}$ in its interior. Hence, the negative extension of $\mathbf{b}_0$ intersects a unique facet of Δ . Let $\mathbf{b}_s$ be the vertex opposite to that facet; then s is the desired index. Consider now a general system $\{\mathbf{b}_i\}_{i=0}^{n+1}$, which is admissible, i.e., satisfies the assumptions of the lemma. It can be slightly perturbed to become nondegenerate and still stay admissible. Hence, there exists a sequence of nondegenerate admissible systems $\{\mathbf{b}_i^{(k)}\}_{i=0}^{n+1}$ that converges pointwise to $\{\mathbf{b}_i\}_{i=0}^{n+1}$. By what we have proved above, for every k , there exist a unique index $s = s(k)$ such that the simplex with vertices $\{\mathbf{b}_i^{(k)}\}_{i \neq s}$ contains the origin. Passing to a subsequence, it can be assumed that $s(k)$ is the same for all k . Now by passing to the limit we conclude that the simplex with vertices $\{\mathbf{b}_i\}_{i \neq s}$ contains the origin.

4.2. The scheme of Algorithm 1

Here we give a brief description of the algorithm. The details and the pseudocode will follow in subsection 4.3. We assume that there exist points $t_i \in K$ such that the system of moment vectors $\{\mathbf{u}(t_i)\}_{i=1}^{n+1}$ is nondegenerate. This is a part of the regularity assumption formulated in the next subsection.

Input. A system $\Phi = \{\varphi_1, \dots, \varphi_n\}$, a function $f \in C(K)$, and $\varepsilon > 0$.

Output. Numbers b, B and a polynomial $\hat{p} \in \mathcal{P}$ such that $B - b < \varepsilon$, $b \leq \text{dist}(f, \mathcal{P}) \leq B$ and $b \leq \|\hat{p} - f\| \leq B$.

Initialization. We start with arbitrary points $t_1, \dots, t_{n+1}$ for which the set of vectors $\{\mathbf{u}(t_i)\}_{i=1}^{n+1}$ is nondegenerate and choose the signs σ_i so that the convex hull of the set $\{\sigma_i \mathbf{u}(t_i)\}_{i=1}^{n+1}$ contains the origin. Denote $\mathbf{a}_i = \sigma_i \mathbf{u}(t_i)$ and consider the linear system

$$(\mathbf{p}, \mathbf{u}(t_i)) = f(t_i) + \sigma_i d, \quad i = 1, \dots, n+1, \quad (3)$$

where $\mathbf{p} = (p^1, \dots, p^n)^T \in \mathbb{R}^n$ and $d \in \mathbb{R}$ are the variables. Since $\{\mathbf{u}(t_i)\}_{i=1}^{n+1}$ is nondegenerate, the system $\{\mathbf{a}_i\}_{i=1}^{n+1}$ is also nondegenerate. In view of this and the fact that $\mathbf{O} \in \text{co}\{\mathbf{a}_i\}_{i=1}^{n+1}$, Lemma 7.5 (see Appendix) implies that the system (3) has a unique solution $(p^1, \dots, p^n, d)^T$, where $(p^1, \dots, p^n)^T$ are coefficients of $p \in \mathcal{P}$. If $d < 0$, then change all the signs σ_i . Thus, $d \geq 0$. Set $b_1 = d$, $B_1 = \|p - f\|_{C(K)}$.

Main loop. After $k - 1$ iterations we have $n + 1$ points (for the sake of simplicity, denote them as above $t_1, \dots, t_{n+1}$) and the corresponding vectors $\mathbf{a}_1, \dots, \mathbf{a}_{n+1}$, whose convex hull contains the origin as an interior point (we assume that the conditions of Lemma 4.1 are fulfilled at each iteration of the algorithm). Then we take a point t_0 of maximum of the function $|p(t) - f(t)|$ on K and replace one of the vectors $\mathbf{a}_s$ with $\mathbf{a}_0 = \sigma_0 \mathbf{u}(t_0)$ so that the new system still contains the origin inside its convex hull, see Fig.1.

This means that the negative extension of $\mathbf{a}_0$ intersects the facet of the simplex $\text{co}\{\mathbf{a}_i\}_{i=1}^{n+1}$ opposite to the vertex $\mathbf{a}_s$. If the extended system $\{\mathbf{a}_i\}_{i=0}^{n+1}$ is nondegenerate, then the index s is uniquely defined. Then we solve the system (3) and set $b_{k+1} = d$, $B_{k+1} = \min\{B_k, \|p_k - f\|\}$.

Termination. The algorithm halts when $B_N - b_N < \varepsilon$. Then we set $b = b_N$, $B = B_N$ and $\hat{p} = p_k$, where $k \leq N$ is the number of iteration, for which $\|p_k - f\| = B$.

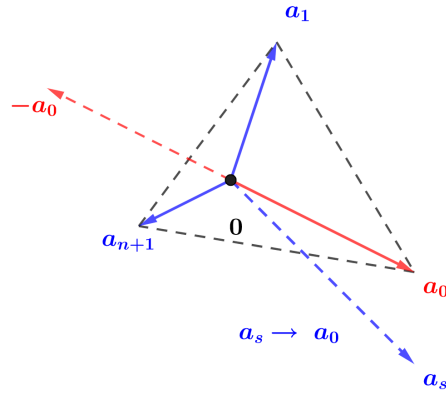


Figure 1

The regularity assumption. We say that the *regular case* takes place for Algorithm 1 if in each iteration, the extended system $\{a_i\}_{i=0}^{n+1}$ is nondegenerate and the coefficient α_0 in the convex combination $O = \sum_i \alpha_i a_i$ (the summation is over all $i = 0, \dots, n+1$ except for s) is not less than a prescribed number $\mu > 0$. This number will be referred to as a *regularity parameter*. Recall that in the convex combination $\sum_i \alpha_i a_i$, all coefficients α_i are supposed to be nonnegative and $\sum_i \alpha_i = 1$.

4.3. The routine of Algorithm 1

Below, we give a detailed description of all steps.

Algorithm 1: Regular Case

Require: A function $f \in C(K)$; a system of linearly independent real-valued functions $\Phi = \{\varphi_i\}_{i=1}^n$ from $C(K)$; real number $\varepsilon > 0$

Ensure: Real numbers b, B and a polynomial $\hat{p} \in \mathcal{P} \equiv \text{span}\{\varphi_1, \dots, \varphi_n\}$ such that $B - b < \varepsilon$, $b \leq \text{dist}(f, \mathcal{P}) \leq B$ and $b \leq \|\hat{p} - f\| \leq B$

Initialization: Take an arbitrary set $\mathcal{T} = \{t_i\}_{i=1}^{n+1}$ of $n+1$ different points on K such that the system of moment vectors $\{\mathbf{u}(t_i)\}_{t_i \in \mathcal{T}}$ is nondegenerate (see Section 4.1). Then the system of n linear equations $\sum_{i=1}^{n+1} x_i \mathbf{u}(t_i) = 0$ possesses a unique, up to scalar multiple, solution $\mathbf{x} = (x_1, \dots, x_{n+1})$ with all nonzero components. If σ_i is a sign of x_i , $\alpha_i = |x_i|$ and $\mathbf{a}_i = \sigma_i \mathbf{u}(t_i)$, then

$$\sum_{i=1}^{n+1} x_i \mathbf{u}(t_i) = \sum_{i=1}^{n+1} |x_i| \sigma_i \mathbf{u}(t_i) = \sum_{i=1}^{n+1} |x_i| \mathbf{a}_i = \sum_{i=1}^{n+1} \alpha_i \mathbf{a}_i = 0.$$

Now find a vector $\mathbf{p} \in \mathbb{R}^n$ and a number d from the linear system

$$(\mathbf{p}, \mathbf{u}(t_i)) = f(t_i) + \sigma_i d, \quad i = 1, \dots, n+1.$$

The corresponding polynomial p possesses the property $p(t_i) = f(t_i) + \sigma_i d$, $i = 1, \dots, n+1$. If $d < 0$, then change all signs σ_i . Thus, $d \geq 0$. Set $k = 1$, $\mathcal{T}_1 = \mathcal{T}$, $p_1 = p$, $b_1 = d$, $B_1 = \|p_1 - f\|$.

Main loop:

while $B_k - b_k > \varepsilon$ **do**

At k -th iteration we have a polynomial $p_k \in \mathcal{P}$, a set of $n+1$ points $\mathcal{T}_k = \{t_i^{(k)}\}_{i=1}^{n+1}$ (for the sake of simplicity, we denote $t_i^{(k)} = t_i$) and two numbers $B_k \geq b_k \geq 0$ such that:

1) $p_k(t_i) = f(t_i) + \sigma_i b_k$, where $\sigma_i \in \{-1, 1\}$, $i = 1, \dots, n+1$;

2) the simplex $\Delta_k = \text{co}\{\mathbf{a}_i\}_{i=1}^{n+1}$, where $\mathbf{a}_i = \sigma_i \mathbf{u}(t_i)$ contains the origin inside: $\mathbf{O} = \sum_{i=1}^{n+1} \alpha_i^{(k)} \mathbf{a}_i$ with all $\alpha_i^{(k)}$ positive and $\sum_{i=1}^{n+1} \alpha_i^{(k)} = 1$.

Denote $r_k = \|p_k - f\|$, $\alpha_i = \alpha_i^{(k)}$, and find a point t_0 for which $|p_k(t_0) - f(t_0)| = r_k$. Clearly, $r_k \geq b_k$. Denote by σ_0 the sign of $p_k(t_0) - f(t_0)$ and $\mathbf{a}_0 = \sigma_0 \mathbf{u}(t_0)$.

Now find the facet Δ_k^s of Δ_k intersected by the negative extension of $\mathbf{a}_0$ (see Section 4.1). To this end, we construct the linear system $\mathbf{a}_0 = \sum_{i=1}^{n+1} x_i \mathbf{a}_i$. By Kronecker–Capelli theorem (see [8]), there exists a solution $\mathbf{x} = (x_1, \dots, x_{n+1})$ (not unique, we take any of them) to this system. Denote by $\tau = \frac{x_s}{\alpha_s}$, the maximal ratio $\frac{x_i}{\alpha_i}$. Then $\mathbf{a}_0 = \sum_{i=1}^{n+1} (x_i - \tau \alpha_i) \mathbf{a}_i$, and the coefficient $x_s - \tau \alpha_s$ vanishes, while all other coefficients are negative. Indeed, if at least one other coefficient $x_j - \tau \alpha_j$ were equal to zero, then the system of n vectors $\{\mathbf{a}_i\}_{i=0, i \neq j, s}^{n+1}$ would be linearly dependent, which contradicts the regularity assumption. Hence, the negative extension of $\mathbf{a}_0$ intersects the facet Δ_k^s opposite to $\mathbf{a}_s$.

Thus, replacing $\mathbf{a}_s$ by $\mathbf{a}_0$, we obtain a new simplex Δ_{k+1} containing the origin. Replacing t_s by t_0 , we obtain a new system of points $\mathcal{T}_{k+1}$. Redenote the points of the set $\mathcal{T}_{k+1}$ by $\{t_i\}_{i=1}^{n+1}$ in an arbitrary order. Then we find a polynomial p_{k+1} and a number $b_{k+1} > 0$ from the linear system $p_{k+1}(t_i) = f(t_i) + \sigma_i b_{k+1}$, $i = 1, \dots, n+1$. The system has a unique solution because the new system of moment vectors $\{\mathbf{u}(t_i)\}$, $t_i \in \mathcal{T}_{k+1}$, is also nondegenerate. Set $k = k+1$, $B_{k+1} = \min\{r_k, B_k\}$.

end while

return Set $b = b_N$, $B = B_N$ and $\hat{p} = p_k$, where $k \leq N$ is the number of iteration, for which $\|p_k - f\| = B$.

Remark 4.2. The procedure of the Algorithm 1 stays well-defined also without the regularity assumption. Lemma 4.1 provides the existence of the required point $\mathbf{a}_s$ even if the simplex Δ_k is degenerate. However, in this case, the difference $B_k - b_k$ may not decrease and the termination criterion $B_N - b_N < \varepsilon$ may not be reached within finite time.

4.4. The rate of convergence of Algorithm 1

In the regular case, Algorithm 1 converges with a linear rate, which follows from Theorem 4.3 below.

Theorem 4.3. In the k th iteration of Algorithm 1, we have

$$B_{k+1} - b_{k+1} \leq \left(1 - \alpha_0^{(k)}\right) (B_k - b_k), \quad (4)$$

where $\alpha_0^{(k)} > 0$ is the coefficient in the convex combination $\mathbf{O} = \sum_i \alpha_i^{(k)} \mathbf{a}_i$ in the k th iteration (the summation is over all $i = 0, \dots, n+1$ except for s). Recall that in the convex combination $\sum_i \alpha_i^{(k)} \mathbf{a}_i$, all coefficients $\alpha_i^{(k)}$ are supposed to be nonnegative and $\sum_i \alpha_i^{(k)} = 1$.

Proof. Without loss of generality it can be assumed that $s = n+1$, i.e., the k th iteration replaces $\mathbf{a}_{n+1}$ with $\mathbf{a}_0$. Denote $\alpha_i = \alpha_i^{(k)}$. We have $(\mathbf{p}_k, \mathbf{a}_i) = \sigma_i f(t_i) + b_k$, for all $i = 1, \dots, n$, and $(\mathbf{p}_k, \mathbf{a}_0) = |\mathbf{p}_k(t_0)| = r_k$. On the other hand, $(\mathbf{p}_{k+1}, \mathbf{a}_i) = \sigma_i f(t_i) + b_{k+1}$, $i = 0, \dots, n$. Taking the difference, we obtain $(\mathbf{p}_k - \mathbf{p}_{k+1}, \mathbf{a}_i) = b_k - b_{k+1}$, for all $i = 1, \dots, n$, and $(\mathbf{p}_k - \mathbf{p}_{k+1}, \mathbf{a}_0) = r_k - b_{k+1}$. Multiplying these equalities by α_i , computing the sum over all $i \leq n$, and taking into account that $\sum_{i=0}^n \alpha_i = 1$ we get

$$\begin{aligned} 0 &= \sum_{i=0}^n \alpha_i (\mathbf{p}_k - \mathbf{p}_{k+1}, \mathbf{a}_i) = \alpha_0 (r_k - b_{k+1}) + \sum_{i=1}^n \alpha_i (b_k - b_{k+1}) = \\ &= \alpha_0 r_k - b_{k+1} + (1 - \alpha_0) b_k. \end{aligned}$$

Thus, $\alpha_0 r_k - b_{k+1} + (1 - \alpha_0) b_k = 0$. Adding and subtracting $(1 - \alpha_0) B_k$ we obtain

$$(1 - \alpha_0) B_k + \alpha_0 r_k - b_{k+1} = (1 - \alpha_0) (B_k - b_k). \quad (5)$$

Since $\alpha_0 \in [0, 1]$, we get $B_{k+1} = \min \{B_k, r_k\} \leq (1 - \alpha_0) B_k + \alpha_0 r_k$. Thus, we see that the left-hand side of (5) is bigger than or equal to $B_{k+1} - b_{k+1}$, which concludes the proof. $\square$

Let us recall the regularity assumption: in all iterations we have $\alpha_0 \geq \mu$, where $\mu > 0$ is fixed number.

Theorem 4.4. In the regular case, Algorithm 1 converges linearly with the regularity parameter μ :

$$B_{k+1} - b_{k+1} \leq (1 - \mu)^k (B_1 - b_1), \quad k \in \mathbb{N}.$$

Remark 4.5. Equation (5) and inequality $B_{k+1} - b_{k+1} \leq (1 - \alpha_0) B_k + \alpha_0 r_k - b_{k+1}$ show that if α_0 is small, then the k th iteration only slightly changes the difference $B_k - b_k$. In this case, the algorithm slows down. This happens if the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ are almost linearly dependent and the linear system $\mathbf{a}_0 = \sum_{i=0}^{n+1} x_i \mathbf{a}_i$ is ill-conditioned. Those problems can be resolved by the regularization procedure presented in the next section.

Remark 4.6. From the proof of Theorem 4.3, we see that it is not actually necessary to find the point t_0 of maximum of $|\mathbf{p}_k(t) - f(t)|$ on K at each iteration. It suffices to choose an arbitrary point $\hat{t}$ where the value $|\mathbf{p}_k(\hat{t}) - f(\hat{t})|$ is significantly bigger than that at the point t_k . This may be done, for instance, by a random search. This strategy can be applied when finding the point of maximum of $|\mathbf{p}_k(t) - f(t)|$ is a difficult problem. For example, this situation occurs in the multivariate case.

4.5. Algorithm in the general case. Regularization

The regularization procedure is needed when the simplex Δ_k becomes close to degenerate. In this case, we apply an essentially different method for choosing the next point $t \in K$ at each iteration. This can be realized under the additional assumptions:

Assumptions. $K \subset \mathbb{R}^N$ is a convex body, and all the functions φ_i are nonzero analytic on K .

We begin with auxiliary results. In this section, all lengths and distances in $\mathbb{R}^d$ are measured in the Euclidean norm, which is denoted by $\|\cdot\|_2$.

Lemma 4.7. Let vectors $\mathbf{u}(t_1), \dots, \mathbf{u}(t_{n-1})$, where $\mathbf{u}(t) = (\varphi_1(t), \dots, \varphi_n(t))^T$, span a hyperplane $H \subset \mathbb{R}^n$. Then there exists a unique, up to the sign, polynomial $q \in \mathcal{P}$ normalized by the condition $\|q\|_2 = 1$, which vanishes at the points $t_1, \dots, t_{n-1}$. For every $t \in K$, the distance from $\mathbf{u}(t)$ to H is equal to $|q(t)|$.

Proof. If $\mathbf{q}$ is a unit normal vector to H , then for the corresponding polynomial q (polynomial from $\mathcal{P}$ with coefficients $\mathbf{q}$), we have $q(t_i) = (\mathbf{q}, \mathbf{u}(t_i)) = 0$. If another polynomial $y(\cdot) \in \mathcal{P}$, $\|y\|_2 = 1$ vanishes at the points t_i , then its vector $\mathbf{y}$ is orthogonal to all $\mathbf{u}(t_i)$ and hence, is orthogonal to H , i.e., collinear to $\mathbf{q}$. Thus, the polynomial q is unique up to the sign. The distance from $\mathbf{u}(t)$ to H is equal to $|(\mathbf{q}, \mathbf{u}(t))| = |q(t)|$. $\square$

Lemma 4.7 expresses the distance from a moment vector to a hyperplane in terms of a special polynomial q . This trick will be applied in the regularization method.

Let a set of points $\mathcal{T} = \{t_i\}_{i=1}^{n+1}$, where $n \geq 2$, and polynomial $p \in \mathcal{P}$ be such that the system of vectors $\{\mathbf{u}(t_i)\}_{i=1}^{n+1}$ is nondegenerate (consequently, $\{\mathbf{a}_i\}_{i=1}^{n+1}$ is nondegenerate), $p(t_i) - f(t_i) = \sigma_i d$, $\sigma_i \in \{-1, 1\}$ for all i , where $d \geq 0$, $r = \|p - f\|_{C(K)}$ and $r > d$. For each pair i_1, i_2 such that $1 \leq i_1 < i_2 \leq n+1$, let $q_{i_1 i_2} \in \mathcal{P}$ be the polynomial such that $\|q_{i_1 i_2}\|_2 = 1$ and $q_{i_1 i_2}(t_s) = 0$ for every $s \notin \{i_1, i_2\}$. By Lemma 4.7, this polynomial exists and is unique up to the sign. For a given $v \in [0, 1)$, we denote by t_v arbitrary solution of the following univariate optimization problem.

$$\begin{cases} g(t) = \sum_{i_1 < i_2} \frac{1}{|q_{i_1 i_2}(t)|^2} \rightarrow \min, \\ |p(t) - f(t)| \geq (1-v)r + v d, \quad t \in K. \end{cases} \quad (6)$$

A solution always exists, and the optimal value is finite. Indeed, the objective function $g(t)$ is lower semicontinuous on K (if we set $g(t) = +\infty$ at the points where it is not defined). Moreover, $g(t)$ is equal to $+\infty$ only at a finite number of points because the functions φ_i are analytic, while the feasible set is an infinite compact set.

Theorem 4.8. For $v \in [0, 1)$, let $\rho(v)$ be the maximal number such that there exists $\hat{t} \in K$ for which $|p(\hat{t}) - f(\hat{t})| \geq (1-v)r + v d$ and the point $\mathbf{u}(\hat{t})$ is bounded away by the distance of at least $\rho(v)$ from each of $\frac{n(n+1)}{2}$ hyperplanes spanned by $n-1$ vectors from $\{\mathbf{a}_i\}_{i=1}^{n+1}$. Then

$$\frac{1}{\sqrt{g(t_v)}} < \rho(v) \leq \sqrt{\frac{n(n+1)}{2}} \frac{1}{\sqrt{g(t_v)}}, \quad (7)$$

where t_v is an arbitrary solution of problem (6). Moreover, the system $\{\mathbf{a}_1, \dots, \mathbf{a}_{n+1}, \mathbf{u}(t_v)\}$ is nondegenerate.

Proof. By Lemma 4.7, the distance from $\mathbf{u}(\hat{t})$ to the hyperplane spanned by all vectors $\mathbf{u}(t_i)$ except for $\mathbf{u}(t_{i_1}), \mathbf{u}(t_{i_2})$, where $1 \leq i_1 < i_2 \leq n+1$, is equal to $|q_{i_1 i_2}(\hat{t})|$. Thus, by the definition of the function $g(t)$, we have $g(\hat{t}) \leq \frac{n(n+1)}{2\rho^2}$. Hence, $\rho \leq \sqrt{\frac{n(n+1)}{2}} \frac{1}{\sqrt{g(\hat{t})}} \leq \sqrt{\frac{n(n+1)}{2}} \frac{1}{\sqrt{g(t_v)}}$, which proves the upper bound in (7). Furthermore, $g(t_v) > \max_{i_1 < i_2} \frac{1}{|q_{i_1 i_2}(t_v)|^2}$, hence the minimal number $|q_{i_1 i_2}(t_v)|$ over all i_1, i_2 is greater than $\frac{1}{\sqrt{g(t_v)}}$.

In view of Lemma 4.7, this means that the minimal distance from $\mathbf{u}(t_v)$ to those $\frac{n(n+1)}{2}$ hyperplanes is greater than $\frac{1}{\sqrt{g(t_v)}}$, which completes the proof of (7).

If the system $\{\mathbf{a}_1, \dots, \mathbf{a}_{n+1}, \mathbf{u}(t_v)\}$ is degenerate, then it contains a subset of n linearly dependent vectors, one of which must be $\mathbf{u}(t_v)$, because the system $\{\mathbf{a}_i\}_{i=1}^{n+1}$ is nondegenerate. Hence, $\mathbf{u}(t_v)$ must belong to the

linear span of other $n - 1$ vectors of this subset, in which case $g(t_v) = +\infty$. This contradicts the fact that the optimal value of problem (6) is finite. $\square$

Now we present Algorithm 2 based on the aforementioned regularization. It coincides with Algorithm 1 with one crucial modification: we fix $\nu \in [0, 1)$, in each iteration solve problem (6) for the set of points $\mathcal{T} = \mathcal{T}_k$ and take the optimal point t_ν instead of t_0 (the point of maximum of $|p(t) - f(t)|$). Respectively, r_k is replaced by $(1 - \nu)r_k + \nu b_k$.

Algorithm 2: General Case

Require: A function $f \in C(K)$; a system of linearly independent real-valued functions $\Phi = \{\varphi_i\}_{i=1}^n$ from $C(K)$; real numbers $\nu \in [0, 1)$, $\varepsilon > 0$

Ensure: Real numbers b, B and a polynomial $\hat{p} \in \mathcal{P} \equiv \text{span}\{\varphi_1, \dots, \varphi_n\}$ such that $B - b < \varepsilon$, $b \leq \text{dist}(f, \mathcal{P}) \leq B$ and $b \leq \|\hat{p} - f\| \leq B$

Initialization: Take an arbitrary set $\mathcal{T} = \{t_i\}_{i=1}^{n+1}$ of $n + 1$ different points on K such that the system of moment vectors $\{\mathbf{u}(t_i)\}_{t_i \in \mathcal{T}}$ is nondegenerate (see Section 4.1). Then the system of n linear equations $\sum_{i=1}^{n+1} x_i \mathbf{u}(t_i) = 0$ possesses a unique, up to scalar multiple, solution $\mathbf{x} = (x_1, \dots, x_{n+1})$ with all nonzero components. If σ_i is a sign of x_i , $\alpha_i = |x_i|$ and $\mathbf{a}_i = \sigma_i \mathbf{u}(t_i)$, then

$$\sum_{i=1}^{n+1} x_i \mathbf{u}(t_i) = \sum_{i=1}^{n+1} |x_i| \sigma_i \mathbf{u}(t_i) = \sum_{i=1}^{n+1} |x_i| \mathbf{a}_i = \sum_{i=1}^{n+1} \alpha_i \mathbf{a}_i = 0.$$

Now find a vector $\mathbf{p} \in \mathbb{R}^n$ and a number d from the linear system

$$(\mathbf{p}, \mathbf{u}(t_i)) = f(t_i) + \sigma_i d, \quad i = 1, \dots, n + 1.$$

The corresponding polynomial p possesses the property $p(t_i) = f(t_i) + \sigma_i d$, $i = 1, \dots, n + 1$. If $d < 0$, then change all signs σ_i . Thus, $d \geq 0$. Set $k = 1$, $\mathcal{T}_1 = \mathcal{T}$, $p_1 = p$, $b_1 = d$, $B_1 = \|p_1 - f\|$.

Main loop:

while $B_k - b_k > \varepsilon$ **do**

At k -th iteration we have a polynomial $p_k \in \mathcal{P}$, a set of $n + 1$ points $\mathcal{T}_k = \{t_i^{(k)}\}_{i=1}^{n+1}$ (for the sake of simplicity we denote $t_i^{(k)} = t_i$) and two numbers $B_k \geq b_k \geq 0$ such that:

1) $p_k(t_i) = f(t_i) + \sigma_i b_k$, where $\sigma_i \in \{-1, 1\}$, $i = 1, \dots, n + 1$;

2) the simplex $\Delta_k = \text{co}\{\mathbf{a}_i\}_{i=1}^{n+1}$, where $\mathbf{a}_i = \sigma_i \mathbf{u}(t_i)$ contains the origin inside: $\mathbf{O} = \sum_{i=1}^{n+1} \alpha_i^{(k)} \mathbf{a}_i$ with all $\alpha_i^{(k)}$ positive and $\sum_{i=1}^{n+1} \alpha_i^{(k)} = 1$.

Let $r_k = \|p_k - f\|$, $\alpha_i = \alpha_i^{(k)}$. Solve the problem (6) for the set of points $\mathcal{T} = \mathcal{T}_k$ and polynomial $p = p_k$ and take the optimal point t_ν . Let σ_ν be the sign of $p_k(t_\nu) - f(t_\nu)$ and $\mathbf{a}_\nu = \sigma_\nu \mathbf{u}(t_\nu)$.

Now find the facet Δ_k^s of Δ_k intersected by the negative extension of $\mathbf{a}_\nu$ (see Section 4.1). To this end, we construct the linear system $\mathbf{a}_\nu = \sum_{i=1}^{n+1} x_i \mathbf{a}_i$. By Kronecker–Capelli theorem (see [8]) there exists a solution $\mathbf{x} = (x_1, \dots, x_{n+1})$ (not unique, we take any of them) to this system. Denote by $\tau = \frac{x_s}{\alpha_s}$, the maximal ratio $\frac{x_i}{\alpha_i}$. Then $\mathbf{a}_\nu = \sum_{i=1}^{n+1} (x_i - \tau \alpha_i) \mathbf{a}_i$, and the coefficient $x_s - \tau \alpha_s$ vanishes, while all other coefficients are negative. Indeed, if at least one other coefficient $x_j - \tau \alpha_j$ were equal to zero, then the system of n vectors $\{\mathbf{a}_\nu\} \cup \{\mathbf{a}_i\}_{i=1, i \neq j, s}^{n+1}$ would be linearly dependent. Consequently, the system $\{\mathbf{a}_1, \dots, \mathbf{a}_{n+1}, \mathbf{a}_\nu\}$ would be degenerate, which contradicts Theorem 4.8. Hence, the negative extension of $\mathbf{a}_\nu$ intersects the facet Δ_k^s opposite to $\mathbf{a}_s$.

Thus, replacing $\mathbf{a}_s$ by $\mathbf{a}_\nu$ we obtain a new simplex Δ_{k+1} containing the origin. Replacing t_s by t_ν we obtain a new system of points $\mathcal{T}_{k+1}$. Redenote the points of the set $\mathcal{T}_{k+1}$ by $\{t_i\}_{i=1}^{n+1}$ in an arbitrary order. Then we find a polynomial p_{k+1} and a number $b_{k+1} > 0$ from the linear system $p_{k+1}(t_i) = f(t_i) + \sigma_i b_{k+1}$, $i = 0, \dots, n + 1$, $i \neq s$. The system has a unique solution because the new system of moment vectors $\{\mathbf{u}(t_i)\}$, $t_i \in \mathcal{T}_{k+1}$, is also nondegenerate. Set $k = k + 1$, $B_{k+1} = \min\{(1 - \nu)r_k + \nu b_k, B_k\}$.

end while

return Set $b = b_N$, $B = B_N$ and $\hat{p} = p_k$, where $k \leq N$ is the number of iteration, for which $\|p_k - f\| = B$.

Remark 4.9. The key difference between Algorithms 1 and 2 is the choice of the new point at each iteration. Algorithm 1 takes the point t_0 of the maximal distance $r_k = |p_k(t_0) - f(t_0)|$. Algorithm 2, instead, considers all points t , where this distance is big enough, i.e., $|p_k(t) - f(t)| \geq (1 - \nu)r_k + \nu d_k$. On this set we find the point of minimum t_ν of the function $g(t)$. In other words, Algorithm 2 aims to guarantee the nongeneracy of the newly constructed simplex Δ_{k+1} provided that the value $|p_k(t_\nu) - f(t_\nu)|$ was significantly larger than b_k . The nondegeneracy, in turn, will guarantee a sufficient descent of the value $B_k - b_k$ on the next iterations.

The value $g(t_\nu)$ can be estimated from above. Under our assumptions (all φ_i are analytic and linearly independent, K is a convex body, f is Lipschitz), it has the power growth in $1/|r_k - d_k|$.

Proposition 4.10. We have $g(t_\nu) = O(|r_k - d_k|^{-2m})$, where m is the maximal multiplicity of zeros of polynomials from $\mathcal{P}$ (this multiplicity is finite because all the functions φ_i are analytic).

Proof. Without loss of generality, we set $\nu = 0.5$. Suppose $M = \max_{p \in \mathcal{P}, \|p-f\| \leq r_k} \|p'\|_{C(K)}$ and L is the Lipschitz constant of f on K , then

$$|p(t) - f(t)| \geq |p(t_0) - f(t_0)| - |p(t) - p(t_0)| - |f(t) - f(t_0)| \geq r_k - (L + M)|t - t_0| \quad \forall t \in K.$$

Thus, the inequality $|p(t) - f(t)| \geq \frac{1}{2}(r_k + d_k)$ is valid for all t from the ball $B(t_0, R)$ of the radius $R = \frac{1}{2(M+L)}(r_k - d_k)$ centered at t_0 . On the other hand, there are constants $C, R_0 > 0$ such that $|q_{i_1 i_2}(t)| > C|t - t_0|^m$, whenever $|t - t_0| < R_0$. Then for every $t \in K$ such that $|t - t_0| < \min\{R, R_0\}$, we have $g(t) = O(|r_k - d_k|^{-2m})$, which is required. $\square$

Practical issue. In all numerical results, we set $\nu = 0.5$. In practice, it often suffices to apply a simple procedure, without solving any optimisation problem. We fix a small $\delta > 0$ and, at every iteration, compute n values: $|q_{s,j}(t_0)|$, $j \neq s, j = 1, \dots, n+1$. Here t_0 is the point of maximum of $|p_k(t) - f(t)|$ on K , and s is the index of the point t_s that would be replaced by t_0 in the k th iteration if Algorithm 1 were applied. Let us recall that $q_{s,j}$ is the polynomial with zeros at the points t_i , $i \notin \{s, j\}$. If $|q_{s,j}(t_0)| < \delta$ for some j , we find a point $\bar{t}$ for which $|p_k(\bar{t}) - f(\bar{t})| \geq \frac{1}{2}(r_k + d_k)$ and $|q_{s,j}(\bar{t})| > \delta$. This point can be found, for example, by the penalty function method and we then set $t_\nu = \bar{t}$.

Remark 4.11. The regularization moves the point a_0 away from all hyperplanes spanned by the other $(n+1)$ vectors a_i . Hence, it always keeps the simplices Δ_k nondegenerate. This, however, does not imply that they are degenerate uniformly for all k and that the parameter $\alpha_0^{(k)}$ can be bounded away from zero by some positive constants. Although, a vast majority of the numerical experiments demonstrate the uniform nondegeneracy of those simplices, and, respectively, linear convergence of Algorithm 2. Nevertheless, we do not have a proof of its convergence for any initial data and can only formulate the corresponding conjecture. According to the numerical experiments, Algorithm 1 handles from 20% to 80% (depending on the problem) cases without suffering by degeneracy. The remaining cases need the regularization and are always handled by Algorithm 2. We report the numerical results in Section 7.

5. Approximation under linear constraints

Unlike the classical Remez algorithm, the method introduced in Sections 3 and 4 is easily extended to the constrained approximation. For given linear functionals ℓ_j and numbers b_j , $j = 1, \dots, r$, we consider the problem (2). As before, f is continuous on a compact set K , $\mathcal{P} = \text{span}\{\varphi_1, \dots, \varphi_n\}$. Clearly, it can be assumed that all ℓ_j are linearly independent and, therefore, $r \leq n-1$.

A Chebyshev system may lose its properties after imposing a linear constraint (Example 7.2). However, the concept of generalized alternance works for the constrained problems as well.

We denote $\tilde{L} = \{\ell_j\}_{j=1}^r$, $L = \text{span}\{\ell_j\}_{j=1}^r$, $L^\perp = \{p \in \mathcal{P} : \ell_1(p) = \dots = \ell_r(p) = 0\}$ and, as usual, identify $\mathcal{P}$ with $\mathbb{R}^n$. Every functional ℓ_j is associated to the vector $\ell_j \in \mathbb{R}^n$ such that $\ell_j(p) = (\ell_j, p)$. The orthogonal complement to the vectors $\ell_1, \dots, \ell_r$ will be denoted by the same symbol $L^\perp$. Since ℓ_j are linearly independent, it follows that $\dim L^\perp = n - r$. By projection, we always mean an orthogonal projection. The projections of φ_i onto $L^\perp$ are assumed to be independent, otherwise, we remove redundant functions from $\Phi = \{\varphi_i\}_{i=1}^n$.

Definition 5.1. Let $\mathcal{P}$ be the space of polynomials by a system $\Phi = \{\varphi_i\}_{i=1}^n$, $\tilde{L} = \{\ell_j\}_{j=1}^r$, $r < n$, be a set of linearly independent linear functionals on $\mathcal{P}$. For a given polynomial $p \in \mathcal{P}$, a set of points $\{\tau_1, \dots, \tau_m\} \subset K$, $m \leq n - r + 1$, is called *L-alternance* if

- 1) for each $i = 1, \dots, m$, we have $p(\tau_i) - f(\tau_i) = \sigma_i \|p - f\|$, $i = 1, \dots, m$, where $\sigma_i \in \{-1, 1\}$.
- 2) the convex hull of the projections of the vectors $\{\sigma_i \mathbf{u}(\tau_i)\}_{i=1}^m$ onto $L^\perp$ contains the origin.

In the case $\|p - f\| = 0$, in the Definition 5.1, for each $\tau \in K$, instead of taking a single vector $\sigma(\tau)\mathbf{u}(\tau)$, we take a pair of vectors $\mathbf{u}(\tau), -\mathbf{u}(\tau)$.

The proof of the following generalization of Theorem 3.2 is literally the same.

Theorem 5.2. A polynomial $p \in \mathcal{P}$ is the best polynomial approximation for a function f under the constraints $\ell_j(p) = b_j$, $j = 1, \dots, r$, if and only if it has an *L-alternance*.

In case of non-uniqueness, all polynomials of best approximation are characterised by a common *L-alternance*. The proof of the following theorem is the same as for the corresponding Theorem 3.4:

Theorem 5.3. For every $f \in C(K)$, all polynomials of best approximation from $\mathcal{P}$ under the constraints $\ell_j(p) = b_j$, $\ell_j \in \tilde{L}$, have the same *L-alternance* and the same values on it.

Algorithms 1 and 2 are transferred to the constrained problem (2) merely by replacing all vectors $\mathbf{a}_i$ by their projections onto $L^\perp$. Respectively, the dimension n is replaced by $n - r$ and $\mathbb{R}^n$ is replaced by $L^\perp$. In Algorithm 2 for the constrained problem, the polynomial $q_{i_1 i_2}$ is defined by the set of its zeros t_i , $i \in \{1, \dots, n - r + 1\} \setminus \{i_1, i_2\}$ and by the conditions $\ell_j(q_{i_1 i_2}) = 0$, $j = 1, \dots, r$. We refer to those algorithms as Algorithms 1c and 2c, respectively (“c” is from “constrained”).

Remark 5.4. It is possible to obtain some generalizations of Theorem 5.2 to nonlinear smooth constraints ℓ_j . In this case, the subspace L will be spanned by the derivatives $\ell'_j(p)$. The analogue of Theorem 5.2 is valid in the direction of “necessity”: every optimal polynomial do possess an *L-alternance*, but the converse is, in general, not true. It becomes true if all ℓ_j are convex and all equality constraints are replaced by inequalities $\ell_j(p) \leq b_j$. In this case, we have a complete analogue of Theorem 5.2 with the subspace L spanned by the derivatives $\ell'_j(p)$ of active constraints (computed at the point p).

However, for Algorithms 1c and 2c, the linearity of functionals ℓ_j is significant. The reason is that if some of the functionals $\ell_j(p)$ are nonlinear, then the subspace L depends on p . Therefore, the subspaces $L(p_k)$ and $L(p_{k+1})$ in Algorithms 1c can be different. The key requirement of the algorithm that the projections of the simplex Δ_k onto $L^\perp$ contains the origin may not be preserved in k th iteration.

Remark 5.5. All our results are valid for arbitrary closed bodies K (and convex in Algorithms 2 and 2c), not necessarily bounded, provided that all φ_i and f tend to zero as $t \in K \rightarrow \infty$. Indeed, let us take any $p_0 \in \mathcal{P}$ and consider the set $\mathcal{P}_0 = \{p \in \mathcal{P} : \|p - f\|_{C(K)} \leq \|p_0 - f\|_{C(K)}\}$. This set is compact in $\mathcal{P}$. Hence, all polynomials $p \in \mathcal{P}_0$ uniformly tend to zero on this set as $t \rightarrow \infty$. Thus, there exists a closed ball B such that the best polynomial approximation of f on B coincides with that on K . This reduces our problem to the compact set B . The criteria of the best approximation in terms of the generalized alternance and *L-alternance*, as well as the algorithm are extended to closed unbounded domain.

6. Examples and numerical results

The code can be found on the website

<https://gitlab.com/users/rinatik7.kamalov.7/projects>.

6.1. Approximation by shifts of Gaussian functions

Unconstrained problems. We approximate the signal

$$f(t) = \frac{1}{10}(t-5)^2 + \frac{1}{2}(t-4) + \sin(0.4 \cdot t^2 \cos(0.5t)),$$

with polynomials by the system $\Phi = \{e^{\frac{-(t-1)^2}{9}}, e^{\frac{-(t-5)^2}{9}}, e^{\frac{-(t-7)^2}{9}}\}$ on the segment $K = [0, 8]$ with $\varepsilon = 10^{-6}$. Eight iterations of the Algorithm 2 yield the required accuracy (Fig.2). The polynomial of best approximation $p(t)$ has corresponding coefficients $\mathbf{p} = (1.902091, -2.453699, 3.842463)^T$ with the distance

$\|f - p\| = 1.254985$. The generalized alternance consists of points $\tau_1 = 0.517919$, $\tau_2 = 4.430493$, $\tau_3 = 5.992115$, $\tau_4 = 7.942944$.

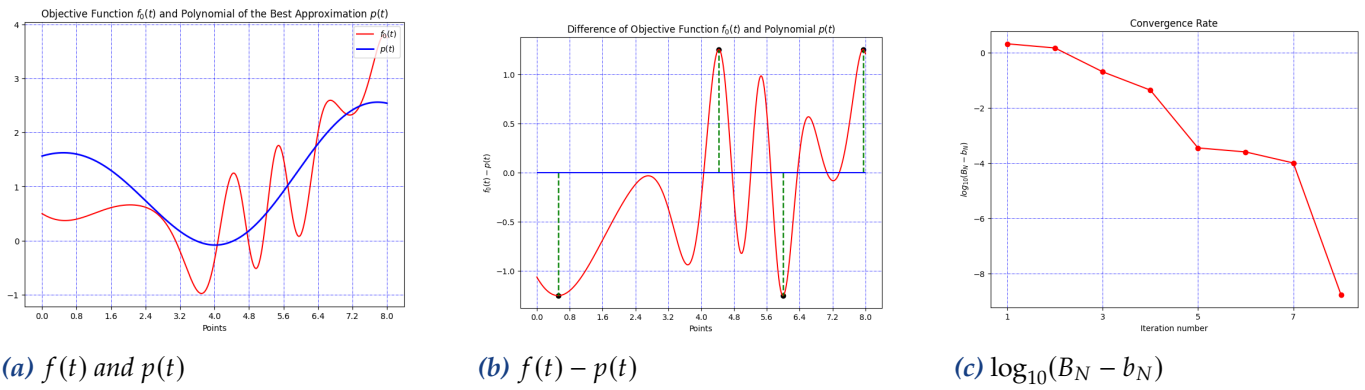


Figure 2

A problem with linear constraints. In the previous example, the value of the approximated function $f(t)$ at the point $t = 6.4$ is 1.999... while for the polynomial of best approximation, is $p(6.4) = 1.78...$ If we need to approximate f and keep its value at $t = 6.4$, which we round to 2 for simplicity, we come to the constrained problem

$$\begin{cases} \|f - p\|_{C[0,8]} \rightarrow \min \\ l(p) = p(6.4) = 2 \\ p \in \mathcal{P} \end{cases} \quad (8)$$

The results of Algorithm 2c are shown in the Fig. 3.

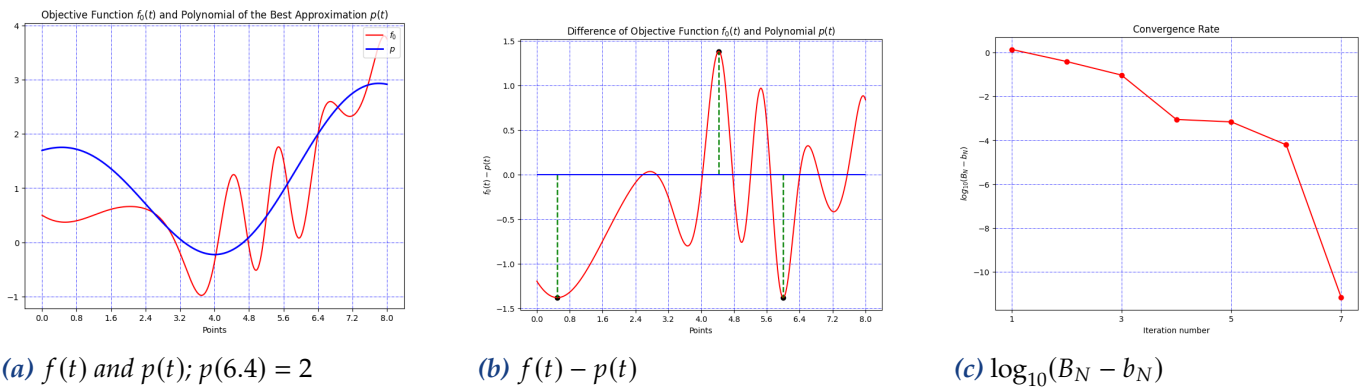


Figure 3

The polynomial of best approximation $p(t)$ has corresponding coefficients $p_1 = 2.078450$, $p_2 = -2.939696$, $p_3 = 4.457802$ with the minimal distance $\|f - p\| = 1.3807$. The L -alternance comprises $\tau_1 = 0.500162$, $\tau_2 = 4.427931$, $\tau_3 = 5.998317$.

Now add an extra condition: the value of the derivative of $p(t)$ at the same point $t = 6.4$ coincides with that of f , which is approximately 4.47. Thus, we add the constraint $l_2(p) = p'(6.4) = 4.47$.

The Algorithm 2c produces the following results (see Fig. 4). The coefficients of the best polynomial approximation are $p_1 = 7.407235$, $p_2 = -12.84065$, $p_3 = 12.52896$ with the minimal distance $\|f - p\| = 5.614225$. The L -alternance consists of two points $\tau_1 = 4.430836$, $\tau_2 = 0.386453$.

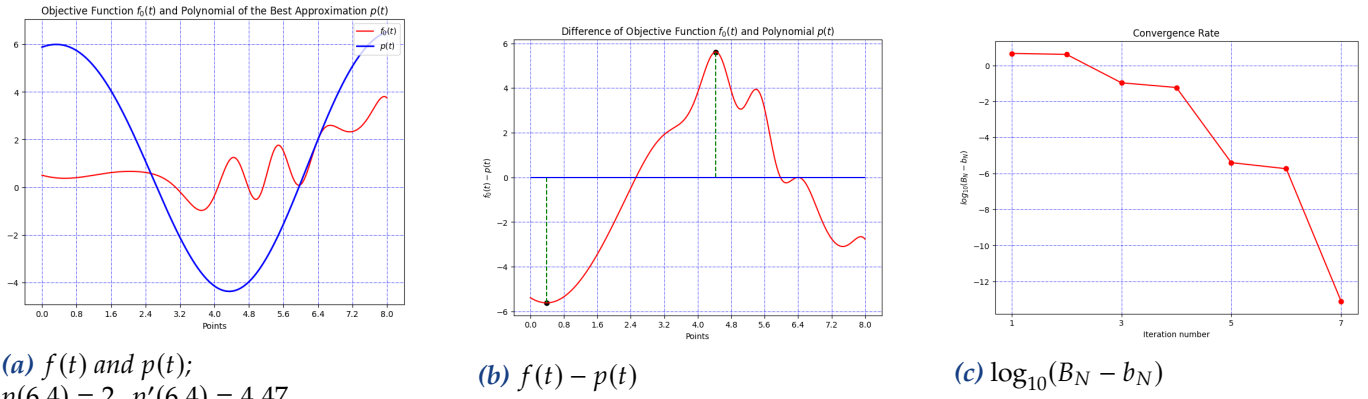


Figure 4

6.2. Signal recovery by complex exponentials

Unconstrained problems. We consider the space $\mathcal{P}$ of exponential polynomials on $[0, +\infty)$:

$$p(t) = e^{-0.1t} (a_3 \cos(0.2t) + a_4 \sin(0.2t) + a_5 \cos(0.3t) + a_6 \sin(0.3t)) + e^{-0.5t} (a_1 \cos(0.4t) + a_2 \sin(0.4t)) + e^{-0.9t} (a_7 \cos t + a_8 \sin t) + a_9 e^{-0.3t}$$

and a signal f , which is a sum of these exponentials with coefficients:

$$(a_k)_{k=1}^9 = (1, 1, 4, -7, -3, -2, 1, 5, 6)$$

and of the noise $8e^{\frac{-|t-7|}{2}}$. Algorithm 2 finds the polynomial of best approximation $p \in \mathcal{P}$ with the precision $\varepsilon = 10^{-8}$ within $k = 31$ iterations, Fig. 5. The distance $\|p - f\|_{C(\mathbb{R}_+)} = 1.318352$ (we keep six digits after the decimal point). The algorithm converges slower than in the previous example (subsection 6.1) because of the unbounded domain and a slow decrease of the functions $\varphi_3, \varphi_4, \varphi_5, \varphi_6$ (with the exponential part $e^{-0.1t}$).

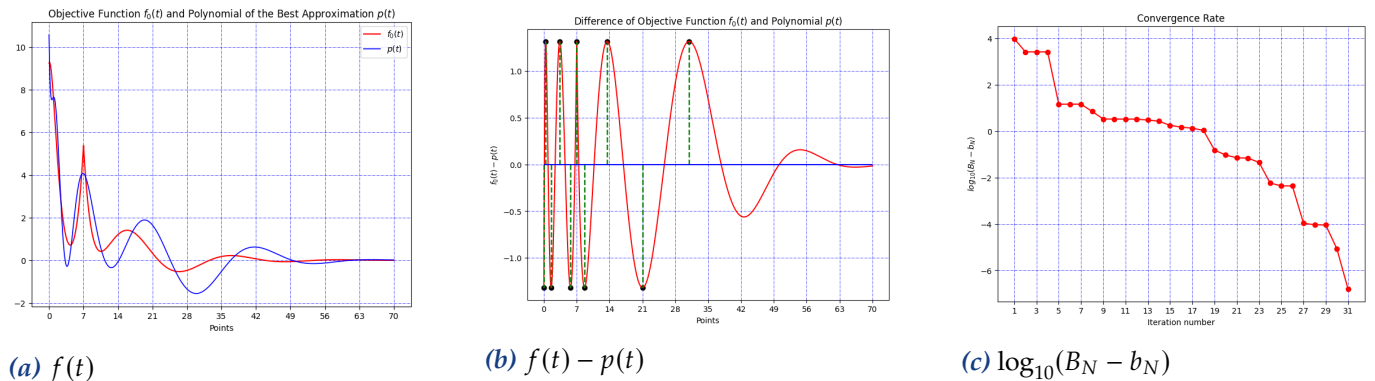


Figure 5

Problems with linear constraints. The same problem under the constraint $\int_{\mathbb{R}_+} p dt = 1$ gives the following results (see Fig. 6). The distance to the best approximation $\|p - f\|$ increases in more than 1.5 times, from 1.318352 to 2.104564. The best polynomial approximation p now possess a degenerate L -alternance that consists of 5 points instead of $n + 1 - r = 9$. Because of the degeneracy, Algorithm 1c stacks on this problem and we have to apply the regularization. Algorithm 2c finds the solution within 43 iterations with the precision $\varepsilon = 10^{-6}$.

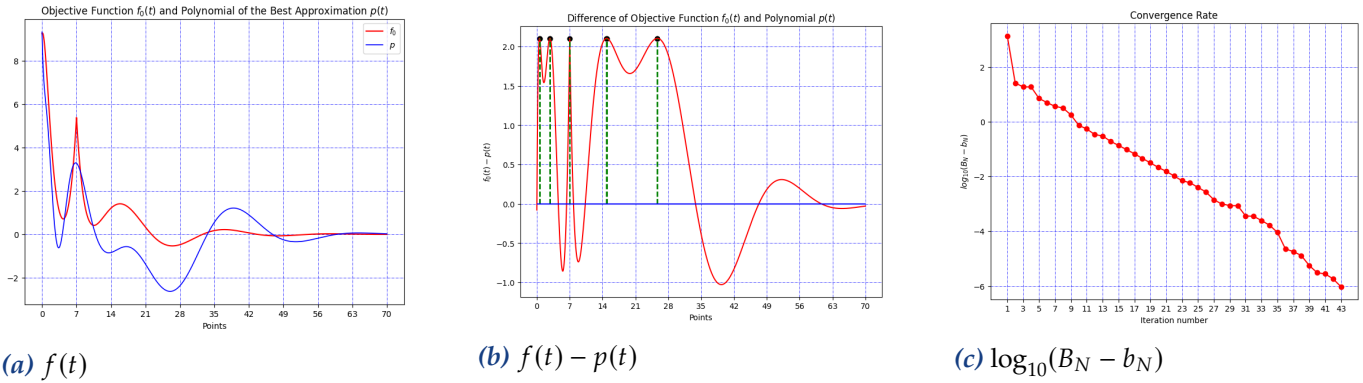


Figure 6

6.3. Recovery of highly non-stationary signal

The following signal on the segment $[0, 1]$ was analysed in [23]:

$$f(t) = \underbrace{\cos(4\pi\lambda(t)t)}_{\varphi_1(t)} + \underbrace{2 \sin(4\pi t)}_{\varphi_2(t)}, \quad (9)$$

where

$$\lambda(t) = \begin{cases} 4 + 32t & , \quad 0 \leq t \leq 0.5, \\ 4 + 32(1 - t) & , \quad 0.5 < t \leq 1. \end{cases}$$

The polynomial p of best approximation by the system $\Phi = \{\varphi_1, \varphi_2\}$ is, of course, equal to f , i.e., $\|p - f\|_{C[0,1]} = 0$. Algorithm 2 finds this solution in two iterations.

Approximating the same signal by the system $\phi_1 = 1, \phi_2 = \cos(4\pi t), \phi_3 = \sin(4\pi t)$ is shown in Fig. 7. The best approximation polynomial is $p(t) = 2 \sin(4\pi t)$, Algorithm 2 finds it within three iterations.

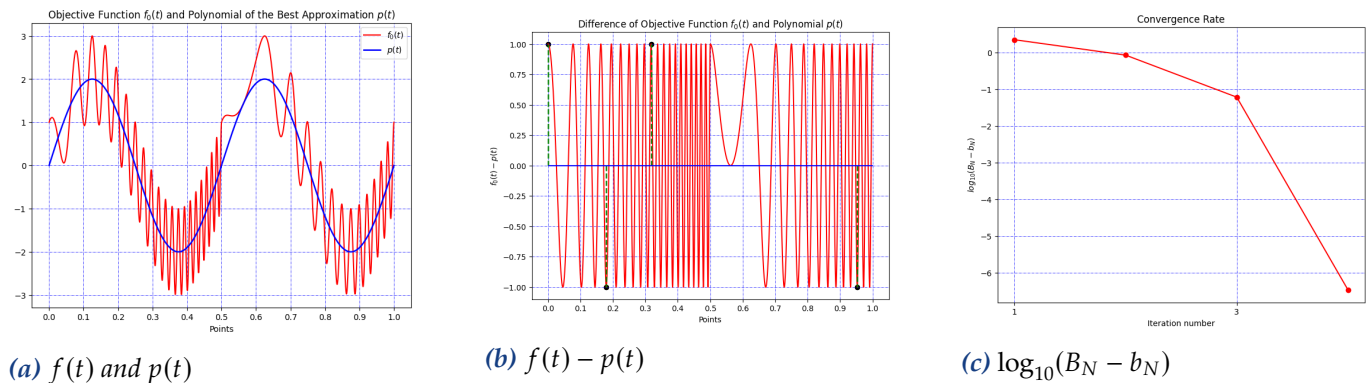


Figure 7

In fact, it decomposes the signal $f(t)$ into the sum of $2 \sin(4\pi t)$ and $f(t) - 2 \sin(4\pi t) = \varphi_1(t)$ as the Empirical Mode Decomposition (EMD) using a non-uniform mask [32, Sec. Numerical Experiments].

6.4. Markov-Bernstein type inequalities

We apply Algorithm 2c to find the sharp constant in the Markov-Bernstein inequality (subsection 7.3.2) for the space of quasipolynomials $\mathcal{P} = \{e^{-t}(a_1 \cos t + a_2 \sin t + a_3) \mid a_1, a_2, a_3 \in \mathbb{R}\}$. Algorithm 2c applied to the problem $\|p\|_{C(\mathbb{R}_+)} \rightarrow \min, p'(0) = 1$ returns the polynomial p with coefficients $a_1 = 1.006772, a_2 = 0.884983, a_3 = 1.121789$. By Proposition 7.6, the corresponding Markov-Bernstein constant is $C_1 = \frac{1}{\|p(t)\|_{C(\mathbb{R}_+)}} = 8.694367$. The alternance is nondegenerate and consists of $m = 3$ points, the precision $\varepsilon = 10^{-6}$ is achieved within 8 iterations of Algorithm 2c. The algorithm runs 1.7 sec. in a standard laptop.

6.5. How often does the degeneracy occur?

In our numerical experiments with polynomials by complex exponentials, shifts of Gaussian and Cauchy functions, lacunary algebraic and trigonometric polynomials, in 12 % of non-constrained problems degenerate simplices occur in Algorithm 1, after which the algorithm suffers and we had to apply the regularization and Algorithm 2. Moreover, 8 % of those cases have degenerate alternances, which consists of less than $n + 1$ points. For problems with linear constraints, there number of degenerate cases is much larger: in 82 % of cases degenerate simplices appear in Algorithm 2c, and 73 % of cases produce degenerate alternances.

In this section, we show the statistics over random system of functions, and it turns out to be quite similar. For this purpose, we generate m random independent uniformly distributed points (knots) on $[-1, 1]$, order them as $-1 < t_1 < \dots < t_m < 1$, then generate other m random points $v_1, v_2, \dots, v_m$ (the values of the function) on the same segment $[-1, 1]$, interpolate the points $(t_1, v_1), \dots, (t_m, v_m)$ with a cubic spline $\varphi_1(t)$. This is a unique interpolating $C^2[-1, 1]$ spline with the “not-a-knot” boundary conditions, i.e., the two extreme polynomials coincide on each side of the segment.

In the same way, we generate functions $\varphi_2, \dots, \varphi_n$. The boundary conditions are chosen so that the first and second segment at a curve end are the same polynomial. Having generated the system Φ on the segment $[-1, 1]$, we test Algorithm 2 with finding the best approximations for two functions: 1) $f(t) = |t|$ and 2) a random function f generated as above (in the same way as all φ_i). Then we test Algorithm 2c with the problem $\|p\|_{C[-1,1]} \rightarrow \min$ under the constraint $\ell(p) = \sum_{i=1}^n p^i = 1$. For each pair (m, n) we run 100 random systems Φ and then calculate the proportion of degenerate cases, average number of iterations and average program runtime for degenerate and non-degenerate cases. We call a case degenerate if at least in one iteration we have $q_{i_1, i_2}(t_0) < 0.05$ for some $i_1 \neq i_2$, where t_0 is the point of maximum of $|p_k(t) - f(t)|$ on the segment $[-1, 1]$. The polynomial q_{i_1, i_2} is normalized as $\|q_{i_1, i_2}\|_2 = 1$. The results are given in Table 1.

Table 1. The performance of Algorithms 2 and 2c for random polynomial systems

Problem	Parameters	Non-degenerate cases			Degenerate cases		
		share	iter.	time (sec)	share	iter.	time (sec)
$f = 0, \ell(p) = 1$	$m = 10, n = 3$	0.83	5.66	1.75	0.17	19.41	6.85
	$m = 10, n = 5$	0.59	12.02	4.14	0.41	18.95	7.26
	$m = 5, n = 7$	0.27	11.04	3.75	0.73	24.79	10.24
$f = t $	$m = 10, n = 3$	0.33	10.81	8.27	0.67	17.27	14.51
	$m = 10, n = 5$	0.17	15.5	10.38	0.83	22.12	21.79
	$m = 5, n = 7$	0.37	11.47	10.56	0.63	17.29	20.18
f is random	$m = 10, n = 3$	0.63	8.16	5.51	0.37	19.24	16.21
	$m = 10, n = 5$	0.36	13.74	8.13	0.64	26	17.85
	$m = 5, n = 7$	0.39	12.47	6.76	0.61	24.31	15.39

Several empirical conclusions can be made from Table 1. The share of degenerate cases increases in the degree n of the system. Also approximation of a non-smooth function (in our case $f = |t|$) admits more degenerate cases. The number of iterations in degenerate cases is, on average, from 1.5 to 2 times greater than in non-degenerate ones.

7. Appendix

7.1. Examples and an Auxiliary Lemma

The first example demonstrates that, if the system Φ is not a Chebyshev system, the polynomial of best approximation may be non-unique.

Example 7.1. Let $K = [-1, 1]$, $\Phi = \{t, t^2, \dots, t^n\}$. For the function $f = 1$, each of the functions t^k , $k = 1, \dots, n$, is a best approximation polynomial. Indeed, $p(0) = 0$ for all $p \in \mathcal{P}$, hence, the distance from f to $\mathcal{P}$ is at least one. On the other hand, it is equal to one for every $p_k(t) = t^k$. It is interesting that for each p_k , the generalized alternance is a singleton $\{0\}$. On the other hand, if n is even and $T_n = \cos(n \arccos t)$ is the Chebyshev polynomial, then

$p(t) = 1 - (-1)^{n/2}T_n$ is also a best approximation polynomial, and its alternance $\tau_k = \cos \frac{\pi k}{n}$, $k = 0, \dots, n$, is the same as for the Chebyshev polynomial.

If we impose a linear constraint on the problem of the best polynomial approximation, the classical alternation property (the Chebyshev - Vallée-Poussin theorem) may be lost even if the system $\{\varphi_i\}_{i=1}^n$ is Chebyshev, as demonstrated in the following example.

Example 7.2. Let $\mathcal{P}$ be the space of algebraic polynomials of degree n on the segment $[-1, 1]$. The problem of approximating the identical zero $f \equiv 0$ under the constraint $\ell(p) = p_0 = -1$ (the constant coefficient of p is equal to -1) is equivalent to the problem from Example 7.1. Hence, each of the functions t^k , $k \geq 1$, is a best approximation polynomial with the one-point alternance $\{0\}$. We see that the Chebyshev system $\{t^k\}_{k=0}^n$ loses its properties after imposing the linear constraint.

In the non-Chebyshev case, the polynomial of best approximation may be non-unique; however, as proved in Theorem 2, all these polynomials share a common (generalized) alternance. Example 3 shows that this alternance may be singleton.

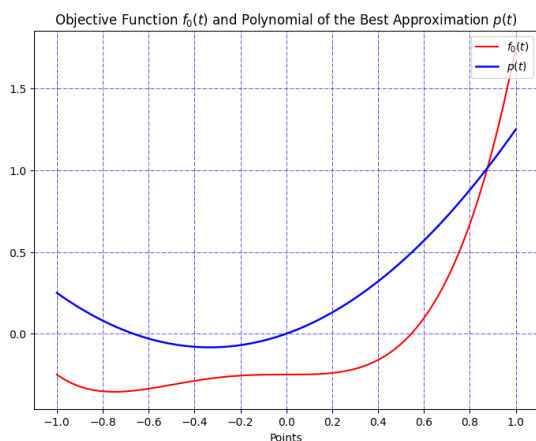
Example 7.3. (A one-point alternance). In Example 7.1, all polynomials $p(t) = \sum_{k=1}^n p^k t^k$ of the best approximation for $f \equiv 1$ on the segment $[-1, 1]$ have the same alternance $\{0\}$. On the other hand, some of them have other alternances. For instance, the polynomial $p(t) = 2t^2$ also gives the best approximation and it has an alternance $\{-1, 0, 1\}$; the polynomial $p(t) = 1 - (-1)^{n/2}T_n$, where n is an even number and T_n is the Chebyshev polynomial, have the alternance $\cos \frac{\pi k}{n}$, $k = 0, \dots, n$ (the same as T_n has).

In contrast to the classical alternance, the generalized one may not possess the sign alternation property, as demonstrated in the following example.

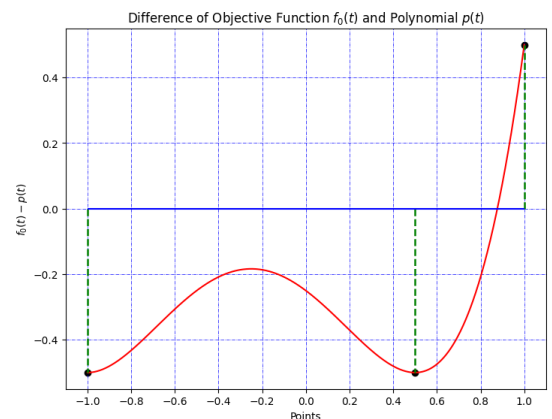
Example 7.4. (An alternance with non-alternating signs). Consider the space of quadratic polynomials without the constant term: $\Phi = \{\varphi_1, \varphi_2\} = \{t^2, t\}$ and approximate the function $f(t) = t^4 + t^3 - \frac{1}{4}$ on the segment $[-1, 1]$. Let us show that $p(t) = \frac{3}{4}t^2 + \frac{1}{2}t$ is a unique polynomial of the best approximation. We have

$$p(t) - f(t) = \frac{1}{2} - \left(t + 1\right)^2 \left(t - \frac{1}{2}\right)^2.$$

The modulus of this function reaches its maximum on $[-1, 1]$ at three points: $\tau_1 = -1$, $\tau_2 = \frac{1}{2}$ (at those points $p - f = \frac{1}{2}$), and $\tau_3 = 1$, where $p - f = -\frac{1}{2}$, see Fig. 8.



(a) $f(t)$ and $p(t)$



(b) $f(t) - p(t)$

Figure 8

This is an alternance because $\mathbf{a}_1 = (-1, 1)^T$; $\mathbf{a}_2 = (\frac{1}{2}, \frac{1}{4})^T$, $\mathbf{a}_3 = -(1, 1)^T$ and $\mathbf{0} \in \text{co}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. Thus, p is a polynomial of the best approximation. It is unique by Theorem 3.4. Indeed, p has a unique alternance, since τ_i are the only points of maximum of the function $|p - f|$ and there is no subset of those points for which the convex hull of the corresponding vectors $\mathbf{a}_i$ contains the origin. Consequently, another best approximation polynomial, if it exists, has the same alternance and the same values on it, which is impossible since there is only one quadratic function passing through three fixed points. Note that the difference $f - p$ has the same sign at τ_1 and τ_2 .

Lemma 7.5. *Suppose that the system of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{n+1}\}$ is nondegenerate (see subsection 4.1) and $\mathbf{O} \in \text{co}\{\mathbf{a}_1, \dots, \mathbf{a}_{n+1}\}$. Then the linear system*

$$(\mathbf{p}, \mathbf{a}_i) - d = f_i, \quad i = 1, \dots, n+1, \quad (10)$$

where $\mathbf{p} = (p^1, \dots, p^n)^T \in \mathbb{R}^n$ and $d \in \mathbb{R}$ are the variables, has a unique solution for each right-hand side $\mathbf{f} = (f_1, \dots, f_{n+1})^T \in \mathbb{R}^{n+1}$.

Proof. Consider points $\{\mathbf{a}_i\}_{i=1}^{n+1}$ satisfying the conditions of the lemma. If they are all contained in some affine hyperplane $L \subset \mathbb{R}^n$, then, since $\mathbf{O} \in \text{co}\{\mathbf{a}_i\}_{i=1}^{n+1}$, it follows that L is a linear subspace of $\mathbb{R}^n$, which contradicts the nondegeneracy of the vectors $\{\mathbf{a}_i\}_{i=1}^{n+1}$. Thus, the points $\{\mathbf{a}_i\}_{i=1}^{n+1}$ are affinely independent, which is equivalent to the fact that the linear system

$$(\mathbf{p}, \mathbf{a}_i) - d = 0, \quad i = 1, \dots, n+1,$$

with unknowns $\mathbf{p} = (p^1, \dots, p^n)^T \in \mathbb{R}^n$ and $d \in \mathbb{R}$, admits only the trivial solution. Therefore, by finite-dimensional Fredholm alternative, the system (10) has a unique solution $(p^1, \dots, p^n, d)^T$ for each right-hand side.

7.2. Algorithm scheme

Below, we provide a compact outline of Algorithms 1 and 2, which can be used for implementation. □

Algorithm: Pseudocode for Regular and General Cases**Require:** $f, \varphi_1, \dots, \varphi_n \in C(K)$; $\varepsilon > 0$ **Ensure:** $b, B \in \mathbb{R}$; $\hat{p} \in \mathcal{P} \equiv \text{span}\{\varphi_1, \dots, \varphi_n\}$ such that $B - b < \varepsilon$, $b \leq \text{dist}(f, \mathcal{P}) \leq B$ and $b \leq \|\hat{p} - f\|_{C(K)} \leq B$ **Initialization:**

- 1: Set $k = 1$ and take an array $\mathcal{T}_1 = [t_1, \dots, t_{n+1}]$ of $n + 1$ different points in K and a nonzero vector $v \in \mathbb{R}^{n+1}$ such that the matrix

$$U_1 = \begin{pmatrix} - & \mathbf{u}(t_1) & - & | \\ & \vdots & & \mathbf{v} \\ - & \mathbf{u}(t_{n+1}) & - & | \end{pmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}$$

is invertible. Usually, if one takes any $v \neq 0$ and $t_1, \dots, t_{n+1} \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}[0, 1]$, where $\mathcal{U}[0, 1]$ stands for uniform distribution on $[0, 1]$, this condition holds.

- 2: Take a solution $x = (x_1, \dots, x_{n+1})^T \in \mathbb{R}^{n+1}$ of the linear system

$$U_1^T x = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}, \quad \mathbf{0} = (0, \dots, 0)^T \in \mathbb{R}^n,$$

and set arrays $\Sigma_1 = [\sigma_1, \dots, \sigma_{n+1}]$, $\mathcal{A}_1 = [\alpha_1, \dots, \alpha_{n+1}]$, where $\sigma_i = \text{sign } x_i$, $\alpha_i = \frac{|x_i|}{\sum_{j=1}^{n+1} |x_j|}$.

- 3: Introduce the matrix (here $k = 1$)

$$A_k = \begin{pmatrix} - & \sigma_1 \mathbf{u}(t_1) & - & | \\ & \vdots & & -\mathbf{1}_{n+1} \\ - & \sigma_{n+1} \mathbf{u}(t_{n+1}) & - & | \end{pmatrix} \in \mathbb{R}^{(n+1) \times (n+1)},$$

where $-\mathbf{1}_{n+1} = (-1, \dots, -1)^T \in \mathbb{R}^{n+1}$ and solve the linear system

$$A_k \begin{pmatrix} p \\ d \end{pmatrix} = \sigma f, \quad p \in \mathbb{R}^n, d \in \mathbb{R}, \quad (11)$$

where $\sigma f = (\sigma_1 f(t_1), \dots, \sigma_{n+1} f(t_{n+1}))^T \in \mathbb{R}^{n+1}$. If $d < 0$, set $d = -d$ and $\Sigma_1 = -\Sigma_1$. Set $p_1 = p$, $b_1 = d$, $B_1 = \|p_1 - f\|_{C(K)}$.

Main loop:

- 4: **while** $B_k - b_k > \varepsilon$ **do**

- 5: In the regular case, find a point $t_0 \in K$ for which $|p_k(t_0) - f(t_0)| = \|p_k - f\|_{C(K)} \stackrel{\text{def}}{=} r_k$. In the general case (see subsection 4.5), solve problem (6) for the set of points $\mathcal{T} = \mathcal{T}_k$ and the polynomial $p = p_k$, and denote the solution by t_0 . In both cases, denote the sign of $p_k(t_0) - f(t_0)$ by σ_0 and $a_0 = \sigma_0 \mathbf{u}(t_0)$.

- 6: Take a solution $x = (x_1, \dots, x_{n+1})^T \in \mathbb{R}^{n+1}$ of the linear system

$$A_k^T x = \begin{pmatrix} a_0 \\ -1 \end{pmatrix}.$$

Let $\tau = \frac{x_s}{\alpha_s}$ be the maximal ratio $\frac{x_i}{\alpha_i}$, $i = 1, \dots, n + 1$. Set $\mathcal{T}_{k+1} = [t_1, \dots, \underbrace{t_0}_{s\text{-th index}}, \dots, t_{n+1}]$,

$$\Sigma_{k+1} = [\sigma_1, \dots, \underbrace{\sigma_0}_{s\text{-th index}}, \dots, \sigma_{n+1}], \quad \mathcal{A}_{k+1} = [\alpha_1 - \frac{x_1}{\tau}, \dots, \underbrace{\frac{1}{\tau}}_{s\text{-th index}}, \dots, \alpha_{n+1} - \frac{x_{n+1}}{\tau}].$$

- 7: Update the matrix A_{k+1} as follows: take the matrix A_k and replace its s -th row with the row vector $(\sigma_0 \mathbf{u}(t_0), -1) \in \mathbb{R}^{n+1}$.

- 8: Solve the linear system (11) for the matrix A_{k+1} and the vector $\sigma f = (\sigma_1 f(t_1), \dots, \underbrace{\sigma_0 f(t_0)}_{s\text{-th index}}, \dots, \sigma_{n+1} f(t_{n+1}))^T \in \mathbb{R}^{n+1}$ and update $p_{k+1} = p$, $b_{k+1} = d$, $k = k + 1$. In

the regular case, set $B_{k+1} = \min\{r_k, B_k\}$; in the general case, $B_{k+1} = \min\{(1 - \nu)r_k + \nu b_k, B_k\}$.

- 9: **end while**

return Set $b = b_N$, $B = B_N$ and $\hat{p} = p_k$, where $k \leq N$ is the number of iteration, for which $\|p_k - f\| = B$.

7.3. Applications

7.3.1. Structural recovery of signal

The problem of decomposing a continuous time signal $f(t)$ into the form

$$f(t) = \sum_{i=1}^n c_i \varphi(t, z_i),$$

where $\varphi(t, z_i)$ are functions from a certain class, $z_1, \dots, z_n$ are distinct complex or real numbers, $c_1, \dots, c_n$ are real numbers arises in a number of applications such as sparse array of arrival estimation [27], system theory for parametrized model reduction [24], radar [1], ultrasound [47], etc. For the case when $\varphi(t, z_i) = e^{2\pi z_i t}$ and z_i, c_i are unknown complex and real parameters, respectively, there are many methods. A popular approach is related to Prony's method [3, 4, 38]. If $\varphi(t, z_i) = \varphi(t - z_i)$, for example $\varphi(t, z_i) = e^{\frac{-(t-z_i)^2}{\sigma^2}}$ or $\varphi(t, z_i) = \frac{1}{1 + \frac{(t-z_i)^2}{\sigma^2}}$, where $\sigma > 0$, the matrix pencil method is efficient [22]. Most of those methods are unstable in the presence of deterministic or random noise. In this case, the decomposition problem is replaced by that of approximation and sometimes requires different tools. Algorithms 1 and 2 find the best uniform approximation of a signal in the case when all the decomposition functions $\varphi(t, z_i)$ are known, and the problem is to find the numbers $c_1, \dots, c_n$.

Consider a continuous time attenuating signal $f \in C^1(\mathbb{R}_+)$ on the positive half-line, which tends to zero as $t \rightarrow +\infty$. One needs to find its best uniform polynomial approximation by a finite sum $p(t) = \sum_{k=1}^n C_k(t) e^{z_k t}$, where $z_k = a_k + b_k i$, $a_k < 0$, are known complex frequencies and $C_k(t)$ are unknown algebraic polynomials of the prescribed powers (they correspond to multiple frequencies). We specialize our investigation to real-valued signals, when p can be rewritten as a quasipolynomial $p(t) = \sum_{k=1}^n P_k(t) e^{a_k t} \cos b_k t + Q_k(t) e^{a_k t} \sin b_k t$. Here P_k, Q_k are real algebraic polynomials of the prescribed powers. In most of applications, all the polynomials are constants, and one has to realize an optimal recovery in the uniform metric in $C(\mathbb{R}_+)$ by finding coefficients c_k .

The optimal recovery problem is also considered with a sampling. For example, we know the values of $p(t)$ at some points or other linear parameters (derivatives, Fourier coefficients, etc.) In this case, we have a problem with linear constraints which can be solved with Algorithms 1c and 2c. Numerical examples are given in Section 6.

7.3.2. Markov-Bernstein type inequalities

The Markov-Bernstein constants are the sharp upper bounds $C_j = C_j(\Phi)$ in the inequalities between the functions and their derivatives: $\|p^{(j)}\|_{C(K)} \leq C_j \|p\|_{C(K)}$, $p \in \mathcal{P}$, where $p^{(j)}$ denotes the j -th derivative of the function p . The classical Bernstein theorem asserts that for trigonometric polynomials $p(t) = \sum_{k=0}^N (a_k \cos kt + b_k \sin kt)$ on the period $[-\pi, \pi]$, the sharp constant $C_1 = N$ (respectively, $C_j = N^j$) is attained on the functions $p(t) = \cos Nt$ and $p(t) = \sin Nt$. The Markov inequality establishes that for algebraic polynomials of degree N on $K = [-1, 1]$, the sharp constant is $C_1 = N^2$ and it is attained on the Chebyshev polynomial $T_N = \cos(N \arccos t)$. See [20] for a survey and extensions to entire functions. Many works are concerned with the generalization of the Markov-Bernstein type inequalities for polynomials over other functional systems. In [52, 34] this is done for the system $\{t^{k-1}e^{-t}\}_{k=1}^n$ (algebraic polynomials with the Laguerre weight), in [18] it is done for algebraic polynomials with a Gaussian weight. Systems of real negative exponents $\{e^{a_k t}\}_{k=1}^n$, $a_k < 0$, $k = 1, \dots, n$, have been analysed in [6, 7, 37, 41, 43]. All these systems are Chebyshev. Generalizations of Markov-Bernstein inequalities to non-Chebyshev systems, apart from the theoretical interest, are important in analytic geometry, ODEs, and dynamical systems [16, 41]. The last application deals with the systems of complex exponentials in $\mathbb{R}_+$ [43].

We propose a method for computing Markov-Bernstein constants for real-valued polynomials of complex exponentials (quasipolynomials). We assume that all frequencies z_k are of multiplicity one.

Proposition 7.6. *For system $\varphi_k = e^{z_k t}$, $\operatorname{Re} z_k < 0$, $k = 1, \dots, n$, on the half-line $\mathbb{R}_+$, the constant C_j is the reciprocal to the objective value of the problem*

$$\begin{cases} \|p\|_{C[0,+\infty]} \rightarrow \min, \\ p^{(j)}(0) = 1, \\ p \in \mathcal{P}. \end{cases} \quad (12)$$

Proof. If the minimal norm $\|p\|_{C[0,+\infty]}$ under the assumption $\|p^{(j)}\|_{C[0,+\infty]} = 1$ is attained at some polynomial p , for which $\|p^{(j)}\|_{C[0,+\infty]} = |p^{(j)}(a)|$, then for the polynomial $\bar{p}(\cdot) = p(\cdot + a)$, we have $\|\bar{p}\|_{C[0,+\infty]} \leq \|p\|_{C[0,+\infty]}$ and $\|\bar{p}^{(j)}\|_{C[0,+\infty]} = |\bar{p}^{(j)}(0)| = 1$. Hence, up to the sign, $\bar{p}$ coincides with the solution of the problem (12). $\square$

7.3.3. Bounded trajectories of linear ODEs

We consider a linear ODE of n th order: $x^{(n)} = \sum_{k=0}^{n-1} c_k x^{(k)}$ on $[0, +\infty)$ with constant coefficients c_k under the asymptotic stability assumption (every solution tends to zero as $t \rightarrow +\infty$). The problem is how large the derivative $\dot{x}(0)$ can be if $|x(t)| \leq 1$ for all $t \geq 0$. Each solution $x(\cdot)$ is a linear combination (polynomial p) of n elementary solutions $t^s e^{zt}$, where z is a root of the characteristic polynomial $P(z) = z^n - \sum_{k=0}^{n-1} c_k z^k$ and s does not exceed its multiplicity. Hence, by duality, the problem is equivalent to minimizing $\|p\|_{C(\mathbb{R}_+)}$ under the constraint $\dot{p}(0) = 1$. It is solved as in subsection 6.2 (computation of Markov-Bernstein constants).

7.3.4. Linear switching dynamical systems

Linear switching system is a linear ODE $\dot{x}(t) = A(t)x(t)$ on the vector-function $x : \mathbb{R}_+ \rightarrow \mathbb{R}^n$ with the initial condition $x(0) = x_0$ such that the matrix function $A(t)$ takes values from a given compact set $\mathcal{A}$ called a *control set*. The *switching law* is an piecewise-constant function $A : \mathbb{R}_+ \rightarrow \mathcal{A}$. The *switching interval* is an interval between two consecutive switching points. This is the time of a continuous action of one matrix. The *dwell time assumption* means that each interval is not less than a fixed time m .

Linear switching systems have found many applications, mostly in electronic engineering, robotics, and planning [2, 19, 31]. One of the main issues is the stability of the system [5, 10, 35]. A system is called *asymptotically stable* if all its trajectories tend to zero. Deciding stability is usually hard [21]. One of the problems in this direction is to establish the minimal stability interval [25]. This is the smallest number M with the following property: if all trajectories with lengths of switching intervals at most M tend to zero, then the system is stable. In other words, it is always sufficient to check the stability only for switching laws with short intervals. It is known [40] that $M = m + T$, where T is a minimal number such that for every $A \in \mathcal{A}$, the value $p(T)$ of the problem:

$$\begin{cases} \|p\|_{C(\mathbb{R}_+)} \rightarrow \min, \\ p(T) = 1, \\ p \in \mathcal{P}_A \end{cases} \quad (13)$$

is bigger than 1. Here $\mathcal{P}_A$ denotes the space of polynomials defined by the system $\Phi = \{t^k e^{zt} : z \in \text{sp}(A), k \leq d(z)\}$, where $\text{sp}(A)$ is a spectrum of A and $d(z)$ is the size of the largest Jordan block of the eigenvalue z .

Thus, to find M one needs to solve the problem (13) for each $A \in \mathcal{A}$. Then by bisection we find the minimal number T for which the value of all problems is bigger than one. In turn, problem (13) is solved as the approximation problem (2) with one constraint $\ell(p) = p(T) = 1$.

7.4. Discussion. Recursive approximations vs convex programming

Both constrained and unconstrained problems of uniform polynomial approximation can be solved by tools of convex optimisation. However, in this paper, we prefer to use Remez-like recursive methods for the following reasons. First, Remez-type algorithms provide an *a posteriori* estimate and hence, usually reach a desired accuracy in much fewer iterations than theoretically prescribed. Another reason is that convex optimization algorithms may suffer from the structure of the objective function in this problem. The problem (1) belongs to the class of minimax problems $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} F(x, y)$:

$$\min_{p_1, \dots, p_n \in \mathbb{R}} \max_{t \in K} |f(t) - \sum_{i=1}^n p_i \varphi_i(t)|.$$

A popular way to solve minimax problems is the gradient (or the subgradient) descent-ascent method (GDA) with the following iteration:

$$x_{k+1} = x_k - \alpha_k \nabla_x F(x_k, y_k), \quad y_{k+1} = y_k + \beta_k \nabla_y F(x_k, y_k); \quad \alpha_k, \beta_k \in \mathbb{R}_{\geq 0}, \quad (14)$$

Usually GDA is quite efficient for minimax problems, where the objective function $F(x, y)$ is convex in x and concave in y (see [15, 30, 54]). However, in our case, the objective function $F(p, t) = |f(t) - \sum_{i=1}^n p_i \varphi_i(t)|$ is not necessarily concave in t , and, moreover, is non-smooth, which significantly slows down the convergence of Algorithm 14.

Acknowledgments: The authors are grateful to the anonymous Referees for many useful remarks and suggestions.

Declarations:

Conflict of interest. The authors declare no conflicts of interest.

References

- [1] O. Bar-Ilan and Y.C. Eldar, *Sub-nyquist radar via doppler focusing*, IEEE Transactions on Signal Processing, 62 (2014), no 7, 1796–1811.
- [2] C. Basso, *Switch-mode power supplies spice simulations and practical designs*, McGraw-Hill, Inc., New York, NY, USA, 1 edition, 2008.
- [3] J. Berent, P.L. Dragotti, T. Blu, *Sampling piecewise sinusoidal signals with finite rate of innovation methods*, IEEE Trans. Signal Process. 58 (2010), no 2, 613–625.
- [4] G. Beylkin and L. Monzón, *On approximation of functions by exponential sums*, Appl. Comput. Harmon. Anal. 19 (2005), 17–48.
- [5] F. Blanchini, D. Casagrande, S. Miani, *Modal and transition dwell time computation in switching systems: a set-theoretic approach*, Automatica J. IFAC, 46 (2010), no 9, 1477–1482.
- [6] P.B. Borwein and T. Erdélyi, *Upper bounds for the derivative of exponential sums* Proc. Amer. Math. Soc. 123 (1995), 1481–1486.
- [7] P.B. Borwein and T. Erdélyi, *A sharp Bernstein-type inequality for exponential sums*, J. Reine Angew. Math. 476 (1996), 127–141.
- [8] A. Capelli, *Sopra la compatibilità o incompatibilità di più equazioni di primo grado fra picci incognite*, Revista di Matematica, 2 (1892), 54–58.
- [9] E.W. Cheney and H.L. Loeb, *Generalized Rational Approximation*, SIAM J. Numer. Anal., Series B, (1964), 11–25.
- [10] Y. Chitour, N. Guglielmi, V. Yu. Protasov, M. Sigalotti, *Switching systems with dwell time: computing the maximal Lyapunov exponent*, Nonlinear Anal. Hybrid Syst. 40 (2021), Paper No. 101021, 21 pp
- [11] L. Collatz and W. Krabs, *Tschebyscheffsche Approximation mit Anwendungen*, Teubner Studienbücher Mathematik (TSBMA), Stuttgart, 1973.
- [12] V.F. Dem'yanov and V.N. Malozemov, *Introduction to Minimax*, Courier Corporation, Chelmsford, MA, USA: 1990.
- [13] V.K. Dzyadyk and I.A. Shevchuk, *Theory of uniform approximation of functions by polynomials*, Walter de Gruyter, 2008.
- [14] R.E. Edwards, *Functional analysis*, Dover Publications, Inc., New York, 1995, Theory and applications, Corrected reprint of the 1965 original. MR 1320261
- [15] A. Fallah, A. Ozdaglar, S. Pattathil, *An optimal multistage stochastic gradient method for minimax problems*, in 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, 2020, 3573–3579.
- [16] J.P. Francoise, *Bernstein inequalities and applications to analytic geometry and differential equations*, Journal of functional analysis 146 (1997), 185–205.

- [17] W. Fraser, *A Survey of methods of computing minimax and near-minimax polynomial approximations for functions of a single independent variable*, J. ACM 12 (1965), no 295.
- [18] G. Freud, *On two polynomial inequalities*, Acta Math. Acad. Sci. Hungar., 22 (1971), no 1–2, 109–116.
- [19] J. C. Geromel and P. Colaneri, *Stability and stabilization of continuous-time switched linear systems*, SIAM J. Control Optim., 45 (2006), no 5, 1915–1930.
- [20] D.V. Gorbachev, *Sharp Bernstein–Nikolskii inequalities for polynomials and entire functions of exponential type*, Chebyshevskii Sb., 22 (2021), no 5, 58–110.
- [21] L. Gurvits and A. Olshevsky, *On the NP-Hardness of checking matrix polytope stability and continuous-time switching stability*, IEEE Trans. Autom. Control, 54 (2009), no 2, 337–341.
- [22] Y. Hua and T.K. Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, Acoustics, Speech and Signal Processing, IEEE Transactions on, 38 (1990), no 5, 814–824.
- [23] J.F. Huang and L. Yang, *Empirical mode decomposition based on locally adaptive filters*, preprint.
- [24] A.C. Ionita and A.C. Antoulas, *Data-driven parametrized model reduction in the Loewner framework*, SIAM J. Sci. Comput. 36 (2014), no 3, A984–A1007.
- [25] R. Kamalov and V.Yu. Protasov, *On the length of switching intervals of a stable dynamical system*, Proc. Steklov Inst. Math. 321 (2023), no 1, 149–157.
- [26] S. Karlin and W.J. Studden, *Tchebycheff systems: with applications in analysis and statistics*, Interscience, New York, 1966
- [27] F. Knaepkens, A. Cuyt, W.S. Lee, de Villiers, I.L. Dirk, *Regular sparse array direction of arrival estimation in one dimension*, IEEE Trans. Antennas Propag. 68 (2020), no 5, 3997–4006.
- [28] A.N. Kolmogorov, *A remark on the polynomials of P. L. Chebyshev deviating the least from a given function*, Uspehi Mat. Nauk., 3 (1948), 216–221 (Russian).
- [29] M.G. Krein and A.A. Nudelman, *The Markov moment problem and extremal problems: ideas and problems of P.L. Chebyshev and A.A. Markov and their further development*, Translations of mathematical monographs, Providence, R.I., v. 50 (1977).
- [30] T. Liang and J. Stokes, *Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks*, in The 22nd International Conference on Artificial Intelligence and Statistics, PMLR (2019), 907–915.
- [31] D. Liberzon, *Switching in systems and control*, Systems & Control: Foundations and Applications. Birkhäuser Boston, Inc., Boston, MA, 2003.
- [32] L. Lin, Y. Wang, H. Zhou, *Iterative filtering as an alternative algorithm for empirical mode decomposition*, Adv. Adapt. Data Anal. 1 (2009), no 4, 543–560.
- [33] G.G. Magaril-Ilyayev and V.M. Tikhomirov, *Convex Analysis: Theory and Applications*, Translations of Mathematical Monographs, Amer. Mathematical Society (2003).
- [34] L. Milev and N. Naidenov, *Exact Markov inequalities for the Hermite and Laguerre weights*, J. Approx. Theory, 138 (2006), no 1, 87–96.
- [35] A. P. Molchanov and Y. S. Pyatnitskiy, *Criteria of asymptotic stability of differential and difference inclusions encountered in control theory*, Systems Control Lett., 13 (1989), no 1, 59–64.
- [36] Y. Nakatsukasa, O. Sete, L.N. Trefethen, *The AAA algorithm for rational approximation*, SIAM J.Sci. Comput. 40 (2018), no 3, A1494–A1522.
- [37] D.J. Newman, *Derivative bounds for Müntz polynomials*, J. Approx. Theory 18 (1976), 360–362.

- [38] G. Plonka and M. Tasche, *Prony methods for recovery of structured functions*, GAMM Mitt. 37 (2014), no 2, 239–258.
- [39] M. Petz, G. Plonka, N. Derevianko, *Exact reconstruction of sparse non-harmonic signals from their Fourier coefficients*, Sampling Theory, Signal Processing, and Data Analysis 10 (2021), no 7.
- [40] V.Yu. Protasov and R. Kamalov, *How do the lengths of switching intervals influence the stability of a dynamical system?*, Automatica 171 (2025), art. no 111929.
- [41] V.Yu. Protasov and R.M. Jungers, *Is switching systems stability harder for continuous time systems?*, Proc. IEEE Conf. Decision and Control (2013), art. no. 6759964, 704–709.
- [42] V.Yu. Protasov and R.M. Jungers, *Analysing the stability of linear systems via exponential Chebyshev polynomials*, IEEE Trans. Autom. Control, 61 (2016), no 3, art. no. 7131458, 795–798.
- [43] V.Yu. Protasov, *Generalized Markov–Bernstein inequalities and stability of dynamical systems*, Proc. Steklov Inst. Math. 319 (2022) no 1, 237–252.
- [44] E.Ya. Remez, *Sur le calcul effectiv des polynomes d’approximation des Tschebyscheff*, Compt. Rend. Acade. Sc. 199, 337 (1934).
- [45] T.J. Rivlin and H.S. Shapiro, *A unified approach to certain problems of approximation and minimization*, J. Society for Industrial and Appl. Math., 9 (1961), no 4 , 670–699,
- [46] R.T. Rockafellar, *Convex Analysis*, Princeton University Press (1979).
- [47] R. Tur, Y.C. Eldar, Z. Friedman, *Innovation rate sampling of pulse streams with application to ultrasound imaging*, Signal Processing, IEEE Trans. on, 59 (2011), no 4, 1827–1842.
- [48] P. Schwartz, *Characterization and uniqueness of best uniform vector-valued linear approximation*, J. Approx. Theory, 14 (1975), 218–224.
- [49] V.P. Sklyarov, *The sharp constant in Markov’s inequality for the Laguerre weight*, Sb. Math., 200 (2009), no 6, 887–897.
- [50] G.S. Smirnov, *Best uniform approximation of complex-valued functions by generalized polynomials having restricted ranges*, J. Approx. Theory, 100 (1999), 284–303.
- [51] G.S. Smirnov and R.G. Smirnov, *Best uniform approximation of complex-valued functions by generalized polynomials having restricted ranges*, J. Approx. Theory 100 (1999), 284–303.
- [52] G. Szegő, *On some problems of approximations*, Magyar Tud. Akad. Mat. Kutató Int. Közl., 9 (1964), 3–9.
- [53] N. Sukhorukova and J. Ugon, *A generalisation of de la Vallée Poussin procedure to multivariate approximations*, Advanc. Comput. Math., 48 (2022), no. 5
- [54] G. Zhang, Y. Wang, L. Lessard, R. B. Grosse, *Near-optimal local convergence of alternating gradient descent-ascent for minimax optimization*, in Int. Conf. on Art. Intel. Statistics, PMLR, 2022, pp. 7659–7679.