Chebyshev approximation for multivariate functions

Nadezda Sukhorukova, Swinburne University of Technology, Australia, nsukhorukova@swin.edu.au,
Julien Ugon, Centre for Informatics and Applied Optimization, Federation University Australia, j.ugon@federation.edu.au,
David Yost, Centre for Informatics and Applied Optimization, Federation University Australia, d.yost@federation.edu.au

In this paper, we derive optimality conditions (Chebyshev approximation) for multivariate functions. The theory of Chebyshev (uniform) approximation for univariate functions is very elegant. The optimality conditions are based on the notion of alternance (maximal deviation points with alternating deviation signs). It is not very straightforward, however, how to extend the notion of alternance to the case of multivariate functions. There have been several attempts to extend the theory of Chebyshev approximation to the case of multivariate functions. We propose an alternative approach, which is based on the notion of convexity and nonsmooth analysis.

Keywords: polynomial splines, Chebyshev approximation, free knots, quasidifferential, best approximation conditions
Math subject classification[2010]: 49J52, 90C26, 41A15, 41A50

1 Introduction

In this paper, we obtain Chebyshev (uniform) approximation optimality conditions for multivariate functions. The theory of Chebyshev approximation for univariate functions was built in the late nineteenth (Chebyshev) and twentieth century (just to name a few [3, 6, 9]). In most cases, the authors were working on polynomial and polynomial spline approximations, however, other types of functions (e.g., trigonometric polynomials) have also been used. This theory is very elegant. In most cases, the optimality conditions are based on the notion of alternance (that is, maximal deviation points with alternating deviation signs).
There have been several attempts to extend this theory to the case of multivariate functions. One of them is [7]. In this paper the author underlines the fact that the main difficulty is to extend the notion of alternance to the case of more than one variable. The main obstacle is that it is not very easy to extend the notion of monotonicity to the case of several variables. Also, several studies have been done in the area of multivariate interpolation [2, 4], where triangulation based approaches were used to extend the notion of polynomial splines to the case of multivariate function.

The objective function, appearing in Chebyshev approximation optimisation problems is nonsmooth (minimisation of the maximal absolute deviation). Therefore, it is logical to use nonsmooth optimisation techniques to tackle this problem. In this paper we propose an approach, which is based on the notion of convex function subdifferentials [8]. Subdifferentials can be considered as generalisation of the notion of gradients for convex nondifferential functions.

Another powerful nonsmooth analysis tool, quasidifferentiability, has been successfully applied to improve the existing optimality conditions in the case of free knot polynomial spline approximation [12]. In particular, these techniques allowed the researchers to overcome the difficulties highlighted in [1] as “the existing optimisation tools are not adapted to this problem, due to its nonconvex and nonsmooth nature”. Quasidifferentiability is one of the modern nonsmooth optimisation approaches, which is not well-known outside of the optimisation research community. This technique, however, is very powerful and enables one to work efficiently with nonsmooth and nonconvex functions. In particular, it has been successfully applied to improve the existing [3] necessary optimality conditions for the case of free knots polynomial splines [11, 12].

Quasidifferentiability can be treated as a generalisation of subdifferentials to the case of nonconvex functions. In this study, however, the objective function is convex and therefore, it is sufficient to use subdifferentials, since this tool is simpler and easier to use and understand.

The paper is organised as follows. In section 2 we present the most relevant results from the theory of convex and nonsmooth analysis, that are essential to obtain our optimality conditions. Then, in the same section, we investigate the extremum properties of the objective function, appearing in Chebyshev approximation problems, from the points of view of convexity and nonsmooth analysis. In section 2 we obtain our main results. In section 3 we demonstrate the relation between the obtained results and the classical ones (alternance-based). Then in section 4 we demonstrate the relation with other optimality results (multivariate case), obtained by J. Rice [7]. Finally, in section 5 we draw our conclusions and underline further research directions.
2 Optimality conditions

2.1 Convexity of the objective

Let us now formulate the objective function. Suppose that a continuous function \( f(x) \) is to be approximated by a function

\[
L(A, x) = a_0 + \sum_{i=1}^{n} a_i g_i(x),
\]

(2.1)

where \( g_i(x) \) are the basis functions and the multipliers \( A = (a_1, \ldots, a_n) \) are the corresponding coefficients. At a point \( x \) the deviation between the function \( f \) and the approximation is:

\[
d(A, x) = |f(x) - L(A, x)|.\]

(2.2)

Then we can define the uniform approximation error over the set \( Q \) by

\[
\Psi(A) = \sup_{x \in Q} \max \{ f(x) - a_0 - \sum_{i=1}^{n} a_i g_i(x), a_0 + \sum_{i=1}^{n} a_i g_i(x) - f(x) \}.\]

(2.3)

The approximation problem can be formulated as follows.

\[
\text{minimise } \Psi(A) \text{ subject to } A \in \mathbb{R}^{n+1}.\]

(2.4)

We will consider two cases.

**Continuous case** the set \( Q \) is a hyperbox, such that \( c_i \leq x_i \leq d_i, i = 1, \ldots, d \).

**Discrete case** the set \( Q \) is a finite set of points.

Since the function \( L(A, x) \) is linear in \( A \), the approximation error function \( \Psi(A) \), as the supremum of affine functions, is convex. Furthermore, its subdifferential at a point \( A \) is trivially obtained using the active affine functions in the supremum:

\[
\partial \Psi(A) = \text{co} \left\{ \left( \begin{array}{c} \frac{1}{g_1(x)} \\ \frac{1}{g_2(x)} \\ \vdots \\ \frac{1}{g_n(x)} \end{array} \right) : x \in E^+, \quad \left( \begin{array}{c} -1 \\ g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{array} \right) : x \in E^- \right\},\]

(2.5)

where \( E^+ \) and \( E^- \) are respectively the points of maximal positive and negative deviation:

\[
E^+ = \left\{ x \in Q : L(A, x) - f(x) = \max_{y \in Q} d(A, y) \right\}
\]

\[
E^- = \left\{ x \in Q : f(x) - L(A, x) = \max_{y \in Q} d(A, y) \right\}.
\]
2.2 Optimality conditions: general case

Theorem 2.1. The convex hulls of the vectors \((g_1(x), \ldots, g_n(x))^T\), built over positive and negative maximal deviation points intersect.

Proof. The necessary and sufficient condition for the optimality of a vector \(A^*\) for the convex problem (2.3) is

\[0_{n+1} \in \partial \Psi(A^*).\]

Note that due to Caratheodory’s theorem, \(0\) can be constructed as a convex combination of a finite number of points. Namely, since the dimension of the corresponding space is \(n + 1\), it can be done using at most \(n + 2\) points.

By the formulation of the subdifferential of \(\Psi \partial \Psi(A)\) (2.5), there exists a nonnegative number \(\gamma \leq 1\) and two vectors

\[g^+ \in \co \begin{pmatrix} 1 \\ g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{pmatrix} : x \in E^+ \]

and \(g^- \in \co \begin{pmatrix} 1 \\ g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{pmatrix} : x \in E^- \)

such that \(0 = \gamma g^+ - (1 - \gamma)g^-\). Noticing that the first coordinates \(g_1^+ = g_1^- = 1\), we see that \(\gamma = \frac{1}{2}\). This means that \(g^+ - g^- = 0\). This happens if and only if

\[\co \begin{pmatrix} 1 \\ g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{pmatrix} : x \in E^+ \cap \co \begin{pmatrix} 1 \\ g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{pmatrix} : x \in E^- \neq \emptyset.\]

As noted before, the first coordinates of all these vectors are the same, and therefore the theorem is true. \(\square\)

2.3 Optimality conditions for multivariate linear functions

In the case of linear functions (multidimensional case), \(n = d\) and the functions \(g_i = x_i\), \(i = 1, \ldots, n\). Then theorem 2.1 can be formulated as follows.

Theorem 2.2. The convex hull of the maximal deviation points with positive deviation and convex hull of the maximal deviation points with negative deviation have common points.

Theorem 2.2 can be considered as an alternative formulation to the necessary and sufficient optimality conditions that are based on the notion of alternance. Clearly, theorem 2.2 can be used in univariate cases, since the location of the alternance points insures the common points for the corresponding convex hulls, constructed over the maximal deviation points with positive and negative deviations respectively.

Note that in general \(d \leq n\).
2.4 Optimality conditions for multivariate polynomial (non-linear) functions

We start from introducing the following definitions and notation.

**Definition 2.1.** An exponent vector

\[ e = (e_1, \ldots, e_d) \in \mathbb{R}^d, \quad e_i \in \mathbb{N}, \quad i = 1, \ldots, d \]

for \( x \in \mathbb{R}^d \) defines a monomial

\[ x^e = x_1^{e_1} x_2^{e_2} \ldots x_d^{e_d}. \]

**Definition 2.2.** A product \( cx^e \), where \( c \neq 0 \) is called the term, then a multivariate polynomial is a sum of a finite number of terms.

**Definition 2.3.** The degree of a monomial \( x^e \) is the sum of the components of \( e \):

\[ \deg(x^e) = \sum_{i=1}^{d} e_i. \]

**Definition 2.4.** The degree of a polynomial is the largest degree of the composing it monomials.

Let us consider some essential properties of polynomials and monomials.

1. For any exponential \( e = (e_1, \ldots, e_d) \), such that \( \deg x^e = m \) the degree of the monomial \( x^e = m + 1 \), where the exponential \( \tilde{e} \) has been obtained from \( e \) by substituting one of its components \( e_k \) by \( e_k + 1, \quad k = 1, \ldots, n \). Any monomial of degree \( m + 1 \) can be obtained in such a way. In general, there may be more than one way to do this.

2. For any exponential \( e = (e_1, \ldots, e_d) \), such that \( \deg x^e = m \) the degree of the monomial \( x^e = m - 1 \), where the exponential \( \tilde{e} \) has been obtained from \( e \) by substituting one of its components positive components \( e_k > 0 \) by \( e_k - 1, \quad k = 1, \ldots, n \). Any monomial of degree \( m - 1 \) can be obtained in such a way. In general, there may be more than one way to do this.

In general, a polynomial of degree \( m \) can be obtained as follows:

\[ P^m(x) = a_0 + \sum_{i=1}^{n} a_i M_i(x, e), \quad (2.7) \]

where \( a_i \) are the coefficients and \( g_i = M_i \) are the monomials, such that \( \deg M_i \leq m \) and there exists a monomial \( M_k \), such that \( \deg(M_k) = m \). Any polynomials \( P^m \) from (2.7) can be presented as the sum of a lower degree polynomials \( (m - 1 \text{ or less}) \) and a finite number of terms that correspond to the monomials of degree \( m \). The following lemma holds.
Lemma 2.1. Consider two sets of non-negative coefficients

- \( \alpha_i \geq 0, \quad i = 1, \ldots, n \) such that \( \sum_{i=1}^{n} \alpha_i = 1 \);
- \( \beta_i \geq 0, \quad i = 1, \ldots, n \) such that \( \sum_{i=1}^{n} \beta_i = 1 \).

If

\[
\sum_{i=1}^{n} \alpha_i a_i x_i = \sum_{i=1}^{n} \beta_i b_i y_i \quad (2.8)
\]

and

\[
\sum_{i=1}^{n} \alpha_i a_i = \sum_{i=1}^{n} \beta_i b_i \quad (2.9)
\]

then for any scalar \( \delta \) the following equality holds

\[
\sum_{i=1}^{n} \alpha_i a_i (x_i - \delta) = \sum_{i=1}^{n} \beta_i b_i (y_i - \delta). \quad (2.10)
\]

Proof.

\[
\sum_{i=1}^{n} \alpha_i a_i (x_i - \delta) = \sum_{i=1}^{n} \alpha_i a_i x_i - \delta \sum_{i=1}^{n} \alpha_i a_i = \sum_{i=1}^{n} \beta_i b_i - \delta \sum_{i=1}^{n} \beta_i b_i y_i = \sum_{i=1}^{n} \beta_i b_i (y_i - \delta).
\]

Theorem 2.1 can be used to formulate our necessary and sufficient optimality conditions for multivariate polynomial approximations. Each \( g_i \) corresponds to a monomial and, in general, we need to use all the possibilities to construct the monomials, keeping the corresponding monomial degree at most \( m \).

Note that due to lemma 2.1 one can assume that all the \( x_i \) in the monomials are non-negative, since \( \delta \) can be chosen as

\[
\min \{ \min_{i=1,\ldots,d} x_i, \min_{i=1,\ldots,d} y_i \}.
\]

Then a necessary and sufficient optimality condition can be formulated as follows.

Theorem 2.3. A polynomial of degree \( m \) is optimal if and only if there exists a pair of sets of non-negative coefficients (with at least one positive coefficient in each set)

\[
\alpha_{i+}, \quad \alpha_{i-}, \quad i+ = 1, \ldots, n+2, \quad \sum_{i+}^{n+2} \alpha_{i+} = \sum_{i-}^{n+2} \alpha_{i-} = 1,
\]
such that for any monomial $M_j$, $j = 1, \ldots, n$ of degree at most $m$ the following equality holds

$$\sum_{i=1}^{n+2} \alpha_i M_j(x^i) = \sum_{i=1}^{n+2} \alpha_i M_j(x^{i^-}),$$

(2.11)

where $x^+$ and $x^-$ are maximal deviation points with positive and negative deviation sign respectively.

Proof. The proof is obvious, since this is a reformulation of theorem 2.1 for the case of polynomials. Similar to theoremrefthm:main, Caratheodorys theorem is used to prove that it is possible to do all the necessary constructions using at most $n + 2$ points. □

Note that any monomial $\prod_{i=1}^{l} x_i^{e_i}$ of degree $m \geq 1$, such that $m = \sum_{i=1}^{l} e_i$ can be presented as a product of a lower degree monomial and $x_i$ (assume that $e_i \geq 1$). Therefore, theorem 2.3 can be also formulated as follows.

**Theorem 2.4.** A polynomial of degree $m$ is optimal if and only if there exists a pair of sets of non-negative coefficients (with at least one positive coefficient in each set)

$$\alpha_{i+}, \alpha_{i-}, i+, i- = 1, \ldots, n + 2, \sum_{i=1}^{n+2} \alpha_i = 1,$$

such that for any monomial $M$ of degree at most $m - 1$ the following equality holds

$$\sum_{i=1}^{n+2} \alpha_i M(x^i)x^i = \sum_{i=1}^{n+2} \alpha_i M(x^{i^-})x^{i^-},$$

(2.12)

where $x^+$ and $x^-$ are maximal deviation points with positive and negative deviation sign respectively.

Theorem 2.4 tells us that the convex hulls built over positive and negative maximal deviation points intersect. Moreover, the linear combinations of these two sets with the coefficients $\alpha_i M(x_i^+)$ and $\alpha_i M(x_i^-)$, where $M(x)$ is a monomial of degree at most $m - 1$.

Note that these linear combinations are not necessary convex, since some of the coefficients $M(x_i^+)$ and $M(x_i^-)$ may be negative. One can make these coefficients non-negative by applying lemma 2.1. This can be achieved in a number of ways. For example, for a monomial

$$M(x) = x_1^{e_1} \times x_2^{e_2} \times \cdots \times x_d^{e_d},$$

where $e_1, \ldots, e_d$ are non-negative integers, representing the degree, such that

$$\sum_{j=1}^{d} e_j \leq m - 1,$$
apply lemma 2.1 where $\delta$ is chosen as follows

$$\delta = \min_{j=1,\ldots,d,e_j>0} x_j$$

(2.13)

or

$$\delta = -\max_{j=1,\ldots,d,e_j>0} x_j.$$  

(2.14)

Necessary and sufficient optimality conditions formulated in theorems 2.1-2.4 are not very easy to verify. In this paper we develop a necessary optimality condition that is more practical.

Assume that our polynomial approximation is optimal and consider all the monomials of degree $m-1$

$$M^{m-1} = x_1^{e_1} x_2^{e_2} \cdots x_l^{e_l},$$

such that $\sum_{i=1}^d e_i = m - 1.$  

(2.15)

Then all the monomials of degree $m$ can be obtained by multiplying one of the monomials of degree $m-1$ by one of the coordinates of $\mathbf{x} = (x_1, \ldots, x_d)^T$. Therefore, there exists a pair of sets of positive coefficients (remove zero coefficients for simplicity)

$$\alpha_{i+}, \ i = 1, \ldots, N^+, \ \alpha_{i-} = 1, \ldots, N^-,$$

(2.16)

such that

$$\sum_{i=1}^{N^+} \alpha_{i+} M^{m-1}(1, \mathbf{x}_{i+})^T = \sum_{i=1}^{N^-} \alpha_{i-} M^{m-1}(1, \mathbf{x}_{i-})^T$$

(2.17)

Assume that in the monomial $M^{m-1}$ there exists $j$, such that $e_j > 0$. Consider

$$\delta^j_{\min} = \min \{ \min_{i=1,\ldots,N^+} x_j^{i+}, \ \min_{i=1,\ldots,N^-} x_j^{i-} \}$$

and

$$\delta^j_{\max} = \max \{ \max_{i=1,\ldots,N^+} x_j^{i+}, \ \max_{i=1,\ldots,N^-} x_j^{i-} \},$$

where $x_j^{i+}$ is the $j$–th coordinate of the maximal deviation points $\mathbf{x}_{i+}$ and $x_j^{i-}$ is the $j$–th coordinate of the maximal deviation points $\mathbf{x}_{i-}$. Apply lemma 2.1 with $\delta = \delta^j_{\min}$ or $\delta = \delta^j_{\max}$ and remove all the maximal deviation points with the minimal (maximal) value for the $j$–th coordinate (there may be more than one point), since the corresponding monomial is zero. At the end of this process, the convex hulls of the remaining maximal deviation points should intersect or all the maximal deviation points are removed.

The following algorithm can be used to verify this necessary optimality condition.

**Necessary optimality conditions verification for best polynomial Chebyshev approximation: degree $m > 1$, multivariate case.**
Step 1 Identify maximal deviation points that correspond to positive and negative deviation:

\[ E^+ = \{ x^{i+}, i = 1, \ldots, N^+ \}; \quad E^- = \{ x^{i-}, i = 1, \ldots, N^- \}. \]

Step 2 For each dimension \( k : k = 1, \ldots, l \) identify

\[ \delta = \min \left\{ \min_{j=1, \ldots, l} x^{i+j}_j, \min_{j=1, \ldots, l} x^{i-j}_j \right\} \]

or

\[ \delta = \max \left\{ \max_{j=1, \ldots, l} x^{i+j}_j, \max_{j=1, \ldots, l} x^{i-j}_j \right\} \]

where \( x^{i+j}_j \) and \( x^{i-j}_j \) are the \( j \)-th coordinates of \( x^{i+} \) and \( x^{i-} \) respectively.

Step 3 Apply the following coordinate transformation (to transform the coordinates of the maximal deviation points to non-negative numbers):

\[ \tilde{x}^{i+j}_j = x^{i+j}_j - \delta; \quad \tilde{x}^{i-j}_j = x^{i-j}_j - \delta. \]

Step 4 Points reduction Remove maximal deviation points whose updated coordinates have a zero at the corresponding coordinate and assign \( m_{\text{new}} = m - 1 \).

If \( m_{\text{new}} > 1 \) and the remaining sets of maximal deviation (positive and negative) are non-empty GO TO Step 1 for the corresponding lower degree polynomials approximation optimality verification, \( m = m_{\text{new}} \).

Otherwise GO TO the final step of the algorithm.

Step 5 If the remaining maximal deviation sets are non-empty and the conditions of theorem [2,2] are not satisfied then the original polynomials is not optimal.

There are two main advantages of this procedure.

1. It demonstrates how the concept of alternance can be generalised to the case of multivariate functions.

2. It is based on the verification whether two convex sets are intersecting or not, but since \( l \leq n \) it is much easier to verify it after applying the algorithm.

However, the this condition is a necessary optimality condition, but not sufficient. In section sec:counterexample we will present an interesting counterexample, that demonstrates that even more general alternating conditions still remain only necessary optimality conditions for multidimensional functions.

Note that theorem [2,1] can be also applied to verify optimality (necessary and sufficient optimality condition). In this case one needs to check if two convex sets are intersecting in \( \mathbb{R}^n \). The above algorithm requires to check if two convex sets are intersecting in \( \mathbb{R}^d \) (considerably lower dimension), however, it only verifies a necessary optimality condition.
3 Counterexample

Consider new notation.

$M^m$ is a vector that contains all the monomials of degree $m$ or less; 

$\alpha_x$ is a nonnegative coefficient associated with $x$.

It is clear that the system:

$$
\sum_{x \in E^+} \alpha_x M^m(x) = \sum_{x \in E^-} \alpha_x M^m(x)
$$

is equivalent to:

$$
\sum_{x \in E^+} \alpha_x M^{m-1}(x) = \sum_{x \in E^-} \alpha_x M^{m-1}(x) \quad (3.1(0))
$$

$$
\sum_{x \in E^+} \alpha_x x_i M^{m-1}(x) = \sum_{x \in E^-} \alpha_x x_i M^{m-1}(x) \quad i = 1, \ldots, d. \quad (3.1(i))
$$

Consider any vector $u \in \mathbb{R}^d$ and scalar $a \in \mathbb{R}$, and let

$$
H^+ = \{x \in \mathbb{R}^d : \langle u, x \rangle - a > 0\}
$$

$$
H^- = \{x \in \mathbb{R}^d : \langle u, x \rangle - a < 0\}
$$

and consider the equation $\sum_{i=1}^d u_i \times (3.1(i)) - a \times (3.1(0))$. We find that

$$
\sum_{x \in E^+} \alpha_x (\langle u, x \rangle - a) M^{m-1}(x) = \sum_{x \in E^-} \alpha_x (\langle u, x \rangle - a) M^{m-1}(x)
$$

Define

$$
A^+ = \sum_{y \in (E^+ \cap H^+) \cup (E^- \cap H^-)} \alpha_y (\langle u, y \rangle - a)
$$

$$
A^- = \sum_{y \in (E^+ \cap H^-) \cup (E^- \cap H^+)} \alpha_y (\langle u, y \rangle - a)
$$

and

$$
\hat{\alpha}_x = \frac{\alpha_x}{A^+} \quad \text{if} \quad x \in (E^+ \cap H^+) \cup (E^- \cap H^-)
$$

$$
\hat{\alpha}_x = \frac{\alpha_x}{A^-} \quad \text{if} \quad x \in (E^- \cap H^+) \cup (E^+ \cap H^-)
$$

Then, we have:

$$
\sum_{x \in (E^+ \cap H^+) \cup (E^- \cap H^-)} \hat{\alpha}_x M^{m-1}(x) = \sum_{x \in (E^- \cap H^+) \cup (E^+ \cap H^-)} \hat{\alpha}_x M^{m-1}(x) \quad (3.2)
$$

Note that Formula (3.2) is similar to Formula (2.11) with degree $m - 1$. Thus, the above result means that if one runs any hyperplane and inverts the signs on one side of this hyperplane, the formula holds for degree $m - 1$. This implies the following corollaries:
1. Since any \( k \leq d \) points define (not necessarily uniquely) a hyperplane, by setting the pair \((u, a)\) to be defining this hyperplane, the result for degree \( m - 1 \) applies, after setting the signs accordingly.

2. In particular, for if one chooses these \( k \) points as the vertices defining a \( k - 1 \)-face of the polytope \( P = \text{co}\{x \in E\} \), then all remaining points lie on the same side of the hyperplane. Therefore, by removing any \( k \) face (and facets in particular) of the polytope \( P \), the result holds for degree \( m - 1 \).

3. By iteratively removing any \( m - 1 \) facets as described above, the remaining polytopes \( P^+ \) and \( P^- \) must intersect (that is, the result holds for degree 1).

4. Similarly, it is possible to remove any \((m - 1)d\) points and, updating the signs accordingly, the remaining polytopes \( P^+ \) and \( P^- \) must intersect.

Remark 3.1. It is easy to verify that Condition 3 is exactly equivalent to the alternation criterion in the univariate case. Indeed, after removing \( m - 1 \) points (which are also facets in the univariate case), there needs to remain at least 3 alternating points to ensure intersection of the remaining polytopes. This means that there must be at least \( m + 2 = m - 1 + 3 \) points, which can be shown to alternate by removing the adequate \( m - 1 \) points.

Similarly, Condition 3 is trivial for degree \( m = 1 \) polynomial approximation, since all these conditions are equivalent to the intersection between \( E^+ \) and \( E^- \) being nonempty.

Generally, however, the condition (3.2) is not sufficient, as is illustrated in the following example:

Example 3.1. Consider the quadratic approximation case and assume \( E^+ = \text{co}\{(0, 0), (1, 2), (2, 0)\} \) and \( E^- = \{(0, 1), (1, -1), (2, 1)\} \). These extreme deviation points do not satisfy the condition (2.11). Yet, they satisfy the condition (3.2). The sets are illustrated in the figure below:

and below we show that for any hyperplane cutting through the polytope \( \text{co}(E^+ \cup E^-) \) the intersection exists (obvious symmetries are omitted).
4 Relation with classical (univariate) polynomial approximation results

4.1 Linear approximation

For univariate linear approximation the alternance-based optimality conditions are as follows.

**Theorem 4.1.** There exist three maximal deviation points, such that the sign of maximal deviation is alternating.

Clearly, this theorem is a special case of theorem 2.2 for \( n = 1 \), since one of these three maximal deviation points is located between two others and the deviation sign of this middle point is opposite to the sign at the far left and right points. Therefore, the convex hulls of positive and negative maximal deviation points intersect.

4.2 Nonlinear approximation

In the case of higher degree polynomials, the proposed algorithm, at each iteration reduces the degree of the polynomial. If we assume that our dimension is one (univariate approximation), then at each iteration we remove one point (minimal or maximal value) and the remaining \( m+2-(m-1) = 3 \) points satisfy the univariate (alternance) condition.

4.3 Uneven distribution of maximal deviation points

In the case of univariate approximation the distribution of positive and negative maximal deviation points, involved in optimality verification, is (almost) even. Indeed, if \( n \) is even, then each set contains \( \frac{n}{2} + 1 \) points, otherwise, one set contains \( \frac{n+1}{2} \) points and another one \( \frac{n+1}{2} + 1 \) points. The following example demonstrates, that this is not always the case for multivariate approximations.

**Example 4.1.** Consider a discrete function defined at four isolated points \((1, 1), (-1, 1), (0, -1)\) and \((0, 0)\):

\[
f(1, 1) = f(-1, 1) = f(0, -1) = 0, \quad \text{and} \quad f(0, 0) = 2.
\]

Find a best linear approximation to this function.

It is clear that the plane \( L(x, y) = 1 \) is optimal, since the maximal deviation is attained at \((1, 1), (-1, 1), (0, -1)\) and \((0, 0)\). Namely, the maximal deviation magnitude is equal
to 1 and the deviation signs are negative at the points $(1, 1), (-1, 1), (0, -1)$ and positive at $(0, 0)$. Therefore, $N$ contains three points, while $P$ only one.

### 5 Relation with existing multivariate results

In [7] Rice gives necessary and sufficient optimality conditions for multivariate approximation. These results are obtained for a very general class of functions, not necessary polynomials. These conditions are fundamentally important, however, it is not very easy to verify them (even in the case of polynomials). Also their relation with the notion of alternance is not very clear. Before formulating Rice’s optimality conditions, we need to introduce the following notation and definitions ([7]).

Recall that the set of extremal (maximal deviation points) $E$ is divided into two parts as follows:

$$E^+ = \{ x | x \in E, f(x) - L(A^*, x) = \| f(x) - L(A^*, x) \|_\infty \}$$

$$E^- = \{ x | x \in E, f(x) - L(A^*, x) = -\| f(x) - L(A^*, x) \|_\infty \},$$

where $A^*$ is a vector of the parameters and $L(A^*, x)$ is the corresponding approximation, defined as in (2.1). The elements of $E^+$ and $E^-$ are positive and negative extremal points.

**Definition 5.1.** The point sets $P$ and $N$ are said to be isolable if there is an $A$, such that $L(A, x) > 0 \ x \in P$, $L(A, x) < 0 \ x \in N$.

**Definition 5.2.** $\Gamma(A)$ is is called an isolating curve if

$$\Gamma(A) = \{ x | L(A, x) = 0 \}.$$

Therefore, the sets $P$ and $N$ are isolable, if they lie on opposite sides of an isolating curve $\Gamma(A)$.

**Definition 5.3.** A subset of extremal points is called a critical point set if its positive and negative parts $P$ and $N$ are not isolable, but if any point is deleted then $P$ and $N$ are isolable.

Rice formulated his necessary and sufficient optimality conditions as follows.

**Theorem 5.1.** (Rice [7]) $L(A^*, x)$ is a best approximation to $f(x)$ if and only if the set of extremal points of $L(A^*, x) - f(x)$ contains a critical point set.

Note that $L(A, x)$ is linear with respect to $A$ (due to (2.1)). Then $\Gamma(A)$ can be interpreted as a linear function (hyperplane). If two convex sets (convex hulls of positive and negative points) are not intersecting, then there is a separating hyperplane, such that these two convex sets lie on opposite sides of this hyperplane.

Note that in our necessary and sufficient optimality conditions we only consider finite subsets of $P$ and $N$, namely, we only consider the set of at most $n + 2$ points from the corresponding sundifferential that are used to form zero on their convex hull. Generally,
there are several ways to form zero, but if we choose the one with the minimal number
of maximal deviation points, then, indeed, the removal of any of the extremal points
will lead to a situation where zero can not be formed anymore and the corresponding
subsets of positive and negative points are isolable (their convex hulls do not intersect).

Therefore, our necessary and sufficient optimality conditions are equivalent to Rice’s
conditions. The main advantages of our formulations are as follows. First of all, our
condition is much simpler, easier to understand and connect with the classical theory of
univariate chebyshev approximation. Second, it is much easier to verify our optimality
conditions, which is especially important for the construction of of a Remez-like algo-

rithm, where necessary and sufficient optimality conditions need to be verified at each
iteration.

6 Conclusions and further research directions

In this paper we obtained necessary and sufficient optimality conditions for best poly-
nomial Chebyshev approximation (characterisation theorem). The main obstacle was to
modify the notion of alternance to the case of multivariate polynomials. This has been
done using nonsmooth calculus and quasidifferentiability. We also propose an algorithm
for optimality verification.

For the future we are planning to proceed in the following directions.

1. Find a necessary and sufficient optimality condition that is easy to verify in practice
   (current, we only have a necessary condition, but not a sufficient one).

2. Extend these results to the case of variable polynomial degrees for each dimension.

3. Develop similar optimality conditions for multivariate trigonometric polynomials
   and polynomial spline Chebyshev approximations.

4. Develop an approximation algorithm to construct best multivariate approximations
   (similar to the famous Remez algorithm, developed for univariate polynomials [5]
   and extended to polynomial splines [10]).

References

[1] P. Borwein, I. Daubechies, V. Totik, and G. Nürnberger, Bivariate segment approxi-
mation and free knot splines: Research problems 96-4, Constructive Approximation 12
(1996), no. 4, 555–558.

[2] O. V. Davydov, G. Nurnberger, and F. Zeilfelder, Approximation order of bivariate
spline interpolation for arbitrary smoothness, 1998.

[3] G. Nürnberg, Approximation by spline functions, Springer-Verlag, 1989.
[4] G. Nurnberger and F. Zeilfelder, *Interpolation by spline spaces on classes of triangulations*, Journal of Computational and Applied Mathematics **119** (2000), no. 1-2, 347 – 376.

[5] E.Ya Remez, *General computational methods of chebyshev approximation*, Atomic Energy Translation **4491** (1957).

[6] J. Rice, *Characterization of chebyshev approximation by splines*, SIAM Journal on Numerical Analysis **4** (1967), no. 4, 557–567.

[7] John Rice, *Tchebycheff approximation in several variables*, Trans. Amer. Math. Soc. **109** (1963), 444–466.

[8] R.T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[9] L. Schumaker, *Uniform approximation by chebyshev spline functions. II: free knots*, SIAM Journal of Numerical Analysis **5** (1968), 647–656.

[10] Nadezda Sukhorukova, *Vallée poussin theorem and remez algorithm in the case of generalised degree polynomial spline approximation*, Pacific Journal of Optimization **6** (2010), no. 1, 103–114.

[11] Nadezda Sukhorukova and Julien Ugon, *Characterization theorem for best linear spline approximation with free knots*, Dynamics of Continuous, Discrete & Impulsive Systems **17** (2010), no. 5, 687–708.

[12] ———, *Characterization theorem for best polynomial spline approximation with free knots*, arXiv:1412.2323 (2013), Submitted.