Chebyshev multivariate polynomial approximation: alternance interpretation

Nadezda Sukhorukova and Julien Ugon and David Yost

Abstract In this paper, we derive optimality conditions for Chebyshev approximation of multivariate functions. The theory of Chebyshev (uniform) approximation for univariate functions was developed in the late nineteenth and twentieth century. The optimality conditions are based on the notion of alternance (maximal deviation points with alternating deviation signs). It is not clear, 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.

1 Introduction

The theory of Chebyshev approximation for univariate functions was developed 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 (for example, trigonometric polynomials) have also been used. In most cases, the optimality conditions are based on the notion of alternance: maximal deviation points with alternating deviation signs.

Nadezda Sukhorukova
Swinburne University of Technology, John St, Hawthorn VIC 3122, Australia and Federation University Australia, Postal address: PO Box 663. Ballarat VIC 3353 e-mail: nsukhorukova@swin.edu.au

Julien Ugon
Federation University Australia, Postal address: PO Box 663. Ballarat VIC 3353 e-mail: j.ugon@federation.edu.au

David Yost
Federation University Australia, Postal address: PO Box 663. Ballarat VIC 3353 e-mail: d.yost@federation.edu.au

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, since $\mathbb{R}^d$, unlike $\mathbb{R}$, is not totally ordered. There have been also several studies in a slightly different direction, namely, 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 functions.

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

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. Finally, in Section 3 we draw our conclusions and underline further research directions.

2 Optimality conditions

2.1 Convexity of the objective

Let us now define the objective function. Suppose that $Q \in \mathbb{R}^d$ is a compact set and a continuous function $f : Q \to \mathbb{R}$ is to be approximated on Q by a function

$$L(\mathbf{A},\mathbf{x}) = a_0 + \sum_{i=1}^{n} a_i g_i(\mathbf{x}), \tag{1}$$

where g_i are the basis functions and the multipliers $\mathbf{A} = (a_1,\dots,a_n)$ are the corresponding coefficients. At a point $\mathbf{x}$ the deviation between the function f and the approximation is defined as follows

$$p(\mathbf{A},\mathbf{x}) = |f(\mathbf{x}) - L(\mathbf{A},\mathbf{x})|. \tag{2}$$

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

$$\Psi(\mathbf{A}) = \|f - a_0 - \sum_{i=1}^{n} a_i g_i\|_\infty, \tag{3}$$

where

$$\|f - a_0 - \sum_{i=1}^{n} a_i g_i\|_\infty = \max_{\mathbf{x} \in Q} \max\{f(\mathbf{x}) - a_0 - \sum_{i=1}^{n} a_i g_i(\mathbf{x}), a_0 + \sum_{i=1}^{n} a_i g_i(\mathbf{x}) - f(\mathbf{x})\}.$$

The approximation problem can be formulated as follows.

$$\text{minimise } \Psi(\mathbf{A}) \text{ subject to } \mathbf{A} \in \mathbb{R}^{n+1}. \tag{4}$$

Since the function $L(\mathbf{A}, \mathbf{x})$ is linear in $\mathbf{A}$, the approximation error function $\Psi(\mathbf{A})$, as the supremum of affine functions, is convex. Convex analysis tools [8] can be applied to study this function.

Define by $E^+(\mathbf{A})$ and $E^-(\mathbf{A})$ the points of maximal positive and negative deviation:

$$E^+(\mathbf{A}) = \left\{\mathbf{x} \in Q : L(\mathbf{A}, \mathbf{x}) - f(\mathbf{x}) = \max_{\mathbf{y} \in Q} p(\mathbf{A}, \mathbf{y})\right\}$$
$$E^-(\mathbf{A}) = \left\{\mathbf{x} \in Q : f(\mathbf{x}) - L(\mathbf{A}, \mathbf{x}) = \max_{\mathbf{y} \in Q} p(\mathbf{A}, \mathbf{y})\right\}$$

and the corresponding sets $G^+(\mathbf{A})$ and $G^-(\mathbf{A})$ as

$$G^+(\mathbf{A}) = \left\{(1, g_1(\mathbf{x}), \ldots, g_n(\mathbf{x}))^T : \mathbf{x} \in E^+(\mathbf{A})\right\}$$
$$G^-(\mathbf{A}) = \left\{-(1, g_1(\mathbf{x}), \ldots, g_n(\mathbf{x}))^T : \mathbf{x} \in E^-(\mathbf{A})\right\}$$

Then the subdifferential of the approximation error function $\Psi(\mathbf{A})$ at a point $\mathbf{A}$ can be obtained using the active affine functions in the supremum [11, Theorem 2.4.18]:

$$\partial\Psi(\mathbf{A}) = \operatorname{co}\left\{G^+(\mathbf{A}) \cup G^-(\mathbf{A})\right\}. \tag{5}$$

2.2 Optimality conditions: general case

In the case of univariate polynomial approximation, the optimality conditions are based on the notion of an alternating sequence.

Definition 1. A sequence of maximal deviation points whose deviation signs are alternating is called an alternating sequence or alternance.

The following theorem holds

Theorem 1. *(Chebyshev, [1]) A degree n polynomial approximation is optimal if and only if there exist $n+2$ alternating points.*

In the case of multivariate approximation there is no natural order and therefore the notion of alternance, as a base for optimality verification, has to be modified. The following theorem holds.

Theorem 2. *A vector* $\mathbf{A}^*$ *is an optimal solution to problem (4) if and only if the convex hulls of the vectors* $(g_1(\mathbf{x}),\dots,g_n(\mathbf{x}))^T$*, built over corresponding positive and negative maximal deviation points, intersect, that is*

$$\operatorname{co}\left\{G^+(\mathbf{A})\right\} \cap \operatorname{co}\left\{-G^-(\mathbf{A})\right\} \neq \emptyset. \tag{6}$$

Proof. The vector $\mathbf{A}^*$ is an optimal solution to the convex problem (4) if and only if

$$\mathbf{0}_{n+1} \in \partial\Psi(\mathbf{A}^*),$$

where Ψ is defined in (3). Note that due to Carathéodory's theorem, $\mathbf{0}$ can be represented as a convex combination of a finite number of points (one more than the dimension of the corresponding space). Since the dimension of the corresponding space is $n+1$, it can be done using at most $n+2$ points.

Assume that in this collection of $n+2$ points k points $(h_i,\ i=1,\dots,k)$ are from $G^+(\mathbf{A})$ and $n+2-k$ $(h_i,\ i=k+1,\dots,n+2)$ points are from $G^-(\mathbf{A})$. Note that $0<k<n+2$, since the first coordinate is either 1 or -1 and therefore $\mathbf{0}_{n+1}$ can only be formed by using both sets ($G^+(\mathbf{A})$ and $-G^-(\mathbf{A})$). Then

$$\mathbf{0}_{n+1} = \sum_{i=1}^{n+2} \alpha_i h_i,\ 0 \le \alpha \le 1.$$

Let $0<\gamma=\sum_{i=1}^{k}\alpha_i$, then

$$\mathbf{0}_{n+1} = \sum_{i=1}^{n+2}\alpha_i h_i = \gamma\sum_{i=1}^{k}\frac{\alpha_i}{\gamma}h_i + (1-\gamma)\sum_{i=k+1}^{n+2}\frac{\alpha_i}{1-\gamma}h_i = \gamma h^+ + (1-\gamma)h^-,$$

where $h^+\in G^+$ and $h^-\in -G^-$. Therefore, it is enough to demonstrate that $\mathbf{0}_{n+1}$ is a convex combination of two vectors, one from $G^+(\mathbf{A})$ and one from $-G^-$.

By the formulation of the subdifferential of Ψ given by (5), there exists a non-negative number $\gamma\le 1$ and two vectors

$$g^+ \in \operatorname{co}\left\{\begin{pmatrix}1\\ g_1(\mathbf{x})\\ g_2(\mathbf{x})\\ \vdots\\ g_n(\mathbf{x})\end{pmatrix} : \mathbf{x}\in E^+(\mathbf{A})\right\},\ \text{and } g^- \in \operatorname{co}\left\{\begin{pmatrix}1\\ g_1(\mathbf{x})\\ g_2(\mathbf{x})\\ \vdots\\ g_n(\mathbf{x})\end{pmatrix} : \mathbf{x}\in E^-(\mathbf{A})\right\}$$

such that $\mathbf{0}=\gamma g^+ - (1-\gamma)g^-$. 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

$$\operatorname{co}\left\{\begin{pmatrix}1\\ g_1(\mathbf{x})\\ g_2(\mathbf{x})\\ \vdots\\ g_n(\mathbf{x})\end{pmatrix} : \mathbf{x}\in E^+(\mathbf{A})\right\}\cap\operatorname{co}\left\{\begin{pmatrix}1\\ g_1(\mathbf{x})\\ g_2(\mathbf{x})\\ \vdots\\ g_n(\mathbf{x})\end{pmatrix} : \mathbf{x}\in E^-(\mathbf{A})\right\}\neq\emptyset. \qquad (7)$$

As noted before, the first coordinates of all these vectors are the same, and therefore the theorem is true, since if γ exceeds one, the solution where all the components are divided by γ can be taken as the corresponding coefficients in the convex combination.

In the rest of this section we show how Theorem 2 can be used to formulate necessary and sufficient optimality conditions for the case of multivariate polynomial approximation. We also demonstrate how the notion of alternance can be extended to multidimensional cases. Equivalent results have been obtained in [7], however, the conditions of Theorem 2 are easier to verify. Rice's optimality verification is based on separation of positive and negative maximal deviation points by a polynomial of the same degree as the degree of the approximation m: there exists no polynomial of degree m that separates positive and negative maximal deviation points, but the removal of any maximal deviation point results in the ability to separate the remaining points by a polynomial of degree m.

2.3 Optimality conditions for multivariate linear functions

In the case of multivariate linear functions (that is $g_i(\mathbf{x}) = x_i,\ i = 1,\dots,n$) Theorem 2 can be formulated as follows.

Theorem 3. *A multivariate linear approximation is optimal if and only if 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 3 can be considered as an alternative formulation to the necessary and sufficient optimality conditions that are based on the notion of alternance. Clearly, Theorem 3 can be used in univariate cases, since the location of the alternance points ensures the common points for the corresponding convex hulls, constructed over the maximal deviation points with positive and negative deviations respectively.

Note that in general $d \leq n$. Non-linear multivariate polynomial approximation is one of our future research priorities.

3 Conclusions and further research directions

In this paper we obtained necessary and sufficient optimality conditions for best polynomial Chebyshev approximation (characterisation theorem). The main obstacle was to extend the notion of alternance to the case of multivariate polynomials. This has been done using nonsmooth calculus.

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 (currently, 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 [3, 10])

Acknowledgements This paper was inspired by the discussions during a recent MATRIX program “Approximation and Optimisation'' that took place in July 2016. We are thankful to the MATRIX organisers, support team and participants for a terrific research atmosphere and productive discussions.

References

1. P.L Chebyshev, *The theory of mechanisms known as parallelograms*, Selected Works, Publishing House of the USSR Academy of Sciences, Moscow, 1955, (In Russian), pp. 611–648.
2. O. V. Davydov, G. Nurnberger, and F. Zeilfelder, *Approximation order of bivariate spline interpolation for arbitrary smoothness*, Journal of Computational and Applied Mathematics **90** (1998), no. 2, 117–134.
3. G. Nürnberger, *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. C. Zalinescu, *Convex analysis in general vector spaces*, World Scientific, 2002.