A generalisation of Vallée-Poussin procedure to multivariate polynomials

Nadia Sukhorukova and Julien Ugon

Chebyshev (uniform) approximation

The theory of Chebyshev approximation for univariate functions was developed in the late nineteenth (Chebyshev) and twentieth century (just to name a few Vallée-Poussin, Rice, Nurnberger, Schumaker). 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 (that is, maximal deviation points with alternating deviation signs).

There have been several attempts to extend this theory to the case of multivariate functions. The main obstacle in extending these results to the case of multivariate functions is that it is not very easy to extend the notion of monotonicity to the case of several variables.

We propose an alternative approach, which is based on the notion of convexity and nonsmooth analysis.

Chebyshev (uniform) approximation

Approximation theory is concerned with the approximation of a function f, defined on a (continuous or discrete) domain Ω, by another function s taken from a family $\mathfrak{S}$. At any point $t \in \Omega$ the difference

$$d(t) \triangleq s(t) - f(t)$$

is called the *deviation* at t, and the *maximal absolute deviation* is defined as

$$\|s - f\| \triangleq \sup_{t \in \Omega} |s(t) - f(t)|.$$

The problem of best Chebyshev approximation is to find a function $s^* \in \mathfrak{S}$ minimising the maximal absolute deviation over $\mathfrak{S}$. Such a function s^* is called a *best approximation* of f.

Chebyshev's Theorem

The seminal result of approximation theory is Chebyshev's alternation theorem which can be stated as follows. Let $\mathcal{P}_n$ be the set of polynomials of degree at most n with real coefficients.

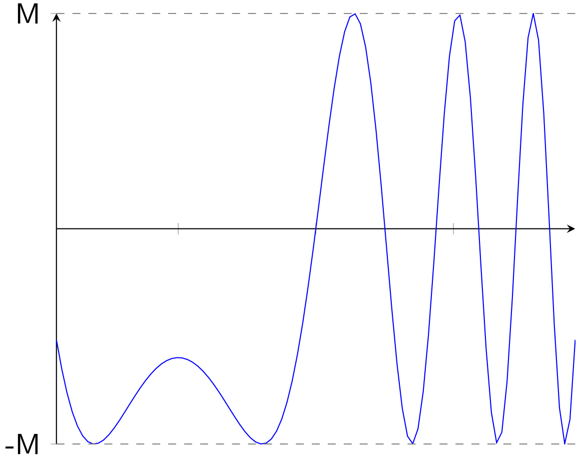
Theorem

(1854) A polynomial $p^* \in \mathcal{P}_n$ is a best approximation to a continuous function f on an interval $[a, b]$ if and only if there exist $n+2$ points $a \leq t_1 < \ldots < t_{n+2} \leq b$ and a number $\sigma \in \{-1, 1\}$ such that

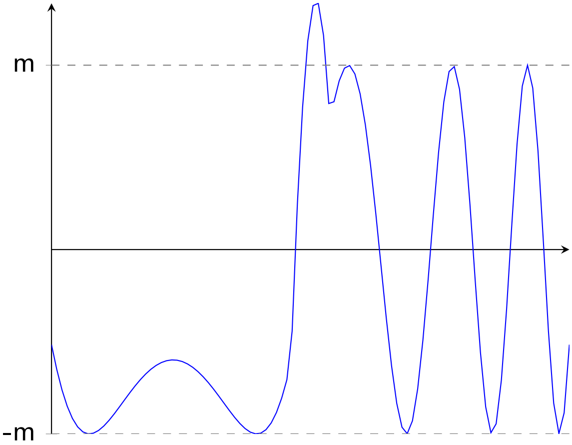
$$(-1)^i \sigma(f(t_i) - p^*(t_i)) = \|f - p^*\|, \forall i = 1 \ldots, n+2.$$

The sequence of points $(t_i)_{i=1,\ldots,n+2}$ is called an *alternating sequence*.

Alternance: example



Alternance: example



Vallée-Poussin procedure to univariate polynomials

Definition

Any $n+2$ points form a basis.

1. For any basis there exists a unique polynomial, such that the absolute deviation at the basis points is the same and the deviation sign is alternating (Chebyshev interpolation polynomial).
2. If there is a point (outside of the current basis), such that the absolute deviation at this point is higher than at the basis points then this point can be included in the basis by removing one of the current basis points and the deviation signs are deviating.
3. The absolute deviation of the new Chebyshev interpolating polynomial is at least as high as the absolute deviation for the original basis.

Multivariate polynomials: definitions and notations

Definition

An exponent vector

$$\mathbf{e} = (e_1, \ldots, e_l) \in \mathbb{R}^l,\ e_i \in \mathrm{N},\ i = 1, \ldots, l$$

for $\mathbf{x} \in \mathbb{R}^l$ defines a *monomial*

$$\mathbf{x}^{\mathbf{e}} = x_1^{e_1} x_2^{e_2} \ldots x_l^{e_l}.$$

Definition

A product $c\mathbf{x}^e$, where $c \neq 0$ is called the term, then a multivariate polynomial is a sum of a finite number of terms.

Multivariate polynomials: definitions and notations

Definition

The degree of a monomial $\mathbf{x}^e$ is the sum of the components of $\mathbf{e}$:

$$\deg(\mathbf{x}^e) = \sum_{i=1}^{l} e_i.$$

Definition

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

Polynomials

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

$$P^m(x) = \sum_{i=0}^{n} a_i M_i(\mathbf{x}, e), \tag{1}$$

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 (1) can be presented as the sum of a lower degree polynomials ($m - 1$ or less) and a finite number of terms that correspond to the monomials of degree m.

Dimension

The dimension of the parameter space of a polynomial of degree m is d. Note that in the case of linear functions and univariate function (that is, $l = 1$) $d = l + 1$. If $l \geq 2$ and $m \geq 2$ then d (the total number of possible monomials of degree at least m) is increasing very fast.

Example

Let $l = 2$ (variables x and y) and $m = 2$. Then the total number of all possible monomials of degree zero is one, of degree one is $l = 2$ (x and y) and of degree two is three (x^2, y^2 and xy).

Convexity of the objective function

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

$$L(\mathbf{A}, \mathbf{x}) = a_0 + \sum_{i=1}^{n} a_i g_i(\mathbf{x}), \tag{2}$$

where $g_i(\mathbf{x}),\ i = 1, \ldots, n$ are the basis functions (not the degree!!!) and the multipliers $\mathbf{A} = (a_0, a_1, \ldots, a_n)$ are the corresponding coefficients. At a point $\mathbf{x}$ the deviation between the function f (also referred as approximation function) and the approximation is:

$$d(\mathbf{A}, \mathbf{x}) = |f(\mathbf{x}) - L(\mathbf{A}, \mathbf{x})|. \tag{3}$$

Objective function reformulation

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

$$\Psi(\mathbf{A}) = \sup_{\mathbf{x}\in Q} \max\{f(\mathbf{x}) - a_0 - \sum_{i=1}^{n} a_i g_i(\mathbf{x}), a_0 + \sum_{i=1}^{n} a_i g_i(\mathbf{x}) - f(\mathbf{x})\}. \tag{4}$$

The approximation problem is

$$\text{minimise } \Psi(\mathbf{A}) \text{ subject to } \mathbf{A} \in \mathbb{R}^{n+1}. \tag{5}$$

Subdifferential

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. Furthermore, its subdifferential at a point $\mathbf{A}$ is trivially obtained using the active affine functions in the supremum:

$$\partial\Psi(\mathbf{A}) = \mathrm{co}\left\{ \begin{pmatrix} 1 \\ g_1(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{pmatrix} : \mathbf{x} \in E_+, - \begin{pmatrix} 1 \\ g_1(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{pmatrix} : \mathbf{x} \in E_- \right\}, \quad (6)$$

where E_+ and E_- are respectively the points of maximal positive and negative deviation:

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

$$E_- = \left\{ \mathbf{x} \in Q : -f(\mathbf{x}) + L(\mathbf{A}, \mathbf{x}) = \max_{\mathbf{y} \in Q} d(\mathbf{A}, \mathbf{y}) \right\}.$$

Necessary and sufficient optimality conditions

In the case of multivariate polynomial approximation, $g_i(\mathbf{x}), i = 1, \dots, n$ are monomials. Denote

$$g(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_n(\mathbf{x})) \in \mathbb{R}^n. \tag{7}$$

The following theorem holds.

Theorem

$\mathbf{A}^*$ *is an optimal solution to problem (5) if and only if the convex hulls of the vectors*

$$g(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_n(\mathbf{x}))^T,$$

built over corresponding positive and negative maximal deviation points, intersect.

Another formulation

Assume that $\mathbf{y}_i \in \mathbb{R}^l,\ i = 1, \ldots, N_+$ are positive deviation points and $\mathbf{z}_j \in \mathbb{R}^l$, $j = 1, \ldots, N_-$ are negative deviation points. Also assume that $N_+ + N_- = n + 2$ and construct the following sets:

$$\mathcal{Y} = \mathrm{co}\left\{g(\mathbf{y}_i),\ i = 1, \ldots, N_+\right\} \tag{8}$$

and

$$\mathcal{Z} = \mathrm{co}\left\{g(\mathbf{z}_j),\ j = 1, \ldots, N_-\right\}. \tag{9}$$

Radon's theorem and Basis

Published by Johann Radon in 1921.

Theorem

Any set of $d + 2$ points in $\mathbb{R}^d$ can be partitioned into two disjoint sets whose convex hulls intersect.

Definition

A point in the intersection of these convex hulls is called a Radon point of the set.

What do we call basis in multivariate case?

Definition

Any set of $d + 2$ points in $\mathbb{R}^d$ is called *basis*.

Relative interior and Basis

Definition

The relative interior of a set S (denoted by $\mathrm{relint}(S)$) is defined as its interior within the affine hull of S. That is,

$$\mathrm{relint}(S) = \{\mathbf{x} \in S : \exists \varepsilon > 0, B_\varepsilon(x) \cap \mathrm{aff}(S) \subseteq S\},$$

where $B_\varepsilon(x)$ is a ball of radius ε centred in x and $\mathrm{aff}(S)$ is the affine hull of S.

Definition

Consider a set $\mathcal{S}$ of $n + 2$ points partitioned into two sets, the sets $\mathcal{Y}$ of points with positive deviation and $\mathcal{Z}$ of points with negative deviation. These points are said to form a *basis* if the convex hulls of $\mathcal{Y}$ and $\mathcal{Z}$ intersect. Furthermore, if the relative interiors of the convex hulls intersect then the basis is said to be *non-singular*.

Useful property

A nice property of relative interiors of convex hulls of finite number of points is formulated in the following lemma.

Lemma

Any relative interior point of a convex combination of a finite number of points can be presented as a convex combination of all these points with strictly positive convex combination coefficients and vice versa.

Affine independent systems and Basis

Definition

Consider a set $\mathcal{S}$ of $n+2$ points partitioned into two sets, the sets $\mathcal{Y}$ of points with positive deviation and $\mathcal{Z}$ of points with negative deviation. These points are said to form a *basis* if the convex hulls of $\mathcal{Y}$ and $\mathcal{Z}$ intersect. Furthermore, if there exists an $(n+1)$-point subset forms an affine independent system in $\mathbb{R}^n$ then the basis is said to be *non-singular*.

Optimality conditions and Basis

Ideal

Definition

Consider a set $\mathcal{S}$ of $n+2$ points partitioned into two sets, the sets $\mathcal{Y}$ of points with positive deviation and $\mathcal{Z}$ of points with negative deviation. These points are said to form a *basis* if the convex hulls of $\mathcal{Y}$ and $\mathcal{Z}$ intersect. Furthermore, if the removal of any point will disconnect the convex hulls then the basis is said to be *non-singular*.

Our definition

Definition

Consider a set $\mathcal{S}$ of $n+2$ points partitioned into two sets, the sets $\mathcal{Y}$ of points with positive deviation and $\mathcal{Z}$ of points with negative deviation. These points are said to form a *basis* if the convex hulls of $\mathcal{Y}$ and $\mathcal{Z}$ intersect. Furthermore, if relative interiors of these sets are intersecting and there exists an $(n+1)$-point subset forms an affine independent system in $\mathbb{R}^n$ then the basis is said to be *non-singular*.

STEP 1: Chebyshev interpolation polynomial

Theorem

Assume that a system of points $\mathbf{y}_i, \ i = 1, \dots, N_+$ *and* $\mathbf{z}_i, \ i = 1, \dots, N_-$ *forms a non-singular basis. Then there exists a* **unique** *polynomial deviating from* f *at the points* $\mathbf{y}_i, \ i = 1, \dots, N_+$ *and* $\mathbf{z}_i, \ i = 1, \dots, N_-$ *by the same value and the deviation signs are opposite for* $\mathbf{y}_i$ *and* $\mathbf{z}_i$.

STEP 1: Proof

$$\begin{pmatrix} 1 & g(\mathbf{y}_1) & 1 \\ 1 & g(\mathbf{y}_2) & 1 \\ \vdots & \vdots & \vdots \\ 1 & g(\mathbf{y}_{N_+}) & 1 \\ 1 & g(\mathbf{z}_1) & \text{-1} \\ 1 & g(\mathbf{z}_2) & \text{-1} \\ \vdots & \vdots & \vdots \\ 1 & g(\mathbf{z}_{N_-}) & \text{-1} \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \sigma \end{pmatrix} = \begin{pmatrix} f(\mathbf{y}_1) \\ f(\mathbf{y}_2) \\ \vdots \\ f(\mathbf{y}_{N_+}) \\ f(\mathbf{z}_1) \\ f(\mathbf{z}_2) \\ \vdots \\ f(\mathbf{z}_{N_-}) \end{pmatrix}, \qquad (10)$$

where $\mathbf{A}$ represents the parameters of the polynomial, while σ is the deviation. If $\sigma = 0$, there exists a polynomial passing through the chosen points (interpolation).

STEP 1: Proof: comments

The proof is relying on the fact that the basis is non-singular (any $n+1$ subset forms an affine independent system). However, the optimality conditions are the same for singular and non-singular basis, therefore we will have to reconsider the notion of singular basis (or, ideally, not use it at all).

$$\begin{aligned}
\mathrm{sign}(\det \tilde{M}) &= 2(-1)^{l+2+i}\,\mathrm{sign}(\det M_i^{+}) && (11)\\
&= 2(-1)^{l+2+N_{+}+j+1}\,\mathrm{sign}(\det M_j^{-}). && (12)\\
& && (13)
\end{aligned}$$

Let $\Delta_k = \det(M_k)$. If now we evaluate the determinant of M directly, then

$$\det M = \sum_{i=1}^{N_{+}} (-1)^{l+2+i}\Delta_i + \sum_{j=N_{+}+1}^{N_{+}+N_{-}} (-1)^{l+2+j+1}\Delta_j. \tag{14}$$

Each component in the right hand side of (14) has the same sign. Moreover, since none of the vertices can be removed without disconnecting the sets, the determinant of M is not zero.

STEP 2: Basis update

Theorem

Consider two intersecting sets constructed as in (8) and (9). Assume now that the deviation sign at $\mathbf{y}$ *is the same as at the vertices of (8) and opposite to the deviation sign at the vertices of (9). There exists a point in the combined collection of vertices of* $\mathcal{Y}$ *and* $\mathcal{Z}$*, that can be removed while* $\mathbf{y}$ *is included in* $\mathcal{Y}$*, such that the updates sets* $\tilde{\mathcal{Y}}$ *and* $\tilde{\mathcal{Z}}$ *are intersecting.*

Proof.

If there exists a point whose removal will not disconnect the convex hulls, this points can be removed and the new points can be included without disconnecting the updated convex hulls. Otherwise, there exist strictly positive coefficients α_i, $i = 1, \ldots, N_+$ and β_i, $j = 1, \ldots, N_-$, such that $\sum_{i=1}^{N_+} \alpha_i = 1$ and $\sum_{j=1}^{N_-} \beta_j = 1$. Find

$$\gamma = \min \left\{ \min_{i=1,\ldots,N_+} \frac{\tilde{\alpha}_i}{\alpha_i}, \min_{j=1,\ldots,N_-} \frac{\tilde{\beta}_j}{\beta_j} \right\}. \tag{15}$$

First, assume that $\gamma = \frac{\tilde{\alpha}_1}{\alpha_1}$. Note that $\alpha_1 \neq 0$, then

$$\mathbf{y}_1 = \frac{1}{\alpha_1} \left(\sum_{j=1}^{N_-} \beta_j g(\mathbf{z}_j) - \sum_{i=2}^{N_+} \alpha_i g(\mathbf{y}_i) \right).$$

Then, the convex hull with the new point $\mathbf{y}$ is

$$\alpha g(\mathbf{y}) + \frac{\tilde{\alpha}_1}{\alpha_1} \left(\sum_{j=1}^{N_-} \beta_j g(\mathbf{z}_j) - \sum_{i=2}^{N_+} \alpha_i g(\mathbf{y}_i) \right) + \sum_{i=2}^{N_+} \alpha_i \tilde{g}(\mathbf{y}_i) = \sum_{j=1}^{N_-} \tilde{\beta}_j g(\mathbf{z}_j)$$

(16)

STEP 3: Deviation update

Theorem

Assume that a point with a higher absolute deviation is included in the basis instead of one of the points of the original basis. The absolute deviation of the Chebyshev interpolation polynomial that corresponds to the new basis is higher than the one of the Chebyshev interpolation polynomial on the original basis.

Conclusions

By replacing the classical definition of "basis" by the proposed one, the following generalisations can be obtained.

1. All the steps of Vallée-Poussin procedure has been generalised to the case of multivariate polynomials for non-singular basis.
2. The corresponding proofs and formulations are very similar to the univariate case.
3. In the case of a singular basis one has a dimension reduction.