Communications in Optimization Theory

Available online at http://cot.mathres.org

MATHRES

APPLICATIONS AND ISSUES IN ABSTRACT CONVEXITY

R. DÍAZ MILLÁN[1], NADEZDA SUKHORUKOVA[2,*], JULIEN UGON[1]

[1]Faculty of Science, Engineering and Built Environment, Deakin University, Australia
[2]Department of Mathematics, Swinburne University of Technology, Australia

Dedication to the memory of Professor Hoang Tuy

Abstract. The theory of abstract convexity, also known as convexity without linearity, is an extension of classical convex analysis. There are a number of remarkable results, mostly concerning duality, and some numerical methods, however, this area has not found many practical applications yet. In this paper, we study the application of abstract convexity to function approximation. Another important research direction addressed in this paper is the connection with the so-called axiomatic convexity.
Keywords. Abstract convexity; Axiomatic convexity; Chebyshev approximation; Carathéodory number; quasi-affine functions.
2020 Mathematics Subject Classification. 90C26, 90C47.

1. INTRODUCTION

Abstract convexity appears as a natural extension of the classical convex analysis when linear and affine functions (elementary functions) are "replaced" by other types of functions. This is why abstract convexity is also known as "convexity without linearity".

It was demonstrated that many results from the classical convex analysis: conjugation, duality, subdifferential related issues can be extended to "non-linear" settings [26, 32]. In [26] the authors provide a very detailed historical review of the theory and it turned out that the origins of this environment comes back to early 70s, while some specific issues were already in place in the 50s of the twentieth century.

Despite very productive work done in the development of abstract convexity, there are still several directions for improvement. From the onset of abstract convex theory, much effort has been devoted to studying duality, leading to very elegant generalisations of classical convexity

results. On the other hand, the geometric aspects of convexity have not received as much attention. Additionally, the theory remains "abstract" and needs practical applications. One of the goals of this paper is to provide some applications and develop a framework on the properties that the elementary functions have to satisfy to be efficient in applications. The application we study is the generalised Chebyshev approximation problem, which consists in uniformly approximating a function f using a certain family of functions. In the classical (and convex) case of polynomial approximation, the approximating function is linear in its coefficients (the decision variables of the problem). We extend this classical case to more general cases. We formulate the problem and show how it can be approached using abstract convexity. Furthermore, when the approximating function is quasiaffine in the decision variables, we propose an algorithm for solving the approximation problem.

There are a number of difficulties to develop efficient computational methods. The first step is to choose the set of elementary functions. This set should be simple enough to construct fast and efficient methods. On another hand, this set has to generate accurate approximations to the functions from the applications. In some cases, the choice of elementary functions is reasonably straightforward (see section 3), but this is rather exceptional and in most applications, the choice of a suitable class of elementary functions is a difficult task.

A deep and detailed study on abstract convexity and related issues can be found in [32] and also in [30]. A detailed survey of methods of abstract convex programming can be found in [2]. Many more contributions and developments and results on duality and methods are not mentioned in this paper, but the most recent state of the developments in abstract convexity and a comprehensive literature review can be found in [6].

The paper is organised as follows. In section 2 we provide the essential background of abstract convexity and axiomatic convexity. In the same section, we underline connections and possible cross-feeding between these two types of convexity. Then, in section 3 we illustrate how abstract convexity and axiomatic convexity can be applied to function approximation. Section 4 illustrates the results of numerical experiments. Finally, section 5 provides conclusions and future research directions.

2. ABSTRACT CONVEXITY

2.1. **Definitions and preliminaries.** Considering the set of all real-valued functions acting on the domain $X \subseteq \mathbb{R}^n$, denoted by $\mathscr{F}$ and defined by $\mathscr{F} := \{f : X \to \mathbb{R}\}$. We called the set of *abstract linear functions* to any subset of functions $L \subseteq \mathscr{F}$. Following, we introduce some known definitions to state the abstract convexity concept. Suppose that we defined the set of abstract linear functions $L \subseteq \mathscr{F}$.

Definition 2.1. The vertical closure of L, $H := \{l + c : l \in L, c \in \mathbb{R}\}$ is the set of *abstract affine functions*.

Definition 2.2 (Abstract Convexity [30]). A function f is said to be *L-convex* if there exists a set $U \subset H$ such that for any $x \in X$, $f(x) = \sup_{u \in U} u(x)$. Such functions are also called *abstract convex with respect to L*.

In this paper, we denote the set of all *L*-convex functions by $\mathscr{F}_L$.

Definition 2.3 (Support Set [30])**.** The *L-support set* of a function $f \in F$ is defined as: $\text{supp}_L f := \{l \in L : l \leq f\}$, where $l \leq f$ means that $l(x) \leq f(x), \forall x \in X$. The support sets of *L*-convex functions are called *L-convex sets*.

Therefore, informally, convexity without linearity simply means that the "role" of linear functions of the classical convex analysis is held by functions from a certain class. The first (and for some applications, this is the most important) question is how to choose the set of elementary functions.

The next definition, which will be detailed and used in the next subsection, "helps" the chosen of a "better" set of abstract linear functions.

Definition 2.4 (Supremal generator [14, 30])**.** Let $\mathscr{F}$ be a set of functions defined on a set X. A set $L \subset \mathscr{F}$ is called a *supremal generator* of $\mathscr{F}$ if each $f \in \mathscr{F}$ is abstract convex with respect to L.

2.2. **Quasiconvexity.** The notion of quasiconvexity plays an essential role in abstract convexity. We are now going to explain why.

The notion of quasiconvexity was originally introduced in [11], where the author studied the behaviour of functions with convex sublevel sets, but the term quasiconvexity was introduced much later.

Definition 2.5. Let D be a convex subset of $\mathbb{R}^n$. A function $f : D \to \mathbb{R}$ is *quasiconvex* if and only if its sublevel set $S_\alpha = \{x \in D \,|\, f(x) \leq \alpha\}$ is convex for any $\alpha \in \mathbb{R}$.

There are several equivalent definitions of quasiconvex functions, but the one we use is convenient for our study.

Definition 2.6. A function $f : D \to \mathbb{R}$, where D is a convex subset of $\mathbb{R}^n$, is called quasiconcave if $-f$ is quasiconvex.

Definition 2.7. Functions that are both quasiconvex and quasiconcave are called quasiaffine (quasilinear).

Note the following important observations.
- Quasiconvex functions do not need to be continuous.
- In the case of univariate functions, quasiaffine functions are monotone functions.

If a function is quasiaffine on $\mathbb{R}^n$ (unconstrained problems) then, from the definition, its level sets must be half-spaces, and it is clear that the hyperplanes defining these half-spaces need to be parallel. In the presence of constraints, this observation is not valid and we provide an example in section 4.

There are many studies dedicated to quasiconvex functions and quasiconvex optimisation [5, 9, 10, 14, 30, 31]. In these studies, the notion of quasiconvexity appears as one of the possible generalisations of convexity.

2.3. **Quasiconvex functions and supremal generators.** Theorem 7.13 from [14] states the following.

Theorem 2.8. *The set of all lower semicontinuous quasiaffine functions forms a supremal generator of the set of all lower semicontinuous quasiconvex functions.*

Essentially, this theorem states that in the case of quasiconvex minimisation problems, quasi-affine functions "replace" the role of linear functions in classical convex settings. Therefore, the choice of elementary functions is clear (all quasiaffine functions). One main advantage is that the sublevel sets of "new" elementary functions are half-spaces and this is similar to classical convex analysis.

In the rest of this section, we clarify the "roles" of quasiaffine functions in the new abstract convex settings. Then section 4 contains the results of computational experiments and illustrations.

2.4. **Relation to Axiomatic Convexity.** *Axiomatic Convexity* aims at generalising the notion of convexity using only set-theoretic definitions. Good reviews can be found in [33, 34]. Given a set X, a family $\mathscr{C}$ of subsets of X is called a *convexity* if it follows the following axioms:

(1) $\emptyset$ and V belong to $\mathscr{C}$;
(2) Any arbitrary intersection of sets from $\mathscr{C}$ is in $\mathscr{C}$: if $D \subset \mathscr{C}$,

$$\bigcap_{A \in D} A \in \mathscr{C}$$

(3) The nested union of sets from $\mathscr{C}$ is in $\mathscr{C}$: given a family of sets $D \subset \mathscr{C}$ that is totally ordered by inclusion then

$$\bigcup_{A \in D} A \in \mathscr{C}.$$

The pair $(X, \mathscr{C})$ is called a *convex structure* (or *(abstract) convex space*).

These axioms (in particular axioms 1 and 2) ensure the existence of the convex hull of any set $S \subset X$, namely the smallest convex set containing S:

$$\mathrm{co}_{\mathscr{C}} S = \bigcap \{A \in \mathscr{C} : S \subset A\}.$$

If F is finite, then $\mathrm{co}_{\mathscr{C}} F$ is called a *polyhedron*.

Axiomatic Convexity, also called *abstract convexity* by [34], has found applications in combinatorial geometry [13, 18], among others. A family $\mathscr{C}$ that satisfies the first two axioms above is called a *closure space*, or $\cap$-stable [15].

There are clear connections between abstract convexity as defined by Definition 2.3 and axiomatic convexity. Indeed, the notion of abstract convex function (Definition 2.2) first appeared in [15] in the context of closure spaces. The families of sets considered in [15] were the closures of the set of sublevel sets of a given family of functions.

Abstract convex functions (Definition 2.2) and Abstract convex sets (Definition 2.3) were then formally introduced by [17], who showed the equivalence between these sets and the ones considered by [15]. Below we formalise some results on the correspondence between abstract convex sets and closure spaces. Let us start by noting that the support set of any (not necessarily L-convex) function is L-convex. Indeed, the support set of f is also the support set of its L-convex envelope, which is convex.

Proposition 2.9. *Consider a family of functions L. Then, the set of L-convex sets form a closure space. Furthermore, every closure space is isomorphic to such a set.*

Proof. Consider a family of functions L and let $\mathscr{C}$ be the set of L-convex sets. We first show that L is closed with respect to inclusion. Indeed, let $D \subset \mathscr{C}$ be an arbitrary set of L-convex sets. For each $A \in D$, let $f_A(x) = \sup_{l \in A} l(x)$ and consider the function $f(x) = \inf_{A \in D} f_A(x)$. We

will show that $\cap_{A\in D}A = \operatorname{supp} f$, which implies that $\cap_{A\in D}A$ is L-convex (note that $\operatorname{supp} f$ is the support set of the lower convex envelope of f).

$$
\begin{aligned}
\operatorname{supp} f &= \{l \in L : l(x) \le \inf_{A\in D} f_A(x) \forall x \in X\} \\
&= \{l \in L : l(x) \le f_A(x) \forall A \in D, x \in X\} \\
&= \cap_{A\in D}\{l \in L : l(x) \le f_A(x), \forall x \in X\} = \cap_{A\in D}A.
\end{aligned}
$$

The first axiom is clearly verified by $\mathscr{C}$ from the fact that $\emptyset$ is the support set of $\sup_{l\in\emptyset} l$ and L is the support set of $\sup_{l\in L} l$.

The second part of the statement is obtained by considering the set L of indicator functions of the sets in the closure space. More precisely, let $\mathscr{C}$ be a closure space, and consider the set $L = \{i_A : A \in \mathscr{C}\}$ where

$$
i_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{otherwise.} \end{cases}
$$

Then, L is also the set of L-convex functions. Indeed, let $D \subset \mathscr{C}$. Define $U = \{i_A : A \in D\} \subset L$ and $f = \sup_{l\in U} l = \sup_{A\in D} i_A$. Then $f(x) = 0$ if and only if $x \in A$ for all $A \in D$. That is, $f = i_{\cap_{A\in D}A}$. The isomorphism between the set of support sets of functions from L and $\mathscr{C}$ is evident. $\qquad\square$

Note that in practice it is not necessary to consider the entire set of indicator functions, but only the set of indicator functions of a *basis* of $\mathscr{C}$ (i.e., a set of functions whose closure is $\mathscr{C}$), which corresponds to a suppremal generator of the corresponding abstract functions. It can also easily be seen that the set of domains of the L-abstract functions defined as in the above proof, as well as the set of 0-sublevel sets, is precisely $\mathscr{C}$.

On the other hand, not all families of abstract convex sets are convexity structures.

Example 2.10. Let L be the set of linear functions on $\mathbb{R}$, and consider the functions f_α defined by $f_\alpha(x) = \alpha|x|$, for $\alpha > 0$. These functions are L-convex, and induce a nested sequence of L-convex sets $S_\alpha = \operatorname{supp}_L f_\alpha = \{x \to \beta x : \beta \in [-\alpha, \alpha]\}$.

However,

$$
U = \bigcup_{\alpha<1} S_\alpha = \{x \to \beta x : \beta \in (-1, 1)\}
$$

is not a L-convex set. Indeed, $f_1 = \sup_{l\in U} l$ is the smallest L-convex function whose support set contains U, but the function $x \to x \in \operatorname{supp} f_1 \setminus U$, which shows that U cannot be the support set of f_1 and therefore of any L-convex function. This shows that the set of L-convex sets is not a convex structure.

Let us introduce the notion of a strict support set, which will enable us to address this issue. The L-strict support set of a function f is defined as:

$$
\underline{\operatorname{supp}} f = \{l \in L : l(x) < f(x), \forall x \in \operatorname{dom}(f)\}.
$$

We define the family of sets $\mathscr{C}_L$ as follows.

Definition 2.11 (convexity extension)**.**

$$\mathscr{C}_L^f := \{A \in L : \underline{\text{supp}} f \subseteq A \subseteq \text{supp} f\}$$
$$\mathscr{C}_L := \bigcup_{f \in L} \mathscr{C}_L^f.$$

We call the set $\mathscr{C}_L$ the *convexity extension* of the set of support sets of L-convex functions.

Proposition 2.12. *For any family of functions L, the convexity extension of L, $\mathscr{C}_L$, forms a convexity structure.*

Proof. Since $\mathscr{C}_L$ contains all L-convex sets, it contains a closure space and therefore by Proposition 2.9, it contains the sets $\emptyset$ and X.

To see why the second axiom is satisfied consider an arbitrary family of sets $D \subset \mathscr{C}_L$. For each $A \in D$, there exists a L-function f_A such that $\underline{\text{supp}} f_A \subseteq A \subseteq \text{supp} f_A$. Define $f = \text{co}_L \inf_{A \in D} f_A$. Since $A \subseteq \text{supp} f_A$ for any A, we have that

$$\bigcap_{A \in D} A \subset \bigcap_{A \in D} \text{supp} f_A = \text{supp} f,$$

where the last equality was shown in the proof of Proposition 2.9. Additionally, we have:

$$\underline{\text{supp}} f = \{l \in L : l(x) < \inf_{A \in D} f_A(x) \forall x \in X\}$$
$$\subset \{l \in L : l(x) < f_A(x) \forall A \in D, x \in X\}$$
$$= \cap_{A \in D} \{l \in L : l(x) < f_A(x) \forall x \in X\} \subset \cap_{A \in D} A.$$

Therefore $\underline{\text{supp}} f \subseteq \cap_{A \in D} A \subseteq \text{supp} f$, we conclude that $\cap_{A \in D} A$ is in $\mathscr{C}_L$.

Finally, to see that the third axiom is also satisfied, let D be an ordered family of sets from $\mathscr{C}$. For each $A \in D$, we define f_A as above. The nested nature of the sets in D implies that for A, A' in D, if $A \leq_D A'$, then

$$f_A \leq f_{A'}. \tag{2.1}$$

Let $S = \cup_{A \in D} A$ and $f = \sup_{A \in D} f_A$.

It is clear that $S \subset \text{supp} f$, since for any $u \in S$ there exists $A \in D$ such that $u \in \text{supp} f_A$. Then $u \leq f_A \leq f$. Now, consider $u \notin S$. Then $u \notin \cup_{A \in D} \text{supp} f_A$, and so there exists $x \in X$ such that $u(x) \geq f_A(x)$ for any $A \in D$, and therefore that $\overline{u(x)} \geq \sup_{A \in D} f_A(x) = f(x)$. Therefore $u \notin \underline{\text{supp}} f$. This implies that $\underline{\text{supp}} f \subseteq S \subseteq \text{supp} f$, and therefore S is in $\mathscr{C}_L$ and $\mathscr{C}_L$ is a convexity structure. $\square$

Remark 2.13. Propositions 2.9 and 2.12 imply an equivalence relation between sets of abstract linear functions according to the convexity structure (axioms 1-3) or closure space (axioms 1,2) the induced abstract convex sets are isomorphic to.

Axiomatic convexity has been applied to obtain generalisations of well-known geometric results in convexity theory. Of particular interest to this paper are generalisations of Carathéodory's theorems. Much research has been devoted to investigating this topic. We refer to [34] for a review of classical results in the area.

Definition 2.14 ([34, Definition 1.5]). *In a given Convexity structure $\mathscr{C}$, a set F is Carathéodory dependent if $\mathrm{co}_{\mathscr{C}} F \subseteq \bigcup_{a \in F} \mathrm{co}_{\mathscr{C}}(F \setminus \{a\})$, and Carathéodory independent otherwise.*

Then we can define the *Carathéodory number* of a convex structure $\mathscr{C}$ as the largest cardinality of a Carathéodory independent set (i.e., any set F with cardinality greater than $c(x)$ is Carathéodory dependent). The Carathéodory numbers of several important classes of convexity structures are known and discussed in the literature [34].

Generalisations of Helly's and Radon's theorems were obtained similarly. The Carathéodory number enables the following generalisation of Carathéodory's classical result:

Proposition 2.15 ([34, Theorem 1.7]). *Consider a convexity structure $(\mathscr{C}, X)$ with Carathéodory number $c(x)$, and let $A \subset \mathscr{C}$ be a convex set from $\mathscr{C}$. Then, for any $x \in A$ there exists a set $F \subset A$ such that $|F| \leq c(x)$ and $x \in \mathrm{co}F$. This is the best possible.*

3. APPLICATION TO APPROXIMATION

3.1. **Problem formulation.** We are working with uniform (Chebyshev) approximation. Define a function $g : \mathbb{R}^k \times X \to \mathbb{R}$, where $X \subset \mathbb{R}^n$ is a compact set and $k > 0, n > 0$ are integers. The approximation problem is:

$$\min_{A \in \mathbb{R}^k} \sup_{\mathbf{x} \in X} |f(\mathbf{x}) - g(A, \mathbf{x})|. \tag{3.1}$$

In (3.1) A is the set parameters of approximation (e.g., polynomial coefficients), that are the decision variables for the optimisation problems. For example, in the case of polynomial approximation, we choose the polynomial coefficients in such a way that the maximal deviation over X is minimal.

In many computer-based applications, this set is a finite grid defined on a convex compact set. If we define $L = \{A \to g(A, \mathbf{x}) : x \in X\}$, then it is clear from formulation (3.1) that the objective function of this problem is H_L-convex.

In the case when the approximation function $g(A, t)$ is a polynomial (and its coefficients are subject to optimisation), the optimality conditions are based on maximal deviation points (Chebyshev theorem, see [7]). In the case of univariate function approximation, the conditions are based on the notion of alternating sequence.

Theorem 3.1 (Chebyshev). *A polynomial of degree at most n is optimal in uniform (Chebyshev) norm if and only there are $n + 2$ alternating points.*

Essentially, the approximations here are presented as linear combinations of known univariate functions. These functions are called basis functions. In the case of polynomials, the basis functions are monomials, but other classes of basis functions can be used. The decision variables are the coefficients (parameters) of the linear combinations of basis functions. The corresponding optimisation problems are convex as the supremum of linear forms. Theorem 3.1 can be proved using convex analysis and the number of alternative points $(n + 2)$ obtained through Carathéodory's theorem. In other words, the following conjecture is true, in the case when L is a set of linear functions:

Conjecture 3.2. *For a function from L to be an optimal approximation in the uniform norm, it is enough that it is optimal at $c(\mathscr{C}_{H_L}) + 1$ extreme points.*

We also conjecture that this is best possible, in the sense that the statement is generally not true if we replace $c(\mathscr{C}_{H_L}) + 1$ with $c(\mathscr{C}_{H_L})$.

If true, this conjecture could form the basis for algorithms for best approximation, such as a generalisation of Vallée-Poussin's procedure [12]. Later in this section, we discuss an example of families L in the context of this conjecture.

One possible generalisation to polynomial approximation is rational approximation, that is, approximation by the ratio of two polynomials, whose coefficients are subject to optimisation. Note that in some cases, the degree of the denominator and numerator can be reduced without compromising the accuracy:

$$\frac{\sum_{j=0}^{n} a_j t_i^j}{\sum_{k=0}^{m} b_k t_i^k} = \frac{\sum_{j=0}^{n-\nu} a_j t_i^j}{\sum_{k=0}^{m-\mu} b_k t_i^k},$$

where $d = \min\{\nu, \mu\}$ is called *the defect*. Then the necessary and sufficient optimality conditions are as follows [1].

Theorem 3.3. *A rational function in $\mathscr{R}_{n,m}$ with defect d is the best polynomial approximation of a function $f \in C^0(I)$ (the space of continuous functions over I) if and only if there exists a sequence of at least $n + m + 2 - d$ points of maximal deviation were the sign of the deviation at these points alternates.*

Therefore, similar to polynomial approximation, these conditions are based on the number of alternating points.

In connection to Conjecture 3.2, $c(\mathscr{C}_{H_L})$ is $n + m + 2$. If n and m are the degrees of the corresponding polynomials, then the total number of the decision variables is $n + 1 + m + 1 = n + m + 2$. However, one of the decision variables can be fixed (otherwise one can divide the numerator and denominator by the same number). Therefore, there are $n + m + 1$ "free" variables and we add one more for $c(\mathscr{C}_{H_L})$ (similar to Carathéodory's theorem from classical convex analysis). The proof of this conjecture is out of the scope of the current paper: this result is non-trivial due to the effect of the defect.

Rational approximation was a very popular research area in the 1950s-70s [1, 4, 21, 28, 29] (just to name a few).

If the basis functions are not restricted to monomials, the approximations are called generalised rational approximations. The term is due to Cheney and Loeb [8].

One way to approach rational approximation is through constructing "near optimal" solutions [25]. This approach is very efficient and therefore very popular. The extension of this approach to non-monomial basis functions and multivariate approximation remains open (an extension to complex domains can be found in [24]).

An alternative way is based on modern optimisation techniques. This approach is preferable when the basis of functions is not restricted to univariate monomials and when a deeper theoretical study is required. The corresponding optimisation problems are quasiconvex and can be solved using general quasiconvex optimisation methods.

Rational and generalised rational functions are quasiaffine as ratios of linear forms [5, 19, 27]. The corresponding optimisation problems can be solved efficiently by applying a simple, but robust approach called the bisection method for quasiconvex functions [5].

In [27], the authors used a well-known bisection method for quasiconvex functions (see [5]) to solve these problems. In [23], the authors used a projection-type algorithm for solving these problems.

3.2. **Bisection method.**

Bisection method for quasiconvex functions [5] can be used efficiently for rational and generalised rational approximation, including multivariate settings [22, 27] when the compact set X is finite. Can we use abstract convexity to extend this method to a wider class of approximations? The answer is positive. This method can be extended to any type of approximation if these approximations are quasiaffine with respect to the approximation parameters (that is, the decision variables of the corresponding optimisation problem), if we know how to find their sub-level sets. The reasoning is provided in the rest of this section.

The problem can be formulated as follows:

$$\text{minimise } \tilde{z} \tag{3.2}$$

subject to

$$f(\mathbf{x}) - \frac{\mathbf{A}^T \mathbf{G}(\mathbf{x})}{\mathbf{B}^T \mathbf{H}(\mathbf{x})} \leq \tilde{z}, \ \mathbf{x} \in X, \tag{3.3}$$

$$\frac{\mathbf{A}^T \mathbf{G}(\mathbf{x})}{\mathbf{B}^T \mathbf{H}(\mathbf{x})} - f(\mathbf{x}) \leq \tilde{z}, \ \mathbf{x} \in X, \tag{3.4}$$

$$\mathbf{B}^T \mathbf{H}(\mathbf{x}) \geq \delta, \ \mathbf{x} \in X, \tag{3.5}$$

where $\tilde{z}$ is the maximal deviation and δ is a small positive number. The functions $\mathbf{G}$ and $\mathbf{H}$ are vector-valued functions, such that their components are basis functions for the numerator and denominator respectively. In other words, the functions $g(\mathbf{A}, \mathbf{x}) = \mathbf{A}^T \mathbf{G}(\mathbf{x})$ and $h(\mathbf{B}, \mathbf{x}) = \mathbf{B}^T \mathbf{H}(\mathbf{x})$ are linear functions of the decision variables of the problem. Problem (3.2)-(3.5) is not linear.

The idea of this method is based on the fact that all the sublevel sets of quasiconvex functions are convex and therefore the sublevel sets of quasiaffine functions are half-spaces. Essentially, this means that the constraint set (3.3)-(3.5) for a fixed $\tilde{z}$ is an intersection of a finite number of half-spaces (X is a finite grid) and therefore it is a polytope.

The algorithm starts with an upper and lower bound (u and l) for the maximal deviation, then the sublevel set for the maximal deviation at the level $z = \frac{u+l}{2}$ is a convex set and the algorithm checks if this set is empty. If it is empty, then the upper bound is updated to z, otherwise, the lower bound is updated to z. The algorithm terminates when the upper and lower bounds are within the specified precision.

In general, checking if the convex set (sublevel set of the maximal deviation) is empty or not may be a difficult task (convex feasibility problems). There are a number of efficient methods ([3, 20, 35, 36, 37] just to name a few), but there are still several open problems here. The discussion of these problems is out of the scope of the current paper.

In the case of multivariate generalised rational approximation, however, this problem (convex feasibility) can be reduced to solving a linear programming problem [22]. The denominator of the approximation does not change the sign. Assume, for simplicity, that it is positive, then the problem of checking the feasibility is equivalent to solving the following problem:

$$\text{minimise } \tilde{u} \tag{3.6}$$

subject to

$$f(\mathbf{x})\mathbf{B}^T\mathbf{H}(\mathbf{x}) - \mathbf{A}^T\mathbf{G}(\mathbf{x}) \leq z\mathbf{B}^T\mathbf{H}(\mathbf{x}) + \tilde{u}, \ \mathbf{x} \in X, \tag{3.7}$$

$$\mathbf{A}^T\mathbf{G}(\mathbf{x}) - f(\mathbf{x})\mathbf{B}^T\mathbf{H}(\mathbf{x}) \leq z\mathbf{B}^T\mathbf{H}(\mathbf{x}) + \tilde{u}, \ \mathbf{x} \in X, \tag{3.8}$$

$$\mathbf{B}^T\mathbf{H}(\mathbf{x}) \geq \delta, \ \mathbf{x} \in X, \tag{3.9}$$

where $z = \frac{1}{2}(u + l)$ is the current bisection point (bisecting the possible values of maximal deviation).

If an optimal solution $\tilde{u} \leq 0$, the corresponding sublevel set of the maximal deviation function is not empty (update the upper bound), otherwise the set is empty (update the lower bound). If X is a finite grid, then (3.6)-(3.9) is a linear programming problem and can be solved efficiently at each step of the bisection method.

To summarise, an efficient implementation of the bisection method for quasiconvex functions can be extended to approximation by any type of quasiaffine function. Indeed, if rational approximations are replaced by any other quasiaffine functions (with respect to the approximation parameters, which are also the decision variables), then each constraint in (3.3)-(3.4) represents a sublevel set of a quasiaffine functions, which is a half-space. Therefore, the problem is reduced to solving linear programming problems. The main problem is how to find this polytope (that is, how to find sublevel sets of an arbitrary quasiaffine function). In the case of rational and generalised rational approximation, this problem is simple, but not for some other types of quasiaffine approximations. In Section 4, we give more examples where the construction of this polytope is straightforward (composition of monotone and affine or ratios of affine functions).

One possibility for constructing sublevel sets for smooth quasiaffine approximation is to use their gradients (if it is not vanishing). In the rest section, we give an example demonstrating that this approach may be complicated if there are constraints (even simple linear constraints). In particular, this approach will not work even in the case of rational approximation due to the requirement for the denominator to be strictly positive.

3.3. Approximation by a quasiaffine function.
Suppose that instead of a generalised rational approximation one needs to approximate by a function from a different class, but the approximations are quasiaffine functions with respect to the decision variables. Then the problem is as follows:

$$\text{minimise } \tilde{z} \tag{3.10}$$

subject to

$$f(\mathbf{x}) - g(A, \mathbf{x}) \leq \tilde{z}, \ \mathbf{x} \in X, \tag{3.11}$$

$$g(A, \mathbf{x}) - f(\mathbf{x}) \leq \tilde{z}, \ \mathbf{x} \in X, \tag{3.12}$$

where $g(A, \mathbf{x})$ is a quasiaffine function with respect to A. Since the sublevel sets of quasiaffine functions are half-spaces, the constraint set (3.11)-(3.12) is a polytope, since X is a finite grid.

Remark 3.4. A finite number of linear constraints may be added to (3.10)-(3.12), while the constraint set remains a polytope.

Therefore, the constraint set (for any fixed $\tilde{z}$), with or without additional linear constraints, is a polytope and we only need to check if this polytope is non-empty. However, we still need an efficient approach to finding this polytope. In the case when the approximations are smooth functions, one can use the gradient as a possible normal vector to the hyperplanes, providing that

it is not a zero vector. In section 4 we study the examples where the quasiaffine approximations can be decomposed as a composition of strictly monotone and affine (quasiaffine) functions. Other situations may be harder.

Note that if a function is quasiaffine on the whole space, then the sublevel sets are half-spaces, whose boundary hyperplanes are parallel. Otherwise, two hyperplanes will intersect. If there are additional (even linear) constraints, this observation may not be valid anymore. The following example illustrates that this may happen even in a very simple case of two variables.

Example 3.5. Consider $f(x,y) = \frac{x}{y}$, where $y > 0$. The sublevel sets are

$$S_\alpha = \{(x,y) : \frac{x}{y} \le \alpha, y > 0\},$$

where α is a given real number. Then S_α can be described as

$$\{x - \alpha y \le 0, \quad y > 0\}.$$

Each sublevel set is still a half-space, but the corresponding hyperplanes (in this example they correspond to level sets) are not parallel. These hyperplanes intersect at $(0,0)$, but this point is excluded from the domain due to the requirement for the denominator y to be strictly positive.

This example demonstrates that the computation of sublevel sets for quasiaffine functions may be challenging. In the next section, we give some examples of the classes of approximations that can be handled efficiently. We also provide the results of the numerical experiments.

4. NUMERICAL EXPERIMENTS

4.1. **Strictly monotone function in composition with an affine function.** In this example we approximate the function

$$f(x,y) = (-x + y^3 + x^4)^4 \tag{4.1}$$

by a quasiaffine function in the form

$$g(A,x,y) = (a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 y^2 + a_6 xy)^3, \tag{4.2}$$

where $A = (a_1, a_2, a_3, a_4, a_5, a_6)$ are the decision variables, x and y form a grid on $[-1,1]$ with the step-size 0.1. Function $g(A,x,y)$ is quasiaffine as a composition of a monotonous (cubic) function and an affine function (see [5, 30] for details). The optimal coefficients are (rounded to two decimal places):

$$A = (-1.88, -0.75, 0.31, 3.29, 0.98, -0.74),$$

the maximal absolute deviation is 8.01 (to two decimal places).

Figure 1 represents the function $f(x,y)$, figure 2 depicts the best found approximation and figure 3 contains the deviation function $f(x,y) - g(A,x,y)$, which is the error of approximation.

Function $f(x,y)$ is almost flat with an abrupt increase around point $(-1,1)$. Visually, the approximation resembles the shape of the original function and the magnitude of the deviation confirms that the approximation is reasonably good.
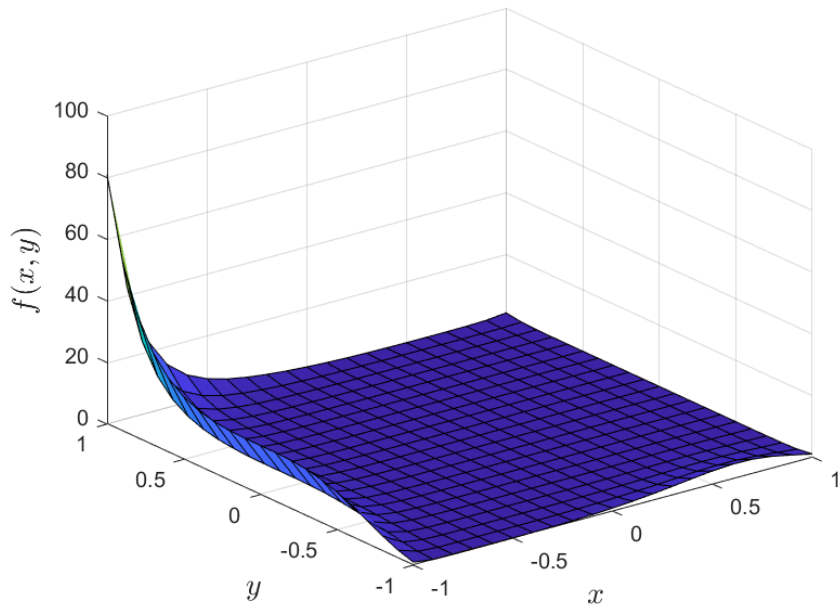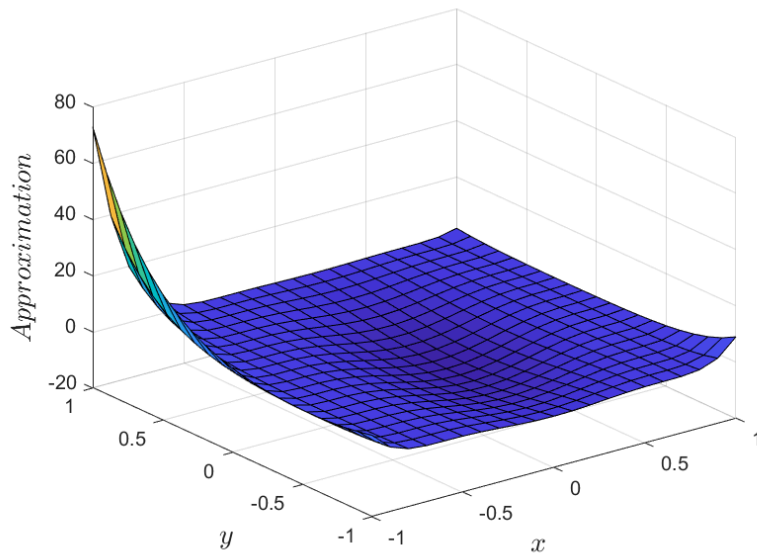
FIGURE 1. Function $f(x, y) = (-x + y^3 + x^4)^4$.

FIGURE 2. Approximation $g(A, x, y) = (a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 y^2 + a_6 xy)^3$.

4.2. Strictly monotone function in composition with a rational function.
In this example, we approximate the same function
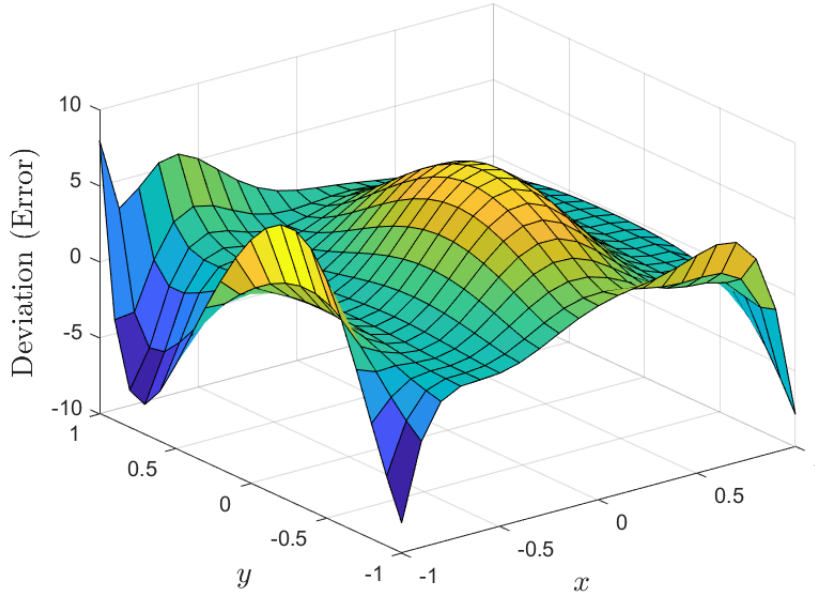
$$f(x, y) = (-x + y^3 + x^4)^4 \qquad (4.3)$$

FIGURE 3. Deviation (error) function $f(x,y) - g(A,x,y) = (-x + y^3 + x^4)^4 - (a_1 + a_2x + a_3y + a_4x^2 + a_5y^2 + a_6xy)^3$.

by a quasiaffine function in the form

$$g_1(A,x,y) = \left(\frac{a_1 + a_2x + a_3y + a_4x^2 + a_5y^2}{1 + a_6xy} \right)^3, \tag{4.4}$$

where, similar to the previous example, $A = (a_1, a_2, a_3, a_4, a_5, a_6)$ are the decision variables (the number of decision variables is the same as in the previous example), x and y are on a grid $[-1, 1]$ with the step-size 0.1. Function $g_1(A,x,y)$ is quasiaffine as a composition of a monotonous (cubic) function and an affine function (see [5] for details). The purpose of approximation is to use lower degree polynomials in the composition. The optimal coefficients are (rounded to two decimal places):

$$A = (-1.66, -0.93, 1.05, 2.87, 0.98, 1),$$

the maximal absolute deviation is 11.48 (to two decimal places).

Figure 4 contains the best found approximation and figure 5 depicts the deviation function $f(x,y) - g_1(A,x,y)$, which is the error of approximation.

The new approximation still resembles the shape of the original function, but the maximal deviation is higher.

4.2.1. *Deep learning applications.* At first glance, the class of a composition of a monotone univariate function with an affine or rational function is restrictive. However, it has many practical applications, including deep learning, where compositions of univariate activation functions and affine transformations are the main components of the models [16]. Most common choices for activation functions are sigmoid functions, ReLU and Leaky ReLU (the last two functions are simply non-decreasing piecewise linear functions with two linear pieces).
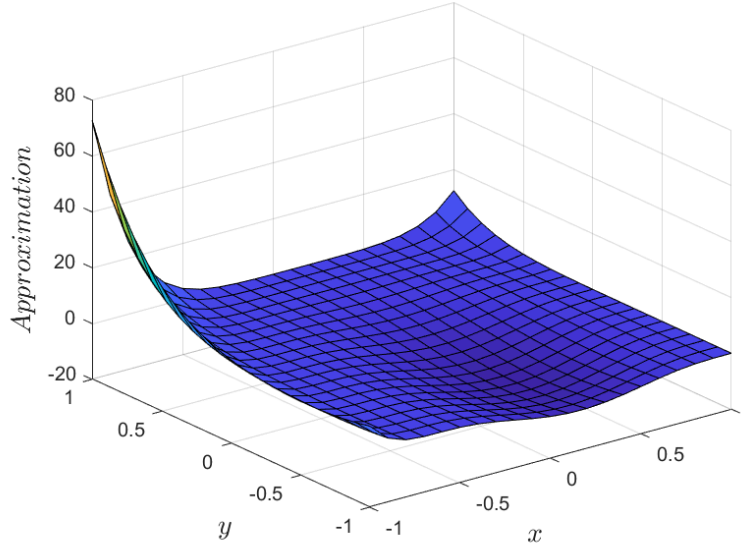
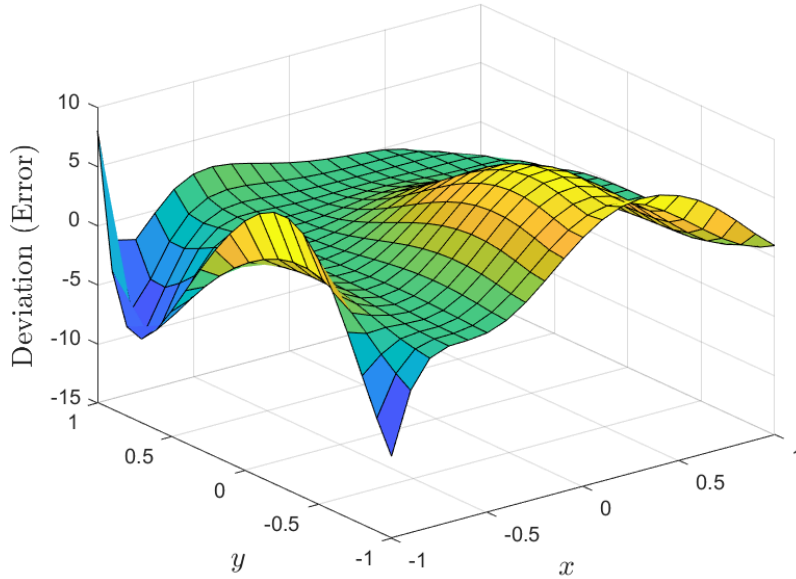FIGURE 4. Approximation $g_1(A,x,y) = \left(\frac{a_1+a_2x+a_3y+a_4x^2+a_5y^2}{1+a_6xy}\right)^3$.

FIGURE 5. Deviation (error) function $f(x,y) - g(A,x,y) = (-x+y^3+x^4)^4 - \left(\frac{a_1+a_2x+a_3y+a_4x^2+a_5y^2}{1+a_6xy}\right)^3$.

5. CONCLUSIONS

The goal of this study is to demonstrate that abstract convexity (in the sense of "convexity without linearity") has several applications in different branches of mathematics, including approximation theory. This application appears naturally in the new settings.

The results of the numerical experiments demonstrate that our approach is computationally efficient. The applications lead to practical applications as well. One potential application is data science and deep learning.

We also touched on the connections between "abstract convexity" and "axiomatic convexity". These areas have many overlappings that can be seen as a new way of looking at the problems, in particular, function approximation.

Acknowledgements

REFERENCES

[1] N.I. Achieser, Theory of Approximation, Frederick Ungar, New York, 1956.

[2] M. Andramonov, A survey of methods of abstract convex programming, J. Stat. Manag. Sys. 5 (2002) 21–37.

[3] H.H. Bauschke, A.S. Lewis, Dykstras algorithm with bregman projections: A convergence proof, Optimization 48 (2000) 409–427.

[4] B. Boehm, Functions whose best rational chebyshev approximation are polynomials, Numer. Math. 6 (1964) 235-242.

[5] S. Boyd, L. Vandenberghe. Convex Optimization. Cambridge University Press, New York, USA, seventh edition, 2009.

[6] H.T. Bui, R.S. Burachik, A.Y. Kruger, D.T. Yost, Zero duality gap conditions via abstract convexity, Optimization 71 (2022) 811-847.

[7] P.L. Chebyshev, The theory of mechanisms known as parallelograms, In Selected Works, pages 611–648. Publishing Hours of the USSR Academy of Sciences, Moscow, 1955. In Russian.

[8] E.W. Cheney, H.L. Loeb, Generalized rational approximation, J. Soc. Indust. Appl. Math. 1 (1964) 11–25.

[9] J.P. Crouzeix, Conditions for convexity of quasiconvex functions, Math. Oper. Res. 5 (1980) 120–125.

[10] A. Daniilidis, N. Hadjisavvas, J.-E. Martinez-Legaz, An appropriate subdifferential for quasiconvex functions, SIAM J. Optim. 12 (2002) 407–420.

[11] B. de Finetti, Sulle stratificazioni convesse, Ann. Mat. Pura Appl. 30 (1949) 173–183.

[12] C.J. de la Vallee Poussin, Sur la methode de l'approximation minimum, Ann. Soc. Sci. Bruxelles, 35 (1911) 1–16.

[13] P. Duchet, Convexity in combinatorial structures, In Proceedings of the 14th Winter School on Abstract Analysis, pp. 261–293, Circolo Matematico di Palermo, 1987.

[14] J. Dutta, A.M. Rubinov. Abstract convexity. Handbook of generalized convexity and generalized monotonicity, 76 (2005) 293–33.

[15] K. Fan, On the Krein-Milman theorem. In Convexity: Proceedings of the Seventh Symposium in Pure Mathematics of the American Mathematical Society, pp. 211–219. American Mathematical Society, 1963.

[16] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

[17] S.S. Kutateladze, A M Rubinov, Minkowski duality and its applications, Russ. Math. Surv. 27 (1972) 137–191.

[18] F.W. Levi, On helly's theorem and the axioms of convexity, J. Indian Math. Soc. 15 (1951) 65–76.

[19] H.L. Loeb, Algorithms for chebyshev approximations using the ratio of linear forms, J. Soc. Indust. Appl. Math. 8 (1960) 458–465.

[20] S.Y. Matsushita, L. Xu, On the finite termination of the douglas-rachford method for the convex feasibility problem, Optimization 65 (2016) 2037–2047.

[21] G. Meinardus, Approximation of Functions: Theory and Numerical Methods, Springer, Berlin, 1967.

[22] R.D. Millan, V. Peiris, N. Sukhorukova, J. Ugon, Multivariate approximation by polynomial and generalised rational functions, Optimization 71 (2022) 1171-1187.

[23] R.D. Millan, N. Sukhorukova, J. Ugon, An algorithm for best generalised rational approximation of continuous functions, Set-Valued Var. Anal. 30 (2022), 923–941.

[24] Y. Nakatsukasa, L.N. Trefethen, An algorithm for real and complex rational minimax approximation, SIAM J. Sci. Comput. 42 (2020) A3157–A3179.

[25] Y. Nakatsukasa, O. Sete, L.N. Trefethen, The aaa algorithm for rational approximation, SIAM J. Sci. Comput. 40 (2018) A1494–A1522.

[26] D. Pallaschke, S. Rolewicz, Foundations of Mathematical Optimization: Convex Analysis without Linearity, volume 388, Springer, 1997.

[27] V. Peiris, N. Sharon, N. Sukhorukova, J. Ugon, Generalised rational approximation and its application to improve deep learning classifiers, Appl. Math. Comput. 389 (2021) 125560.

[28] A. Ralston, Rational chebyshev approximation by remes' algorithms, Numer. Math. 7 (1965) 322–330.

[29] T.J. Rivlin, Polynomials of best uniform approximation to certain rational functions, Numer. Math. 4 (1962) 345–349.

[30] A.M. Rubinov, Abstract Convexity and Global Optimization, volume 44, Springer, 2013.

[31] A. M. Rubinov, B. Simsek, Conjugate quasiconvex nonnegative functions, Optimization 35 (1995) 1–22.

[32] I. Singer, Abstract Convex Analysis / Wiley-Interscience and Canadian Mathematics Series of Monographs and Texts, Wiley, New York, 1997.

[33] V.V. Soltan, Introduction to the Axiomatic Theory of Convexity, Sthiinka, Kishinev, 1984. In Russian.

[34] M.L.J. van de Vel, Theory of Convex Structures, North-Holland, Amsterdam, New York, 1993.

[35] Y. Yang, Q. Yang, Some modified relaxed alternating projection methods for solving the two-sets convex feasibility problem, Optimization 62 (2013) 509–525.

[36] A.J. Zaslavski, Subgradient projection algorithms and approximate solutions of convex feasibility problems, J. Optim. Theory Appl. 157 (2013) 803–819.

[37] X. Zhao, M.A. Kobis, On the convergence of general projection methods for solving convex feasibility problems with applications to the inverse problem of image recovery, Optimization 67 (2018) 1409–1427.