On the approximation of strongly convex functions by an upper or lower operator

Osama Yousif Mohammed Alabdali\textsuperscript{a}, Allal Guessab\textsuperscript{a,*}

\textsuperscript{a}Laboratoire de Mathématiques et de leurs Applications, UMR CNRS 4152, Université de Pau et des Pays de l’Adour, 64000 Pau, France, Tel.: +33 559674418

Abstract

The aim of this paper is to find a convenient and practical method to approximate a given real-valued function of multiple variables by linear operators, which approximate all strongly convex functions from above (or from below). Our main contribution is to use this additional knowledge to derive sharp error estimates for continuously differentiable functions with Lipschitz continuous gradients. More precisely, we show that the error estimates based on such operators are always controlled by the Lipschitz constants of the gradients, the convexity parameter of the strong convexity and the error associated with using the quadratic function, see Theorems 3.1 and 3.3. Moreover, assuming the function, we want to approximate, is also strongly convex, we establish sharp upper as well as lower refined bounds for the error estimates, see Corollaries 3.2 and 3.4. As an application, we define and study a class of linear operators on an arbitrary polytope, which approximate strongly convex functions from above. Finally, we present a numerical example illustrating the proposed method.

Keywords: Approximation, Convexity, Error estimates, Lipschitz gradients, Strongly convex functions.

1. Some background and motivation

Let $\Omega \subset \mathbb{R}^d$ be a nonempty compact convex set and let $\phi : \Omega \to \mathbb{R}$ be a given function. We would like to find an easier and good approximation
to compute $\phi$. We sometimes know beforehand that the function $\phi$ satisfies various known structural and regularity properties. For example, it may be known that $\phi$ has some additional kind of convexity, therefore we would wish to use this information in order to get a good approximation of $\phi$. Approximating an arbitrary function is, in general, very difficult, but if we restrict our attention to the class of strongly convex functions and if the linear operator, we wish to use, approximates all strongly convex functions from above (or from below) then things become simpler. The strongly convex functions are used widely in economic theory (see [17]), and are also central to optimization theory (see [13]). Indeed, in the framework of function minimization, this convexity notion has important and well-known implications. As we will see, the key advantage of dealing with such an operator is that an estimate of its approximation error is always controlled by the error associated with using the quadratic function.

In order to illustrate this idea more precise, we start by describing briefly a specific one-dimensional example, since its simplicity allows us to analyze all the necessary steps through very simple computation. Suppose that $\sigma$ is a fixed nonnegative real number. In the univariate approximation, say on an interval $[a, b]$, a simple way of approximating a given real $\sigma$-strongly function $\phi : [a, b] \to \mathbb{R}$ is first to choose a partition $P := \{x_0, x_1, \ldots, x_n\}$ of the interval $[a, b]$, such that $a = x_0 < x_1 < \ldots < x_n = b$, and then to fit to $\phi$ using a linear interpolant $B_n$ at these points in such a way that:

1. The domain of $B_n$ is the interval $[a, b]$;
2. $B_n$ is a linear polynomial on each subinterval $[x_i, x_{i+1}]$;
3. $B_n$ is continuous on $[a, b]$ and it interpolates the data, that is, $B_n(x_i) = \phi(x_i)$, $i = 0, \ldots, n$.

This is a convenient class of interpolants because every such interpolant can often be written for all $i = 0, \ldots, n - 1$ in a barycentric form:

$$B_n[\phi](x) = \frac{x_{i+1} - x}{x_{i+1} - x_i} \phi(x_i) + \frac{x - x_i}{x_{i+1} - x_i} \phi(x_{i+1}), \quad (x \in [x_i, x_{i+1}]). \tag{1}$$

One of the main features of the usual linear interpolant, in its simplest form (1), is that the error in approximating the quadratic function $(.)^2$ by $B_n$ is simply given by:

$$B_n[(.)^2](x) - x^2 = (x - x_i)(x_{i+1} - x), \quad (x \in [x_i, x_{i+1}]), \quad (x \in [x_i, x_{i+1}])$$
and also that \( B_n \) approximates all \( \sigma \)-strongly convex functions from above. More precisely, \( B_n \) satisfies for any \( \sigma \)-strongly convex function the following estimates:

\[
\frac{\sigma}{2} (x - x_i)(x_{i+1} - x) \leq B_n[f](x) - f(x), \quad (x \in [x_i, x_{i+1}]).
\]

Moreover, as it can be derived from our multivariate general results, see Remark 4.5, if we also know that the first derivative of \( f \) is a Lipschitz function with a local Lipschitz constant \( L_i(f') \) in the subintervals \([x_i, x_{i+1}]\), then the error \( B_n[f] - f \) can often be estimated at any \( x \in [x_i, x_{i+1}] \) as:

\[
\frac{\sigma}{2} (x - x_i)(x_{i+1} - x) \leq B_n[f](x) - f(x) \leq \frac{L_i(f')}{2} (x - x_i)(x_{i+1} - x). \tag{2}
\]

Hence, the lower and upper bounds of the approximation error for this class of functions can be controlled by the Lipschitz constants of the first derivatives, the convexity parameter (of the strong convexity) and the error associated with using the quadratic function. It should be noted that equalities in (2) are attained for all \( \sigma \)-strongly convex functions of the form

\[
f(x) = a(x) + \frac{\sigma}{2} x^2, \tag{3}
\]

where \( a(\cdot) \) is any affine function. Therefore, in this sense, the error estimates (2) are sharp for the class of \( \sigma \)-strongly convex functions having Lipschitz continuous first derivatives. This provides the starting point of the forthcoming results.

This paper deals with the problem of approximation of functions of multiple variables by using linear operators, which approximate from above (or from below) all strongly convex functions with Lipschitz-continuous gradients. Geometrically, if a function \( f \) belongs to such a class, then its gradient \( \nabla f \) cannot change too quickly and it cannot change too slowly either. Functions satisfying these conditions are widely used in the optimization literature, we refer to Nesterov’s book [13]. A natural question is: can these functions be well approximated by simpler functions and how?

There are several papers investigating various methods to approximate arbitrary functions, very little research has been done subject to some kind of additional convexity assumption. For instance, if some smoothness is allowed for the function, which is to be approximated, say \( C^2(\Omega) \), this will play a crucial role in the determination of the “best” (or “optimal”) cubature formulas,
see [1–10]. This article builds on the previous work [3, 4], where a theoretical framework for approximating $C^2(\Omega)$—convex functions was developed.

The motivation for such an approach is that the general sharp error estimates, that we derive, permit us to study a multivariate version defined on an arbitrary (convex) polytope of the univariate interpolation operators given by (1). Throughout the paper, a linear operator is said to be upper (resp. lower) operator for strongly convex functions, if it approximates from above (resp. from below) strongly convex functions.

The paper is organized as follows: In Section 2 we state some definitions and some of the basic properties and facts about strongly convex functions. The main theorems of Section 3 establish, in terms of sharp error estimates, simple and elegant characterizations of upper or lower approximation operators for strongly convex functions with Lipschitz-continuous gradients. In this way, we offer sharp error estimates which only depend on the Lipschitz constants of the gradients, the convexity parameter (of the strong convexity), and the error associated with using the quadratic function, see Theorems 3.1 and 3.3. This also allows us to recover and extend the simple approach of [3], which presented in the case where the functions are only convex. A particularly interesting situation arises, when the function, we want to approximate, is also strongly convex. In this special case, we establish sharp upper as well as lower refined bounds for the error estimates, see Corollaries 3.2 and 3.4. In Section 4, we will introduce and study a multivariate version defined on an arbitrary polytope of the univariate interpolation operators given by (1). Finally, Section 5 will provide a numerical example to illustrate the efficiency of this approach.

2. Some inequalities involving strongly convex functions

In this section, we state some definitions and properties of strongly convex functions, which are very useful in the proofs of our characterizations of upper (resp. lower) operators. Before proceeding, we shall now recall some definitions and results which will be needed in the sequel. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we will denote the standard inner product of $\mathbf{x}$ and $\mathbf{y}$ by $\langle \mathbf{x}, \mathbf{y} \rangle$ and the Euclidean vector norm of $\mathbf{x} \in \mathbb{R}^d$ by $\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Recall that a function $f : \Omega \to \mathbb{R}$ is called convex if for all $\mathbf{x}, \mathbf{y} \in \Omega, \lambda \in [0, 1]$

$$f((1 - \lambda)\mathbf{x} + \lambda \mathbf{y}) \leq (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{y})$$
holds. Finally, further standard definitions or notations in convex analysis which are not explicitly mentioned here can be found in the classical book [16].

We now present the notion of strong convexity, which generalizes the classical definition of convexity.

**Definition 2.1.** A convex function \( f \) is \( \sigma \)-strongly convex if, and only if, there exists a constant \( \sigma \geq 0 \) such that the function \( f - \frac{\sigma}{2} \| \cdot \| \) is convex.

Obviously, the case \( \sigma = 0 \) merely implies usual convexity of \( f \). The constant \( \sigma \) is called the modulus of strong convexity of \( f \). We shall denote by \( S_\sigma(\Omega) \) the set of all strongly convex functions with parameter \( \sigma \). Observe that \( S_0(\Omega) \) is just the set of convex functions.

For later purposes we can observe the following simple facts:

**Remark 2.2.** For any positive real number \( \sigma \) the following functions are \( \sigma \)-strongly convex functions:

1. \( \frac{\sigma}{2} \| \cdot \|^2 \).

2. Addition of a convex function to a strongly convex function gives a strongly convex function with the same modulus of strong convexity. Therefore, adding a convex function to \( \frac{\sigma}{2} \| \cdot \|^2 \) does not affect \( \sigma \)-strong convexity.

As usually, we say that \( f \) is continuously differentiable on \( \Omega \) if it is continuously differentiable on an open set containing \( \Omega \).

**Definition 2.3.** A differentiable function \( f \) is said to have a Lipschitz continuous gradient, if there exists a constant \( \rho(\nabla f) \), such that

\[
\| \nabla f(x) - \nabla f(y) \| \leq \rho(\nabla f) \| x - y \|, \quad (x, y \in \Omega).
\]

(4)

For any differentiable \( f \) with Lipschitz continuous gradient, there exists a smallest possible \( L(\nabla f) \) such that (4) holds. The smallest constant \( L(\nabla f) := \text{Lip}(\nabla f) \) satisfying the inequality (4) is called the Lipschitz constant for \( \nabla f \). While the Lipschitz constant provides an upper bound for the ‘curvature’ of the function, the convexity parameter determines a lower bound. By \( C^{1,1}(\Omega) \) we will denote the subclass of all functions \( f \) which are
continuously differentiable on $\Omega$ with Lipschitz continuous gradients.

If $f$ is differentiable, then there is an equivalent, and perhaps easier, definition of strong convexity of $f$.

**Lemma 2.4.** Let $f$ be a continuously differentiable function on $\Omega$. Then, the following conditions are equivalent:

1. $f$ is $\sigma$-strongly convex.
2. $\sigma \|x - y\|^2 \leq \langle \nabla f(x) - \nabla f(y), y - x \rangle, \forall x, y \in \Omega.$

**Proof.** The proof is straightforward and done in detail in Nesterov’s book [13, Pg. 64, Theorem 2.1.9].

The Lipschitz continuity of $\nabla f$ will play a crucial role in our analysis. Hence, we state here a very important result derived from this property (see [15, Pg. 24, Lemma 2]):

**Lemma 2.5.** Let $f : \Omega \to \mathbb{R}$ be a continuously differentiable convex function with Lipschitz continuous gradient with constant $\rho(\nabla f)$. Then $\nabla f$ satisfies the following property:

$$\|\nabla f(y) - \nabla f(x)\|^2 \leq \rho(\nabla f)\langle \nabla f(y) - \nabla f(x), y - x \rangle, \forall x, y \in \Omega. \quad (5)$$

This property is known as co-coercivity of $\nabla f$. In what follows $S^{1,1}_\sigma(\Omega)$ will denote the set of $\sigma$-strongly convex continuously differentiable functions with Lipschitz-continuous gradients. Note that for any $f \in S^{1,1}_\sigma(\Omega)$ we always have $\sigma \leq L(\nabla f)$. It is quite easy to see that for a convex quadratic $f(x) = \frac{1}{2}x^T H x$, the Lipschitz constant of the gradient is given by the maximal eigenvalue of the Hessian $H$ while the parameter of strong convexity is given by its minimal eigenvalue. Hence, for any nonnegative $\sigma$, the function $f(x) = \frac{\sigma}{2} \|x\|^2$ defines a $\sigma$-strongly convex function with a Lipschitz gradient constant $L(\nabla f)$ equal to $\sigma$.

The following Theorem will be essential to our general characterization results.

**Theorem 2.6.** Let $\sigma$ be a nonnegative real number and let $f : \Omega \to \mathbb{R}$ be a continuously differentiable function with Lipschitz continuous gradient and Lipschitz constant $L(\nabla f)$. Then the two functions defined by

$$\tilde{f}_\pm := \frac{\sigma + L(\nabla f)}{2} \|z\|^2 \pm f$$
belong to $S^1_2(\Omega)$. If in addition $f$ is $\sigma$-strongly convex, then $L(\nabla \tilde{f}_-) \leq L(\nabla f)$.

Proof. It is easily seen that $\tilde{f}_\pm$ belong to $C^{1,1}(\Omega)$. Indeed, they are obviously differentiable and a simple calculation shows that

$$\|\nabla \tilde{f}_\pm(y) - \nabla \tilde{f}_\pm(x)\| = \| (\sigma + L(\nabla f))(y - x) \pm (\nabla f(y) - \nabla f(x))\| \quad (6)$$

which implies, using the triangle inequality,

$$\|\nabla \tilde{f}_\pm(y) - \nabla \tilde{f}_\pm(x)\| \leq (\sigma + L(\nabla f))\|y - x\| + \|\nabla f(y) - \nabla f(x)\| \leq (\sigma + 2L(\nabla f))\|y - x\|.$$ 

Hence, $\tilde{f}_\pm \in C^{1,1}(\Omega)$ and obviously we have $L(\nabla \tilde{f}_\pm) \leq \sigma + 2L(\nabla f)$.

We now show that $\tilde{f}_\pm$ are both $\sigma$-strongly convex. First observe that functions $\tilde{f}_\pm$ may be written as follows:

$$\tilde{f}_\pm = \frac{\sigma}{2} \| . \|^2 + \tilde{g}_\pm,$$

where

$$\tilde{g}_\pm := \frac{L(\nabla f)}{2} \| . \|^2 \pm f.$$

Clearly from Remark 2.2 it is enough to check that $\tilde{g}_\pm$ are convex. To this end, it is immediate to verify that

$$\langle \nabla \tilde{g}_\pm(y) - \nabla \tilde{g}_\pm(x), y - x \rangle = L(\nabla f) \|y - x\|^2 \pm \langle \nabla f(y) - \nabla f(x), y - x \rangle. \quad (7)$$

Taking in account that $f$ has a Lipschitz continuous gradient with constant $L(\nabla f)$, then by the Cauchy-Schwartz inequality, we come to

$$\pm \langle \nabla f(y) - \nabla f(x), y - x \rangle \leq L(\nabla f) \|y - x\|^2, \quad (8)$$

and so

$$L(\nabla f) \|y - x\|^2 \pm \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0. \quad (9)$$

From this and (7), it immediately follows

$$\langle \nabla \tilde{g}_\pm(y) - \nabla \tilde{g}_\pm(x), y - x \rangle \geq 0, \quad (x, y \in \Omega),$$

which means that $\tilde{g}_\pm$ are convex, and concludes that $\tilde{f}_\pm$ are $\sigma$-strongly convex.
Next, we show that $L(\nabla \tilde{f}_-) \leq L(f)$. Since function $\tilde{f}_-$ has a Lipschitz continuous gradient, then from Lemma 2.5, applied to $\tilde{f}_-$, it follows

$$\|\nabla \tilde{f}_-(y) - \nabla \tilde{f}_-(x)\|^2 \leq L(\nabla \tilde{f}_-)(\nabla \tilde{f}_-(y) - \nabla \tilde{f}_-(x), y - x).$$

(11)

It is straightforward to check that for any $x, y \in \Omega$, $\nabla \tilde{f}_-(y) - \nabla \tilde{f}_-(x)$ is given by

$$\nabla \tilde{f}_-(y) - \nabla \tilde{f}_-(x) = (\sigma + L(f))(y - x) - (\nabla f(y) - \nabla f(x)).$$

(12)

On the other hand, by Lemma 2.4, the assumption of $\sigma$-strong convexity of $f$ implies that

$$-\langle \nabla f(y) - \nabla f(x), y - x \rangle \leq -\sigma \|y - x\|^2.$$  

(13)

Now, an obvious combination of the preceding equations (11), (12) and (13) leads to

$$\|\nabla \tilde{f}_-(y) - \nabla \tilde{f}_-(x)\|^2 \leq L(\nabla \tilde{f}_-)(\nabla f)\|y - x\|^2.$$  

(14)

This allows us to conclude that $L(\nabla \tilde{f}_-) \leq \sqrt{L(\nabla \tilde{f}_-)(\nabla f)}$, since $L(\nabla \tilde{f}_-)$ is the smallest possible Lipschitz constant for $\nabla f$. Hence, $L(\nabla \tilde{f}_-)$ satisfies $L(\nabla \tilde{f}_-) \leq L(\nabla f)$. Thus $\tilde{f}_-$ has a Lipschitz continuous gradient, and moreover, the Lipschitz constant $L(\nabla \tilde{f}_-) \leq L(\nabla f)$. This completes the proof of Theorem 2.6.

3. Characterizations of upper or lower approximation operators

In this section, the first main results, Theorems 3.1 and 3.3 are on simple characterizations, in terms of sharp error estimates, of approximation operators, which approximate from below or above strongly convex functions with Lipschitz continuous gradients. It is shown that the error estimates using these operators can often be controlled by the Lipschitz constants of the gradients, the convexity parameter (of the strong convexity) and the error associated with using the quadratic function. The second ones, which are their Corollaries 3.2 and 3.4, are on the establishment of sharp upper as well as lower refined bounds for the error estimates, assuming that the function, which is to be approximated, is also strongly convex. Here, we continue to assume that the strong convexity parameter $\sigma$ is given (possibly null). Our characterization of linear operators, which approximate all $\sigma$-strongly convex functions from above, is as follows:
Theorem 3.1. Let $\sigma$ be a nonnegative real number and let $A : C^1(\Omega) \to C(\Omega)$ be a linear operator. The following statements are equivalent:

(i) For every $\sigma$-strongly convex function $g \in C^{1,1}(\Omega)$, we have
$$g(x) \leq A[g](x), \quad (x \in \Omega).$$

(ii) For every function $f \in C^{1,1}(\Omega)$, we have
$$|A[f](x) - f(x)| \leq \frac{\sigma + L(\nabla f)}{2} (A[\|\cdot\|^2](x) - \|x\|^2), \quad (x \in \Omega).$$

Proof. Let us fix $f \in C^{1,1}(\Omega)$ and suppose that (i) holds. Introduce the functions $\tilde{f}_+^\pm$ and $\tilde{f}_-^\pm$ as:
$$\tilde{f}_+^\pm := \frac{\sigma + L(\nabla f)}{2} \|\cdot\|^2 \pm f.$$

According to Theorem 2.6, we know that $\tilde{f}_+^\pm$ are in $C^{1,1}(\Omega)$ and are both $\sigma$-strongly convex. Then, since $A$ is linear, statement (i) implies that
$$\frac{\sigma + L(\nabla f)}{2} \|\cdot\|^2 \pm f \leq \frac{\sigma + L(\nabla f)}{2} A[\|\cdot\|^2] \pm A[f],$$
or equivalently
$$\mp(A[f] - f) \leq \frac{\sigma + L(\nabla f)}{2} \left( A[\|\cdot\|^2] - \|\cdot\|^2 \right).$$

This obviously implies the error estimate given in (16). The case of equality is easily verified.

Conversely, let $g \in C^{1,1}(\Omega)$ be a $\sigma$-strongly function, and suppose that statement (ii) holds. Let the function $\tilde{f}_-^-$ be defined by
$$\tilde{f}_-^- := \frac{\sigma + L(\nabla g)}{2} \|\cdot\|^2 - g.$$

Again Theorem 2.6 tells us that $\tilde{f}_-^-$ is convex, it also belongs to $C^{1,1}(\Omega)$ and satisfies $L(\nabla \tilde{f}_-) \leq L(\nabla g)$. Now, the error estimate given in statement (ii), applied to $\tilde{f}_-^-$, implies that
$$A[\tilde{f}_-^-] - \tilde{f}_-^- \leq \frac{\sigma + L(\nabla \tilde{f}_-) - \|\cdot\|^2}{2} (A[\|\cdot\|^2] - \|\cdot\|^2) \leq \frac{\sigma + L(\nabla g)}{2} \left( A[\|\cdot\|^2] - \|\cdot\|^2 \right).$$
or equivalently
\[ g \leq A[g], \]
and statement \((i)\) is proved. This establishes Theorem 3.1.

\[ \square \]

Theorem 3.1 extends a result given in [3, Theorem 2.3] for convex functions to the case of strongly convex functions.

We already know how one can estimate the approximation error \(A[f] - f\) for a function possessing Lipschitz continuous gradient; what happens if we know in advance that the function is, moreover, strongly convex?

The answer is given by the following Corollary, which is a direct consequence of Theorem 3.1.

**Corollary 3.2.** Let \(\sigma\) be a nonnegative real number and let \(A : C^1(\Omega) \to C(\Omega)\) be a linear operator. Assume that for every convex function \(g \in C^{1,1}(\Omega)\), we have
\[ A[g](x) \leq g(x), \quad (x \in \Omega). \]
Then the following error estimates hold for every \(\sigma\)-strongly convex function \(f \in S^1_\sigma(\Omega)\):
\[ \frac{\sigma}{2} E_+[\|\cdot\|^2](x) \leq A[f](x) - f(x) \leq \frac{L(\nabla f)}{2} E_+[\|\cdot\|^2](x), \quad (x \in \Omega), \]
where \(E_+[\|\cdot\|^2] := A[\|\cdot\|^2] - \|\cdot\|^2\). Equalities in (19) are attained for all functions of the form
\[ f(x) = a(x) + \frac{\sigma}{2} \|x\|^2, \]
where \(a(\cdot)\) is any affine function.

**Proof.** The error upper bound is a direct consequence of Theorem 3.1. So it remains to check that the error lower bound holds, too. Assume that the statement \((i)\) is true for every convex function and let us fix a \(\sigma\)-strongly convex function \(f\). Then, since \(g = f - \frac{\sigma}{2} \|\cdot\|^2\) is convex, statement \((i)\) and the linearity of \(A\) imply
\[ f - \frac{\sigma}{2} \|\cdot\|^2 \leq A[f] - \frac{\sigma}{2} A[\|\cdot\|^2], \]
or equivalently
\[
\frac{\sigma}{2} (A[\| \cdot \|^2] - \| \cdot \|^2) \leq A[f] - f.
\]
This shows that \( \frac{\sigma}{2} E_+[\| \cdot \|^2] \) estimates \( A[f] - f \) from below and completes the proof of the lower bound. Finally, since \( A \) reproduces linear and constant functions, the case of equality can be confirmed by a little algebra. \( \square \)

According to the error estimates (16) and (19), Corollary 3.2 provides a better error lower bound than Theorem 3.1 for strongly convex functions with Lipschitz continuous gradients. A slight modification of Theorem 3.1 given below addresses the case in which the linear operator \( A \), we wish to use, approximates all convex functions from below. Indeed, in this setting our characterization of those operators can be stated as follows:

**Theorem 3.3.** Let \( \sigma \) be a nonnegative real number and let \( A : C^1(\Omega) \to C(\Omega) \) be a linear operator. The following statements are equivalent:

(i) For every \( \sigma \)-strongly convex function \( g \in C^{1,1}(\Omega) \), we have
\[
A[g](x) \leq g(x), \ (x \in \Omega).
\]
(ii) For every function \( f \in C^{1,1}(\Omega) \), we have
\[
|A[f](x) - f(x)| \leq \frac{\sigma + L(\nabla f)}{2} \left( \|x\|^2 - A[\| \cdot \|^2](x) \right), \ (x \in \Omega).
\]

Note that in the error estimates (ii), established in Theorems 3.1 and 3.3, are valid for all functions in \( C^{1,1}(\Omega) \), as long as statements (i) hold for the class of \( \sigma \)-strongly convex functions. We remark here that similar arguments to those used in Corollary 3.2 will derive the following refined error estimates:

**Corollary 3.4.** Let \( \sigma \) be a nonnegative real number and let \( A : C^1(\Omega) \to C(\Omega) \) be a linear operator. Assume that for every convex function \( g \in C^{1,1}(\Omega) \), we have
\[
A[g](x) \leq g(x), \ (x \in \Omega).
\]
Then the following error estimates hold for every \( \sigma \)-strongly convex function \( f \in S^{1,1}_\sigma(\Omega) \):
\[
\frac{\sigma}{2} E_-[\| \cdot \|^2](x) \leq A[f](x) - f(x) \leq \frac{L(\nabla f)}{2} E_-[\| \cdot \|^2](x), \ (x \in \Omega),
\]

11
where \( E_{-[\|\cdot\|^{2}]} := \|\cdot\|^{2} - A[\|\cdot\|^{2}] \). Equalities in (24) are attained for all functions of the form
\[
f(\mathbf{x}) = a(\mathbf{x}) + \frac{\sigma}{2}\|\mathbf{x}\|^{2},
\]
where \( a(\cdot) \) is any affine function.

4. Applications to the barycentric approximation schemes

In this section, we are going to consider a multivariate version different from the tensor product construction that in the univariate case, \( d = 1 \), yields the operator defined by (1). Indeed, the simple univariate operator (1) can be extended to arbitrary higher-dimensional polytopes. To this end, let \( X_{n} = \{\mathbf{x}_{i}\}_{i=0}^{n} \) be a given finite set of pairwise distinct points in \( \Omega \subset \mathbb{R}^{d} \), with \( \Omega = \text{conv}(X_{n}) \) denoting the convex hull of the point set \( X_{n} \). We are interested in approximating an unknown scalar-valued continuous \( \sigma \)-strongly convex function \( f : \Omega \rightarrow \mathbb{R} \) from given function values \( f(\mathbf{x}_{0}), \ldots, f(\mathbf{x}_{n}) \) sampled at \( X_{n} \). In order to obtain a simple and stable global approximation of \( f \) on \( \Omega \), we may consider a weighted average of the function values at data points of the following form:
\[
B_{n}[f](\mathbf{x}) = \sum_{i=0}^{n} \lambda_{i}(\mathbf{x})f(\mathbf{x}_{i}),
\]

or, equivalently, a convex combination of the data values \( f(\mathbf{x}_{0}), \ldots, f(\mathbf{x}_{n}) \). This means that we require the system of functions \( \lambda := \{\lambda_{i}\}_{i=0}^{n} \) to form a partition of unity, that is, for all \( \mathbf{x} \in \Omega \), we have
\[
\lambda_{i}(\mathbf{x}) \geq 0, \quad i = 0, \ldots, n, \quad \sum_{i=0}^{n} \lambda_{i}(\mathbf{x}) = 1.
\]
In addition, we shall also impose the set of functions \( \lambda \) to satisfy the first-order consistency condition:
\[
\mathbf{x} = \sum_{i=0}^{n} \lambda_{i}(\mathbf{x})\mathbf{x}_{i}, \quad (\mathbf{x} \in \Omega).
\]

We will call any set of functions \( \lambda_{i} : \Omega \rightarrow \mathbb{R}, \ i = 0, \ldots, n, \) barycentric coordinates if they satisfy the three properties (27), (28) and (29) for all \( \mathbf{x} \in \Omega \).
In view of these properties, we shall refer to the approximation schemes \( B_n \) as barycentric approximation (schemes). Recall that these coordinates exist for more general types of polytopes. The first result on their existence was due to Kalman [14, Theorem 2]. Let us go back now to the simple case of a univariate function \( f \) for the computation of a barycentric approximation function created in this manner. To do this, we consider a subinterval \([x_i, x_{i+1}]\), then it is easily seen that the barycentric coordinates of a point \( x \) of \([x_i, x_{i+1}]\) with respect to \( v_0 := x_i, v_1 := x_{i+1} \) are given respectively as follows:

\[
\begin{align*}
\lambda_{i,0}(x) &= \frac{x_{i+1} - x}{x_{i+1} - x_i}, \\
\lambda_{i,1}(x) &= \frac{x - x_i}{x_{i+1} - x_i}.
\end{align*}
\]

This shows that in one dimension the barycentric approximation function (26) is nicely reduced to the simple form given in (1). Hence, our proposed method of approximation scheme (26) can be viewed as a multivariate generalization of the approach in the univariate case.

For a \( \sigma \)-strongly convex function \( f \in S^1_\sigma(\Omega) \), the symbol

\[
E_n[f](x) := E_n[f, \lambda](x) = \sum_{i=0}^{n} \lambda_i(x)f(x_i) - f(x), \ (x \in \Omega), \quad (30)
\]

will be reserved exclusively to denote the incurred approximation error between \( f \) and its barycentric approximation \( B_n[f] \).

We begin our analysis by giving general identities, which show simple expressions of the error \( E_n[\|\|^{2}] \) in terms of barycentric coordinates.

**Lemma 4.1.** The error \( E_n[\|\|^{2}] \) when approximating the quadratic function \( \|\|^{2} \) by the barycentric approximation operator \( B_n[\|\|^{2}] \) can be expressed in terms of the barycentric coordinates as:

\[
E_n[\|\|^{2}](\mathbf{x}) = \frac{1}{2} \sum_{i=0}^{n} \sum_{j=0}^{n} \lambda_i(\mathbf{x}) \lambda_j(\mathbf{x}) \|x_i - x_j\|^2 \quad (31)
\]

\[
= \sum_{i=0}^{n} \lambda_i(\mathbf{x}) \|x - x_i\|^2. \quad (32)
\]
Proof. In order to show (31), we use the affine precision property of the barycentric coordinates. Indeed, from (28) and (29) we immediately deduce
\[
\sum_{i=0}^{n} \sum_{j=0}^{n} \lambda_i(x) \lambda_j(x) \|x_i - x_j\|^2
\]
\[
= \sum_{i=0}^{n} \sum_{j=0}^{n} \lambda_i(x) \lambda_j(x) \left( \|x_i\|^2 - 2 \langle x_i, x_j \rangle + \|x_j\|^2 \right)
\]
\[
= \sum_{i=0}^{n} \lambda_i(x) \|x_i\|^2 - 2 \left( \sum_{i=0}^{n} \lambda_i(x) x_i \right) \left( \sum_{j=0}^{n} \lambda_j(x) x_j \right) + \sum_{j=0}^{n} \lambda_j(x) \|x_j\|^2
\]
\[
= 2 \left( \sum_{i=0}^{n} \lambda_i(x) \|x_i\|^2 - \|x\|^2 \right) = 2 E_n[[\|\cdot\|^2]](x).
\]
Moreover, it is easily verified that
\[
\|x_i - x_j\|^2 = \|x - x_i\|^2 + 2 \langle x_i - x, x - x_j \rangle + \|x - x_j\|^2.
\]
Applying this, yields (32) and completes the proof of the Lemma.

The following Lemma shows that the operator $B_n$ approximates every strongly convex function from above. Moreover, it now allows us to prove a sharp lower bound for the error of any strongly convex function.

Lemma 4.2. Let $\sigma$ be a nonnegative real number. Then, the barycentric approximation operator $B_n$ approximates every $\sigma$-strongly convex function from above. Moreover, for every $\sigma$-strongly convex function $f$, it holds
\[
\frac{\sigma}{2} \sum_{i=0}^{n} \lambda_i(x) \|x - x_i\|^2 \leq B_n[f](x) - f(x), (x \in \Omega).
\] (33)

Equality in (33) is attained for all functions of the form
\[
f(x) = a(x) + \frac{\sigma}{2} \|x\|^2,
\] (34)
where $a(\cdot)$ is any affine function.
Proof. Let us fix $f$ a $\sigma$-strongly convex function and define $h := f - \frac{\sigma}{2} \| \cdot \|^2$. Since, $h$ is convex then by the Jensen-convexity of $h$, we get

$$h(x) \leq \sum_{i=0}^{n} \lambda_i(x) h(x_i), (x \in \Omega).$$

or equivalently

$$f(x) - \frac{\sigma}{2} \| x \|^2 \leq \sum_{i=0}^{n} \lambda_i(x) \left( f(x_i) - \frac{\sigma}{2} \| x_i \|^2 \right), (x \in \Omega).$$

Thus, we get

$$\frac{\sigma}{2} \left( \sum_{i=0}^{n} \lambda_i \| x_i \|^2 - \| x \|^2 \right) \leq \sum_{i=0}^{n} \lambda_i(x) f(x_i) - f(x).$$

This inequality, combined with Lemma 4.1, implies that the required identity is satisfied. The case of equality is easily verified.

The following Lemma gives an upper bound for the absolute value of the error of any function possessing Lipschitz continuous gradient:

Lemma 4.3. The following error estimate holds for every function $f \in C^{1,1}(\Omega)$ :

$$|B_n[f](x) - f(x)| \leq \frac{L(\nabla f)}{2} \sum_{i=0}^{n} \lambda_i(x) \| x - x_i \|^2, \ (x \in \Omega). \quad (35)$$

Equality in (35) is attained for all functions of the form

$$f(x) = a(x) + \frac{\sigma}{2} \| x \|^2,$$

where $a(\cdot)$ is any affine function.

Proof. This Lemma is an immediate consequence of Corollary 3.4 and Lemma 4.1. The case of equality is easily verified.

Now everything is set for giving an upper bound and a lower bound for the error estimate $B_n[f] - f$ of any $\sigma$-strongly convex function $f$, having Lipschitz continuous gradient.
Theorem 4.4. Let $\sigma$ be a nonnegative real number. Then, for every $\sigma$-strongly convex function $f \in S_\sigma^{1,1}(\Omega)$ and any $x \in \Omega$, it holds:

$$\frac{\sigma}{2} \sum_{i=0}^{n} \lambda_i(x) \| x - x_i \|^2 \leq B_n[f](x) - f(x) \leq \frac{L(\nabla f)}{2} \sum_{i=0}^{n} \lambda_i(x) \| x - x_i \|^2.$$ \hfill (37)

Equality in (37) is attained for all functions of the form

$$f(x) = a(x) + \frac{\sigma}{2} \| x \|^2,$$ \hfill (38)

where $a(\cdot)$ is any affine function. This inequality is sharp in the sense that the equality is attained for all affine functions.

Proof. This is an immediate consequence of Lemmas 4.2, 4.3 and Corollary 3.4. The case of equality is easily verified. \qed

Remark 4.5. In the univariate case, a simple inspection of the error estimates (37) reveals that (37) is nicely reduced to the simple form given in (2).

5. Numerical experiments

One possible natural approach to construct an interesting class of particular barycentric approximations would be to simply construct a triangulation of the polytope $\Omega$ - the convex hull of the data set $X_n = \{x_i\}_{i=0}^{n}$ into simplices such that the vertices $v_i$ of the triangulation coincide with the data points $x_i$. After that, one can use the standard barycentric coordinates for these simplices. As a result, each triangulation of the data set $X_n$ generates a barycentric approximation. Hence, there exists at least one barycentric approximation of type (26) which is generated by a triangulation. A very natural triangulation $DT(\Omega)$ of $\Omega$ is the one which uses only the points of $X_n$ as triangulation vertices and such that no point in $X_n$ lies inside the circumscribing ball of any simplex of $DT(\Omega)$. Such a triangulation exists and is called a Delaunay triangulation of $\Omega$ with respect to $X_n$.

Let $T(\Omega)$ be any triangulation of the point set $X_n$. Then $\lambda^{T(\Omega)} := \{\lambda_i^{T(\Omega)}\}_{i=0}^{n}$ denotes the set of barycentric coordinates associated with each $x_i$ of $X_n$. We now list the basic properties of $\lambda^{T(\Omega)}$ of which the following are particularly relevant to us:
They are well-defined, piecewise linear and nonnegative real-valued continuous functions.

The function $\lambda^T_i(\Omega)$ has to equal 1 at $x_i$ and 0 at all other points in $X \setminus \{x_i\}$, that is, $\lambda^T_i(x_j) = \delta_{ij}$ ( $\delta$ is the Kronecker delta).

We denote by

$$E^T_n[\Omega](f)(x) := \sum_{i=0}^n \lambda^T_i(x) f(x_i) - f(x).$$

As regard the error estimates (37), it was shown that Delaunay triangulation is the triangulation that minimizes the approximation error $E^T_n[\Omega][\|\cdot\|^2]$ among all triangulations with the same number of vertices, see [3, Theorem 4.10]. This optimality condition also characterizes Delaunay triangulation.

Suppose a set of scattered data $\{(x_i, y_i, f_i)\}_{i=1}^N$, which are assumed to be sampled from a strongly convex function $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$. Taking the $N$ scattered points as nodes, a barycentric approximation is constructed in domain $\Omega$ using Delaunay triangulation. We now illustrate this approach by the following numerical example.

**5.1. An illustrative simple example**

This example is designed to follow the exact steps of the above methodology. We take the following strongly convex function:

$$f(x, y) := 100((x - 0.4)^2 + (y + 0.5)^2) + 400 \exp((x - 0.5)^2 + (y - 0.5)^2),$$

with the restriction of domain $D := [0, 1] \times [0, 1]$. The data is generated from the above function and it is based on 21 equally spaced nodes on each edge of the boundary of square $D$ and 216 nodes in the square $D$. The nodes in the domain are placed randomly selected from $D$ while the nodes on the boundary is equally spaced.

From Figures 1 it is clear that the strong convexity of $f$ has been preserved and there is no visual difference between the test function and its piecewise-linear interpolant.

**6. Conclusion and a final remark**

This paper presents a new and efficient way of approximating a given function of multiple variables by linear operators, which approximate all
Figure 1: The figure on the left shows the graph of $f$ produced by MAPLE, and using MATLAB the graph on the right is for the piecewise-linear interpolation of the data generated from $f$.

*strongly* convex functions from above (or from below). This additional information is used to characterize sharp error estimates for continuously differentiable functions with Lipschitz continuous gradients. All the proposed error estimates are controlled by the Lipschitz constants of the gradients, the convexity parameter of the strong convexity and the error associated with using the quadratic function. Moreover, if the function to be approximated is also strongly convex then we establish sharp upper as well as lower refined bounds for the error estimates.

For future research, it will be interesting to investigate this problem under various abstract convexity notions, for instance the general case of uniformly convex functions also remains an open question.
Acknowledgment

The authors would like to thank financial support from the Volubilis Hubert Curien Program (Grant No. MA/13/286), and Hassan I University, Settat, Morocco for hosting and supporting us during the research stay that led to this collaboration. We would also like to thank the referees for their careful reading of our work.

References


