ON CONVEX QUADRATIC APPROXIMATION

By Dick den Hertog, Etienne de Klerk and Kees Roos

April 2000

ISSN 0924-7815
On convex quadratic approximation

Dick den Hertog\(^*\), Etienne de Klerk\(^\dagger\) and Kees Roos\(^\ddagger\)

April 7, 2000

Abstract

In this paper we prove the counterintuitive result that the quadratic least squares approximation of a multivariate convex function in a finite set of points is not necessarily convex, even though it is convex for a univariate convex function. This result has many consequences both for the field of statistics and optimization. We show that convexity can be enforced in the multivariate case by using semidefinite programming techniques.

Key words: Convex function, least squares, quadratic interpolation, semidefinite programming

1 Introduction

Interpolation and approximation are widely used techniques in many research fields. See [4, 13, 14]. In this paper we investigate whether the quadratic interpolation and quadratic least squares approximation of a convex function in a finite number of points preserves the convexity property or not. We call this the convexity preserving property. We will prove that the quadratic least squares approximation is convexity preserving for the univariate case, but that even the quadratic interpolation function for the multivariate case is not convexity preserving.

These results are counterintuitive and to the best of our knowledge not described in the literature. Our conjecture is that the result for the multivariate case has not been discovered since least squares approximation is mostly used for the univariate case. We also could not find a proof in the literature for the convexity preserving property of quadratic least squares for the univariate case.

The consequences of these results are significant, both in the field of statistics and optimization. Several optimization methods use quadratic interpolation or quadratic least squares approximations to (locally) approximate the objective and/or the constraint functions. See [2, 3, 5, 6, 7, 8, 9, 17, 18, 19, 20, 23, 24]. Due to the absence of the convexity preserving property, it may happen that the resulting optimization is nonconvex. Such a nonconvex problem is not only difficult to solve, but may also be a bad approximation of the original problem.

\(^*\)Tilburg University, Department of Econometrics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

\(^\dagger\)Faculty of Information Technology and Systems, Department TWI/SSOR, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands.

\(^\ddagger\)Faculty of Information Technology and Systems, Department TWI/SSOR, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands.
We show that convexity can be enforced via semidefinite programming formulations. More precisely, the problem of finding the best convex quadratic approximation in the least squares sense, may be formulated as a semidefinite programming problem. Such problems can be solved efficiently nowadays ([1, 10, 15, 16, 21]).

We note that especially in the field of Computer Aided Design much attention has been given to convexity preserving properties for several interpolation and approximation techniques ([11, 12]). However, this research is mostly restricted to splines and to the univariate and bivariate cases.

This paper is organized as follows. After some preliminaries in Section 2, we treat the univariate case in Section 3. We show that the quadratic least squares solution is convexity preserving. In Section 4 we give an example for the bivariate case which shows that the quadratic interpolation function (and thus the least squares function) is not convexity preserving. We show that requiring convexity leads to a semidefinite programming problem, which can efficiently be solved. In Section 5 we suggest some future research.

2 Preliminaries

Let \( n \geq 1 \) and \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) a convex function. Given distinct points \( z_1, z_2, \cdots, z_N \) in \( \mathbb{R}^n \) we consider the problem of finding a quadratic function \( g : \mathbb{R}^n \rightarrow \mathbb{R} \) such that

\[
    f(z_i) = g(z_i), \quad i = 1, 2, \cdots, N. \tag{1}
\]

The function \( g \) being quadratic, we can write it as

\[
    g(z) = z^T Q z + r^T z + \gamma \tag{2}
\]

for some suitable symmetric \( n \times n \) matrix \( Q \), \( n \)-vector \( r \) and some scalar \( \gamma \). Hence, the problem of finding \( g \) such that (1) holds amount to finding \( Q \), \( r \) and \( \gamma \) such that

\[
    z_i^T Q z_i + r^T z_i + \gamma = f(z_i), \quad i = 1, 2, \cdots, N. \tag{3}
\]

This is a linear system of \( N \) equations in the unknown entries of \( Q \), \( r \) and \( \gamma \). The number of unknowns in \( Q \) is equal to \( n + \frac{1}{2}(n^2 - n) \), hence the total number of unknowns is given by

\[
    n + \frac{1}{2}(n^2 - n) + n + 1 = \frac{1}{2}(n + 1)(n + 2).
\]

Let us call the points \( z_1, z_2, \cdots, z_N \) **quadratically independent** if

\[
    z_i^T Q z_i + r^T z_i + \gamma = 0, \quad i = 1, 2, \cdots, N \quad \Rightarrow \quad Q = 0, \quad r = 0, \quad \gamma = 0. \tag{4}
\]

Note that in this case \( N \geq \frac{1}{2}(n + 1)(n + 2) \). Moreover, if \( N = \frac{1}{2}(n + 1)(n + 2) \) then system (3) has a unique solution. We conclude that if the given points \( z_1, z_2, \cdots, z_N \) are quadratically independent and \( N = \frac{1}{2}(n + 1)(n + 2) \) then there exists a unique quadratic function \( g \) such that (1) holds. This is the interpolation case. When \( N > \frac{1}{2}(n + 1)(n + 2) \), the linear system (3) is overdetermined, in which case quadratic least squares can be applied.

3 Quadratic least square solutions for the univariate case

In this section we consider the univariate case \( (n = 1) \). So \( f \) is a one-dimensional convex function. It is obvious that for any three quadratically independent points \( z_1, z_2, z_3 \) the function \( g \) will be convex. In other words, the quadratic interpolation function is convexity
The quadratic least squares solution is convex if and only if
preserving. More precisely, we show that the quadratic least squares approximation \( g \) of \( f \)
with respect to a set of points
\[
Z := \{ z_1, z_2, \ldots, z_N \}
\]
is convex.

**Theorem 3.1** Let \( z_1 < z_2 < \ldots < z_N \) and \( y_i = f(z_i) \) (\( i = 1, \ldots, N \)) be given, where \( f \) is a
univariate convex function. The quadratic least squares approximation to this data set, i.e. \( g \), is a convex quadratic function.

**Proof:** Let \( y_i \) denote the value of \( f \) in \( z_i \) and let \( g \) be given by \( g(t) = qt^2 + rt + \gamma \). Then
the coefficients \( q, r \) and \( \gamma \) follow from the least squares solution of the system
\[
y = \begin{pmatrix}
y_1 \\
\vdots \\
y_N 
\end{pmatrix} = \begin{pmatrix}
1 & z_1 & z_1^2 \\
\vdots & \vdots & \vdots \\
1 & z_N & z_N^2
\end{pmatrix} \begin{pmatrix}
\gamma \\
r \\
q
\end{pmatrix}
\]
By using Gram-Schmidt we can reformulate this system as
\[
y = \begin{pmatrix}
y_1 \\
\vdots \\
y_N 
\end{pmatrix} = \begin{pmatrix}
1 & z_1 - \bar{x} & z_1^2 - \bar{x}^2 - b(z_1 - \bar{x}) \\
\vdots & \vdots & \vdots \\
1 & z_N - \bar{x} & z_N^2 - \bar{x}^2 - b(z_N - \bar{x})
\end{pmatrix} \begin{pmatrix}
\gamma' \\
r' \\
q
\end{pmatrix}, \tag{5}
\]
where
\[
\gamma = \gamma' - \bar{x}q - (\bar{x}^2 - b\bar{x})q
\]
and
\[
r = r' - bq
\]
and
\[
\bar{x} = \frac{\sum_{i=1}^{N} z_i}{N}, \quad \bar{x}^2 = \frac{\sum_{i=1}^{N} z_i^2}{N}, \quad b = \frac{\sum_{i=1}^{N} (z_i - \bar{x})(z_i^2 - \bar{x}^2)}{\sum_{i=1}^{N} (z_i - \bar{x})^2}.
\]
Let \( X \) denote the matrix of coefficients of the linear system (5). Then the least squares solution of (5) is given by \((X^T X)^{-1} X^T y\). Using that the columns of \( X \) are orthogonal, and hence the inverse of \( X^T X \) is a diagonal matrix, one easily finds
\[
q = \frac{\sum_{i=1}^{N} y_i \left( z_i^2 - \bar{x}^2 - b(z_i - \bar{x}) \right)}{\sum_{i=1}^{N} \left( z_i^2 - \bar{x}^2 - b(z_i - \bar{x}) \right)^2}.
\]
The quadratic least squares solution is convex if and only if \( q \geq 0 \), i.e. if and only if
\[
\sum_{i=1}^{N} y_i \left( z_i^2 - \bar{x}^2 - b(z_i - \bar{x}) \right) \geq 0. \tag{6}
\]
Without loss of generality we may assume that
\[
z_1 < z_2 < \cdots < z_N.
\]
Define
\[
p_i := z_i^2 - \bar{x}^2 - b(z_i - \bar{x}), \quad 1 \leq i \leq N,
\]
\( e \) as the all-one vector, \( z = (z_1, \ldots, z_N)^T \), and \( z^2 = (z_1^2, \ldots, z_N^2)^T \). It is obvious that \( p^T e = 0 \), since \( e^T (z - \overline{z} e) = 0 \) and \( e^T \left( z^2 - \overline{z^2} e \right) = 0 \), due to the definition of \( \overline{z} \) and \( \overline{z^2} \). One can also prove that \( p^T z = 0 \). To this end, recall that
\[
\begin{align*}
    b &= \frac{\sum_{i=1}^N (z_i - \overline{z}) (z_i^2 - a)}{\sum_{i=1}^N (z_i - \overline{z})^2} = \frac{(z - \overline{z} e)^T (z^2 - \overline{z^2} e)}{\|z - \overline{z} e\|^2} = \frac{z^T (z^2 - \overline{z^2} e)}{\|z - \overline{z} e\|^2}.
\end{align*}
\]
Hence
\[
\begin{align*}
p^T z &= z^T \left( z^2 - \overline{z^2} e - b (z - \overline{z} e) \right) \\
&= z^T \left( z^2 - \overline{z^2} e \right) - \frac{z^T (z^2 - \overline{z^2} e)}{\|z - \overline{z} e\|^2} z^T (z - \overline{z} e).
\end{align*}
\]
Since
\[
z^T (z - \overline{z} e) = (z - \overline{z} e)^T (z - \overline{z} e) = \|z - \overline{z} e\|^2
\]
it follows that \( p^T z = 0 \).

To decide whether condition (6) is always satisfied, we consider the problem of minimizing the expression at the left in (6) under the condition that for each \( i, y_i = f(z_i) \), where \( f \) is a convex function.

Assume that \( z \) is given we therefore consider the linear optimization (LO) problem:
\[
(LO) : \quad \min_y \left\{ \sum_{i=1}^N y_i p_i : \quad y_i \leq \frac{(z_i - z_{i-1}) y_{i+1} + (z_{i+1} - z_i) y_{i-1}}{z_{i+1} - z_{i-1}}, \quad 2 \leq i \leq N - 1 \right\}.
\]
Note that the constraints enforce the convexity requirement on the \( y_i \)-values. Since the constraints are homogeneous in \( y \), condition (6) is satisfied if and only if the LO problem has optimal value zero.

We proceed to show that problem (LO) has optimal value zero. Assume that a feasible solution \( \bar{y} \) to problem (LO) is given such that \( p^T \bar{y} < 0 \). Each feasible solution of (LO) corresponds to some convex function, in the sense that we can assume \( \bar{y}_i = \bar{f}(x_i) \) (\( i = 1, \ldots, N \)) for some convex function \( \bar{f} \).

Now define the convex quadratic function
\[
p(t) := t^2 - \overline{z^2} - b(t - \overline{z}), \quad t \in \mathbb{R},
\]
and note that \( p(z_i) = p_i \) (\( i = 1, \ldots, N \)). Since \( p^T e = 0 \), we know that the function value of \( p \) changes sign at least once in the interval \([z_1, z_N]\). By the convexity of \( p \), the level set
\[
\{ t : p(t) \leq 0, \quad t \in [z_1, z_N] \}
\]
is a closed sub-interval of \([z_1, z_N]\). All the nonpositive \( p_i \)'s correspond to \( z_i \) values in this interval. The index set
\[
I := \{ i : p_i \leq 0 \}
\]
is therefore a set of consecutive indices (or of one index). Assume for the moment that \( I \) is not a singleton. Let the first index in \( I \) be \( i_1 \) and the last \( i_2 > i_1 \).

Construct the chord between the points \((z_{i_1}, \bar{y}_{i_1})\) and \((z_{i_2}, \bar{y}_{i_2})\) in the \( z-y \) plane. By the convexity of \( \bar{f} \), all the points \((z_i, \bar{y}_i)\) lie on or below this chord for \( i \in I \) (see Figure 1). Now replace the \( \bar{y}_i \) values for \( i \in I \) with the values on the chord. Note that this does not increase the objective function of (LO) evaluated at the new \( y \)-values, since we are increasing the \( \bar{y}_i \) values for \( i \in I \) and \( p_i \leq 0 \) for \( i \in I \).

Next, extend the chord over the entire interval \([z_1, z_N]\). By the convexity of \( \bar{f} \), the points \((z_i, \bar{y}_i) \) (\( i \in \{1, \ldots, N\} \setminus I \)) lie on or above the chord (see Figure 1). We again replace the \( \bar{y}_i \) values by the corresponding values on the chord.
Again, this does not increase the objective function of (LO), since we are decreasing the $\tilde{g}_i$ values ($i \in \{1, \ldots, N\} \setminus I$), while the corresponding $p_i$'s are nonnegative.

We have thus constructed a new feasible solution of (LO), say $\tilde{y}$, with a negative objective value $p^T \tilde{y} < 0$. Note that $\tilde{y}$ is feasible since it corresponds to a linear function, namely the linear function defined by the chord.

In other words, we have:

$$\tilde{g}_i = c_1 z_i + c_2, \quad i = 1, \ldots, N$$

for some constants $c_1$ and $c_2$. This implies:

$$p^T \tilde{y} = c_1 p^T z + c_2 p^T e = 0,$$

since $p^T z = p^T e = 0$, which is a contradiction.

All that remains is to analyse the case where $I$ is a singleton ($i_1 = i_2$). In this case, it is easy to see that we can replace the chord in the above construction with a line defined by any subgradient of $f$ at $z_{i_1}$. This completes the proof. □

4 Quadratic approximation for the multivariate case

As already said in the previous section, it is obvious that if $n = 1$ (univariate case) then for any three quadratically independent points $z_1, z_2, z_3$ the function $g$ will be convex. Surprisingly enough the analogous property does not hold if $n$ is larger than 1 (multivariate case). This means that quadratic interpolation in the multivariate case is not convexity preserving. Consequently, also quadratic approximation in all norms (1-norm, 2-norm (least squares), $\infty$-norm) is not convexity preserving. In this section we will first give a bivariate example for which the quadratic interpolation is not convexity preserving. Then we will show that convexity can be preserved by using semidefinite programming techniques.
4.1 A counter-example for the bivariate case

The following (bivariate) example shows the counterintuitive fact that quadratic interpolation is not convexity preserving in multivariate cases.

Example 4.1 Consider the case where $f$ is given by

$$f(x) = -\ln x_1 x_2, \quad x_1 > 0, x_2 > 0,$$

which is clearly a convex function, and the points are the 6 columns of the matrix $Z$ given by

$$Z = \begin{pmatrix} 1 & 2 & 3 & 2 & 4 & 6 \\ 2 & 1 & 2 & 3 & 4 & 6 \end{pmatrix}.$$

These points are quadratically independent since the coefficient matrix of the linear system (4), and hence also of (3), is given by

$$\begin{pmatrix} 1 & 2 & 4 & 1 & 2 & 1 \\ 4 & 2 & 1 & 2 & 1 & 1 \\ 9 & 6 & 4 & 3 & 2 & 1 \\ 4 & 6 & 9 & 2 & 3 & 1 \\ 16 & 16 & 16 & 4 & 4 & 1 \\ 36 & 36 & 36 & 6 & 6 & 1 \end{pmatrix},$$

and this matrix is nonsingular. The (unique, but rounded) solution of (3) is given by

$$Q = \begin{pmatrix} -0.2050 & 0.2628 \\ 0.2628 & -0.2050 \end{pmatrix}, \quad r = \begin{pmatrix} -0.7804 \\ -0.7804 \end{pmatrix}, \quad \gamma = 1.6219.$$

The eigenvalues of $Q$ are $-0.4677$ and $0.0578$, showing that $Q$ is indefinite. Hence the quadratic approximation $g$ of $f$ determined by the given points $z_1, z_2, \ldots, z_6$, is not convex. Figure 4.1 shows some of the level curves of $f$ (dashed) and $g$ (solid) as well as the points $z_i, \ i = 1, 2, \ldots, 6$.

The level sets of $g$ are clearly not convex and differ very much from the corresponding level sets of $f$.

In many cases it is important to have a convex quadratic approximation of $f$. In the next section we show how this can be achieved.

4.2 Convex quadratic approximations for the multivariate case

Our aim is to obtain a good convex quadratic approximation $g$ of $f$ on the points in the finite set

$$Z := \{z_1, z_2, \ldots, z_N\}.$$

Convexity of $g$ is equivalent to the matrix $Q$ in (2) being positive semidefinite, yielding the condition

$$Q \succeq 0. \quad (7)$$

It is clear from the above example that it is impossible to guarantee convexity if we want $g$ to coincide with $f$ on $Z$. Therefore, to achieve a convex quadratic approximation we need
Figure 2: Level curves of $f$ and $g$ and the points where they coincide

to relax the condition (1). This can be done in several ways. Here we will treat the infinity norm, the 1-norm and the 2-norm.

First one may want to minimize the infinity norm of $f - g$ at $Z$, yielding the objective

$$\min_{z \in Z} \max \{ |f(z) - g(z)| \}. \quad (8)$$

It will be convenient to use the notation

$$s(z) = f(z) - z^T Q z - r^T z - \gamma, \quad z \in Z.$$  

With the above objective we can find $g$ by solving the problem

$$\min (t : -t \leq s(z) \leq t (\forall z \in Z), \ Q \succeq 0). \quad (9)$$

One also might minimize the 1-norm of $f - g$ at $Z$, yielding the objective

$$\min \sum_{z \in Z} |f(z) - g(z)|. \quad (10)$$

Then $g$ can be found by solving

$$\min \left( \sum_{z \in Z} t_z : -t_z \leq s(z) \leq t_z (\forall z \in Z), \ Q \succeq 0 \right). \quad (11)$$

Finally, we can minimize the 2-norm of $f - g$ at $Z$ (least squares), yielding the objective

$$\min \sum_{z \in Z} (f(z) - g(z))^2. \quad (12)$$
and then $g$ can be found by solving

$$
\min \left( t : \sqrt{\sum_{z \in \mathcal{Z}} s(z)^2} \leq t, \ Q \succeq 0 \right).
$$

(13)

For the first two cases the resulting problems (9) and (11) have linear constraints and a semidefinite constraint $Q \succeq 0$. Such a semidefinite programming problem can efficiently be solved ([1, 10, 15, 16, 21, 25]). The third resulting problem (13) again can be efficiently solved, since the new constraint is a second order cone (Lorentz cone) constraint ([21]).

In practice one sometimes want to add the condition that the approximation is exact or an upper- or underestimate in several points in $\mathcal{Z}$. Observe that such additional properties that $f(z) \geq g(z)$, $z \in \mathcal{Z}$ (or $f(z) \leq g(z)$, $z \in \mathcal{Z}$) then we simply add the constraints $s(z) \geq 0$ (respectively $s(z) \leq 0$) to the above minimization problems. The resulting problems can still be formulated as semidefinite programming problems.

**Example 4.2** For the bivariate example given above we calculated the least squares solution while preserving convexity. Using SeDiMu ([22]) we solved problem (13). We obtained the following (rounded) solution:

$$
Q = 0.02750 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad r = -0.7287 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \gamma = 1.2196.
$$

The eigenvalues of $Q$ are 0.55 and 0, showing that $Q$ is positive semidefinite. Hence the quadratic approximation $g$ of $f$ determined by the given points $z_1, z_2, \ldots, z_6$, is convex, but degenerate. Note that $Q$ is not positive definite because the constraint $Q \succeq 0$ is binding at the optimal solution of problem (13). (If we remove the constraint $Q \succeq 0$, then we get the non-convex interpolation function of the previous example.)

Figure 3 shows some of the level curves of $f$ (dashed) and $g$ (solid) as well as the points $z_i$, $i = 1, 2, \ldots, 6$. Comparing with Figure 2 we see that the convex approximation approximates $f$ much better within the convex hull of the six specified points, if the measure of quality is the maximum error or integral of the error function

$$
\text{err}(z) = |f(z) - g(z)|
$$

over the convex hull. (The convex hull defines a natural trust region for the approximation).

5 Future research

In this paper we showed (among other things) that the quadratic 2-norm (least squares) approximation of a convex univariate function in a finite number of points is convex. It is an interesting question whether this is also true for other norms then the 2-norm (e.g. the infinity norm, or the 1-norm).

As already mentioned in the introduction, several optimization methods for solving problems with expensive function evaluations, use quadratic interpolation or approximation. A consequence of this paper is that for convex problems the interpolation or approximation may be nonconvex, which may increase the number of iterations of such optimization methods. In the near future we will investigate how we can improve these methods by exploiting the convex structure.
Figure 3: Level curves of \( f \) and \( g \)

Acknowledgements

We would like to acknowledge Ruud van Lieshout for his fruitful ideas w.r.t. the convexity preserving proof for the univariate case and Wim van Beers for providing us the references [11] and [12], as well as Kees Vuik and Piet Wesseling for discussions on the associated dual problem of problem (L.O).

References
