Affine Lower Bound Functions for Polynomials and Their Use in Global Optimisation (Part 1)

Jürgen Garloff   Andrew P. Smith

University of Applied Sciences /
Fachhochschule Konstanz
Department of Computer Science   Institute for Applied Research
and
University of Konstanz
Department of Mathematics and Statistics

Christian Jansson

Institute of Computer Science III
Technical University Hamburg–Harburg

Affine Lower Bound Functions for Polynomials and Their Use in Global Optimisation (Part 1)

Outline

- Bernstein expansion for polynomials
- Lower bound functions in the univariate case
- The convex hull property and the inclusion isotonicity of the convex hull
- Lower bound functions in the multivariate case, formulated as the solution of linear programming problems
- Examples
- Verification and applications

Given: $f : S \to \mathbf{R}$, $S \subseteq \mathbf{R}^n$, $X \in IS$ Wanted: $f(X) = \{f(x) : \ x \in X\}$
Methods: natural interval extension, mean value / centered / slope forms... Even if the bounds provided by these methods do not overestimate $f(X)$, all information about the **shape** of f over X is lost, since these methods only deliver an **interval** as the result.

Convex-concave extension (C.Jansson, 2000):
A *convex-concave extension* of f is a mapping $[\underline{f}, \overline{f}]$ which provides for each $X \in IS$
a convex function $\underline{f}(x, X)$ (called the *lower bound function*) and
a concave function $\overline{f}(x, X)$ (called the *upper bound function*) such that

$$\underline{f}(x, X) \ \leq \ f(x) \ \leq \ \overline{f}(x, X) \quad \forall x \in X.$$

Interval extensions are special cases of convex-concave extensions.
In our talk: Construction of affine / convex lower and affine / concave upper bound functions for multivariate polynomials.

Bernstein Polynomials

$B_i(x) = \binom{l}{i} x^i (1-x)^{l-i}, \qquad 0 \le i \le l$

given: $p(x) = \sum\limits_{i=0}^{l} a_i x^i$

wanted: $p(I) = \{p(x) \mid x \in I\}$ w.l.o.g. $I = [0,1]^n$

power form $\longrightarrow$ Bernstein form

$p(x) = \sum\limits_{i=0}^{l} b_i B_i(x)$, where

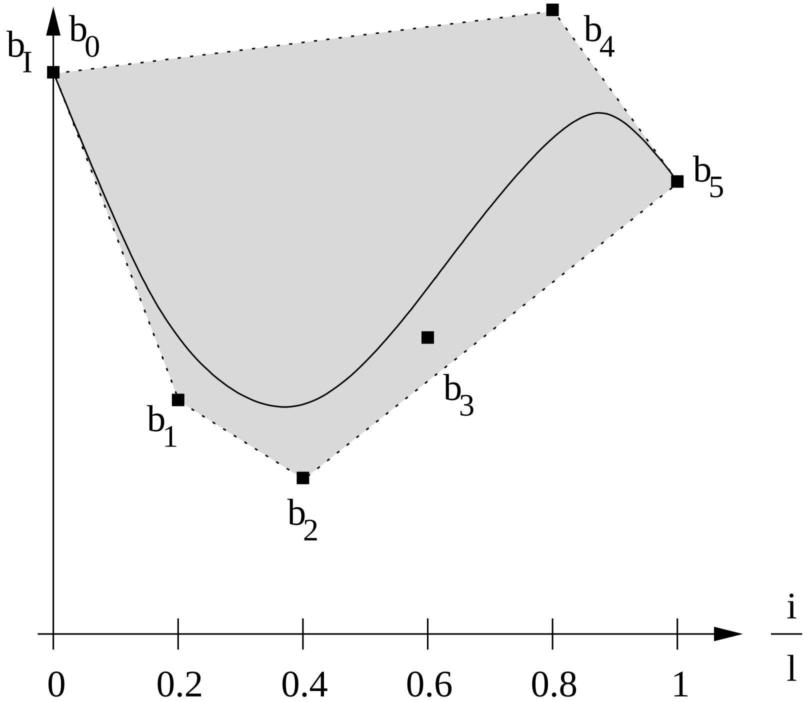
$$b_i = \sum_{j=0}^{i} \frac{\binom{i}{j}}{\binom{l}{j}} a_j, \quad 0 \le i \le l \qquad \textit{Bernstein coefficients}$$

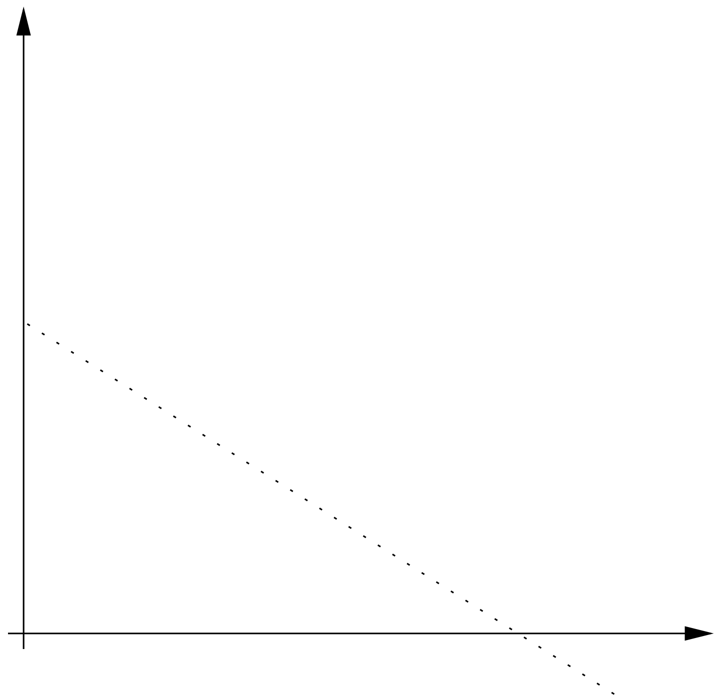
in particular, $b_0 = a_0 = p(0), \quad b_l = \sum\limits_{i=0}^{l} a_i = p(1)$

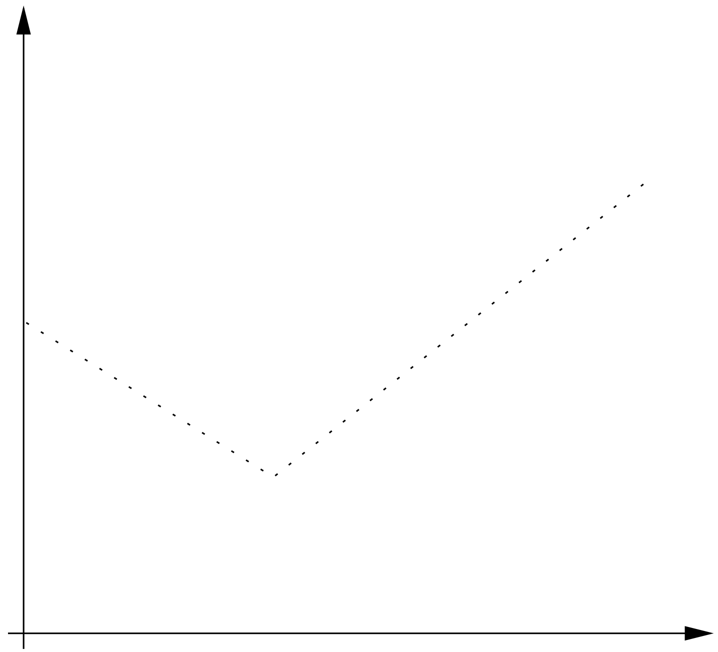
can be calculated economically by difference table method (similarly to the de Casteljau algorithm).

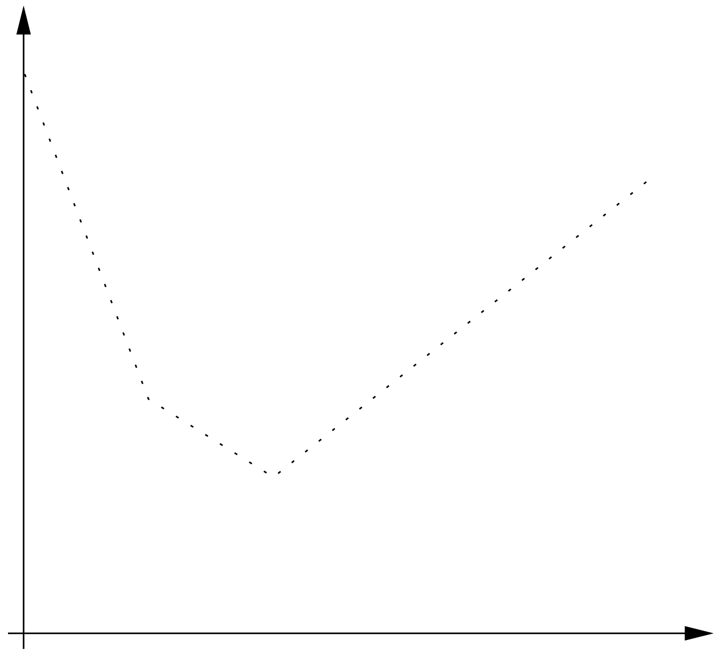
- Convex hull property:

$$\left\{ \binom{x}{p(x)} \;:\; x \in I \right\} \quad \subseteq \quad conv \left\{ \binom{i/l}{b_i} \;:\; 0 \le i \le l \right\}.$$









Range Enclosing Property

$$p(U) \;\subseteq\; B_p \;:=\; \left[\,\min_{i=0}^{n} b_i,\, \max_{i=0}^{n} b_i\,\right]$$

Bounds are sharp iff min./max. is attained at b_0 or b_n.

In fact, $$\begin{array}{rcccl} b_0 & = & a_0 & = & p(0) \\ b_n & = & \sum_{i=0}^{n} a_i & = & p(1) \end{array}$$

A mapping $f : IS \to \mathbf{R}$ is termed *inclusion isotone* if

$$\forall X, Y \in IS : X \subseteq Y \Longrightarrow f(X) \subseteq f(Y).$$

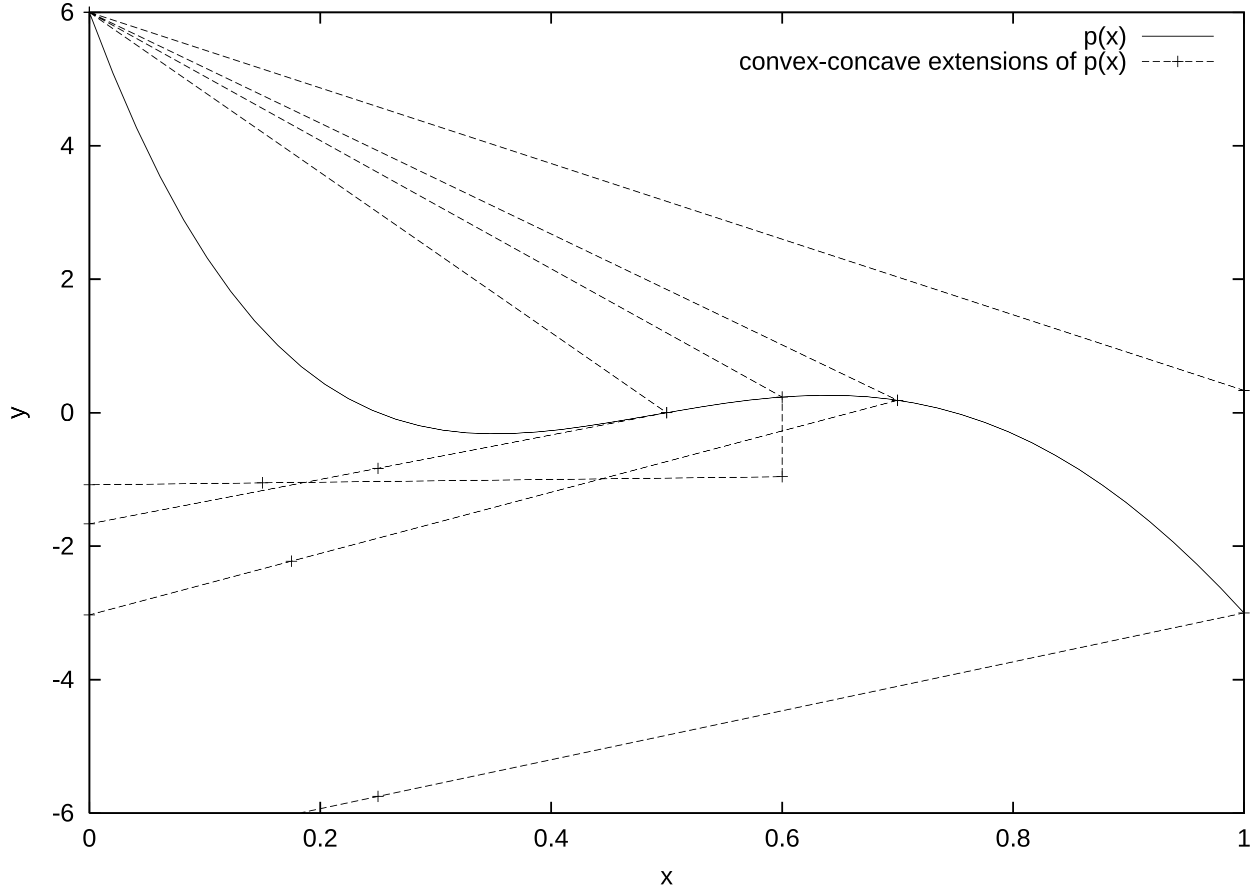


Fig. 1. Failure of inclusion isotonicity with one affine function

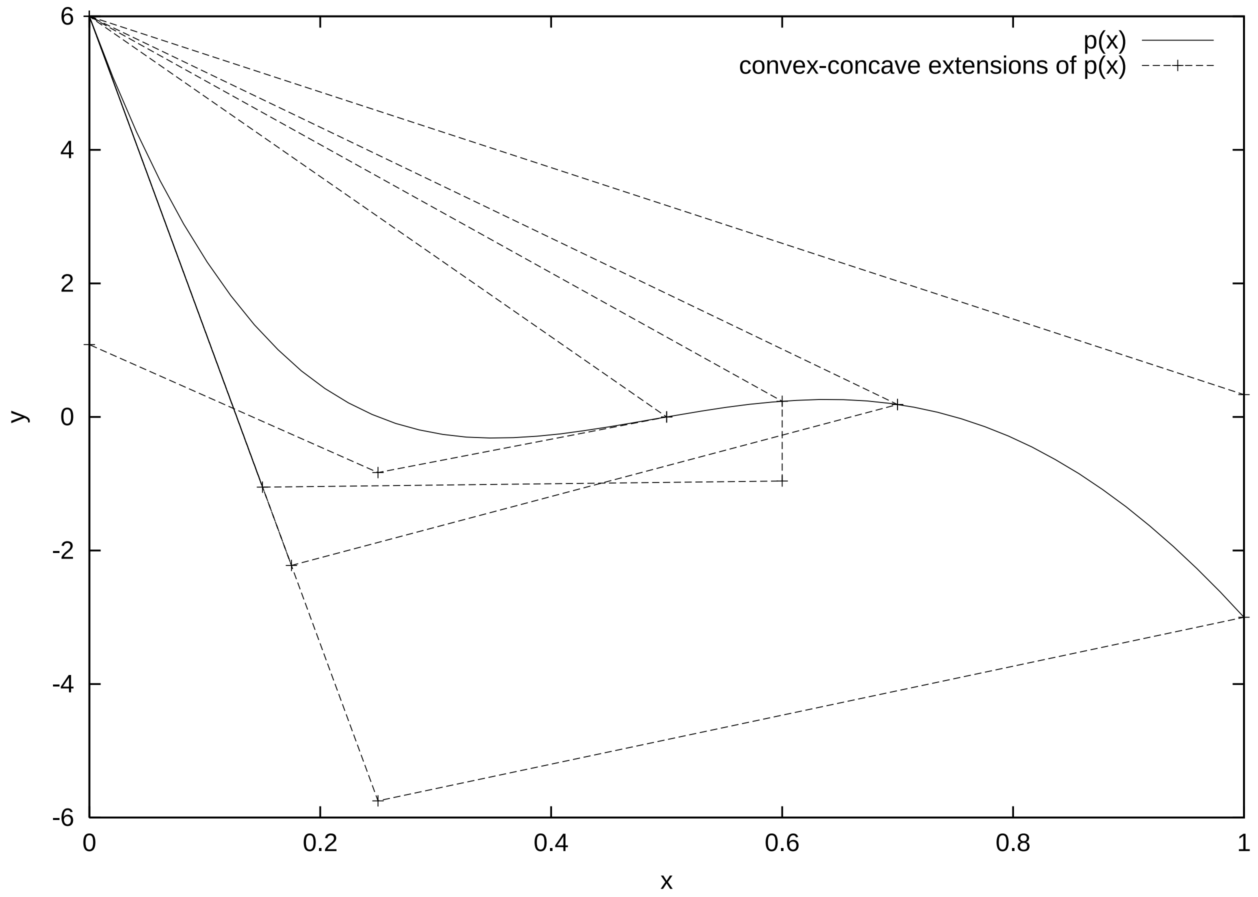


Fig. 2. Failure of inclusion isotonicity with two affine functions

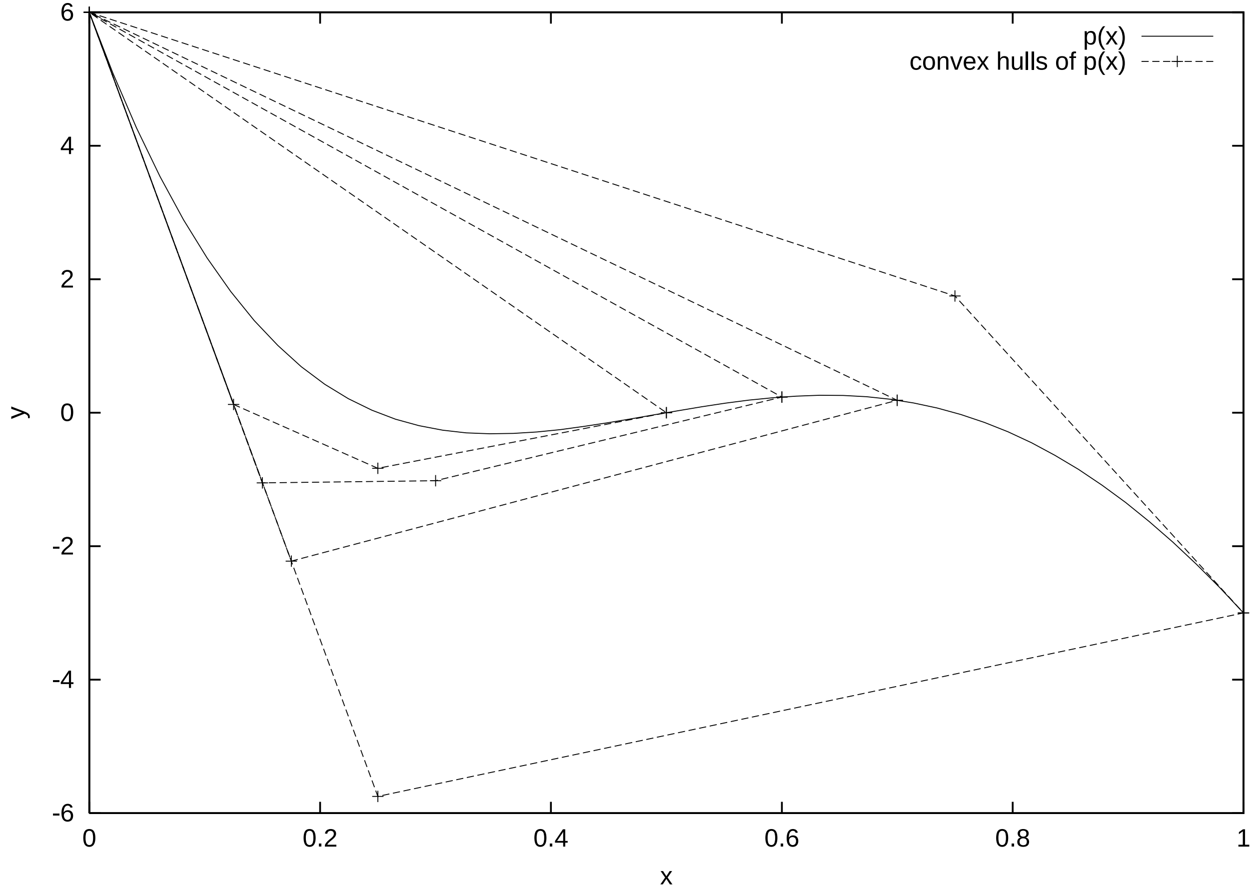


Fig. 3. The convex hull is inclusion isotone.

Fact: The convex-concave extension based on the convex hull of the control points is inclusion isotone.

Proof:

1. It suffices to show that inclusion isotonicity holds if we shrink only one of the two endpoints, say the right-hand one.
2. This shrinking process can be regarded as subdivision with a suitably chosen intersection point.
3. Computation of the Bernstein coefficients on the shrinked interval by an algorithm which is similar to the de Casteljau algorithm in CAGD: By this algorithm, convex combinations of neighbouring entries in a row are formed, giving the entries in the following row. For $k = 1, \ldots, n$ and $\lambda \in (0, 1)$

$$b_i^{(k)} := (1-\lambda) b_{i-1}^{(k-1)} + \lambda b_i^{(k-1)}, \quad i = k, \ldots, n.$$

$$
\begin{array}{ccccccc}
\{b_i\} & \longrightarrow & b_0^{(0)} & b_1^{(0)} & b_2^{(0)} & \cdots & b_n^{(0)} \\
 & & & b_1^{(1)} & b_2^{(1)} & \cdots & b_n^{(1)} \\
 & & & & b_2^{(2)} & \cdots & b_n^{(2)} \\
 & & & & & \ddots & \vdots \\
 & & & & & & b_n^{(n)}
\end{array}
$$

Corollary (H. Hong and V. Stahl, 1995): The Bernstein form (i.e., the form which provides an interval as enclosure for the range) is inclusion isotone.

Notations

Multiindices $i = (i_1, \ldots, i_n)^T$ are vectors, where the components are nonnegative integers. The vectors 0 and 1 denote the multiindices with all components equal to 0 or 1, resp.
Comparisons and the absolute value $|.|$ are used entrywise.
The division of multiindices i, l with $l > 0$ is defined componentwise as $i/l := (i_1/l_1, \ldots, i_n/l_n)^T$. For $x \in \mathbf{R}^n$ its multipowers are $x^i := \prod_{\mu=1}^{n} x_\mu^{i_\mu}$.
We use the notations $\sum_{i=0}^{l} := \sum_{i_1=0}^{l_1} \ldots \sum_{i_n=0}^{l_n}$ and $\binom{l}{i} := \prod_{\mu=1}^{n} \binom{l_\mu}{i_\mu}$.

Theorem 1 *Let* $\{b_i\}_{i=0}^{l}$ *denote the Bernstein coefficients of an* n-*variate polynomial* p *of degree* l. *Let* $\widehat{\imath}$ *be a multiindex such that*

$$b_{\widehat{\imath}} := \min\{b_i : 0 \le i \le l\},$$

and let $\hat{J} \subseteq \{\widehat{\jmath} : \ 0 \le \widehat{\jmath} \le l, \ \widehat{\jmath} \ne \widehat{\imath}\}$ *be a nonempty set of at least* n *multiindices such that*

$$\frac{b_{\widehat{\jmath}} - b_{\widehat{\imath}}}{\|\widehat{\jmath}/l - \widehat{\imath}/l\|} \le \frac{b_i - b_{\widehat{\imath}}}{\|i/l - \widehat{\imath}/l\|} \ \text{ for each } \widehat{\jmath} \in \hat{J}, \ \ 0 \le i \le l, \ i \ne \widehat{\imath}, \ i \notin \hat{J}.$$

Then the affine function

$$c(x) := -\widehat{s}^T(x - \widehat{\imath}/l) + b_{\widehat{\imath}}$$

is a lower bound function for p *on* I, *where* $\widehat{s}$ *is an optimal solution of the LP problem*

$$\max_s \sum_{\widehat{\jmath} \in \hat{J}} c(\widehat{\jmath}/l) \quad \text{subject to} \quad c(i/l) \le b_i \quad \text{for} \quad 0 \le i \le l, i \ne \widehat{\imath}.$$

Theorem 2 *The affine lower bound function c satisfies the a posteriori error bound*

$$0 \leq p(x) - c(x) \leq \max\{b_i - b_{\hat{i}} + (i/l - \hat{i}/l)^T \cdot \hat{s} :\ 0 \leq i \leq l\},\ x \in I,$$

which specifies in the univariate case to $(x \in I)$

$$0 \leq p(x) - c(x) \leq \max\left\{\left(\frac{b_i - b_{\hat{i}}}{i - \hat{i}} - \frac{b_{\hat{j}} - b_{\hat{i}}}{\hat{j} - \hat{i}}\right)(i - \hat{i}) :\ \ 0 \leq i \leq l,\ i \neq \hat{i}\right\}.$$

Remark: If we extend the construction of affine lower bound functions in the univariate case from I to arbitrary intervals $[\underline{a}, \overline{a}]$ with $\underline{a} < \overline{a}$, then we can show that this error bound is quadratic w.r.t. the width of the interval , i.e., the right-hand side can be bounded from above by $C(p)(\overline{a} - \underline{a})^2$, where $C(p)$ is an constant depending only on p. The question whether quadratic convergence holds true in the multivariate case is open, but seems likely also to hold.

Verification

For algorithms for calculating rigorous error bounds for the Bernstein coefficients cf. H.C.Fischer '90 and J.Rokne '77 '82. Using these algorithms rigorous affine / convex lower bound functions in the 1D case can be computed.
In the nD case, we propose the use of the following verified affine lower bound function

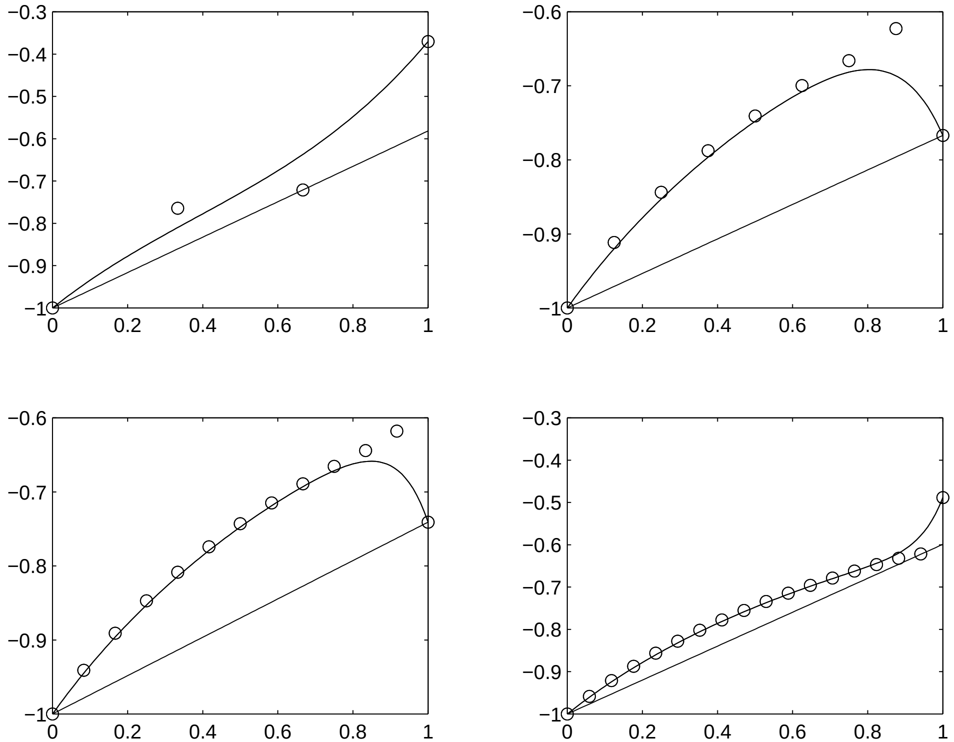
$$c(x) := \hat{s}^T(\hat{i}/l - x) + b_{\hat{i}} + \delta, \quad x \in I,$$

where

$$\delta \leq \min\left\{(i/l - \hat{i}/l)^T \cdot \hat{s} + b_i - b_{\hat{i}} : \ 0 \leq i \leq l, i \neq \hat{i}\right\};$$

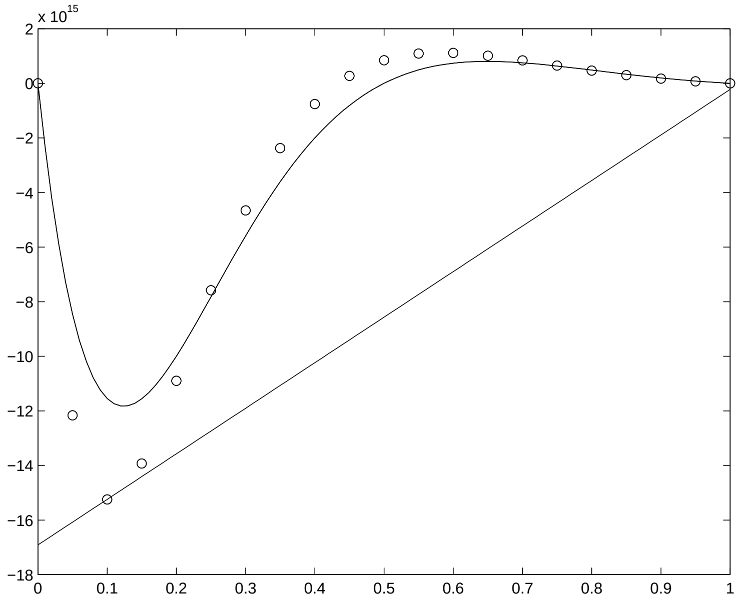
δ should be computed by using interval arithmetic.

Example 1: For $l = 3, 8, 13, 17$: $p(x) = \sum_{i=0}^{l} \frac{(-1)^{i+1}}{i+1} x^i, \quad x \in [0,1]$



l	3	8	12	17
error	0.2113	0,1735	0.1446	0.1103

Example 2: $p(x) = \prod_{i=1}^{20}(x-i), \quad x \in [1,3]$



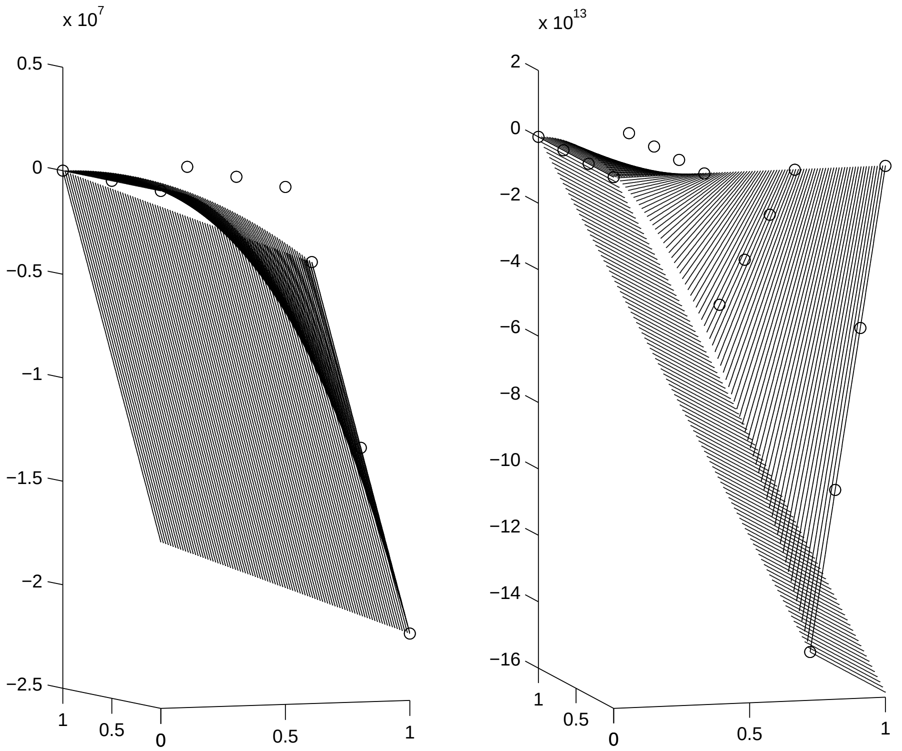
The error bound is 1.6912×10^{16}.

Example 3:

$$
\begin{aligned}
p_1(x_1,x_2) &= \alpha_1 x_1^2 x_2 + \alpha_2 x_1^2 + \alpha_3 x_1 x_2 + \alpha_4 x_1 + \alpha_5 x_2, \\
p_2(x_1,x_2) &= \alpha_6 x_1^2 x_2 + \alpha_7 x_1 x_2^2 + \alpha_8 x_1 x_2 + \alpha_9 x_2^3 + \alpha_{10} x_2^2 + \alpha_{11} x_2 + \alpha_{12},
\end{aligned}
$$

where

$$
\begin{array}{ll}
\alpha_1 = -1.697 \times 10^7 & \alpha_7 = 4.126 \times 10^7 \\
\alpha_2 = 2.177 \times 10^7 & \alpha_8 = -8.285 \times 10^6 \\
\alpha_3 = 0.5500 & \alpha_9 = 2.284 \times 10^7 \\
\alpha_4 = 0.4500 \times 10^7 & \alpha_{10} = 1.918 \times 10^7 \\
\alpha_5 = -1.0000 \times 10^7 & \alpha_{11} = 48.40 \\
\alpha_6 = 1.585 \times 10^{14} & \alpha_{12} = -27.73
\end{array}
$$



The error bounds for p_1 and p_2 are 1.937×10^7 and 1.585×10^4, respectively.