



A New Class of Improved Convex Underestimators for Twice Continuously Differentiable Constrained NLPs

IOANNIS G. AKROTIRIANAKIS and CHRISTODOULOS A. FLOUDAS*

Department of Chemical Engineering, Princeton University, Princeton, NJ 08540, USA
(e-mail: floudas@titan.princeton.edu)

(Received: 7 October 2003; accepted: 10 November 2003)

Abstract. We present a new class of convex underestimators for arbitrarily nonconvex and twice continuously differentiable functions. The underestimators are derived by augmenting the original nonconvex function by a nonlinear relaxation function. The relaxation function is a separable convex function, that involves the sum of univariate parametric exponential functions. An efficient procedure that finds the appropriate values for those parameters is developed. This procedure uses interval arithmetic extensively in order to verify whether the new underestimator is convex. For arbitrarily nonconvex functions it is shown that these convex underestimators are tighter than those generated by the α BB method. Computational studies complemented with geometrical interpretations demonstrate the potential benefits of the proposed improved convex underestimators.

Key words. α BB, convex underestimators, global optimization.

1. Introduction

In this paper we address the class of C^2 constrained nonlinear optimization problems. The mathematical description of these problems is as follows

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h_j(x) = 0, \quad j = 1, 2, \dots, m_1 \\ & g_k(x) \leq 0, \quad k = 1, 2, \dots, m_2 \\ & x \in X = [x^L, x^U], \end{aligned} \tag{1}$$

where $x \in \mathfrak{R}^n$ is the vector of variables, $x^L, x^U \in \mathfrak{R}^n$ are the vectors of the lower and upper bounds of the hyper-rectangular domain $X \subseteq \mathfrak{R}^n$, f is the objective function and $h_j(x), g_k(x)$ are the constraints of the problem. Generally, due to the presence of nonconvex functions, problem (1) possesses many local minima. The larger the number of local minima the more difficult the task of locating the global minimum becomes. Global optimization problems belong to the class of *NP*-hard

* Corresponding author. Tel.: (609) 258-4595; fax: (609) 258-0211.

problems and therefore are generally very difficult to solve. Even proving that a solution of (1) is not a local minimum is an *NP*-complete problem (see (Murty and Kabadi (1987)) and (Pardalos and Schnitger (1988))).

Standard optimization algorithms, like Sequential Quadratic Programming (Boggs and Tolle (1995)) or Interior Point (Methods Wright (1992)), guarantee the location of a local optimum only. However, in many industrial and scientific applications, such as process synthesis, design and control, computational chemistry and molecular biology, a local optimum does not suffice. In the last few decades, a number of global optimization methods have been developed. These methods can be classified into two broad categories: (i) Deterministic and (ii) Stochastic. Deterministic methods (e.g., Adjiman et al. (1998a, b), Al-Khayyal and Falk (1983), Horst and Tuy (1987), Porn et al. (1999), Ryoo and Sahinidis (1996), Sherali and Alameddine (1992), Smith and Pantelides (1996), Tuy (1987)) guarantee to reach an ϵ -neighborhood of the global optimum of problem (1) within a finite number of steps. An extensive coverage of the theory and applications of deterministic methods can be found in (Floudas (2000)). On the other hand, in stochastic methods (e.g., Gelatt et al. (1983), Goldberg (1987), Rinnoy-Kan and Timmer (1987a), Rinnoy-Kan and Timmer (1987b), Schoen (1991)) the probability of finding the global optimum of problem (1) goes to one as the number of steps goes to infinity.

Many deterministic global optimization methods share the following two main ideas: (i) generation of convex underestimators for the nonconvex functions involved in (1) and (ii) partition of the feasible region into smaller ones. The convex underestimators are important since they are used to construct a convex optimization problem that provides a lower bound on the global optimum of the original problem. As the subregions become smaller the convex lower bounding problems approach the original nonconvex problem. This process progressively leads to the determination of the global optimum by repetitively solving convex optimization problems in different sub-domains of the original domain.

The α BB algorithm (Adjiman et al. (1998a, b), Floudas (2000), Maranas and Floudas (1994b)), is a deterministic global optimization algorithm that partitions the continuous feasible region using the Branch and Bound method in the continuous variables. At each node of the tree, convex underestimators are generated by subtracting a parametric separable convex quadratic function (also referred to as *relaxation* function) from each nonconvex function. The values of the parameters in the relaxation function greatly influence the separation distance that exists between the nonconvex function and its convex underestimator. The larger the magnitude of those parameters the larger the separation distance is, and therefore the looser the underestimator becomes. This affects the quality of lower bounds on the global optimum, provided by the convex lower bounding problem. Due to poor fathoming, lower bounds of bad quality can significantly increase the size of the Branch and Bound tree and reduce the efficiency of the convergence rate of the overall global optimization algorithm.

The main objective of this work is to develop methods that provide convex underestimators of arbitrarily nonconvex functions that are tighter than those provided by α BB. This will not only reduce the size of the Branch and Bound tree but also enhance the performance of the overall algorithm. The critical issue regarding the quality of an underestimator is the relaxation function which is subtracted from the nonconvex function it underestimates. We have developed a new relaxation function that shares the same properties as the one in α BB. That is, it is separable, parametric, convex, and non-negative for all $x \in X$. The advantage of the new relaxation function is that its parameters are selected in such a way that it always takes smaller values than the α BB relaxation function in the whole domain X . As a result, when it is subtracted from a nonconvex function it produces a convex underestimator that is tighter than that produced when the α BB relaxation function is subtracted from the same function.

This paper is structured as follows. Section 2 presents an overview of the underestimators used in the α BB global optimization algorithm. Section 3 describes the new relaxation function and its properties. Section 4 presents the new underestimator and its properties. Section 5 discusses an iterative scheme that verifies the convexity of the new underestimators. Section 6 presents several examples where the new underestimators are compared with those produced by α BB. Finally, Section 7 presents the conclusions of this work.

2. Overview of Convex Underestimators of the α BB Method

In α BB, a convex underestimator of a nonconvex function is constructed by decomposing it into a sum of nonconvex terms of special type (e.g., linear, bilinear, trilinear, fractional, fractional trilinear, convex, univariate concave) and nonconvex terms of arbitrary type. The first type is then replaced by very tight convex underestimators which are already known. A complete list of the tight convex underestimators of the above special type nonconvex terms can be found in Floudas (2000).

For the nonconvex terms of arbitrary type, whose convex envelopes are not known, a convex underestimator is generated by adding to them the relaxation function, $\phi(x; \alpha)$:

$$\phi(x; \alpha) = - \sum_{i=1}^n \alpha_i (x_i - x_i^L)(x_i^U - x_i)$$

where $\alpha_i \geq 0$, $i = 1, 2, \dots, n$. That is, if we assume that $f(x)$ is an arbitrarily nonconvex function, then

$$L_{\alpha\text{BB}}(x; \alpha) = f(x) + \phi(x; \alpha) \quad (2)$$

is an underestimator of $f(x)$. Note that since $\phi(x^L; \alpha) = \phi(x^U; \alpha) = 0$ the underestimator $L_{\alpha\text{BB}}(x; \alpha)$ coincides with $f(x)$ at the end-points of X . Also by noting

that the relaxation function $\phi(x; \alpha)$ is separable we can derive the following relationship that exists among the Hessian matrices of $L_{\alpha\text{BB}}(x; \phi)$, $f(x)$ and $\phi(x; \alpha)$

$$\nabla^2 L_{\alpha\text{BB}}(x; \alpha) = \nabla^2 f(x) + 2A \quad (3)$$

where $A = \nabla^2 \phi(x; \alpha) = \text{diag}\{\alpha_1, \alpha_2, \dots, \alpha_n\}$. From the above equation it can be derived that $L_{\alpha\text{BB}}(x; \alpha)$ is convex if and only if $\nabla^2 L_{\alpha\text{BB}}(x; \alpha)$ is positive semi-definite matrix. It is shown in (Adjiman et al. (1998b)) that if the parameters α_i , $i = 1, 2, \dots, n$, have values greater than or equal to the negative one half of the minimum eigenvalue of the Hessian matrix $\nabla^2 f(x)$ in the whole domain $X = [x^L, x^U]$, then the underestimator $L_{\alpha\text{BB}}(x; \alpha)$ is convex function. However, the calculation of the smallest eigenvalue of the Hessian matrix of an arbitrarily nonconvex function is by itself a nonconvex optimization problem and therefore very difficult to solve.

The above difficulty can be alleviated by using interval methods (see for example, (Floudas (2000); Hansen (1992); Neumaier (1990))), and transforming equation (3) into the equation

$$[\nabla^2 L_{\alpha\text{BB}}(x; \alpha)] = [\nabla^2 f(x)] + 2A, \quad (4)$$

where $[\nabla^2 L_{\alpha\text{BB}}(x; \alpha)]$ and $[\nabla^2 f(x)]$ are the interval Hessians of the Hessian matrices $\nabla^2 L_{\alpha\text{BB}}(x; \alpha)$ and $\nabla^2 f(x)$ respectively. If the interval matrix $[\nabla^2 L_{\alpha\text{BB}}(x; \alpha)]$ is positive semi-definite then $\nabla^2 L_{\alpha\text{BB}}(x; \alpha)$ is positive semi-definite for all $x \in X$ and consequently $L_{\alpha\text{BB}}(x; \alpha)$ is convex.

A number of methods have been developed that calculate appropriate values for all α_i , $i = 1, 2, \dots, n$ that ensure the positive semi-definiteness of the interval matrix $[\nabla^2 L_{\alpha\text{BB}}(x; \alpha)]$ and consequently the convexity of the underestimating function $L_{\alpha\text{BB}}(x; \alpha)$. These methods can be classified into two categories. The first category consists of methods that find a common value for every parameter α_i , whereas methods of the second category calculate different values for each α_i .

The most efficient of those methods is the scaled Gerschgorin, which we introduce briefly. Its computational complexity is $O(n^2)$ and the quality of underestimators it produces is fairly good. The value for each parameter α_i is determined by the equation

$$\alpha_i = \max \left\{ 0, -\frac{1}{2} \left(\underline{f}_{ii} - \sum_{j \neq i} \max \{ |\underline{f}_{ij}|, |\bar{f}_{ij}| \} \right) \frac{d_j}{d_i} \right\} \quad (5)$$

where \underline{f}_{ij} and \bar{f}_{ij} are the lower and upper bounds of $\partial^2 f / \partial x_i \partial x_j$ as calculated by interval analysis, and d_i , $i = 1, 2, \dots, n$ are positive parameters. A common choice for those parameters is $d_i = x_i^U - x_i^L$, which reflects the fact that variables with a wider range have a larger effect on the quality of the underestimator than variables with a smaller range. In (Adjiman et al. (1998b)) have proven that the values of α_i defined by (5) are sufficient to ensure that $L_{\alpha\text{BB}}(x; \alpha)$ is a convex underestimator of the nonconvex function $f(x)$. An extensive coverage of the

other methods based on interval analysis can be found in (Adjiman et al. (1998b) and Floudas (2000)).

The effect of the addition of the relaxation function is twofold. The first is to construct an underestimator due to the fact that $\phi(x; \alpha) \leq 0$ for all $x \in X$. The second is to guarantee that the underestimator is convex by assigning appropriate values on the parameters $\alpha_i, i = 1, 2, \dots, n$. Those values convey second order characteristics of the original nonconvex function $f(x)$. That is, the more nonconvex $f(x)$ is, the larger the values of $\alpha_i, i = 1, 2, \dots, n$ are.

A representative measure of the quality of an underestimator is the separation distance between itself and the nonconvex function it underestimates. The smaller the separation distance the tighter the underestimator is. Consequently, the tighter the underestimators are the faster the convergence of the overall Branch and Bound algorithm becomes. The separation distance between $f(x)$ and $L_{\alpha\text{BB}}(x; \alpha)$ is defined by the difference of these functions, that is,

$$d_{\alpha\text{BB}}(x; \alpha) = f(x) - L_{\alpha\text{BB}}(x; \alpha) = \sum_{i=1}^n \alpha_i (x_i - x_i^L)(x_i^U - x_i) \geq 0. \tag{6}$$

In (Maranas and Floudas (1994b)) it has been proved that the distance function, $d_{\alpha\text{BB}}(x; \alpha)$, achieves its maximum at the middle point, x^{mid} , of the interval $X = [x^L, x^U]$ and its value is

$$d_{\alpha\text{BB}}^{\text{max}} = \max_{x^L \leq x \leq x^U} d_{\alpha\text{BB}}(x; \alpha) = \frac{1}{4} \sum_{i=1}^n \alpha_i (x_i^U - x_i^L)^2. \tag{7}$$

As can be seen from (7) the maximum separation distance between a nonconvex function and its αBB underestimator is proportional to the α_i 's. Furthermore, it was shown in (Maranas and Floudas (1994b)) that the number of iterations required by Branch and Bound to achieve ϵ -convergence to the global minimum depends on how large the values of $\alpha_i, i = 1, 2, \dots, n$ as well as the size of the domains $[x_i^L, x_i^U], i = 1, 2, \dots, n$.

3. The New Relaxation Term

In this section we present a new relaxation function. It shares most of the characteristics of the relaxation function $\phi(x; \alpha)$ used in αBB . Moreover, it possesses novel additional properties that enable it to derive tighter convex underestimators. The new relaxation function is defined as follows

$$\Phi(x; \gamma) = - \sum_{i=1}^n (1 - e^{\gamma_i(x_i - x_i^L)})(1 - e^{\gamma_i(x_i^U - x_i)})$$

where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)^T$ is a vector of non-negative parameters. As it will be explained later, these parameters play a similar role as the α_i 's in the αBB method. Note that $\Phi(x; \gamma)$ is not a quadratic function as $\phi(x; \alpha)$.

The gradient of $\Phi(x; \gamma)$ is

$$\nabla\Phi(x; \gamma) = - \begin{pmatrix} -\gamma_1 e^{\gamma_1(x_1 - x_1^L)} + \gamma_1 e^{\gamma_1(x_1^U - x_1)} \\ -\gamma_2 e^{\gamma_2(x_2 - x_2^L)} + \gamma_2 e^{\gamma_2(x_2^U - x_2)} \\ \vdots \\ -\gamma_n e^{\gamma_n(x_n - x_n^L)} + \gamma_n e^{\gamma_n(x_n^U - x_n)} \end{pmatrix}$$

and its Hessian is defined by the diagonal matrix

$$\nabla^2\Phi(x; \gamma) = \text{diag}\{\gamma_i^2 e^{\gamma_i(x_i - x_i^L)} + \gamma_i^2 e^{\gamma_i(x_i^U - x_i)} : i = 1, 2, \dots, n\}$$

Note that $\nabla^2\Phi(x; \gamma)$ is a function of x as opposed to the the Hessian matrix of $\phi(x; \alpha)$, used in αBB , which is constant throughout the domain X .

The new relaxation function $\Phi(x; \gamma)$ has the following important properties.

PROPERTY R1. $\Phi(x; \gamma) \leq 0$, for all $x \in [x^L, x^U]$.

Proof. Since

$$e^{\gamma_i(x_i - x_i^L)} \geq 1 \quad \text{and} \quad e^{\gamma_i(x_i^U - x_i)} \geq 1, \quad \forall x_i \in [x_i^L, x_i^U], \gamma_i \geq 0 \quad (8)$$

we can derive that $-(1 - e^{\gamma_i(x_i - x_i^L)})(1 - e^{\gamma_i(x_i^U - x_i)}) \leq 0$. Taking the sum of the previous inequality for $i = 1, 2, \dots, n$ we conclude that $\Phi(x; \gamma) \leq 0$. \square

PROPERTY R2. $\Phi(x; \gamma)$ becomes zero at the corner points of the interval $[x^L, x^U]$.

Proof. Let $x^C = (x_1^C, x_2^C, \dots, x_n^C)^T$ be a corner point of X , that is, $x_i^C = x_i^L$ or $x_i^C = x_i^U$. Then we have

$$(1 - e^{\gamma_i(x_i^C - x_i^L)})(1 - e^{\gamma_i(x_i^U - x_i^C)}) = 0, \quad i = 1, 2, \dots, n$$

Hence, $\Phi(x^C; \gamma) = 0$. \square

PROPERTY R3. $\Phi(x; \gamma)$ is a convex function.

Proof. Since the Hessian matrix, $\nabla^2\Phi(x; \gamma)$, of $\Phi(x; \gamma)$ is diagonal and for each diagonal element we have

$$\gamma_i^2 e^{\gamma_i(x_i - x_i^L)} + \gamma_i^2 e^{\gamma_i(x_i^U - x_i)} \geq 0, \quad \forall x_i \in [x_i^L, x_i^U], \gamma_i \geq 0$$

we can conclude that $\nabla^2\Phi(x; \gamma)$ is positive semi-definite. Hence $\Phi(x; \gamma)$ is a convex function. \square

PROPERTY R4. $\Phi(x; \gamma)$ achieves its minimum at the middle point, x^{mid} , of X and its maximum at the corner points.

Proof. The solution of the system of equations $\nabla\Phi(x; \gamma) = 0$ is achieved at the middle point, x^{mid} , of the interval $X = [x^L, x^U]$, i.e.,

$$x_i^{\text{mid}} = \frac{x_i^U + x_i^L}{2}, \quad i = 1, 2, \dots, n$$

Since $\Phi(x; \gamma)$ is a convex function, it achieves its minimum at x^{mid} . The value of $\Phi(x; \gamma)$ at its minimum is

$$\min_{x \in X} \Phi(x; \gamma) = \Phi(x^{\text{mid}}; \gamma) = - \sum_{i=1}^n (1 - e^{\frac{1}{2} \gamma_i (x_i^U - x_i^L)})^2$$

Also, since $\Phi(x; \gamma) \leq 0$ for all $x \in X$, we can derive that $\Phi(x; \gamma)$ achieves its maximum at all the corner points of X , that is,

$$\max_{x \in X} \Phi(x; \gamma) = \Phi(x^C; \gamma) = 0. \quad \square$$

PROPERTY R5. The i th diagonal element of $\nabla^2 \Phi(x; \gamma)$ is a convex function and achieves its minimum at the middle point and its maximum at the end points of $[x_i^L, x_i^U]$.

Proof. Consider the i -th diagonal element of $\nabla^2 \Phi(x; \gamma)$

$$\Phi_{x_i x_i}(x; \gamma) = \frac{\partial^2 \Phi(x; \gamma)}{\partial x_i^2} = \gamma_i^2 e^{\gamma_i (x_i - x_i^L)} + \gamma_i^2 e^{\gamma_i (x_i^U - x_i)}.$$

Since the second derivative of $\Phi_{x_i x_i}(x; \gamma)$ is non-negative, that is,

$$\frac{d^2}{dx_i^2} \Phi_{x_i x_i} = \gamma_i^4 e^{\gamma_i (x_i - x_i^L)} + \gamma_i^4 e^{\gamma_i (x_i^U - x_i)} \geq 0, \quad \forall x \in [x_i^L, x_i^U],$$

$\Phi_{x_i x_i}(x; \gamma)$ is convex function. Also, the solution of the equation

$$\frac{d}{dx_i} \Phi_{x_i x_i} = \gamma_i^3 e^{\gamma_i (x_i - x_i^L)} - \gamma_i^3 e^{\gamma_i (x_i^U - x_i)} = 0,$$

is the middle point, x_i^{mid} , of $[x_i^L, x_i^U]$. Hence $\Phi_{x_i x_i}(x; \gamma)$ achieves its minimum at x_i^{mid} , that is,

$$\min_{x_i \in [x_i^L, x_i^U]} \Phi_{x_i x_i}(x; \gamma) = \frac{\partial^2 \Phi(x_i^{\text{mid}}; \gamma)}{\partial x_i^2} = 2\gamma_i^2 e^{\frac{1}{2} \gamma_i (x_i^U - x_i^L)} \tag{9}$$

and its maximum at the two end points of $[x_i^L, x_i^U]$, that is,

$$\begin{aligned} \max_{x_i \in [x_i^L, x_i^U]} \Phi_{x_i x_i}(x; \gamma) &= \frac{\partial^2 \Phi(x_i^L; \gamma)}{\partial x_i^2} = \frac{\partial^2 \Phi(x_i^U; \gamma)}{\partial x_i^2} \\ &= \gamma_i^2 + \gamma_i^2 e^{\gamma_i (x_i^U - x_i^L)}. \end{aligned} \tag{10}$$

□

4. The New Underestimating Function

The new underestimating function is formed by adding $\Phi(x; \gamma)$ to the nonconvex function, that is,

$$L_1(x; \gamma) = f(x) + \Phi(x; \gamma) \tag{11}$$

The Hessian of L_1 is

$$\nabla^2 L_1(x; \gamma) = \nabla^2 f(x) + \nabla^2 \Phi(x; \gamma)$$

The function $L_1(x; \gamma)$ has a number of properties, which derive from the properties of $\Phi(x; \gamma)$. These properties are as follows:

PROPERTY U1. $L_1(x; \gamma)$ is an underestimator of $f(x)$

Proof. This is a consequence of Property R1. Since $\Phi(x; \gamma) \leq 0$ for all $x \in X$ and $L_1(x; \gamma) = f(x) + \Phi(x; \gamma)$ we have $L_1(x; \gamma) \leq f(x)$ for every $x \in X$. \square

PROPERTY U2. $L_1(x; \gamma)$ matches $f(x)$ at all the corner points of X

Proof. This property is a direct result of Property R2. \square

PROPERTY U3. The maximum separation distance between the nonconvex function $f(x)$ and its underestimator $L_1(x; \gamma)$ is bounded.

Proof. This property is a direct result of Property R4. The separation distance between $f(x)$ and $L_1(x; \gamma)$ is

$$d_1(x; \gamma) = f(x) - L_1(x; \gamma) = \sum_{i=1}^n (1 - e^{\gamma_i(x_i - x_i^L)})(1 - e^{\gamma_i(x_i^U - x_i)})$$

The maximum of $d_1(x; \gamma)$ is achieved at the middle point, x^{mid} , of X and is equal to

$$d_1^{\max} = \max_{x^L \leq x \leq x^U} d_1(x; \gamma) = \sum_{i=1}^n (1 - e^{\frac{1}{2}\gamma_i(x_i^U - x_i^L)})^2 \quad (12)$$

Since X is a bounded interval, d_1^{\max} is also bounded. \square

PROPERTY U4. Let $X = [x^L, x^U]$, $Y = [y^L, y^U]$, and $X \subseteq Y \subseteq \mathfrak{R}^n$. Also, let

$$L_1^X(x; \gamma) = f(x) - \sum_{i=1}^n (1 - e^{\gamma_i(x_i - x_i^L)})(1 - e^{\gamma_i(x_i^U - x_i)})$$

and

$$L_1^Y(x; \gamma) = f(x) - \sum_{i=1}^n (1 - e^{\gamma_i(x_i - y_i^L)})(1 - e^{\gamma_i(y_i^U - x_i)})$$

Then the underestimator $L_1^X(x; \gamma)$, is tighter than the underestimator $L_1^Y(x; \gamma)$, i.e.,

$$L_1^X(x; \gamma) \geq L_1^Y(x; \gamma), \quad \forall x \in X \quad (13)$$

Proof. Since $X \subseteq Y$, for every $i = 1, 2, \dots, n$ we have

$$x_i^U \leq y_i^U \quad \text{and} \quad x_i^L \geq y_i^L$$

Taking into consideration that $\gamma_i \geq 0, i=1,2,\dots,n$, and that the exponential function is monotonically increasing, the above two inequalities yield

$$-e^{\gamma_i(x_i^U - x_i)} \geq -e^{\gamma_i(y_i^U - x_i)} \quad \text{and} \quad -e^{\gamma_i(x_i - x_i^L)} \geq -e^{\gamma_i(x_i - y_i^L)} \tag{14}$$

for all $x \in X$. Also, from (8) and (14) we have

$$0 \geq 1 - e^{\gamma_i(x_i^U - x_i)} \geq 1 - e^{\gamma_i(y_i^U - x_i)}, \quad \forall x \in X$$

and

$$0 \geq 1 - e^{\gamma_i(x_i - x_i^L)} \geq 1 - e^{\gamma_i(x_i - y_i^L)}, \quad \forall x \in X.$$

Multiplying the above two inequalities yields

$$-(1 - e^{\gamma_i(x_i - x_i^L)})(1 - e^{\gamma_i(x_i^U - x_i)}) \geq -(1 - e^{\gamma_i(x_i - y_i^L)})(1 - e^{\gamma_i(x_i^U - y_i)}),$$

for all $x \in X$. From the above inequality we can easily derive that (13) holds. \square

Since the function $\Phi(x; \gamma)$ is convex for every $x \in X$ and $\gamma \geq 0$, all non-convexities in the original function $f(x)$ can be eliminated, provided that the parameters γ_i have the appropriate values. We have developed a systematic procedure that determines values for all the parameters γ_i that not only guarantee the convexity of the underestimating function $L_1(x; \gamma)$ but also ensures that $L_1(x; \gamma)$ is tighter than the underestimating function $L_{\alpha\text{BB}}(x; \alpha)$.

The initial value for each γ_i is selected by solving the system of non-linear equations

$$\ell_i + \gamma_i^2 + \gamma_i^2 e^{\gamma_i(x_i^U - x_i^L)} = 0, \quad i = 1, 2, \dots, n \tag{15}$$

where $\ell_i \leq 0, i = 1, 2, \dots, n$. The parameters ℓ_i are the means used to convey second order characteristics of the original nonconvex function into the construction process of the underestimator. Candidate values for these parameters can be provided by the scaled Gerschgorin method (Adjiman et al. (1998b)), that is

$$\ell_i = \max \left\{ 0, -\left(\underline{f}_{ii} - \sum_{j \neq i} \max \{ |\underline{f}_{ij}|, |\bar{f}_{ij}| \} \frac{d_j}{d_i} \right) \right\}, \quad i = 1, 2, \dots, n \tag{16}$$

where $\underline{f}_{ij}, \bar{f}_{ij}$ and d_i are the same quantities as in (5). From (5) and (16) we can derive that

$$\ell_i = \frac{1}{2} \alpha_i, \quad i = 1, 2, \dots, n \tag{17}$$

Note that the term $\gamma_i^2 + \gamma_i^2 e^{\gamma_i(x_i^U - x_i^L)}$ in (15) represent the maximum value of the second derivative of the i th diagonal element of $\nabla^2 \Phi(x; \gamma)$, defined in (10). Note also that if $\ell_i = 0$, then from (15) we can deduce that the corresponding $\gamma_i = 0$ too.

Theorem 4.2 demonstrates the existence of a set of negative numbers ℓ_i , $i=1,2,\dots,n$ such that, if $\gamma_i, i=1,2,\dots,n$ is the corresponding solution of the system of equations (15), then $L_1(x; \gamma)$ is a convex function. First we prove the following lemma that is used in Theorem 4.2.

LEMMA 4.1. *The function $\ell_i: [0, \infty) \rightarrow \Re$ defined by $\ell_i(\gamma_i) = -(\gamma_i^2 + \gamma_i^2 e^{\gamma_i(x_i^U - x_i^L)})$ is monotonically decreasing and*

$$\lim_{\gamma_i \rightarrow \infty} \ell_i(\gamma_i) = -\infty \quad (18)$$

Proof. Let $\gamma_i', \gamma_i'' \in [0, \infty)$ and $\gamma_i' \leq \gamma_i''$. We have $(\gamma_i')^2 \leq (\gamma_i'')^2$ and $e^{\gamma_i'} \leq e^{\gamma_i''}$. From the two previous inequalities we obtain

$$(\gamma_i')^2 + (\gamma_i')^2 e^{\gamma_i'(x_i^U - x_i^L)} \leq (\gamma_i'')^2 + (\gamma_i'')^2 e^{\gamma_i''(x_i^U - x_i^L)}$$

Multiplying the above inequality by -1 yields

$$\ell_i(\gamma_i') \geq \ell_i(\gamma_i''),$$

Hence, $\ell_i(\gamma_i)$ is a monotonically decreasing function. Equation (18) can be derived by considering that $\ell_i(\gamma_i)$ is monotonically decreasing, always takes negative values, and its domain, $[0, \infty)$, is unbounded from above. \square

THEOREM 4.2. *There exists a vector $\ell = (\ell_1, \ell_2, \dots, \ell_n)^T$, such that, if $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)^T$ is the solution of the system (15), then $L_1(x; \gamma)$ is a convex function.*

Proof. Since $f(x) \in C^2$, all its second partial derivatives are continuous and hence bounded in the interval $X = [x^L, x^U]$. Hence, there exist large numbers $M_{ij} > 0$, $i, j = 1, 2, \dots, n$, such that

$$-M_{ij} \leq \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \leq M_{ij}$$

Also from (9), (10) and (18) we can derive that $\partial^2 \Phi(x; \gamma) / \partial x_i^2$ takes arbitrarily large values in $X_i = [x_i^L, x_i^U]$ as γ_i goes to infinity. Consequently, there must exist a number $\ell_i < 0$ such that the corresponding γ_i that solves the i th equation of the system (15) is sufficient to make the convex function $\partial^2 \Phi(x; \gamma) / \partial x_i \partial x_j$ the dominant term in the i th row of the matrix $\nabla^2 L_1(x; \gamma)$. \square

The following theorem examines the relationship between $L_1(x; \gamma)$ and the underestimator obtained by α BB.

THEOREM 4.3. *Let γ be the vector containing the solutions of the system (15), with ℓ_i defined by (16), and α be the vector of the α BB parameters, defined*

by (5). Then, the underestimating function $L_1(x; \gamma)$ is always tighter than the underestimating function $L_{\alpha\text{BB}}(x; \alpha)$. That is,

$$L_1(x; \gamma) \geq L_{\alpha\text{BB}}(x; \alpha), \quad \forall x \in X$$

Proof. It suffices to show that the distance between $L_1(x; \gamma)$ and $L_{\alpha\text{BB}}(x; \alpha)$ is always non-negative, that is,

$$d(x) = L_1(x; \gamma) - L_{\alpha\text{BB}}(x; \alpha) \geq 0, \quad \forall x \in X.$$

The function $d(x)$ that expresses the distance between $L_1(x; \gamma)$ and $L_{\alpha\text{BB}}(x; \alpha)$ can be written as

$$d(x) = \sum_{i=1}^n \alpha_i (x_i - x_i^L) (x_i^U - x_i) - \sum_{i=1}^n (1 - e^{\gamma_i (x_i - x_i^L)}) (1 - e^{\gamma_i (x_i^U - x_i)})$$

The Hessian of $d(x)$ is a diagonal matrix defined as

$$\nabla^2 d(x) = \text{diag} \{ -2\alpha_i + \gamma_i^2 e^{\gamma_i (x_i - x_i^L)} + \gamma_i^2 e^{\gamma_i (x_i^U - x_i)} : i = 1, 2, \dots, n \}$$

Substituting (17) into the i th element of $\nabla^2 d(x)$, and using (10) and (15) we obtain

$$\ell_i + \gamma_i^2 e^{\gamma_i (x_i - x_i^L)} + \gamma_i^2 e^{\gamma_i (x_i^U - x_i)} \leq \ell_i + \gamma_i^2 + \gamma_i^2 e^{\gamma_i (x_i^U - x_i^L)} = 0$$

From the above inequality we can derive that each element of the diagonal matrix $\nabla^2 d(x)$ is non-positive. Therefore $\nabla^2 d(x)$ is negative semi-definite and $d(x)$ is a concave function. Also, evaluating $d(x)$ at any corner point, x^C , of the interval X we have $d(x^C) = 0$. Since $d(x)$ is concave in X and becomes zero at the corner points of X we can conclude that $d(x) \geq 0$ for all $x \in X$. \square

Remark 4.4. The above theorem justifies the use of system (15) in the calculation of the initial values of the parameters γ_i , $i = 1, 2, \dots, n$. The solution of system (15) guarantees that the new underestimator $L_1(x; \gamma)$ is always tighter than the underestimator $L_{\alpha\text{BB}}(x; \alpha)$ of the classical αBB method. Note, however, that the calculated γ_i , $i = 1, 2, \dots, n$ do not guarantee the convexity of the new underestimator. As it will be described in the next section, the convexity of $L_1(x; \gamma)$ will be achieved by an iterative scheme.

5. Convexity of the New Underestimator

In this section we present an iterative scheme that determines whether the underestimating function $L_1(x; \gamma)$ is convex. The scheme is based on interval analysis and consecutive partitions of the domain X . Before we describe the scheme we present two interesting results regarding the relationship between the maximum separation distances among $f(x)$ and its two underestimators, $L_1(x; \gamma)$ and $L_{\alpha\text{BB}}(x; \alpha)$, as well as the parameters γ and α .

THEOREM 5.1. Let $\underline{\gamma} = (\underline{\gamma}_1, \underline{\gamma}_2, \dots, \underline{\gamma}_n)^T$ be the solution of system (15), with ℓ_i defined by (16). Then, the two underestimators $L_1(x; \underline{\gamma})$ and $L_{\alpha\text{BB}}(x; \underline{\alpha})$, where

$$\underline{\alpha} = \left(\frac{4(1 - e^{0.5\underline{\gamma}_1(x_1^U - x_1^L)})^2}{(x_1^U - x_1^L)^2}, \dots, \frac{4(1 - e^{0.5\underline{\gamma}_n(x_n^U - x_n^L)})^2}{(x_n^U - x_n^L)^2} \right)^T \quad (19)$$

have the same maximum separation distance from $f(x)$.

Proof. Substituting $\alpha = \underline{\alpha}$ in (7) and using (12) we obtain

$$\begin{aligned} d_{\alpha\text{BB}}^{\max}(\underline{\alpha}) &= \frac{1}{4} \sum_{i=1}^n \frac{4(1 - e^{0.5\underline{\gamma}_i(x_i^U - x_i^L)})^2}{(x_i^U - x_i^L)^2} (x_i^U - x_i^L)^2 \\ &= \sum_{i=1}^n (1 - e^{0.5\underline{\gamma}_i(x_i^U - x_i^L)/2})^2 \\ &= d_1^{\max}(\underline{\gamma}). \quad \square \end{aligned}$$

THEOREM 5.2. Let $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_n)^T$ be the values of the α parameters as computed by (5). Then, the two underestimators $L_1(x; \bar{\gamma})$ and $L_{\alpha\text{BB}}(x; \bar{\alpha})$, where

$$\bar{\gamma} = \left(\frac{2 \log(1 + \sqrt{\bar{\alpha}_1}(x_1^U - x_1^L)/2)}{x_1^U - x_1^L}, \dots, \frac{2 \log(1 + \sqrt{\bar{\alpha}_n}(x_n^U - x_n^L)/2)}{x_n^U - x_n^L} \right)^T \quad (20)$$

have the same maximum separation distance from $f(x)$.

Proof. Substituting $\gamma = \bar{\gamma}$ in (12) we obtain

$$\begin{aligned} d_1^{\max}(\bar{\gamma}) &= \sum_{i=1}^n \left(1 - e^{\frac{1}{2} \frac{2 \log(1 + \sqrt{\bar{\alpha}_i}(x_i^U - x_i^L)/2)}{x_i^U - x_i^L} (x_i^U - x_i^L)} \right)^2 \\ &= \sum_{i=1}^n (1 - e^{\log(1 + \sqrt{\bar{\alpha}_i}(x_i^U - x_i^L)/2)})^2 \\ &= \sum_{i=1}^n \left(1 - \left(1 + \frac{1}{2} \sqrt{\bar{\alpha}_i}(x_i^U - x_i^L) \right) \right)^2 \\ &= d_{\alpha\text{BB}}^{\max}(\bar{\alpha}). \quad \square \end{aligned}$$

The main result of the above two theorems is that, for any $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ there exists an $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, such that the underestimators $L_1(x; \gamma)$ and $L_{\alpha\text{BB}}(x; \alpha)$ have the same maximum separation distance from the nonconvex function $f(x)$. From all these pairs of underestimators, the only one that is known to be convex *a priori* is $L_{\alpha\text{BB}}(x; \bar{\alpha})$, since this is the one resulting from the classical αBB method. However, as it will be apparent from several examples in the next section, for most arbitrarily nonconvex functions the underestimators $L_{\alpha\text{BB}}(x; \alpha)$ and $L_1(x; \gamma)$ are convex within a large portion of the intervals $[\underline{\alpha}, \bar{\alpha}]$ and $[\underline{\gamma}, \bar{\gamma}]$

respectively. Based on the above observations, it is natural to search for a vector γ in the interval $[\underline{\gamma}, \bar{\gamma}]$ or for a vector α in the interval $[\underline{\alpha}, \bar{\alpha}]$, so that at least one of the underestimators $L_1(x; \gamma)$, $L_{\alpha\text{BB}}(x; \alpha)$ is convex.

5.1. ALGORITHM FOR CONVEXITY VERIFICATION

The main aim of this algorithm is to determine appropriate values for the γ parameters so that the corresponding underestimator is both convex function and tighter than the underestimator used by the classical αBB method. This is very important issue, since the tighter the underestimator is, the better the quality of the lower bound on the global optimum is. Also, the size of the enumeration tree becomes smaller, since a larger number of nodes can be fathomed. Therefore, the overall efficiency of the algorithm is enhanced, because the global optimum of the problem is reached by considering fewer nodes.

The algorithm we have developed searches for a vector $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ so that the corresponding $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, produces an underestimating function $L_{\alpha\text{BB}}(x; \alpha)$ that is convex. The search starts by setting $\gamma = \underline{\gamma}$ and $\alpha = \underline{\alpha}$ and then checking whether $L_{\alpha\text{BB}}(x; \underline{\alpha})$ is convex. This is done by using the scaled Gerschgorin method to determine lower bounds on the eigenvalues of the Hessian matrix $\nabla^2 L_{\alpha\text{BB}}(x; \underline{\alpha})$. For those lower bounds that are negative, we bisect the intervals of the corresponding variables, thereby generating a number of sub-domains that are stored in a list, denoted by Λ_1 . Then the algorithm checks whether $\nabla^2 L_{\alpha\text{BB}}(x; \underline{\alpha})$ is positive semi-definite in each of those sub-domains using again the scaled Gerschgorin method. If the size of the list, Λ_1 , exceeds a certain number of nodes then $\nabla^2 L_{\alpha\text{BB}}(x; \underline{\alpha})$ is most likely *not* positive semi-definite, and the values of all γ_i 's are increased by a prespecified positive quantity, $\eta > 0$, and the corresponding values of the new α_i 's are calculated. The algorithm now tries to verify whether $\nabla^2 L_{\alpha\text{BB}}(x; \alpha)$, with the new increased α , is positive semi-definite. It continues in this manner until the list Λ_1 becomes empty. In that case the corresponding α makes the Hessian matrix, $\nabla^2 L_{\alpha\text{BB}}(x; \alpha)$, positive semi-definite for all $x \in X$ and consequently $L_{\alpha\text{BB}}(x; \alpha)$ a convex underestimator. The main reason for using the underestimator $L_{\alpha\text{BB}}(x; \alpha)$ instead of the underestimator $L_1(x; \gamma)$ is that, it is easier to verify the positive definiteness of the matrix $\nabla^2 L_{\alpha\text{BB}}(x; \alpha)$ than that of the matrix $\nabla^2 L_1(x; \gamma)$. The algorithm in detail is as follows:

ALGORITHM 1. Verification of Convexity

- STEP 1: (*Initialization*) Set $K=1, J=1, J_{\max}=2^n+1, \eta=1.1, X_J=X, \Lambda_1=\{X_J\}$ and $\gamma_{i,K}=\underline{\gamma}_i$.
- STEP 2: Use (19) to calculate the $\alpha_{i,K}, i=1,2,\dots,n$ that correspond to the $\gamma_{i,K}, i=1,2,\dots,n$, and form the underestimator $L_{\alpha\text{BB}}(x; \alpha_K)$.
- STEP 3: If the maximum separation distance of $L_{\alpha\text{BB}}(x; \alpha_K)$ from $f(x)$ is less than the maximum separation distance of $L_{\alpha\text{BB}}(x; \bar{\alpha})$ from $f(x)$ then GoTo Step 4.

Otherwise, adopt as underestimator the classical α BB underestimator, $L_{\alpha\text{BB}}(x; \bar{\alpha})$, and STOP.

STEP 4: Check whether $L_{\alpha\text{BB}}(x; \alpha_K)$ is convex:

Repeat

STEP 4.1: Remove the last element from the list Λ_1 of unexplored sub-domains. Let us name that sub-domain X_{last} .

STEP 4.2: Form the interval Hessian $[\nabla^2 L_{\alpha\text{BB}}(x; \alpha_K)]$ with $x \in X_{\text{last}}$.

STEP 4.3: Use (16) to find lower bounds on each eigenvalue of the interval Hessian $[\nabla^2 L_{\alpha\text{BB}}(x; \alpha_K)]$ in X_{last} .

STEP 4.4: Form the set $I_- = \{i : \ell_i < 0\}$.

STEP 4.5: If $I_- \neq \emptyset$, bisect all intervals $[x_{i,\text{last}}^L, x_{i,\text{last}}^U]$ with $i \in I_-$, and add them at the end of the list Λ_1 .

STEP 4.6: Set $J = J + 2^{|I_-|} - 1$, where $|I_-|$ represents the cardinality of the set I_- (i.e., a total of $2^{|I_-|}$ new sub-domains have been generated and added to the list and one node have been removed).

Until ($\Lambda_1 = \emptyset$ or $J = J_{\text{max}}$)

STEP 5: If $\Lambda_1 = \emptyset$ then STOP. The Hessian $\nabla^2 L_{\alpha\text{BB}}(x; \alpha_K)$ is positive semi-definite for all $x \in X$ and $L_{\alpha\text{BB}}(x; \alpha_K)$ is a convex underestimator. Also the underestimator $L_{\alpha\text{BB}}(x; \alpha_K)$ is tighter than the underestimator $L_{\alpha\text{BB}}(x; \bar{\alpha})$ obtained by the classical α BB method.

Otherwise, increase the values of all $\gamma_{i,K}, i = 1, 2, \dots, n$ by setting $\gamma_{i,K+1} = \eta \gamma_{i,K}$. Set $K = K + 1$ and GoTo Step 2.

Termination of Algorithm 1 is guaranteed by the fact that $L_{\alpha\text{BB}}(x; \bar{\alpha})$ is known *a priori* to be convex underestimator.

6. Examples

In this section we present several examples which demonstrate that for arbitrarily nonconvex functions the method can produce significantly tighter underestimators than those produced by α BB.

It should be mentioned that the new underestimators require more computational effort than the ones used by the classical α BB method. This is because the new underestimators require the solution of the system of nonlinear equations (15), in order to determine the appropriate values of the γ parameters. A detailed computational comparison between the new underestimators and the ones used by the classical α BB method, is provided in Akrotirianakis and Floudas (2004). From that computational study we have concluded that the new underestimators usually perform better than the classical α BB method, both in terms of the overall CPU time and the number of nodes generated by the enumeration tree. From that computational study we can also derive that the new underestimators

perform better when the problem involves many arbitrarily nonconvex terms in the objective or constraints.

EXAMPLE 1. In this example we examine an arbitrarily nonconvex function that describes the molecular conformation of pseudoethane. It is taken from (Floudas et al. (1999) and Maranas and Floudas (1994a)) where the global minimum potential energy conformation of small molecules is studied. The Lennard-Jones potential is expressed in terms of a simple dihedral angle. The potential energy of the molecule is given by the function

$$\begin{aligned}
 f_1(x) = & \frac{588600}{(3r_0^2 - 4\cos(\theta)r_0^2 - 2(\sin^2(\theta)\cos(x - \frac{2\pi}{3}) - \cos^2(\theta))r_0^2)^6} - \\
 & \frac{1079.1}{(3r_0^2 - 4\cos(\theta)r_0^2 - 2(\sin^2(\theta)\cos(x - \frac{2\pi}{3}) - \cos^2(\theta))r_0^2)^3} + \\
 & + \frac{600800}{(3r_0^2 - 4\cos(\theta)r_0^2 - 2(\sin^2(\theta)\cos(x) - \cos^2(\theta))r_0^2)^6} - \\
 & - \frac{1071.5}{(3r_0^2 - 4\cos(\theta)r_0^2 - 2(\sin^2(\theta)\cos(x) - \cos^2(\theta))r_0^2)^3} + \\
 & + \frac{481300}{(3r_0^2 - 4\cos(\theta)r_0^2 - 2(\sin^2(\theta + \frac{2\pi}{3})\cos(x) - \cos^2(\theta))r_0^2)^6} - \\
 & - \frac{1064.6}{(3r_0^2 - 4\cos(\theta)r_0^2 - 2(\sin^2(\theta + \frac{2\pi}{3})\cos(x) - \cos^2(\theta))r_0^2)^3}
 \end{aligned}$$

where r_0 is the covalent bond length ($r_0 = 1.54A$), θ is the covalent bond angle ($\theta = 109.5^\circ$) and x is the dihedral angle ($x \in X = [0, 2\pi]$). Figure 1 shows the graph of $f_1(x)$.

The value of the α parameter computed by the classical α BB method using (5) is

$$\bar{\alpha} = 77.124$$

and the corresponding value for the γ parameter, obtained by (20), is

$$\bar{\gamma} = 1.0673$$

Also by solving (15) for γ we obtain

$$\underline{\gamma} = 0.8521$$

and the corresponding value for the α parameter, obtained by (19), is

$$\underline{\alpha} = 18.579$$

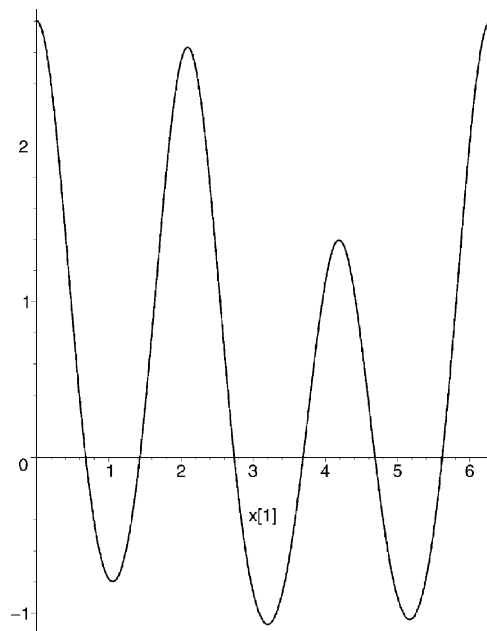


Figure 1. The potential energy function $f_1(x)$.

Algorithm 1 checks whether there exist values of $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ and $\alpha \in [\underline{\alpha}, \bar{\alpha}]$ such that the underestimator $L_{\alpha\text{BB}}(x; \alpha)$ is convex. After 16 iterations it concludes that if $\alpha = \underline{\alpha}$, then $L_{\alpha\text{BB}}(x; \alpha)$ is a convex underestimator of $f_1(x)$. Furthermore, if $\gamma = \bar{\gamma}$ then $L_1(x; \gamma)$ is a convex underestimator of $f_1(x)$ too. Note that the value of γ as well as the value of α did not increase at all.

The minima of the two underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ are

$$\min_{x \in X} L_{\alpha\text{BB}}(x; \bar{\alpha}) = -762.2377 \quad \text{and} \quad \min_{x \in X} L_{\alpha\text{BB}}(x; \alpha) = -184.4244$$

Figure 2 compares the two underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ and shows the improvement. The relative improvement of the new underestimator $L_{\alpha\text{BB}}(x; \alpha)$ compared to the original αBB underestimator, $L_{\alpha\text{BB}}(x; \bar{\alpha})$, is equal to

$$r = 1 - \frac{\min_{x \in X} L_{\alpha\text{BB}}(x; \alpha)}{\min_{x \in X} L_{\alpha\text{BB}}(x; \bar{\alpha})} = 0.758$$

EXAMPLE 2. This example is taken from (Adjiman et al. (1998b)) and examines a two dimensional nonconvex function defined as

$$f_2(x) = \cos(x_1) \sin(x_2) - \frac{x}{y^2 + 1}$$

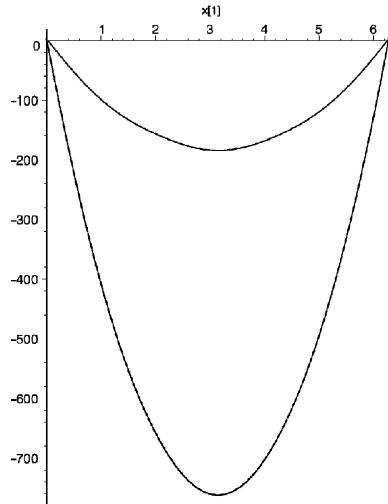


Figure 2. Comparison of the underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ of the nonconvex function $f_1(x)$.

where $x_1 \in [-1, 2]$ and $x_2 \in [1, 1]$. The above function possesses three minima and its graph is shown in Figure 3. The values of the α parameters computed by the classical αBB method using (5) are

$$\bar{\alpha}_1 = 1.921 \quad \text{and} \quad \bar{\alpha}_2 = 10.921$$

Using (20) we can determine the corresponding value for the γ parameters

$$\bar{\gamma}_1 = 0.75 \quad \text{and} \quad \bar{\gamma}_2 = 1.46$$

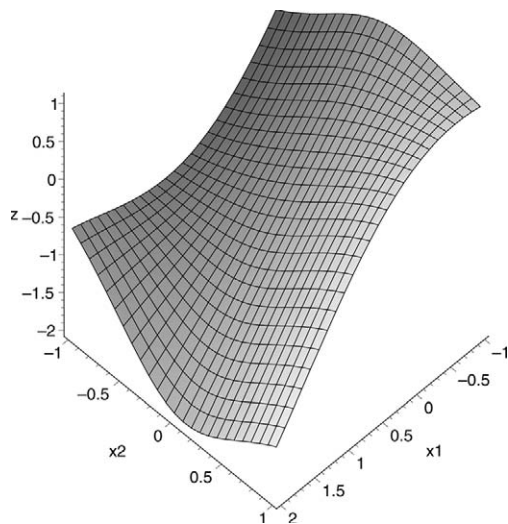


Figure 3. The graph of $f_2(x)$.

Also by solving (15) for γ_i , $i=1,2$, we obtain

$$\underline{\gamma}_1=0.672 \quad \text{and} \quad \underline{\gamma}_2=1.267$$

Using (19) we can determine the corresponding value for the $\underline{\alpha}_i$, $i=1,2$,

$$\underline{\alpha}_1=1.3456 \quad \text{and} \quad \underline{\alpha}_2=6.5$$

Algorithm 1 checks whether there exist values of $\gamma_i \in [\underline{\gamma}_i, \bar{\gamma}_i]$, $i=1,2$ and $\alpha_i \in [\underline{\alpha}_i, \bar{\alpha}_i]$, $i=1,2$, such that the underestimator $L_{\alpha\text{BB}}(x; \alpha)$ is convex. After 8 iterations it concludes that if $\gamma=(0.74, \underline{\gamma}_2)$ and $\alpha=(1.8325, \underline{\alpha}_2)$, then $L_{\alpha\text{BB}}(x; \alpha)$ is a convex underestimator of $f(x)$. Also, if $\gamma=(0.74, \underline{\gamma}_2)$ then $L_1(x; \gamma)$ is a convex underestimator of $f_2(x)$. Note that only the value of γ_1 needed to increase by 10% from its original value.

The minima of the two underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ are

$$\min L_{\alpha\text{BB}}(x; \bar{\alpha}) = -15.88469 \quad \text{and} \quad \min L_{\alpha\text{BB}}(x; \alpha) = -10.22767.$$

Figure 4 compares the two underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ and shows the improvement we get if we use the α_i 's calculated by Algorithm 1.

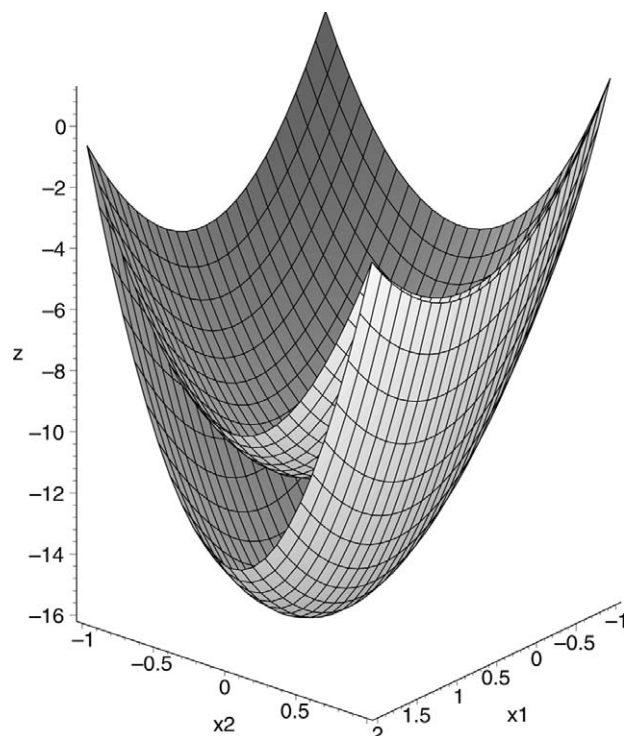


Figure 4. Comparison of the underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ of the nonconvex function $f_2(x)$.

The relative improvement of the new underestimator $L_{\alpha\text{BB}}(x; \alpha)$ compared to the original αBB underestimator, $L_{\alpha\text{BB}}(x; \bar{\alpha})$, is equal to

$$r = 1 - \frac{\min_{x \in X} L_{\alpha\text{BB}}(x; \alpha)}{\min_{x \in X} L_{\alpha\text{BB}}(x; \bar{\alpha})} = 0.3561.$$

EXAMPLE 3. This example is taken from (Floudas et al. (1999)) and (Dixon and Szego (1975)) and examines a two dimensional nonconvex function defined as

$$f_3(x) = -\frac{1}{(x_1-4)^2+(x_2-4)^2+0.1} - \frac{1}{(x_1-1)^2+(x_2-1)^2+0.2} - \frac{1}{(x_1-8)^2+(x_2-8)^2+0.2}$$

where $x_1 \in [0, 10]$ and $x_2 \in [0, 10]$. The above function possesses three minima and its graph is shown in Figure 5. The values of the α parameters computed by the classical αBB method using (5) are

$$\bar{\alpha}_1 = 433000 \quad \text{and} \quad \bar{\alpha}_2 = 433000$$

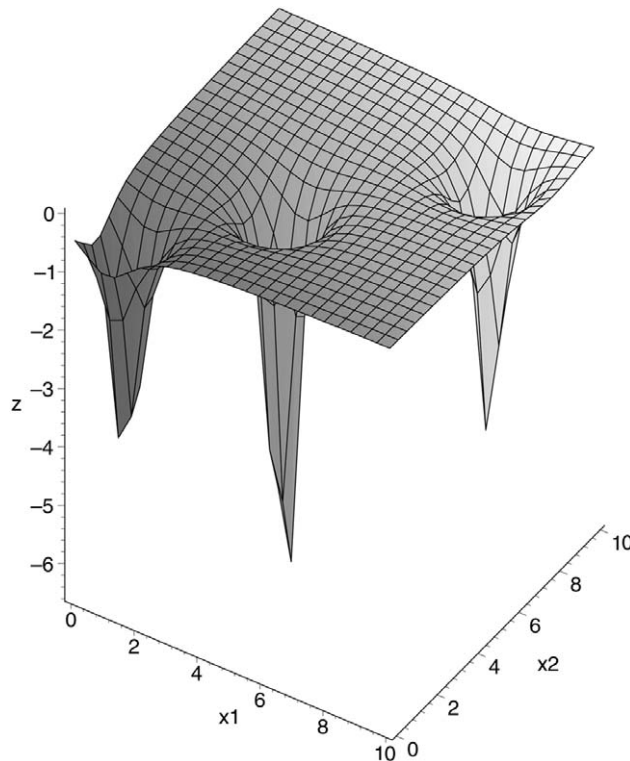


Figure 5. The graph of $f_3(x)$.

Substituting the above values into (20) we can determine the corresponding value for the γ parameters

$$\bar{\gamma}_1 = 1.6198 \quad \text{and} \quad \bar{\gamma}_2 = 1.6198$$

Also, by solving (15) for γ_i , $i = 1, 2$, we obtain

$$\underline{\gamma}_1 = 1.3127 \quad \text{and} \quad \underline{\gamma}_2 = 1.3127$$

and using (19) we can determine the corresponding value for the $\underline{\alpha}_i$, $i = 1, 2$,

$$\underline{\alpha}_1 = 20044.4 \quad \text{and} \quad \underline{\alpha}_2 = 20044.4$$

Algorithm 1 checks whether there exist values of $\gamma_i \in [\underline{\gamma}_i, \bar{\gamma}_i]$, $i = 1, 2$ and $\alpha_i \in [\underline{\alpha}_i, \bar{\alpha}_i]$, $i = 1, 2$, such that the underestimator $L_{\alpha\text{BB}}(x; \bar{\alpha})$ is convex. After 8 iterations it concludes that if $\gamma = (\underline{\gamma}_1, \underline{\gamma}_2)$ and $\alpha = (\underline{\alpha}_1, \underline{\alpha}_2)$, then $L_{\alpha\text{BB}}(x; \alpha)$ is a convex underestimator of $f(x)$. Also, if $\gamma = \underline{\gamma}$ then $L_1(x; \underline{\gamma})$ is a convex underestimator of $f_3(x)$. Note that none of the γ_i 's or α_i 's needed to increase in order to obtain a convex underestimator.

The minima of the two underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ are

$$\min L_{\alpha\text{BB}}(x; \bar{\alpha}) = -2.165 \times 10^7 \quad \text{and} \quad \min L_{\alpha\text{BB}}(x; \alpha) = -1.002 \times 10^6.$$

Figure 6 compares the two underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ and shows the improvement we get if we use the α_i 's calculated by Algorithm 1. The relative improvement of the new underestimator $L_{\alpha\text{BB}}(x; \alpha)$ compared to the original αBB underestimator, $L_{\alpha\text{BB}}(x; \bar{\alpha})$, is equal to

$$r = 1 - \frac{\min_{x \in X} L_{\alpha\text{BB}}(x; \alpha)}{\min_{x \in X} L_{\alpha\text{BB}}(x; \bar{\alpha})} = 0.9537.$$

EXAMPLE 4. This example is taken from (Goldstein and Price (1971)) and considers the nonconvex function

$$f_4(x) = ((1 + (x_1 + x_2 + 1))^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2)) \times \\ \times (30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2))$$

where $x_1 \in [-2, 2]$ and $x_2 \in [-2, 2]$. The graph of f_4 is shown in Figure 7. The values of the α parameters computed by the classical αBB method using (5) are

$$\bar{\alpha}_1 = 2.2646 \times 10^8 \quad \text{and} \quad \bar{\alpha}_2 = 2.9034 \times 10^7$$

Using (20) we can determine the corresponding value for the γ parameters

$$\bar{\gamma}_1 = 5.1561 \quad \text{and} \quad \bar{\gamma}_2 = 5.2182$$

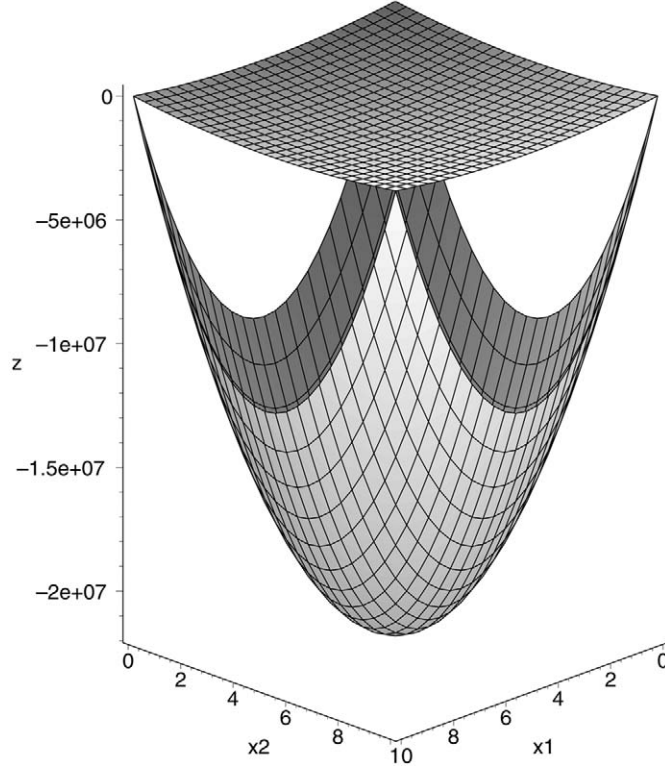


Figure 6. Comparison of the underestimators $L_{\alpha_{BB}}(x; \bar{\alpha})$ and $L_{\alpha_{BB}}(x; \alpha)$ of the nonconvex function $f_3(x)$.

Also by solving (15) for γ_i , $i=1,2$, we obtain

$$\underline{\gamma}_1 = 4.2583 \quad \text{and} \quad \underline{\gamma}_2 = 4.3140$$

Using (19) we can determine the corresponding value for the $\underline{\alpha}_i$, $i=1,2$,

$$\underline{\alpha}_1 = 6.2416 \times 10^6 \quad \text{and} \quad \underline{\alpha}_2 = 7.7976 \times 10^6$$

Algorithm 1 checks whether there exist values of $\gamma_i \in [\underline{\gamma}_i, \bar{\gamma}_i]$, $i=1,2$ and $\alpha_i \in [\underline{\alpha}_i, \bar{\alpha}_i]$, $i=1,2$, such that the underestimator $L_{\alpha_{BB}}(x; \bar{\alpha})$ is convex. After 16 iterations it concludes that if $\gamma = (4.3861, 4.4434)$ and $\alpha = (1.04 \times 10^7, 1.31 \times 10^7)$, then $L_{\alpha_{BB}}(x; \alpha)$ is a convex underestimator of $f(x)$. Furthermore, if $\gamma = (4.3861, 4.4434)$ then $L_1(x; \underline{\gamma})$ is a convex underestimator of $f_4(x)$. Note that both of the γ_i 's have been increased by 5%.

The minima of the two underestimators $L_{\alpha_{BB}}(x; \bar{\alpha})$ and $L_{\alpha_{BB}}(x; \alpha)$ are

$$\min L_{\alpha_{BB}}(x; \bar{\alpha}) = -2.0672 \times 10^9 \quad \text{and} \quad \min L_{\alpha_{BB}}(x; \alpha) = -9.3967 \times 10^7.$$

Figure 8 compares the two underestimators $L_{\alpha_{BB}}(x; \bar{\alpha})$ and $L_{\alpha_{BB}}(x; \alpha)$ and shows the improvement we get if we use the α_i 's calculated by Algorithm 1.

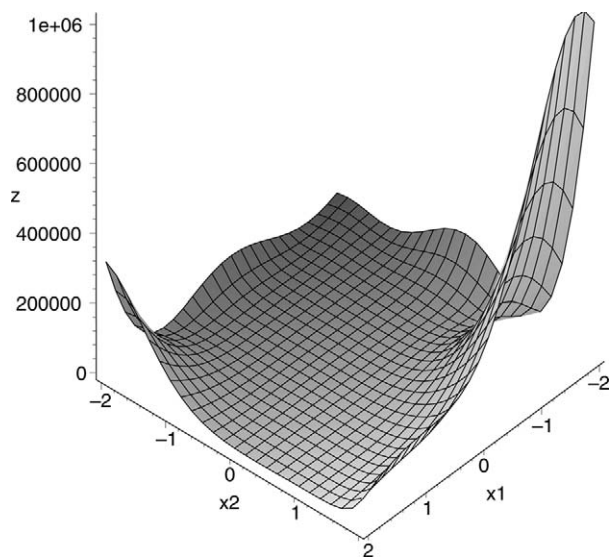


Figure 7. The graph of $f_4(x)$.

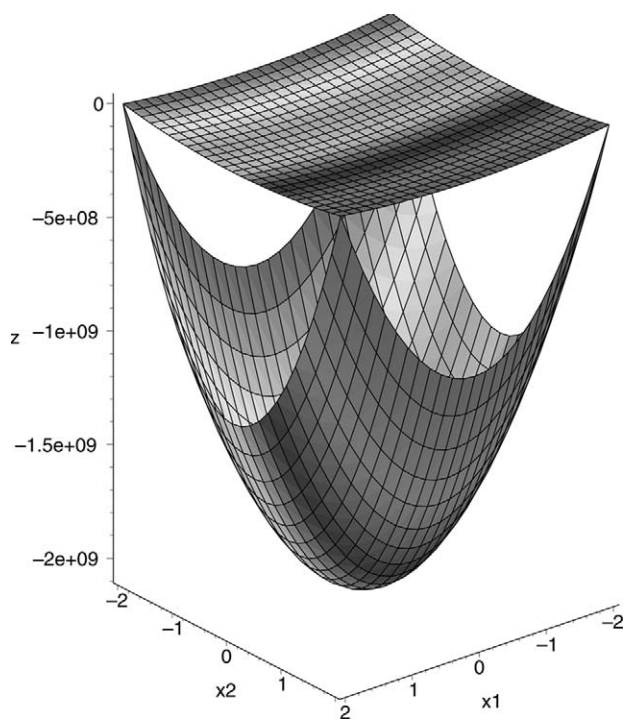


Figure 8. Comparison of the underestimators $L_{\alpha\text{BB}}(x; \bar{\alpha})$ and $L_{\alpha\text{BB}}(x; \alpha)$ of the nonconvex function $f_4(x)$.

The relative improvement of the new underestimator $L_{\alpha\text{BB}}(x; \alpha)$ compared to the original αBB underestimator, $L_{\alpha\text{BB}}(x; \bar{\alpha})$, is equal to

$$r = 1 - \frac{\min_{x \in X} L_{\alpha\text{BB}}(x; \alpha)}{\min_{x \in X} L_{\alpha\text{BB}}(x; \bar{\alpha})} = 0.9545.$$

7. Conclusions

A new method for generating improved convex underestimators for arbitrarily nonconvex functions that are twice continuously differentiable has been presented. A new relaxation function has been introduced and it has been shown that it possesses similar properties to those possessed by the relaxation function used in the αBB method. The underestimators are formed by subtracting the new relaxation function from the original nonconvex function. It is proven that the new underestimators are tighter than the αBB underestimators, because the new relaxation function is always less than the one used in αBB . A rigorous procedure that verifies the convexity of the new improved underestimators has been developed. That procedure is based on interval analysis and partitioning of the feasible region into smaller ones. The ability of the new method to produce tighter underestimators than those used in αBB has been illustrated in several examples. The incorporation of the new underestimators within a Branch-and-Bound framework, algorithmic issues, and computational studies will be reported in a forthcoming paper.

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation and the National Institute of Health (R01 GM52032).

References

- Adjiman, C.S., Androulakis, I. and Floudas, C.A. (1998a), A global optimization method, αBB , for general twice-differentiable constrained NLPs - II. Implementation and computational results, *Computers and Chemical Engineering*, 22, 1159–1179.
- Adjiman, C.S., Dallwig, S., Floudas, C.A. and Neumaier, A. (1998b), A global optimization method, αBB , for general twice-differentiable constrained NLPs - I: Theoretical aspects, *Computers and Chemical Engineering*, 22(9), 1137–1158.
- Akrotirianakis, I.G. and Floudas, C.A. (2004), Computational Experience with a New Class of Convex Underestimators: Box-constrained NLP problems, *Submitted for publication, Journal of Global Optimization*.
- Al-Khayyal, F.H. and Falk, J.E. (1983), Jointly constrained biconvex programming, *Mathematics of Operations Research*, 8, 523.
- Boggs, P.T. and Tolle, J.W. (1995), Sequential quadratic programming, *Acta Numerica*, 4, 1–52.
- Dixon, L.C.W. and Szego, G.P. (1975), Towards global optimization, In: *Proceedings of a Workshop at the University of Cagliari*, Italy: North-Holland.

- Floudas, C.A. (2000), *Deterministic Global Optimization, Theory, Methods and Applications*, Kluwer Academic Publishers.
- Floudas, C.A., Pardalos, P.M., Adjiman C.S., Esposito W.R., Gumus Z.H., Harding S.T., Klepeis J.L., Meyer C.A. and Schweiger, C.A. (1999), *Handbook of Test Problems in Local and Global Optimization*, Dordrecht, The Netherlands, Kluwer Academic Publishers.
- Gelatt, C.D., Kirkpatrick, S. and Vecchi, M.P. (1983), Optimization by simulated annealing, *Science*, 220, 671.
- Goldberg, D.E. (1987), *Genetic Algorithms in Search, Optimization and Machine Learning*, New York, NY, Addison-Welsey.
- Goldstein, A. and Price, J. (1971), On descent from local minima, *Mathematics of Computation*, 25, 569–574.
- Hansen, E. (1992), *Global Optimization using Interval Analysis*, New York, M. Dekker.
- Horst, R. and Tuy, H. (1987), On the convergence of global methods in multiextremal optimization, *Journal of Optimization Theory and Applications*, 54, 283.
- Maranas, C.D. and Floudas, C.A. (1994a), A deterministic global optimization approach for molecular structure determination, *Journal of Chemical Physics*, 100(2), 1247–1261.
- Maranas, C.D. and Floudas, C.A. (1994b), Global minimum potential energy conformations for small molecules, *Journal of Global Optimization*, 4, 135–170.
- Murty, K.G. and Kabadi S.N. (1987), Some NP-complete problems in quadratic and nonlinear programming, *Mathematical Programming*, 39, 117–129.
- Neumaier, A. (1990), *Interval Methods for Systems of Equations*, Cambridge University Press.
- Pardalos, P.M. and Schnitger, G. (1988), Checking local optimality in constrained quadratic programming, *Operations Research Letters*, 7, 33–35.
- Porn, R., Harjunkoski, I. and Westerlund, T. (1999), Convexification of different classes of non-convex MINLP problems, *Computers and Chemical Engineering*, 23, 439–448.
- Rinnoy-Kan, A.H.G. and Timmer, G.T. (1987a), Stochastic global optimization methods. Part I: Clustering methods, *Mathematical Programming*, 39, 27–56.
- Rinnoy-Kan, A.H.G. and Timmer, G.T. (1987b), Stochastic global optimization, Part II: Multi-level Methods, *Mathematical Programming*, 39, 57–78.
- Ryoo, H.S. and Sahinidis, N.V. (1996), A branch-and-reduce approach to global optimization, *Journal of Global Optimization*, 8(2), 107–139.
- Schoen, F. (1991), Stochastic techniques for global optimization: A survey of recent advances, *Journal of Global Optimization*, 1(3), 207–228.
- Sherali, H.D. and Alameddine, A. (1992), A new reformulation linearization technique for bilinear programming problems, *Journal of Global Optimization*, 2(4), 379.
- Smith, E.M.B. and Pantelides C.C. (1996), Global optimization for general process models. In: Grossmann I.E. (ed.), *Global Optimization in Engineering Design*, Kluwer Academic Publishers, pp. 355–386.
- Tuy, H. (1987), Global minimum of the difference of two convex functions, *Mathematical Programming Study*, 30, 150.
- Wright, M.H. (1992), Interior point methods for constrained optimization, *Acta Numerica*, 1, 341–407.