THE ROLE OF THE OFF-DIAGONAL ELEMENTS OF THE HESSIAN MATRIX IN THE CONSTRUCTION OF TIGHT CONVEX UNDERESTIMATORS FOR NONCONVEX FUNCTIONS

I. G. Akrotirianakis, C. A. Meyer and C. A. Floudas
Department of Chemical Engineering, Princeton University
Princeton, NJ 08544

Abstract

In this paper, we propose a new method that produces new forms of tight convex underestimators for twice continuously differentiable nonconvex functions. The algorithm generalizes the main ideas used in aBB. The key idea is to determine a new convexification function that is able to handle the off-diagonal elements of the Hessian matrix of the original nonconvex function. The new convexification function is based on that used in aBB, but it is enhanced with extra convex parametric terms. The Hessian matrix of the new convexification function is a constant non-diagonal matrix. The values of the parameters are determined in such a way that the effect of the off-diagonal elements in the overall underestimating function is minimized. As a result, the new underestimator is tighter than that produced by aBB. We discuss the theoretical properties of the new underestimator and we present several illustrative examples where we demonstrate the improvements of the new convex underestimators over those used in the original aBB method.

Keywords: Global Optimization, convex underestimators, Branch-and-Bound

Introduction

The main objective of this work is to develop convex underestimators of an arbitrarily nonconvex function, $f : X \subset \Re^n \to \Re$, that are tighter than those generated by the aBB method [1,4]. Tight convex underestimators play a very important role in the development of efficient deterministic algorithms for solving global optimization problems [2,3]. 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 aBB. 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 aBB 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 the one produced by the aBB method. Another attractive and distinct property of the relaxation function is that it has a non-diagonal Hessian matrix. This feature allows us to explicitly treat the off-diagonal elements of the Hessian matrix of a nonconvex function and therefore reduce their effect in the construction process of the underestimating function.

Motivation

In aBB a convex underestimator of an arbitrarily twice-continuously differentiable nonconvex function, f(x), is defined as

$$L_{aBB}(x) = f(x) - g(x),$$

where the function $g(x)$ is a parametric and separable function defined as

$$g(x) = \sum_{i=1}^{n} a_i (x_i - x_i^L)(x_i^U - x_i)$$

The values of the parameters $a_i, i = 1,2,...,n,$ are calculated via a variety of methods (see [4] for a complete list of those methods). A very effective method is the scaled Gerschgorin formula defined as $a_i = \max\left\{0, -0.5\left(\underline{f_{ii}} - \sum_{i \neq j} \max\left\{|\underline{f_{ij}}|, |\overline{f_{ij}}|\right\}\frac{d_j}{d_i}\right)\right\}$.

From the above formula it is evident that the contribution of the off-diagonal elements $\underline{f_{ij}}$ and $\overline{f_{ij}}$, of the interval Hessian matrix $\lfloor H_f \rfloor$, can be substantial, giving rise to loose underestimating functions. In the next section we define a new relaxation function that is able to reduce the effect of the off-diagonal elements in the construction process of the underestimating function and thereby produce tighter underestimators.

The new relaxation function

We first define the two complementary index sets $B^i = \left\{j \in N : |\underline{f_{ij}}| \leq |\overline{f_{ij}}|, j \neq i\right\}$ and $\Gamma^i = N - B^i$, where $N = \{1, 2, \ldots n\}$. The new relaxation function is defined as

$$\Phi(x) = \sum_{i=1}^{n-1}\left(\sum_{j \in B^i} \beta_{ij}(x_i - x_j)^2 + \sum_{j \in \Gamma^i} \gamma_{ij}(x_i + x_j)^2)\right).$$

The Hessian matrix, H_Φ, of the relaxation function is a non-diagonal matrix, whose diagonal elements are defined by $\Phi_{ii} = 2\left(\sum_{j \in B^i} \beta_{ij} + \sum_{j \in \Gamma^i} \gamma_{ij}\right)$ and its non-diagonal elements by $\Phi_{ij} = -2\beta_{ij}$, if $j \in B^i$, and $\Phi_{ij} = 2\gamma_{ij}$, if $j \in \Gamma^i$. As we will see in the next section the values of the parameters β_{ij}, γ_{ij} are defined in such a way that the relaxation function is convex.

The new convex underestimating function

The new underestimator is defined in two steps. First the above relaxation function, $\Phi(x)$, is added to the original nonconvex function, f(x), producing an overestimator, that is, $M(x) = f(x) + \Phi(x) > f(x)$. Note that the function $M(x)$ may not be convex and does not match the original function at the corner points of the domain X. Next, a convex underestimator of $M(x)$ is generated by using the aBB methodology. That is, the convex function $L_M(x) = M(x) - \sum_{i=1}^{n} a_i^M (x_i - x_i^L)(x_i^U - x_i)$ is generated, where $a_i^M, i = 1, \ldots n.$ are defined as follows

$$a_i^M = \max\left\{0, -0.5\left(\underline{M_{ii}} - \sum_{j \in B^i \cup \Gamma^i} \max\left\{|\underline{M_{ij}}|, |\overline{M_{ij}}|\right\}\right)\right\}$$

After several algebraic calculations the above expression can be expressed as a function of the parameters β, γ:

$$\begin{aligned} a_i^M = \max\{0, -0.5(&\underline{f_{ii}} + 2\sum_{j \in B} \beta_{ij} + 2\sum_{j \in \Gamma} \gamma_{ij} \\ &- \sum_{j \in B} \max\{|\underline{f_{ij}} - 2\beta_{ij}|, |\overline{f_{ij}} - 2\beta_{ij}|\} \\ &- \sum_{j \in B} \max\{|\underline{f_{ij}} + 2\gamma_{ij}|, |\overline{f_{ij}} + 2\gamma_{ij}|\})\} \end{aligned}$$

Furthermore, the underestimator $L_M(x)$ can be forced to match the original nonconvex function at the corner points of the domain X, by subtracting a linear function from it. The linear function we have decided to subtract is the sum of the convex envelopes of the functions $p_{ij}(x) = -(x_i - x_j)^2$ and $q_{ij}(x) = -(x_i + x_j)^2$, for $i, j = 1, 2, \ldots, n, j > i$. The convex envelopes of those functions are defined as follows: $P_{ij}(x) = \max\left\{r_{ij}x_i + s_{ij}x_j + t_{ij}, r'_{ij}x_i + s'_{ij}x_j + t'_{ij}\right\}$, and

$Q_{ij}(x) = \max\left\{u_{ij}x_i + v_{ij}x_j + w_{ij}, u'_{ij}x_i + v'_{ij}x_j + w'_{ij}\right\}$

The coefficients of the convex envelopes $P_{ij}(x)$ are determined by solving the following systems of linear equations:

$$\begin{aligned} r_{ij}x^L{}_i + s_{ij}x^U{}_j + t_{ij} &= -(x_i^L - x_j^U)^2 \\ r_{ij}x^U{}_i + s_{ij}x^L{}_j + t_{ij} &= -(x_i^U - x_j^L)^2 \\ r_{ij}x^L{}_i + s_{ij}x^L{}_j + t_{ij} &= -(x_i^L - x_j^L)^2 \\ r'_{ij}x^L{}_i + s'_{ij}x^U{}_j + t'_{ij} &= -(x_i^L - x_j^U)^2 \\ r'_{ij}x^U{}_i + s'_{ij}x^L{}_j + t'_{ij} &= -(x_i^U - x_j^L)^2 \\ r'_{ij}x^U{}_i + s'_{ij}x^U{}_j + t'_{ij} &= -(x_i^U - x_j^U)^2 \end{aligned}$$

Similarly, the coefficients of the convex envelopes $Q_{ij}(x)$ are determined by solving the following systems of linear equations:

$$u_{ij}x^L{}_i + v_{ij}x^U{}_j + w_{ij} = -(x_i^L + x_j^U)^2$$
$$u_{ij}x^U{}_i + v_{ij}x^U{}_j + w_{ij} = -(x_i^U + x_j^U)^2$$
$$u_{ij}x^L{}_i + v_{ij}x^L{}_j + w_{ij} = -(x_i^L + x_j^L)^2$$
$$u'_{ij}x^U{}_i + v'_{ij}x^L{}_j + w'_{ij} = -(x_i^U + x_j^L)^2$$
$$u'_{ij}x^U{}_i + v'_{ij}x^U{}_j + w'_{ij} = -(x_i^U + x_j^U)^2$$
$$u'_{ij}x^L{}_i + v'_{ij}x^L{}_j + w'_{ij} = -(x_i^L + x_j^L)^2$$

The solutions of the above systems of linear equations are easy to find. Hence the new underestimator of the original nonconvex function is defined as follows

$$L_1(x) = L_M(x) + \sum_{i=1}^{n-1}\left(\sum_{j\in B^i}\beta_{ij}P_{ij}(x) + \sum_{j\in B^i}\gamma_{ij}Q_{ij}(x)\right)$$

The function $L_1(x)$ has the following important properties:

Property 1: L1 is a convex function

Property 2: L1 matches f(x) at the corner points of X.

Property 3: For every $j \in B^i, 0 \le \beta_{ij} \le \frac{1}{4}\left(\underline{f_{ij} + \overline{f_{ij}}}\right)$

Property 4: For every $j \in \Gamma^i, 0 \le \gamma_{ij} \le \frac{1}{4}\left(\underline{f_{ij} + \overline{f_{ij}}}\right)$

Property 5: L1 achieves its maximum separation distance at the middle point of the domain X.

Calculation of the parameters β and γ

The calculation of the parameters β, γ is done in such a way that the new underestimator $L_1(x)$ is always tighter than that obtained by using the aBB method. More specifically, the values of the parameters β and γ are obtained by solving the following linear programming problem

$$\min_{\beta,\gamma,\zeta,\eta} \left\{ \begin{array}{l} \sum_{i=1}^{n-1}\sum_{j\in B^i}\beta_{ij}\left(x_i^U - x_i^L + x_j^U - x_j^L\right)^2 + \\ \sum_{i=1}^{n-1}\sum_{j\in\Gamma^i}\gamma_{ij}\left(x_i^U - x_i^L + x_j^U - x_j^L\right)^2 - \\ \sum_{i=1}^{n-1}\left(f_{ii} + 2\sum_{j\in B^i}(\beta_{ij} - \zeta_{ij}) + 2\sum_{j\in B^i}(\gamma_{ij} - \eta_{ij})\right) \end{array} \right.$$

subject to

$$-\zeta_{ij} \le \underline{f_{ij}} - 2\beta_{ij} \le \zeta_{ij}, \forall j \in B^i,$$
$$-\zeta_{ij} \le \overline{f_{ij}} - 2\beta_{ij} \le \zeta_{ij}, \forall j \in B^i,$$
$$-\eta_{ij} \le \underline{f_{ij}} + 2\eta_{ij} \le \eta_{ij}, \forall j \in \Gamma^i,$$
$$-\eta_{ij} \le \overline{f_{ij}} + 2\eta_{ij} \le \eta_{ij}, \forall j \in \Gamma^i$$
$$\beta_{ij}, \gamma_{ij}, \zeta_{ij}, \eta_{ij} \ge 0$$

<u>THEOREM 1</u>: If the values of the β, γ parameters are the optimum solution of the above linear programming problem, then the new underestimator $L_1(x)$ is tighter than the underestimator $L_{aBB}(x)$ obtained by applying the aBB method.

Example

Consider the nonconvex function:

$$f(x,x) = x_1^4 + x_2 - (x_1 + x_2^2)^2$$

with $x_1 \in [-1.4, 5]$ and . $x_2 \in [-1.6, 5]$. The graph of that function is shown in Figure 1.

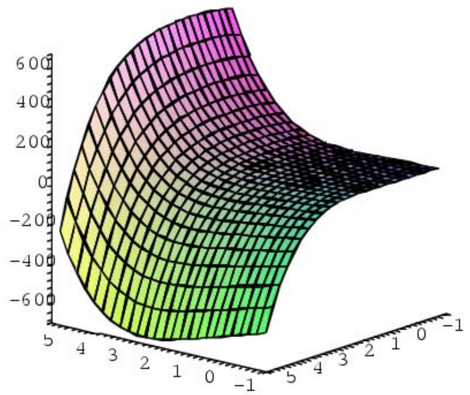

Figure 1: Graph of f(x)

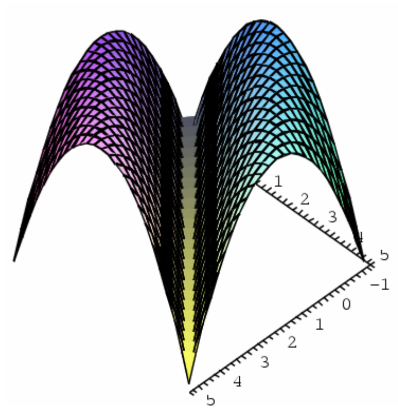

Figure 3: Improvements obtained by the new underestimator $L_1(x)$ over the underestimator obtained by the aBB method

The underestimator obtained by using the methodology described in this paper is shown in Figure 2.

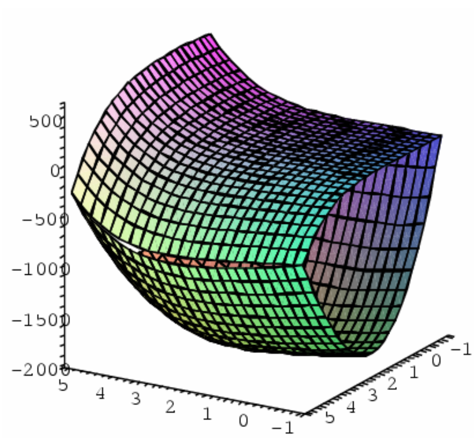

Figure 2: The new underestimator $L_1(x)$

The difference between the new underestimator and the one generated by aBB is shown in Figure 3.

As can be seen in the above figure the improvements in the underestimating function take place mostly at its sides.

Acknowledgments: Financial support from NSF and NIH (R01 GM52032) is gratefully acknowledged.

References

[1]. Adjiman C.S., Dallwig S., Floudas C.A., and Neumaier A. (1998). A global optimization method, aBB, for general twice-differentiable NLPs-I: Theoretical advances. *Comput. Chem. Eng., 22*, pp.1137-1158.

[2]. Akrotirianakis I. G., and Floudas C. A., A new class of improved convex underestimators for twice differentiable constrained NLP problems, accepted for publication: Journal of Global Optimization.

[3]. Akrotirianakis I. G. and Floudas C. A., Computational experience with a new class of convex underestimators: Box-constrained problems, accepted for publication: Journal of Global Optimization.

[4]. Floudas C.A. (2000). Deterministic Global Optimization: Theory, Methods and Applications, Kluwer Academic Publishers.