On the Functional Form of Convex Underestimators for Twice Continuously Differentiable Functions

Christodoulos A. Floudas[1] and Vladik Kreinovich[2]

[1]Department of Chemical Engineering
Princeton University
Princeton, NJ 08544, USA
floudas@titan.princeton.edu

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu

Abstract

The optimal functional form of convex underestimators for general twice continuously differentiable functions is of major importance in deterministic global optimization. In this paper, we provide new theoretical results that address the classes of optimal functional forms for the convex underestimators. These are derived based on the properties of shift-invariance and sign-invariance.

1 Introduction

Mathematical methods that generate convex underestimators for twice continuously differentiable constrained nonlinear optimization problems are of primary importance in deterministic global optimization [7]. A powerful approach for constructing such convex underestimators is the theory behind the αBB global optimization algorithm [1, 2, 3, 6, 7, 8, 9, 10]. In the αBB framework, an underestimator $L(x) = f(x) + \Phi(x)$ is selected, where

$$\Phi(x) = -\sum_{i=1}^{n} \alpha_i \cdot (x_i - x_i^L) \cdot (x_i^U - x_i). \tag{1}$$

The parameters α_i are selected in such a way that the resulting function $L(x)$ is convex and still not too far away from the original objective function $f(x)$ [1, 2, 3, 6, 7, 9, 10].

In the αBB techniques, for each coordinate x_i, there is a single parameter α_i affecting this coordinate. Changing α_i is equivalent to a linear re-scaling of x_i. Indeed, if we change the unit for measuring x_i to a new unit which is λ_i times smaller, then all the numerical values become λ_i times larger: $x_i \to y_i = g_i(x_i)$, where $g_i(x_i) = \lambda_i \cdot x_i$. In principle, we can have two different re-scalings:

- $x_i \to y_i = g_i(x_i) = \lambda_i \cdot x_i$ on the interval $[x_i^L, x_i]$, and
- $x_i \to z_i = h_i(x_i) = \mu_i \cdot x_i$ on the interval $[x_i, x_i^U]$.

If we substitute the new values $y_i = g_i(x_i)$ and $z_i = h_i(x_i)$ into the formula (1), then we get the following expression

$$\Phi(x) = -\sum_{i=1}^{n} \alpha_i \cdot (g_i(x_i) - g_i(x_i^L)) \cdot (h_i(x_i^U) - h_i(x_i)). \tag{2}$$

For the above linear re-scalings, we get

$$\widetilde{\Phi}(x) = -\sum_{i=1}^{n} \widetilde{\alpha}_i \cdot (x_i - x_i^L) \cdot (x_i^U - x_i),$$

where $\widetilde{\alpha}_i = \alpha_i \cdot \lambda_i \cdot \mu_i$.

From this viewpoint, a natural generalization is to replace *linear* re-scalings $g_i(x_i)$ and $h_i(x_i)$ with *non-linear* ones, that is, to consider convex underestimators of the type $L(x) = f(x) + \Phi(x)$, where $\Phi(x)$ is described by the formula (2) with non-linear functions $g_i(x_i)$ and $h_i(x_i)$. Now, instead of selecting a number α_i for each coordinate i, we have an additional freedom of choosing arbitrary non-linear functions $g_i(x_i)$ and $h_i(x_i)$. The fundamental question is what classes of functional forms are best suited for the convex underestimators.

In recent work [4, 5], several new functional forms that correspond to different non-linear functions have been investigated, and through extensive computational studies it was demonstrated that the best results were achieved for the exponential functions $g_i(x_i) = \exp(\gamma_i \cdot x_i)$ and $h_i(x_i) = -\exp(-\gamma_i \cdot x_i)$. For these functions, the expression (2) can be simplified to:

$$\alpha_i \cdot (g_i(x_i) - g_i(x_i^L)) \cdot (h_i(x_i^U) - h_i(x_i)) = \alpha_i \cdot (e^{\gamma_i \cdot x_i} - e^{\gamma_i \cdot x_i^L}) \cdot (-e^{-\gamma_i \cdot x_i^U} + e^{-\gamma_i \cdot x_i}) =$$

$$\widetilde{\alpha}_i \cdot (1 - e^{\gamma_i \cdot (x_i - x_i^L)}) \cdot (1 - e^{\gamma_i \cdot (x_i^U - x_i)}),$$

where $\widetilde{\alpha}_i \stackrel{\text{def}}{=} \alpha_i \cdot e^{\gamma_i \cdot (x_i^U - x_i^L)}$.

If a selection of the functions $g_i(x_i)$ and $h_i(x_i)$ is "optimal", then the results of using these optimal functions should not change if we simply change the starting point for measuring x_i (i.e., replace each value x_i with a new value $x_i + s$, where s is the shift in the starting point). Otherwise, if the "quality" of the resulting convex underestimators changes with shift, we could apply a shift and get better functions $g_i(x_i)$ and $h_i(x_i)$ – which contradicts our assumption that the selection of $g_i(x_i)$ and $h_i(x_i)$ is already optimal.

Therefore, the "optimal" choices $g_i(x_i)$ and $g_i(x_i)$ can be determined from the requirement that each component $\alpha_i \cdot (g_i(x_i) - g_i(x_i^L)) \cdot (h_i(x_i^U) - h_i(x_i))$ in the sum (2) be invariant under the corresponding shift. Let us describe this requirement in precise terms.

Definition *A pair of smooth functions $(g(x), h(x)))$ from real numbers to real numbers is* shift-invariant *if for every s and α, there exists $\widetilde{\alpha}(\alpha, s)$ such that for every x^L, x, and x^U, we have*

$$\alpha \cdot (g(x) - g(x^L)) \cdot (h(x^U) - h(x)) =$$

$$\widetilde{\alpha}(\alpha, s) \cdot (g(x+s) - g(x^L + s)) \cdot (h(x^U + s) - h(x+s)) \qquad (3)$$

At first glance, shift invariance is a reasonable but weak property. It turns out, that this seemingly weak property almost uniquely determines the optimal selection of exponential functions, as shown in the following Proposition.

Proposition *If a pair of functions $(g(x), h(x))$ is shift-invariant, then this pair is either exponential or linear, that is, each of the functions $g(x)$ and $h(x)$ has the form $g(x) = A + C \cdot \exp(\gamma \cdot x)$ or $g(x) = A + k \cdot x$.*

Proof of Proposition For $\alpha = 1$, condition (3) takes the form

$$(g(x) - g(x^L)) \cdot (h(x^U) - h(x)) =$$

$$C(s) \cdot (g(x+s) - g(x^L + s)) \cdot (h(x^U + s) - h(x+s)), \tag{4}$$

where we denoted $C(s) \stackrel{\text{def}}{=} \widetilde{\alpha}(1, s)$. To simplify this equation, let us separate the variables:

- let us move all terms containing x^L to the left-hand side – by dividing both sides by $(g(x+s) - g(x^L + s))$, and
- let us move all terms containing x^U to the right-hand side – by dividing both sides by $(h(x^U) - h(x))$.

As a result, we arrive at the following equation:

$$\frac{g(x) - g(x^L)}{g(x+s) - g(x^L + s)} = C(s) \cdot \frac{h(x^U + s) - h(x+s)}{h(x^U) - h(x)}. \tag{5}$$

Let us denote the left-hand side of this equation by A. By definition, the value A depends on x, s, and x^L. Since A is equal to the right-hand side, and the right-hand side does not depend on x^L, the expression A cannot depend on x^L, so $A = A(x, s)$, that is,

$$\frac{g(x) - g(x^L)}{g(x+s) - g(x^L + s)} = A(x, s). \tag{6}$$

Multiplying both sides by the denominator, we conclude that

$$g(x) - g(x^L) = A(x, s) \cdot (g(x+s) - g(x^L + s)). \tag{7}$$

Differentiating both sides by x^L, we conclude that

$$-g'(x^L) = -A(x, s) \cdot g'(x^L + s), \tag{8}$$

that is, equivalently,

$$\frac{g'(x^L)}{g'(x^L+s)} = A(x,s). \tag{9}$$

In this equation, the left-hand side does not depend on x, so the right-hand does not depend on x either, that is, $A(x,s) = A(s)$. Thus, the equation (7) takes the form

$$a(s) \cdot (g(x) - g(x^L)) = (g(x+s) - g(x^L+s)), \tag{10}$$

where we denoted $a(s) \stackrel{\text{def}}{=} 1/A(s)$.

The function $g(x)$ is smooth, hence the function $a(s)$ is smooth too – as the ratio of two smooth functions. Differentiating both sides of the equation (10) with respect to s and taking $s = 0$, we get

$$a \cdot (g(x) - g(x^L)) = (g'(x) - g'(x^L)), \tag{11}$$

where $a \stackrel{\text{def}}{=} a'(0)$.

To simplify this equation, let us move all the term depending on x to the right-hand side and all the terms depending on x^L to the left-hand side. As a result, we arrive at the following:

$$g'(x^L) - a \cdot g(x^L) = g'(x) - a \cdot g(x). \tag{12}$$

The right-hand side is a function of x only, but since it is equal to the left-hand side – which does not depend on x at all – it is simply a constant. If we denote this constant by b, we get the following equation:

$$g'(x) - a \cdot g(x) = b, \tag{13}$$

that is,

$$\frac{dg}{dx} = a \cdot g + b \tag{14}$$

and

$$\frac{dg}{a \cdot g + b} = dx. \tag{15}$$

When $a = 0$, integrating both sides of this equation, we get $\frac{1}{b} \cdot g(x) = x + C$, that is, $g(x) = b \cdot x + b \cdot C$. When $a \neq 0$, then for $\widetilde{g}(x) \stackrel{\text{def}}{=} g(x) + \frac{b}{a}$, we get

$$\frac{d\widetilde{g}}{a \cdot \widetilde{g}} = dx. \tag{16}$$

hence $\frac{1}{a} \cdot \ln(\widetilde{g}(x)) = x + C$ thence $\ln(\widetilde{g}(x)) = a \cdot x + a \cdot C$, so $\widetilde{g}(x) = C \cdot \exp(a \cdot x)$ and $g(x) = \widetilde{g}(x) - \frac{b}{a} = C \cdot \exp(a \cdot x) + C_1$ for some constants C, a, and C_1. The proposition is proven.

In addition to shift, another natural symmetry is changing the sign. If we require that the expression (2) remains invariant if we change the sign, that is, replace x by $-x$, then we get the relation between $g(x)$ and $h(x)$: $h(x) = -g(-x)$. Therefore, if a pair $(g(x), h(x)$ is shift-invariant and sign-invariant, then:

- either $g(x) = \exp(\gamma \cdot x)$ and $h(x) = -\exp(-\gamma \cdot x)$,
- or $g(x) = h(x) = x$.

In other words, *the optimal generalized αBB scheme is either the original αBB [1, 2, 3, 6, 7, 8, 9, 10], or the scheme with exponential functions [4, 5]* . Thus, we have established that:

- the exponential functions are indeed optimal, and
- the theoretical explanation of why they are optimal is because they are the only pair of functions which satisfies the condition of symmetry (shift-invariance and sign-invariance) that optimal pairs should satisfy.

In addition to changing the starting point for x, we can also change a unit for measuring x, that is, consider *scaling* transformations $x \to \lambda \cdot x$. Shall we require that the expression (2) be invariant not only w.r.t. shifts but w.r.t scalings as well? Theoretical results on the scale invariance are presented in [11].

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, NSF grants EAR-0225670 and DMS-0532645, Army Research Lab grant DATM-05-02-C-0046, Star Award from the University of Texas System, Texas Department of Transportation grant No. 0-5453, National Science Foundation CTS-0330541, and the National Institutes of Health R01 GM52032.

References

[1] C. S. Adjiman, S. Dallwig, I. Androulakis, C. A. Floudas, and A. Neumaier, "A global optimization method, αBB, for general twice-differentiable constrained NLP I. Theoretical aspects", *Computers and Chemical Engineering*, 1998, Vol. 22, No. 9, pp. 1137–1158.

[2] C. S. Adjiman, I. Androulakis, and C. A. Floudas, "A global optimization method, αBB, for general twice-differentiable constrained NLP II. Implementation and computational results", *Computers and Chemical Engineering*, 1998, Vol. 22, pp. 1159–1179.

[3] C. S. Adjiman and C.A. Floudas, "Rigorous convex underestimators for general tweice-differentiable problems", *Journal of Global Optimization*, 1996, Vol. 9, pp. 23-40.

[4] I. G. Akrotirianakis and C. A. Floudas, "Computational experience with a new class of convex underestimators: box-constrained NLP problems", *Journal of Global Optimization*, 2004, Vol. 29, pp. 249–264.

[5] I. G. Akrotirianakis and C. A. Floudas, "A new class of improved convex underestimators for twice continuously differentiable constrained NLPs", *Journal of Global Optimization*, 2004, Vol. 30, pp. 367-390 .

[6] I. P. Androulakis, C. D. Maranas and C. A. Floudas, "Alpha BB: A global optimization method for general constrained nonconvex problems", *Journal of Global Optimization*, 1995, Vol. 7, pp. 337-363.

[7] C. A. Floudas, *Deterministic global optimization: theory, methods, and applications*, Kluwer Academic Publishers, Dordrecht, 2000.

[8] C.A. Floudas, “Research challenges, opportunities and synergism in systems engineering and computational biology”, *AIChE J.*, 2005, Vol. 51, pp. 1872-1884.

[9] C. A. Floudas, I. G. Akrotirianakis, S. Caratzoulas, C. A. Meyer, and J. Kallrath, “Global optimization in the 21st century: advances and challenges”, *Computers & Chemical Engineering*, 1995, Vol. 29, pp. 1185-1202.

[10] C. D. Maranas and C. A. Floudas, “Global minimal potential energy conformations for small molecules”, *Journal of Global Optimization*, 1994, Vol. 4, pp. 135–170.

[11] C.A. Floudas and V. Kreinovich, “Towards Optimal Techniques for Solving Global Optimization Problems: Symmetry-Based Approach”, Chapter in *Models and Algorithms for Global Optimization*, Eds. A. Torn and J. Zilinskas, Springer, 2006.