GLOBIE: An algorithm for the deterministic global optimization of box-constrained NLPs

N. Kazazakis, C.S. Adjiman*

Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, United Kingdom.
*c.adjiman@imperial.ac.uk

Abstract

The GLObal BarrIer Exclusion (GLOBIE) algorithm for the deterministic global optimization of box-constrained NLPs is presented for the first time. It is based on replacing the objective function with a modified subenergy function. The subenergy function preserves the stationary point of the objective function and possesses interesting properties, such as less negative eigenvalues in some domains. These properties are exploited to create an algorithm that uses αBB relaxations to fathom quickly sub-domains of the solution space that do not contain the global solution. GLOBIE successfully identifies the solution of six small test problems.

Keywords: Subenergy function, eigenvalues, αBB, convex relaxation

1. Introduction

Many engineering problems can be formulated as nonlinear optimization problems. Among solution techniques, deterministic global optimization algorithms are of particular interest for their ability to guarantee the identification of the global optimum (Floudas and Gounaris, 2009), as it is often the only solution of relevance, for instance in problems involving phase equilibrium (Floudas 1997, McDonald & Floudas 1995), parameter estimation (Amaran & Sahinidis 2012), or safety (Huang et al. 2002). The present work addresses general non-convex box-constrained problems of the form:

Problem P: \[ \min_{x \in X} f(x), \]

where \( f \) is a general \( C^2 \) function, \( X = \{ x \in \mathbb{R}^n | x_L \leq x \leq x_U \} \) is a compact set, and \( x_L \) and \( x_U \) are two \( n \)-dimensional vectors.

There are two central concepts in deterministic global optimization, which are typically integrated within a Branch-and-Bound (B&B) algorithm: (i) domain reduction techniques, e.g., Zorn & Sahinidis (2014) and (ii) methods to obtain a lower bound on the objective function over a domain. For the latter, many techniques have been developed (Skjäl et al. 2012), including interval arithmetic, interval slopes, convex relaxations, e.g., Akrotirianakis & Floudas (2004). The review by Floudas and Gounaris (2009) provides an overview of key advances in the area.

Many recent theoretical advances have focused on constructing tighter convex relaxations, e.g., Scott et al. (2011). This is mainly because the global minimum is often found early by a B&B algorithm and most of the CPU time is invested in obtaining a certificate of optimality for this solution, by achieving convergence between the best upper bound (BUB) and the lowest lower bound (LLB). Most algorithms proposed to date make limited use of the BUB. In this paper, a novel Spatial Branch-and-Bound
algorithm is introduced, the GLObal Barrler Exclusion (GLOBIE) algorithm, which effectively exploits the information provided by the BUB in order to fathom large sub-domains of the solution space. Exploiting the BUB is a central concept in tunneling (Barhen 1997) and this idea is integrated within the B&B framework of GLOBIE.

A key feature of GLOBIE is a mapping of the objective function to a subenergy function, a modified form of the subenergy function of Barhen (1997). This mapping presents certain advantages because the subenergy function has special properties which make it more amenable to deterministic global optimization. The Hessian matrix of the subenergy function can be shown to have negative eigenvalues that are of smaller magnitude than the corresponding negative eigenvalues of the Hessian matrix of the original function, over domains where \( f \) is greater than a known upper bound on the global solution. The subenergy function and its properties are introduced in Section 2, the GLOBIE algorithm is briefly stated in Section 3 and test results on six small test problems are given in Section 4.

2. The subenergy function

In the following sections, \( f(\mathbf{x}), \mathbf{x} \in X \) is assumed to be a general \( C^2 \) function \( \mathbb{R}^n \rightarrow \mathbb{R} \) and \( \mu \in \mathbb{R}^+ \) is a parameter. \( f^{\ast} \in F \), where \( F \) is the image of \( X \) through \( f \), refers to an arbitrary value of \( f \); in practice, \( f^{\ast} = \text{BUB} \) is usually chosen. \( f^\dagger \) is a global minimum of \( f \). \( f_1 \) is a parametric value which satisfies the following condition: \( f_1 - f^{\ast} = c \), where \( c \) is an arbitrary positive constant.

The subenergy function \( S(\mathbf{x}; f^{\ast}, \mu) \) is defined as follows:

\[
S(\mathbf{x}; f^{\ast}, \mu) = -\ln \left( 1 + e^{-\mu(f(\mathbf{x})-f^{\ast})} \right)
\]

where \( S \) is parametric in \( f^{\ast} \) and \( \mu \). Parameter \( \mu \) is referred to as the subenergy magnitude. The fundamental property of the subenergy function is that it is a continuous bijective transformation of the original objective function that preserves its monotonicity and stationary points, but not its convexity (i.e., it differs from simply scaling the function). The transformation is illustrated in Figure 1, using the global minimum as BUB so that \( f^{\ast} = f^\dagger \). The new function is “less concave” than the original function in that its maxima have less curvature. This is a very desirable property in deterministic global optimization, as this can lead to tight convex relaxations.

By appropriate choice of the subenergy magnitude \( \mu \), it is possible to impose a number of desirable properties on the subenergy function. Before describing these properties, we introduce two definitions:

Definition 1. A function \( f(\mathbf{x}), \mathbf{x} \in X \), is said to be flattened across \( X_1 \subseteq X \), if there exists a bijective mapping of this function onto a function \( S(\mathbf{x}) \) with the following properties: (i) the eigenvalues of the Hessian matrix \( S(\mathbf{x}) \) are of smaller magnitude than those of the Hessian matrix of \( f(\mathbf{x}) \) for all \( \mathbf{x} \in X_1 \) and (ii) \( S(\mathbf{x}) \) has the same monotonicity and stationary points as \( f(\mathbf{x}) \).

Definition 2. \( X_1 = \{ \mathbf{x} \in X : f(\mathbf{x}) > f_1 \} \) is defined as \( f_1 \)-sub-optimal with respect to problem \( P \).

It can readily be shown that the subenergy function satisfies condition (ii) of Def. 1. In order to motivate our approach, the effect of the choice of \( f^{\ast} \) and \( \mu \) on the subenergy transformation is displayed in Figure 2. It can be seen that increasing \( \mu \) can have two distinct effects on the function, depending on whether \( f^{\ast} \) is the global minimum or not. If \( f^{\ast} \) is not the global minimum (Fig. 2a), then \( f \) is flattened in all \( f_1 \)-sub-optimal sub-
domains and has deep valleys in all other sub-domains. If \( f^* \) is the global minimum, then as \( \mu \) increases, the function is flattened everywhere except for a neighborhood of the global minimum. The radius of this neighborhood (equivalently, the value of \( f_1 \)) depends on the magnitude of \( \mu \).

Figure 1: Plot of function \( f(x) = \sin(4x) + 0.1x^2 + 1.5 \) and its \( \mu \)-subenergy function \( S(x; f^*, \mu) \). Open circle: global minimum.

For a value \( f_1 \) of \( f \) such that \( f_1 - f^* = c \), where \( c \) is an arbitrary positive constant, it can be shown that, for all \( x \in X_1 \), the eigenvalues of the Hessian matrix of \( S(x; f^*, \mu) \) tend to zero as \( \mu \) goes to infinity. In fact, it is possible to derive bounds on these values, as well as on the first and second derivatives of the \( \mu \)-subenergy function, that depend only on \( \mu \) and \( c \), for all \( x \in X_1 \). These results are stated here in a series of theorems but no proofs are provided due to space constraints.

**Theorem 1.** For any \( f_1 > f^* \) and \( \epsilon \in (0, \ln(2)) \), \( S(x; f^*, \mu) \geq -\epsilon, \forall x \in X_1 \) if and only if \( \mu = -\ln(e^\epsilon - 1)/c \).

**Corollary:** \( -\epsilon < S(x; f^*, \mu) < 0, \forall x \in X_1 \), iff the subenergy magnitude satisfies Theorem 1.

Thus, the values of \( S(x; f^*, \mu) \) can be brought arbitrarily close to zero over the sub-optimal domains by judicious choice of the subenergy magnitude.

**Theorem 2.** The magnitude of an arbitrary first derivative of the \( \mu \)-subenergy function for any \( x \in X_1 \) is bounded by \( \left| \frac{\partial^2}{\partial x_i \partial x_j} S(x; f^*, \mu) \right| \leq G_1 \max \left( \left| \frac{\partial f}{\partial x_i} \right|, \left| \frac{\partial f}{\partial x_j} \right| \right) = G_1, \) where the superscripts \( L \) and \( U \) denote lower and upper bounds, respectively, of the relevant quantity for all \( x \in X_1 \).

**Theorem 3.** The magnitude of an arbitrary second derivative of the \( \mu \)-subenergy function is bounded from above by:

\[
\left| \frac{\partial^2}{\partial x_i \partial x_j} S(x; f^*, \mu) \right| \leq G_1 \max \left( \left| \frac{\partial f}{\partial x_i} \right|, \left| \frac{\partial f}{\partial x_j} \right| \right) + \mu^2 e^{\epsilon(\epsilon^2 - 1)} \max \left( \left| \frac{\partial^2 f}{\partial x_i \partial x_j} \right| \right) = G_1, \forall x \in X_1
\]

where \( L \) and \( U \) are as above. Note that the bounds derived in Theorem 3 are only dependent on bounds on the original function and arbitrary constants. These relations can be used to calculate a lower bound on the eigenvalues of the Hessian matrix of the \( \mu \)-subenergy function.
Figure 2: The effect of the parameters $f^*, \mu$ on the $\mu$-subenergy function of $f(x) = \sin(4x) + 0.1x^2 + 1.5$. The dashed vertical line indicates the point used to obtain $f^*$ a) $f^* \neq f^\dagger$, and b) $f^* = f^\dagger$. The horizontal lines correspond to $f_1$. The $X_1$ domains for each plot can be seen between zero and the horizontal line of the same color and style.

Theorem 4. For all $x \in X_1$, the minimum eigenvalue of the Hessian matrix of the $\mu$-subenergy function, $\lambda_{\text{min}}$, is bounded from below by:

$$\lambda_{\text{min}} \geq \min \left( \frac{\mu}{1+e^{\mu}} \min \left[ \frac{f_1^L}{\gamma} \right], 0 \right) = \frac{e^{(f^*/-\mu)}}{(1+e^{-\mu})} \left( \left( \frac{\mu}{\gamma} \right)^j \right)^{\frac{1}{2}} - \sum_{j=1}^{m} G_j^0$$

(3)

where the relevant bounds are taken over $x \in X_1$. The eigenvalue bound is dependent on $\mu$ and $c$. In practical terms, this means that by appropriate selection of $\mu$ it is possible to impose a bound on the minimum eigenvalue of the Hessian matrix of the $\mu$-subenergy function that is arbitrarily close to zero (but negative) in every $f_1$-sub-optimal domain. This property can in principle be exploited during the course of a B&B algorithm to fathom $f_1$-sub-optimal domains. In particular, the choice of a large value of $\mu$ will result in negative eigenvalues of a smaller magnitude.

3. The GLOBIE algorithm

The GLOBIE algorithm makes use of the $\mu$-subenergy function to converge to a global minimum. An initial best upper bound (BUB) $f^*$ is determined. The $\alpha\text{BB}$ algorithm is applied until a new BUB is found. $f^*$ is then updated, and the subenergy function changes. All lower bounds calculated in previous iterations are still valid. However, tighter lower bounds can now be computed, especially in domains where $f(x)$ is greater than the new $f^*$. We test for convergence using the direct relation between the two lower bounds $f^{\text{LLB}} = -\frac{1}{\mu} \ln(e^{-S^{\text{LLB}}} - 1) + f^*$. 
In the description of the algorithm below, a node $i$ in the B&B tree represents the set $\{x \in X_i\} \subset X$ and it is characterized by a lower bound $S^L_i$ on the subenergy function over that set. In this initial version, the value of $\mu$ is set according to Theorem 1, using $\epsilon = 10^{-20}$ and $c = 10^{-6}$. On termination, the solution is given by $f^*$ and the corresponding $x$.

**GLOBIE algorithm**

Set convergence tolerance $t$. Set $\epsilon, c$, node counter $k = 0$, $X_0 = X$.

Calculate $f^*$ as a local solution of $\min_{x \in X_0} f(x)$.

Initialize list of active nodes $L = \emptyset$. Calculate $\mu = \mu(\epsilon, c)$.

Derive convex relaxation $\tilde{S}(x; f^*, \mu)$ of $S(x; f^*, \mu)$ for $x \in X_0$.

Solve $S^L_0 = \min_{x \in X_0} \tilde{S}(x; f^*, \mu)$. Set $S^{LLB} = S^L_0$ and $L = \{0\}$.

Obtain $S^{BBU} = \min_{x \in X_0} S(x; f^*, \mu)$.

while $f^* - S^{LLB} \geq t$

\[ f^{LLB} = -\frac{1}{\mu} \ln \left( e^{-S^{LLB}} - 1 \right) + f^*. \]

Select node $i \in L: S^L_i = S^{LLB}$

Branch on $X_i$ to create new nodes $X_{k+1}, \ldots, X_{k+m}$. Add nodes to $L$.

$L = L \setminus \{i\}$

for $j = 1 \ldots m$

Construct convex relaxation $\tilde{S}_{k+j}(x; f^*, \mu)$ over $X_{k+j}$

$S^L_{k+j} = \min_{x \in X_{k+j}} \tilde{S}_{k+j}(x; f^*, \mu)$

$f^U_{k+j} = \min_{x \in X_{k+j}} f(x), S^U_{k+j} = -\ln \left( 1 + e^{-\mu(f^U_{k+j} - f^*)} \right)$

end for

$f^* = \min\{f^*, f^U_{k+1}, \ldots, f^U_{k+m}\}$

$S^{BBU} = \min\{S^{BBU}, S^U_{k+1}, \ldots, S^U_{k+m}\}$

$S^{LLB} = \min\{S^{LLB}, S^L_{k+1}, \ldots, S^L_{k+m}\}$

$k = k + m$

end

4. Results

The GLOBIE algorithm has been implemented in MATLAB using the INTLAB (Rump 1999) interval arithmetic toolbox. In Table 1, we present preliminary numerical results for some small test functions, based on the number of iterations required to reach an absolute tolerance of $t = 10^{-6}$. All test functions have more than 100 local minima.

We also provide the number of iterations required when using the standard $\alpha$BB algorithm. In both cases, bisection is used and the $\alpha$ values are computed using the Gerschgorin theorem (Adjiman et al. 1998). The GLOBIE algorithm performs very well, especially as the non-linearity of the problem increases. This can be seen clearly in the first four examples. This is because, once a better upper bound is found, domains containing hundreds of local minima can be fathomed readily.

5. Conclusions

The GLOBIE deterministic global optimization algorithm for box-constrained NLPs was introduced in this paper. It is based on a branch-and-bound scheme, and makes use of the best known upper bound to modify the objective function and facilitate the elimination of sub-optimal regions. The $\mu$-subenergy function was introduced; it can
several theoretical properties were presented. The eigenvalues of the new function can be bounded from below arbitrarily, through appropriate tuning of the subenergy magnitude. This property was used to devise the GLOBIE algorithm. Preliminary results show that the new algorithm converges to the global solution successfully in all cases, using a modest number of iterations. Based on these encouraging results, a version of the algorithm for NLPs with nonlinear constraints is under development.

Table 1: Summary of computational results. The Rastrigin function is from Muehlenbein (1991).

<table>
<thead>
<tr>
<th>Function $f(x)$</th>
<th>Variable range</th>
<th>$f^*$</th>
<th>$f_0$</th>
<th>GLOBIE #iter</th>
<th>αBB #iter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sin(50x) + 0.1x^2$</td>
<td>$[-12, 12]$</td>
<td>-0.999901</td>
<td>0</td>
<td>40</td>
<td>345</td>
</tr>
<tr>
<td>$\sin(500x) + 0.1x^2$</td>
<td>$[-12, 12]$</td>
<td>-0.999999</td>
<td>0</td>
<td>53</td>
<td>3479</td>
</tr>
<tr>
<td>$\sin(10000x) + 0.1x^2$</td>
<td>$[-12, 12]$</td>
<td>-1.000000</td>
<td>0</td>
<td>66</td>
<td>34016</td>
</tr>
<tr>
<td>Rastrigin’s function (2 variables)</td>
<td>$[-12, 12]$</td>
<td>0.000000</td>
<td>0</td>
<td>106</td>
<td>67998</td>
</tr>
<tr>
<td>Rastrigin’s function (4 variables)</td>
<td>$[-12, 12]$</td>
<td>0.000000</td>
<td>0</td>
<td>157</td>
<td>1347</td>
</tr>
</tbody>
</table>

Acknowledgements

We gratefully acknowledge funding by the EPSRC [EP/J003840/1].

References


