Global optimization of separable objective functions on convex polyhedra via piecewise-linear approximation

Hao Zhang*, Shuning Wang

Department of Automation, Tsinghua University, Beijing 100084, PR China

Received 10 May 2005; received in revised form 31 October 2005

Abstract

The problem of minimizing a separable nonlinear objective function under linear constraints is considered in this paper. A systematic approach is proposed to obtain an approximately globally optimal solution via piecewise-linear approximation. By means of the new approach a minimum point of the original problem confined in a region where more than one linear piece is needed for satisfactory approximation can be found by solving only one linear programming problem. Hence, the number of linear programming problems to be solved for finding the approximately globally optimal solution may be much less than that of the regions partitioned. In addition, zero-one variables are not introduced in this approach. These features are desirable for efficient computation. The practicability of the approach is demonstrated by an example.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Separable objective function; Nonlinear programming; Piecewise-linear approximation; Global optimization

1. Introduction

Consider a class of nonlinear programming problems of the form:

\[
\begin{align*}
\text{min} & \quad f(x) \\
\text{s.t.} & \quad a_r^T x \leq b_r, \quad 1 \leq r \leq m \\
& \quad x \in D,
\end{align*}
\]

where \( D = \{ x \in \mathbb{R}^n : l_j \leq x_j \leq u_j \ \forall j \} \) is a nonempty hypercube, \( x_j \) denotes the \( j \)th component of \( x \), and \( f \) is a sum of \( n \) univariate functions, i.e., \( f(x) = \sum_{j=1}^{n} f_j(x_j) \). As both the objective function and constraints are separable, the nonlinear programming problem (1) is referred to as a separable one. Problems of this type are widespread in fields such as production, economics, facility layout, optimal design and transportation.

In general, the objective function \( f \) is nonconvex. It is difficult to find a globally optimal solution of problem (1). A natural approach is to approximate \( f \) by linear functions in a number of smaller hypercubes and obtain an approximately globally optimal solution via solving a series of linear programming problems (LPs). It is obvious that decreasing the...
number of subproblems is of crucial importance. A classical method is to formulate the problem as an equivalent mixed-integer program (MIP) by introducing some zero-one variables as well as some auxiliary inequalities, see e.g., [1–4]. Using the branch-and-bound algorithm, many redundant LPs can be eliminated by branching on sets of variables.

In this paper, we are going to achieve the same goal in a different way. The foundation of our method is the following fact. If the continuous piecewise-linear (CPWL) function used to approximate \( f \) is convex in a convex union of many smaller hypercubes, an approximately globally optimal solution of the original problem confined in this union can be determined by solving only one LP. In many cases, the number of such unions may be much less than that of all smaller hypercubes partitioned. Hence, the new approach may decrease the computational effort greatly.

The other parts of this paper are organized as follows. In Section 2, we discuss how to partition \( D \) into convex unions of smaller hypercubes where \( f \) can be approximated by convex CPWL functions. In Section 3, we explain how to find a globally optimal solution of a convex CPWL function on a convex polyhedron by solving one linear programming problem. In Section 4 a numerical example is presented to demonstrate the practicability of the approach. Finally in Section 5 a brief conclusion is made.

2. Piecewise-linear approximation

Since \( f \) is the sum of univariate functions \( f_j \), \( 1 \leq j \leq n \), its CPWL approximation can be obtained based on approximating each \( f_j \) by a univariate CPWL function. This can be achieved by partitioning the interested region of each univariate function into sufficiently many nonoverlapping small intervals.

Let \( g \) be an arbitrary univariate function whose interested region is \([c, d] \subset \mathbb{R}\). Let \( m \) breakpoints \( c < x_1 < \cdots < x_m < d \) be suitably chosen so that \( g \) can be approximated to a satisfactory precision by the linear function \( \ell_k(x) = A_k x + B_k \) in each small interval \( I_k = [x_{k-1}, x_k] \) for any \( 1 \leq k \leq m \), where \( x_0 = c, x_{m+1} = d \). \( A_k \) and \( B_k \) are determined by the linear equations \( g(x_{k-1}) = \ell_k(x_{k-1}), g(x_k) = \ell_k(x_k) \). A CPWL approximation function \( \phi \) of \( g \) on \([c, d]\) can be obtained by connecting these segments. The figure of such a CPWL function is shown in Fig. 1.

For any \( 1 \leq k \leq m \), we call \( x_k \) an inflection point if \( A_k > A_{k+1} \). In Fig. 1, both \( x_1 \) and \( x_2 \) are inflection points while the others are not. Denote by \( x_{\tau(1)} < x_{\tau(2)} < \cdots < x_{\tau(q)} \) all inflection points among the breakpoints \( x_k, 1 \leq k \leq m \). Obviously \( q \leq m \). Let \( \tau(0) = 0, \tau(q + 1) = m + 1 \), and define \( \hat{I}_k = [x_{\tau(k-1)}, x_{\tau(k)}] \) for any \( 1 \leq t \leq q + 1 \). It can be seen that \( \hat{I}_k \) is the union of intervals \( I_k \), \( \tau(t-1) < k \leq \tau(t) \). Based on the above partition we can get a piecewise-convex expression of the CPWL function \( \phi \), which is very useful for global optimization of separable programming problems and is given in the following theorem.

Theorem 1. For any \( 1 \leq t \leq q + 1 \),

\[
\phi(\lambda) = \max_{k \in \hat{s}_t} \ell_k(\lambda) \quad \forall \lambda \in \hat{I}_t, \tag{2}
\]

where \( \hat{s}_t = \{ k, \tau(t-1) < k \leq \tau(t) \} \).

Proof. For any \( 1 \leq t \leq q + 1 \), since \( A_{\tau(t-1)+1} < A_{\tau(t-1)+2} < \cdots < A_{\tau(t)} \), according to [1, Lemma 6.2], the CPWL function \( \phi \) is convex on \( \hat{I}_t \). Arbitrarily choose a \( \hat{\lambda} \in \hat{I}_t \). There should be an integer \( k \in \hat{s}_t \) such that \( \hat{\lambda} \in I_k \) and \( \phi(\hat{\lambda}) = \ell_k(\hat{\lambda}) \). Then, the following relation must be satisfied,

\[
\ell_k'(\hat{\lambda}) \leq \ell_k'(\lambda) \quad \forall k' \in \hat{s}_t - \{ k \}. \tag{3}
\]

Otherwise, i.e., there is a \( k' \in \hat{s}_t - \{ k \} \) with which \( \ell_k'(\hat{\lambda}) > \ell_k'(\lambda) \), and we can choose a \( \hat{\lambda} \in I_{k'} \subset \hat{I}_t \) and a sufficiently small positive number \( \varepsilon \) such that \( \hat{\lambda} = \varepsilon \hat{\lambda} + (1 - \varepsilon) \hat{\lambda} \in I_{k'} \). Because \( \phi \) is convex on \( \hat{I}_t \), we have

\[
\phi(\hat{\lambda}) \leq \varepsilon \phi(\hat{\lambda}) + (1 - \varepsilon) \phi(\hat{\lambda}) = \varepsilon \ell_k'(\hat{\lambda}) + (1 - \varepsilon) \ell_k'(\hat{\lambda}) < \varepsilon \ell_k'(\hat{\lambda}) + (1 - \varepsilon) \hat{\lambda} = \ell_{k'}(\hat{\lambda}), \tag{4}
\]

which contradicts the known relation \( \phi(\hat{\lambda}) = \ell_k(\hat{\lambda}) \). Hence (3) must hold. As the above \( \hat{\lambda} \) is arbitrarily chosen, from (3) we can further get (2). Thus the theorem is proved. \( \square \)

According to the above theorem, and using the separability of the objective function, we can partition \( D \) into a number of smaller nonoverlapping hypercubes and approximate each \( f_j \) on every hypercube by a convex CPWL function like
that in (2). Specifically speaking, let \( \hat{D}_i \subset D, 1 \leq i \leq L \) be these hypercubes, \( l_{j,k}, 1 \leq k \leq m_j \) be the univariate linear functions used to approximate \( f_j \) on \( [l_j, u_j] \) for all \( 1 \leq j \leq n \), and \( s_{j,i} \subset \{1, 2, \ldots, m_j\} \) be the index set of linear functions among \( l_{j,k}, 1 \leq k \leq m_j \) used for approximation of \( f_j \) on \( \hat{D}_i \), then

\[
D = \bigcup_{i=1}^{L} \hat{D}_i, \tag{5}
\]

and the CPWL approximation of \( f \) on \( D \) is

\[
\psi(x) = \sum_{j=1}^{n} \max_{k \in s_{j,i}} \ell_{j,k}(x_j) \quad \forall x \in \hat{D}_i, \quad 1 \leq i \leq L. \tag{6}
\]

As will become clear in the next section, based on this expression an approximately globally optimal solution of the original separable programming problem can be found by solving just \( L \) linear programming problems.

### 3. Optimization model

Once the nonlinear objective function of (1) is approximated by the CPWL function \( \psi \) defined in the last section, an approximately globally optimal solution can be obtained by solving the following mathematical programming problem,

\[
\min \psi(x) \\
\text{s.t.} \quad a_r^T x \leq b_r, \quad 1 \leq r \leq m \\
x \in D. \tag{7}
\]

As \( D \) is the union of \( \hat{D}_i, 1 \leq i \leq L \), a globally optimal solution of the above problem can be found by solving \( L \) convex programming problems like the following one

\[
\min \sum_{j=1}^{n} \max_{k \in s_{j,i}} \ell_{j,k}(x_j) \\
\text{s.t.} \quad a_r^T x \leq b_r, \quad 1 \leq r \leq m \\
x \in \hat{D}_i. \tag{8}
\]
An important reason for adopting the above approach is that an optimal solution of the mathematical programming problem (8) can be obtained by solving the following linear programming problem,

\[
\begin{align*}
\min & \sum_{j=1}^{n} z_j \\
\text{s.t.} & \quad z_j \geq \ell_{j,k}(x_j), \quad k \in s_{j,i}, 1 \leq j \leq n \\
& \quad a_r^T x \leq b_r, \quad 1 \leq r \leq m \\
& \quad x \in \hat{D}_i.
\end{align*}
\]  

(9)

It can be explained as follows. If \( y^* = (x_1^*, \ldots, x_n^*, z_1^*, \ldots, z_n^*) \) is an optimal solution of the above linear programming problem, there must be \( z_j^* = \max_{k \in s_{j,i}} \ell_{j,k}(x_j^*) \) \( \forall 1 \leq j \leq n \). Let \( \hat{x} = (\hat{x}_1, \ldots, \hat{x}_n) \) be an arbitrary feasible solution of the mathematical programming problem (9). It is clear that \( \hat{y} = (\hat{x}_1, \ldots, \hat{x}_n, \hat{z}_1, \ldots, \hat{z}_n) \) with \( \hat{z}_j = \max_{k \in s_{j,i}} \ell_{j,k}(\hat{x}_j) \) is also a feasible solution of the linear programming problem (9). Hence \( \sum_{j=1}^{n} z_j^* \leq \sum_{j=1}^{n} \hat{z}_j \), which means \( (x_1^*, \ldots, x_n^*) \) is just the optimal solution of the mathematical programming problem (8).

It is worth emphasizing that the above approach produces only a global optimal solution of (7), its performance as a solution of the original problem is closely related to the accuracy of CPWL approximation. Obviously, the accuracy can be guaranteed by selecting sufficiently many breakpoints in each \([l_j, u_j]\). As each \( f_j \) is a known univariate function, it is not hard to determine an upper bound on the approximation error utilizing the knowledge of \( f_j \). We may set up a reasonable threshold on the approximation accuracy and choose as less as possible breakpoints accordingly. It should be noted that the number of the LPs to be solved does not increase as more breakpoints are added inside any interval of convexity provided the finer approximation remains so. Hence, a very accurate globally optimal solution may be obtained without solving many LPs. This is very useful for global optimization based on CPWL approximation.

4. A numerical example

Now we use an example to illustrate the practicability of the approach proposed.

Example 1. (Taken from Li [2]).

\[
\begin{align*}
\min & \quad x_1^3 - 4x_1^2 + 2x_1 + x_2^3 - 4x_2^2 + 3x_2 \\
\text{s.t.} & \quad 3x_1 + 2x_2 \leq 11.75, \\
& \quad 2x_1 + 5x_2^{0.5} - x_2 \geq 9, \\
& \quad 0 \leq x_1 \leq 5, 0 \leq x_2 \leq 4.
\end{align*}
\]  

(10)

We also select the same breakpoints for \( f_1(x_1) = x_1^3 - 4x_1^2 + 2x_1 \) and \( f_2(x_2) = x_2^3 - 4x_2^2 + 3x_2 \) as [2]. The corresponding univariate CPWL functions \( \varphi_1 \) and \( \varphi_2 \) are drawn in Fig. 2. Thus, the whole domain is partitioned into 60 smaller hypercubes.

The second constraint in this example is nonlinear, which differs from our model. In [2], this constraint is replaced by

\[
2x_1 + 4x_2 - \frac{3.3334}{2} (|x_2 - 1| + x_2 - 1) \geq 9.
\]  

(11)

We can write the inequality (11) in an equivalent form

\[
2x_1 + 4x_2 - 3.3334 \max \{0, x_2 - 1\} \geq 9.
\]  

(12)
Clearly, the above nonlinear constraint (12) equals two linear constraints

\[
\begin{align*}
2x_1 + 4x_2 - 9 & \geq 0, \\
2x_1 + 4x_2 - 9 & \geq 3.3334(x_2 - 1)
\end{align*}
\]  

which means our approach presented in this paper is suitable for this example.

Since \( \varphi_1 \) is convex on \([0, 0.5], [0.5, 1], [1, 5] \) and \( \varphi_2 \) is convex on \([0, 0.32], [0.32, 4] \), we only need to solve 6 LPs, which is much less than the number of small hypercubes. Finally, we get the same globally optimal solution \( x^* = (2.3833, 2.3000) \) as that obtained in [2].

5. Conclusions

A new approach is proposed for global optimization of separable nonlinear programming problems under linear constraints. It may decrease a large number of redundant LPs when the univariate functions can be approximated by a low number of convex pieces. In addition, this result is quite useful for choosing breakpoints. Obviously, the accuracy of the piecewise-linear approximation heavily depends on the number of breakpoints. As mentioned in Section 3, we may choose many breakpoints in convex regions without increasing the computational effort greatly. This is quite different from other methods.
References