Journal of Global Optimization, , 1–9 (2004)

Global Minimization via Piecewise-Linear Underestimation

O. L. MANGASARIAN olvi@cs.wisc.edu
Computer Sciences Department
University of Wisconsin
Madison, WI 53706
Department of Mathematics
University of California at San Diego
La Jolla, CA 92093

J. B. ROSEN jbrosen@cs.ucsd.edu
Department of Computer Science and Engineering
University of California at San Diego
La Jolla, CA 92093

M. E. THOMPSON thompson@cs.wisc.edu
Computer Sciences Department
University of Wisconsin
Madison, WI 53706

Received June 2003; Revised February 2004

Editor: Panos Pardalos

Abstract. Given a function on R^n with many multiple local minima we approximate it from below, via concave minimization, with a piecewise-linear convex function by using sample points from the given function. The piecewise-linear function is then minimized using a single linear program to obtain an approximation to the global minimum of the original function. Successive shrinking of the original search region to which this procedure is applied leads to fairly accurate estimates, within 0.57%, of the global minima of synthetic nonconvex piecewise-quadratic functions for which the global minima are known exactly.

1. Introduction

Unconstrained minimization of nonconvex functions with multiple minima plays a key role in computational biology problems such as protein docking [6]. Although the problem is NP-hard, computational approaches such as convex global underestimation (CGU) [1, 7, 9] have been quite promising. In these methods, the nonconvex function is successively bounded below by a strictly convex quadratic function whose minimum gives improved estimates of the global minimum of the nonconvex function. In the present approach we use instead a piecewise-linear approximation to underestimate the nonconvex function instead of a strictly convex quadratic function. Our piecewise-linear approximation turns out to be automatically convex, and hence it is very easy to find its global minimum. Initially, both the present approach and that of [9] require the minimization of a concave function on a polyhedral set, in order to obtain an underestimate to the nonconvex function to be minimized. However, enforcing positive definiteness on the matrix defining a quadratic underestimate in [9] changes the polyhedral constraints to convex constraints. In contrast, in the present approach, polyhedrality of the constraints remains unaltered and plays a crucial role in proving finite termination of our algorithm. Related results on piecewise convex maximization appear in [2].

We outline the contents of the paper now. In Section 2 we formulate our piecewise-linear underestimation as a concave minimization problem, or more specifically as the minimization of a piecewise-linear concave function on a polyhedral set. We also establish the existence of a vertex solution to the problem. In Section 3 we state our successive linearization algorithm (SLA) and establish its termination in a finite number of steps at a vertex satisfying a necessary optimality condition. In Section 4 we describe the generation of our synthetic nonconvex piecewise-quadratic function with multiple local minima for which the global minimum solution is known. In Section 5 we give results for numerical testing of our algorithm on various sized nonconvex synthetic test problems including a model for a protein docking problem. In all instances tested, the global minimum value was attained within 0.5%. Section 6 concludes the paper.

A word about our notation and background material. All vectors will be column vectors unless transposed to a row vector by a prime superscript $'$. The scalar product of two vectors x and y in the n-dimensional real space R^n will be denoted by $x'y$. For $x \in R^n$ and $p \in [1, \infty)$, the norm $\|x\|_p$ will denote the p-norm: $(\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$ and $\|x\|_\infty$ will denote $\max_{1 \leq i \leq n} |x_i|$. For $x \in R^n$, $|x|_i = |x_i|$, $i = 1, \ldots, n$. For an $m \times n$ matrix A, A_i will denote the ith row of A, $A_{\cdot j}$ will denote the jth column of A and A_{ij} will denote the element in row i and column j. The identity matrix in a real space of arbitrary dimension will be denoted by I, while a column vector of ones of arbitrary dimension will be denoted by e. Finally, as a matter of terminology, we note that our underestimating function p of (4) is, strictly speaking, a piecewise-*affine* function. However, it is standard practice in mathematical programming to refer to an affine function such as $ax + b$, where a and b are constants, as a linear function instead of an affine function. We have followed this practice here.

2. Piecewise-Linear Underestimation via Successive Linear Approximation

The problem we are interested in is to find the global minimum of a function $f : R^n \longrightarrow R$, given m function evaluations of $f(x)$, that is:

$$y^k = f(x^k), \ k = 1, \ldots, m. \tag{1}$$

In [9] a strictly convex quadratic underestimate:

$$q(\alpha, c, H; x) = \alpha + c'x + \frac{1}{2}x'Hx, \ H \text{ symmetric positive definite}, \tag{2}$$

is first obtained by solving the mathematical program:

$$\begin{array}{lrcl} \displaystyle\min_{\alpha,c,H} & \displaystyle\sum_{k=1}^{m}(y^k - q(\alpha,c,H;x^k)) & & \\ \text{s.t.} & q(\alpha,c,H;x^k) & \leq & y^k,\ k=1,\ldots,m, \\ & H & & \text{symmetric positive definite,} \end{array} \tag{3}$$

where $\alpha \in R$, $c \in R^n$ and $H \in R^{n\times n}$, and then minimizing $q(\alpha,c,H;x)$ over $x \in R^n$. The nonlinear positive definiteness constraint in (3) complicates this otherwise linear formulation. Furthermore a single strongly convex quadratic function, such as $q(\alpha,c,H;x)$, might not closely underestimate $f(x)$. We propose here instead the following piecewise-linear function:

$$p(\alpha,c,A,b;x) = \alpha + c'x + \|Ax+b\|_1, \tag{4}$$

where $\alpha \in R$, $c \in R^n$, $A \in R^{\ell\times n}$ and $b \in R^{\ell}$, where ℓ is the number of linear functions generating our piecewise-linear underestimation. We note immediately that $p(\alpha,c,A,b;x)$ is convex and piecewise-linear in x for fixed (α,c,A,b) and similarly it is convex and piecewise-linear in (α,c,A,b) for fixed x. Our approximation problem for an underestimator of $f(x)$ is then the following concave minimization problem.

$$\begin{array}{lrcl} \displaystyle\min_{\alpha,c,A,b} & \displaystyle\sum_{k=1}^{m}(y^k - p(\alpha,c,A,b;x^k)) & & \\ \text{s.t.} & p(\alpha,c,A,b;x^k) & \leq & y^k,\ k=1,\ldots,m. \end{array} \tag{5}$$

Note that unlike the mathematical program (3), there are absolutely no constraints whatsoever on the matrix A. By using nonlinear perturbation theory of linear programs [5], the minimization problem (5) can be rewritten as the following piecewise-linear concave minimization on a polyhedral set:

$$\begin{array}{lrcl} \displaystyle\min_{\alpha,c,A,b,s^k} & \displaystyle\sum_{k=1}^{m}(y^k - (\alpha + c'x^k + e'|Ax^k+b|) + \epsilon e's^k) & & \\ \text{s.t.} & \alpha + c'x^k + e's^k & \leq & y^k,\ k=1,\ldots,m \\ & -s^k \leq Ax^k + b & \leq & s^k,\ k=1,\ldots,m, \end{array} \tag{6}$$

where $\epsilon \in (0,\bar{\epsilon}]$ for some $\bar{\epsilon} > 0$, and $s^k \in R^{\ell}$, $k=1,\ldots,m$. A principal advantage of this formulation is that a global solution exists at a vertex of the nonempty feasible region as follows.

Proposition 1 Existence of a Vertex Solution *Under the assumption that the feasible region of (6) has no straight lines that go to infinity in both directions, a vertex solution to the minimization problem exists.*

Proof This proposition follows directly from [8, Corollary 32.3.4] by noting that the objective function is concave and bounded below on the nonempty polyhedral feasible region. To show that the feasible region is nonempty, take $c=0$, $A=0$, $b=0$, $s^k=0$, $k=1,\ldots,m$ and $\alpha = \min_{k=1,\ldots,m} y^k$. To show that the objective function is bounded below by zero, note that from the last set of constraints of (6), we have that for $k = 1, \dots, m$:

$$e's^k \geq e'|Ax^k + b|,\ s^k \geq 0, \tag{7}$$

and hence:

$$y^k - (\alpha + c'x^k + e'|Ax^k + b|) + \epsilon e's^k \geq y^k - (\alpha + c'x^k + e's^k) + \epsilon e's^k \geq 0. \tag{8}$$

□

Remark 2 *The assumption that the feasible region does not contain straight lines that go to infinity in both directions is easily implemented by making all the variables of the problem nonnegative. This is achieved by redefining the variables of the problem as the difference of new nonnegative variables and a single nonnegative scalar variable. This increases the dimensionality of the problem by one. However, in actual computational applications, this transformation is not needed in order that our successive linear approximation algorithm terminate in a finite number of steps.*

We shall solve the problem (6) by a successive linearization algorithm described in the next section. Once that is done, the piecewise-linear function (4) can be minimized as a linear program as follows:

$$\begin{array}{lrcl} \displaystyle\min_{x,t} & & \alpha + c'x + e't & \\ \text{s.t.} & -t \leq & Ax + b & \leq t, \\ & a^1 \leq & x & \leq a^2, \end{array} \tag{9}$$

where a^1, a^2 are usually known bounds on the variables of the problem.

We are ready now to turn to our algorithm and establish its finite termination.

3. The Finite Successive Linear Approximation Algorithm (SLA)

We shall use the stepless generalization [4] of the Frank-Wolfe algorithm [3] for minimizing nondifferentiable concave functions on polyhedral sets. Finite termination of the generalized Frank-Wolfe algorithm at a vertex satisfying the minimum principle necessary optimality condition is given in [4, Theorem 3]. To state our SLA we need to evaluate the subgradient of the objective function of (6), which we do now. We shall collectively represent the optimization variables of (6) by $z \in R^{1+n+\ell n+\ell+\ell m}$ as follows:

$$z = \begin{bmatrix} \alpha \\ c \\ {A_i}', i = 1, \dots, \ell \\ b \\ s^k, k = 1, \dots, m \end{bmatrix} \tag{10}$$

If we denote the objective function of (6) by $\theta(z)$ and the feasible region of (6) by Z, then (6) can be written as:

$$\min_{z \in Z} \theta(z). \tag{11}$$

The subgradient of $\theta(z)$ with respect to z as defined in (10), is given by:

$$\partial\theta(z) = \sum_{k=1}^{m} - \begin{bmatrix} 1 \\ x^k \\ \partial|A_i x^k + b_i| x^k, i = 1,\ldots,\ell \\ \partial|A_i x^k + b_i|, i = 1,\ldots,\ell \\ -\frac{1}{m}\epsilon e \end{bmatrix}. \tag{12}$$

where $e \in R^{\ell m}$ and, for $i = 1,\ldots,\ell$, $k = 1,\ldots,m$:

$$\partial|A_i x^k + b_i| = \begin{bmatrix} 1 & \text{if} \quad A_i x^k + b_i > 0 \\ \in [-1,1] & \text{if} \quad A_i x^k + b_i = 0 \\ -1 & \text{if} \quad A_i x^k + b_i < 0 \end{bmatrix}. \tag{13}$$

We are ready now to state our algorithm.

Algorithm 1 Successive Linearization Algorithm (SLA) *Start with a random $z^0 \in R^{1+n+\ell n+\ell+\ell m}$. Having z^t determine z^{t+1} as a vertex solution of the linear program:*

$$\min_{z\in Z} \partial\theta(z^t)'(z - z^t). \tag{14}$$

Stop when $\partial\theta(z^t)'(z^{t+1} - z^t) = 0$.

This algorithm terminates in a finite number of steps as follows.

Proposition 2 Finite Termination of SLA *Under the assumption that the feasible region Z of (6) does not contain lines going to infinity in both directions, the SLA generates a finite sequence of feasible vertices $\{z^1, z^2, \ldots, z^{\bar{t}}\}$ of Z of strictly decreasing objective function values: $\{\theta(z^1), \theta(z^2), \ldots, \theta(z^{\bar{t}})\}$, such that $\theta(z^{\bar{t}})$ satisfies the minimum principle necessary optimality condition:*

$$\partial\theta(z^{\bar{t}})'(z - z^{\bar{t}}) \geq 0, \ \forall\, z \in Z. \tag{15}$$

Proof The proof is a direct consequence of [4, Theorem 3]. □

Once (α, c, A, b) are determined by the SLA Algorithm 1, an estimate of the global minimum solution of $f(x)$ is determined by solving the linear program (9).

We turn now to the task of generating synthetic examples that mimic the nonconvexity of computational biology minimization problems.

4. The Synthetic Nonconvex Piecewise-Quadratic Function

We will generate now a piecewise-quadratic nonconvex function as follows:

$$y(x) = \min_{j\in\{1,\ldots,r\}} h_j(x), \tag{16}$$

where $h_j(x)$, $j = 1,\ldots,r$ are arbitrary strictly convex quadratic functions, such as:

$$h_j(x) = \beta^j + {d^j}'x + \frac{1}{2}x'(0.5I + {M^j}'M^j)x, \ j = 1,\ldots,r, \tag{17}$$

Here, $\beta^j \in R$, $d^j \in R^n$ and $M^j \in R^{n\times n}$ are randomly chosen.

An interesting feature of the piecewise-quadratic function $y(x)$ generated as described above, is that its exact global minimum solution can be computed as follows.

Proposition 1 Exact Global Minimum Solution of (16)-(17) *An exact global minimum of (16)-(17) is given by:*

$$\min_{j\in\{1,\ldots,r\}} \min_{x\in R^n} h_j(x). \tag{18}$$

More specifically,

$$\min_{x\in R^n} y(x) = \min_{j\in\{1,\ldots,r\}} h_j(x^j), \tag{19}$$

where:

$$x^j = -(0.5I + {M^j}' M^j)^{-1} d^j,\ j = 1,\ldots,r. \tag{20}$$

Proof The point $x^{\bar{j}}$ such that:

$$h_{\bar{j}}(x^{\bar{j}}) = \min_{j\in\{1,\ldots,r\}} h_j(x^j), \tag{21}$$

is the desired global minimum solution of (16)-(17), because $(x^{\bar{j}}, h_{\bar{j}}(x^{\bar{j}}))$ lies on the function $y(x)$ of (16) and no other point on $y(x)$ has a lower value. □

An example of a piecewise-quadratic function in R^2, generated by (16) and made up of five pieces, that is $r = 5$, is depicted in Figure 1. Figure 1 also shows the exact global minimum computed using Proposition 1 above together with an approximate solution computed by the SLA Algorithm 1.

We turn now to our numerical implementation.

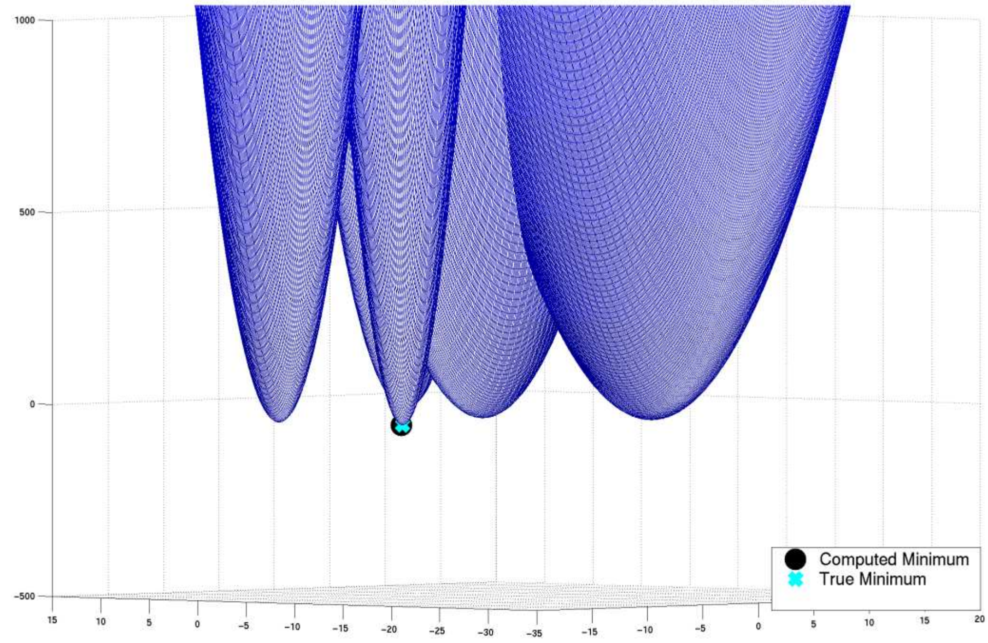

Figure 1. **A 5-piece piecewise-quadratic function $y(x)$ on R^2 defined by (16) and showing the true minimum and the computed minimum by our SLA Algorithm 1.**

5. Numerical Testing

We tested the SLA Algorithm 1 on six nonconvex piecewise-quadratic functions as defined by (16) as well as on a synthetic protein docking (SPD) problem generated from real docking data [6, 7]. For the SPD problem we used the model (16), where each $h_j(x)$ is a strictly convex quadratic function with a pre-determined minimum solution corresponding to local minima of the docking problem energy function. The dimensionality of the spaces in which the test functions were defined varied between one and six. The protein docking problem energy function [6] is defined in R^6.

For each application of the SLA Algorithm 1, we generated a grid of 3^n points spanning the search region using three values from each dimension of R^n. In addition, for each application of SLA, except the first, we included the piecewise-linear minimum solution $\bar{x}$ of the previous application as well as the grid point of the previous application of SLA that resulted in the lowest function value, $\hat{x}$, thus generating a total of $3^n + 2$ points. Each application of the SLA Algorithm 1 results in a piecewise-linear underestimator to the function values at the points in the grid, which is then minimized by the linear program (9) to find an approximate minimum of $f(x)$.

In implementing SLA Algorithm 1 we utilized the stopping criterion: $\partial\theta(z^t)'(z^{t+1}-z^t) < tol$ for some $tol > 0$, instead of $\partial\theta(z^t)'(z^{t+1}-z^t) = 0$. The parameter tol varied from 1e-1 to 1e-10 in our test cases. Also, we found the following initialization procedure helpful in implementing the SLA Algorithm 1 for solving (6). We chose as initial guesses: $\alpha = 0$, $c = 0$, while for each i, $i = 1, \ldots, \ell$, A_i, b_i, were chosen so as the plane $A_i x + b_i$ interpolated $n + 1$ points randomly chosen from the search grid of 3^n or $3^n + 2$ points, and $s_i^k = |A_i x^k + b_i|$, $k = 1, \ldots, m$. This resulted in better solutions than a strictly random initialization of $\alpha, c, A, b, s^1, \ldots, s^m$. Also, it was found that taking ℓ, the number of the linear pieces of the piecewise-linear underestimator (4), to be $3n$, gave a fairly accurate underestimator.

Once $\bar{x}$ was obtained as a solution to (9), we refined the search region. When $\bar{x}$ was strictly interior to the search region, we re-centered the search region around the mean of $\bar{x}$ and $\hat{x}$, as defined above, and decreased the size of each dimension of the search region by a specified refinement rate. When $\bar{x}$ was not strictly interior to the search region, we shifted the search region in the direction of $\bar{x}$ for each $j, j = 1, \ldots, n$ where $\bar{x}_j = a_j^1$ or $\bar{x}_j = a_j^2$. In addition, if the selected refinement resulted in $\bar{x}$ or $\hat{x}$ not being in the new search region, we uniformly expanded the region to include both points.

We continued this process until the change in the 1-norm distance between successive $\bar{x}$ iterates was less than a specified tolerance. The approximate minimum value of the test function $f(x)$ was calculated as $\alpha + c'\bar{x} + \|A\bar{x} + b\|_1$.

Our computations were performed on machines utilizing an 800 Mhz Pentium III processor and 256MB of memory running on Redhat Linux 7.2, with MATLAB 6.5 installed. The results are presented in Table 1.

6. Conclusion

We have proposed a method for finding an accurate estimate of the global minimum of a nonconvex function by underestimating the function by a piecewise-linear convex function and then finding the global minimum of the underestimator. The method gives accurate estimates of the global minima for a class of synthetic nonconvex piecewise-quadratic functions that closely model protein docking problems. An interesting problem for future consideration is that of approximating nonconvex protein docking energy functions by our synthetic piecewise-quadratic function for which the present approach provides an accurate estimate of the global minimum.

Acknowledgments

The research described in this Data Mining Institute Report 03-03, June 2003, was supported by National Science Foundation Grants CCR-0138308, ITR-0082146 and by the Microsoft Corporation.

Table 1. **Numerical test results on six synthetic test problems and the synthetic protein docking (SPD) problem. Note relative errors in minimum values are less than 0.57% and 1-norm relative errors in solution points are less than 0.44%.**

	Synthetic Problems						SPD
n	1	2	3	4	5	6	6
$\ell = 3n$	3	6	9	12	15	18	18
m	3	9	27	81	243	729	729
Refinement rate	0.5	0.5	0.5	0.5	0.25	0.25	0.5
Iterate tolerance	0.1	0.1	0.1	0.1	0.5	0.5	0.1
SLA tolerance *tol*	1e-10	1e-10	1e-10	1e-5	1e-1	1e-1	1e-2
ϵ	1e-10	1e-10	1e-10	1e-10	1e-5	1e-5	1e-5
True min	-1708.1	-1499.9	-3264.7	-705.0	-7351.0	-11375.6	-42.7
Computed min	-1708.1	-1506.6	-3265.7	-705.0	-7392.5	-11422.9	-42.7
Error in min	0.0	-6.7	-1.0	0.0	-41.5	-47.3	0.0
%Error in min	0.000%	-0.446%	-0.031%	0.005%	-0.565%	-0.416%	0.014%
Error in soln (1-norm)	0.002	0.110	0.177	0.114	0.818	1.208	0.062
%Error in soln	0.003%	0.091%	0.110%	0.057%	0.341%	0.431%	0.083%
No. refinements	25	39	54	31	33	171	26
Time (sec)	1.4	12.7	106.4	1032.1	387.3	16045.1	5846.9
Time per refinement	0.1	0.3	1.9	33.1	11.5	93.2	224.4
No. LPs per refinement	2.1	3.4	9.7	11.5	16.5	11.7	15.0

References

1. K. A. Dill, A. T. Phillips, and J. B. Rosen. CGU: An algorithm for molecular structure prediction. In L. T. Biegler et al, editor, *IMA Volumes in Mathematics and its Applications: Large Scale Optimization with Applications III: Molecular Structure and Optimization*, pages 1–22, 1997.
2. D. Fortin and I. Tsevendorj. Piecewise convex maximization problems: Algorithm and computational experiments. *Journal of Global Optimization*, 24:61–77, 2002.
3. M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.
4. O. L. Mangasarian. Solution of general linear complementarity problems via nondifferentiable concave minimization. *Acta Mathematica Vietnamica*, 22(1):199–205, 1997. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/96-10.ps.
5. O. L. Mangasarian and R. R. Meyer. Nonlinear perturbation of linear programs. *SIAM Journal on Control and Optimization*, 17(6):745–752, November 1979.
6. J. C. Mitchell, A. T. Phillips, J. B. Rosen, and L. F. Ten Eyck. Coupled optimization in protein docking. In *Optimization in Computational Chemistry and Molecular Biology*, pages 191–207, Dordrecht, Netherlands, 2000. Kluwer Academic Publishers.
7. A. T. Phillips, J. B. Rosen, and K. A. Dill. Convex global underestimation for molecular stucture prediction. In P. M. Pardalos et al, editor, *From Local to Global Optimization*, pages 1–18, Dordrecht, Netherlands, 2001. Kluwer Academic Publishers.
8. R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.
9. J. B. Rosen and R. F. Marcia. Convex quadratic approximation. *Computational Optimization and Applications*, 8(3), 2004.