Cutting Plane Methods for Nondifferentiable Optimization

Samir Elhedhli
GERAD/Faculty of Management
McGill University, Montreal CA
samir@management.mcgill.ca

Jean-Louis Goffin
GERAD/Faculty of Management
McGill University, Montreal CA
goffin@management.mcgill.ca

Jean-Philippe Vial
Logilab/ Department of Management Studies
Université de Genève, Geneva, Switzerland
jean-philippe.vial@hec.unige.ch

March, 2000

Les Cahiers du GERAD

G–2000–13

Copyright © 2000 GERAD

Abstract

This article was written for the *Encyclopedia of Optimization.*

Key Words: Nondifferentiable Optimization, Nonsmooth Optimization, Cutting Planes

Résumé

Cet article a été écrit pour: *Encyclopedia of Optimization.*

1 Introduction

Cutting plane methods were proposed independently by Kelley [27] and Cheney and Goldstein [5] as a solution technique for constrained convex optimization problems. Although they may not compete with some of the efficient methods for smooth optimization, they are one of the first and fundamental solution approaches for *nondifferentiable optimization*.

The fact that they rely on polyhedral approximations of convex functions, makes the technique suitable for nondifferentiable optimization. It forms with subgradient optimization [39] the two major solution approaches for nondifferentiable problems. The cutting plane algorithms that are being designed over the years are getting more sophisticated and ultimately, showing promising computational results and stable numerical properties.

In this article, we give an overview of cutting plane methods for nondifferentiable optimization problems. We start with a brief introduction to cutting planes. We then give a generic cutting plane algorithm, that will be used to describe some of the main variants.

2 A generic cutting plane algorithm

To describe the general cutting plane approach, we consider the following nondifferentiable problem

$$[NDP]: \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \end{array}$$

where f and g are real-valued, continuous, nondifferentiable, convex functions. Convexity implies that there is at least one *supporting hyperplane* to f at every point x_0 of the domain, whose equation is given by

$$y = f(x_0) + \xi_0^f (x - x_0),$$

where ξ_0^f is any element of the *subdifferential* $\partial f(x_0)$ of f at x_0. (For ease of notation we assume that subnradients are row vectors.) Recalling that a supporting hyperplane gives an under-estimates of f, the *subgradient inequality*

$$f(x_0) + \xi_0^f (x - x_0) \leq f(x) \tag{1}$$

can be used to approximate f by the maximum of a set of piecewise linear functions. Therefore, given a set of points x_i, $i \in I$ and their corresponding subgradients ξ_i^f, f is tangentially approximated by

$$\bar{f}(x) = \max_{i \in I} \left\{ f(x_i) + \xi_i^f (x - x_i) \right\}. \tag{2}$$

Inequality (1) implies that $\bar{f}(x) \leq f(x)$ for any index set I. Larger sets will give better approximations. The same can be said about g, leading to the polyhedral approximation

$$\bar{g}(x) = \max_{j \in J} g(x_j) + \xi_j^g (x - x_j).$$

where ξ_j^g are subgradients of g at a collection of points x_j indexed by J. Thus, $[NDP]$ can be approximated by,

$$\begin{array}{ll} \min & \left\{ \max_{i \in I} f(x_i) + \xi_i^f (x - x_i) \right\} \\ \text{s.t.} & \max_{j \in J} g(x_i) + \xi_j^g (x - x_j) \leq 0, \end{array}$$

which is equivalent to,

$$\begin{array}{lll} \min & v & \\ s.t. & f(x_i) + \xi_i^f (x - x_i) \leq v & \forall i \in I \\ & g(x_j) + \xi_j^g (x - x_j) \leq 0 & \forall j \in J \end{array} \tag{3}$$

Problem (3) is a linear program that is easier to deal with than the original problem. It is to note, however, that this is only an approximation of $[NDP]$, that gets better as more constraints are added. Let us denote by $[MP_k]$ the *relaxed master problem* (3) with index sets I_k and J_k whose optimal solution is denoted by (x_k, v_k).

By transforming $[NDP]$ into (3), nondifferentiability is eliminated at the cost of having a problem with a very large number of constraints. To get around that, cutting plane methods use only a subset of these constraints and generate the rest as needed. In fact, they would handle a series of relaxed master problems $[MP_k]$ and stop when an optimal (satisfactory) solution to $[NDP]$ is reached. $[MP_k]$ is a relaxation of $[NDP]$ as, by convexity of g,

$$\left\{ x : \max_{j \in J_k} \{ g(x_j) + \xi_j^g (x - x_j) \} \leq 0 \right\}$$

contains $\{x : g(x) \leq 0\}$, while, by convexity of f,

$$\max_{i \in I_k} \{ f(x_i) + \xi_i^f (x - x_i) \} \leq f(x)$$

for all $x \in \{x : g(x) \leq 0\}$. This relaxation gets tighter as more points are included in the index sets I and J. This implies that the optimal value v_k is a lower bound on the optimal value $f(x^*)$. This relaxation gets tighter as more points are included in the index sets I and J. The hope is that termination occurs before the sets I and J get enormously big.

The idea described above can be put into the following generic cutting plane algorithm

- *Initialize:*
  1. Get an initial upper bound v_u and a lower bound v_l for the optimal solution $f(x^*)$.
  2. Get an initial relaxed master problem $[MP_0]$.

- *Iterate* (k)

  1. Get a *query point* x_k and a corresponding lower bound v_l^+
  2. If x_k is infeasible to $[NDP]$ $(g(x_k) > 0)$, then the oracle of g generates a *feasibility cut* of type $g(x_k) + \xi_j^g(x - x_k) \leq 0$.
  3. If x_k is feasible to $[NDP]$ $(g(x_k) \leq 0)$, then the oracle of f generates an *optimality cut* of type $f(x_k) + \xi_i^f(x - x_k) \leq v$ and an upper bound v_u^+.
  4. Update the bounds
     (a) if x_k is feasible to $[NDP]$, then $v_u = \max\{v_u, v_u^+\}$.
     (b) $v_l = \min\{v_l, v_l^+\}$.
  5. Update index sets
     (a) If a feasibility cut is added then $I_{k+1} := I_k$ and $J_{k+1} := J_k \cap \{k\}$.
     (b) If an optimality cut is added then $I_{k+1} := I_k \cap \{k\}$ and $J_{k+1} := J_k \cap \{k\}$.
  6. Either *Stop or* $k := k + 1$

In the course of the cutting plane algorithm, upper bounds are usually achieved through the objective evaluation at a feasible point. Lower bounds are either given by the optimal solution of the relaxed master problem, or by evaluating the dual objective at a feasible dual solution. For the stopping criteria, the algorithm may be stopped if the bound on the duality gap drops below a certain threshold. This is given by the difference between the best known upper and lower bounds.

At each iteration of the method, we can construct a bounded polyhedral set, namely the *localization set*, that is guaranteed to contain the optimal solution. Given an upper bound v_u^k to $[MP_k]$, it is given by

$$F_k(v_u^k) = \left\{ \begin{array}{l} (x, v) : v \leq v_u^k; \\ f(x_i) + \xi_i^f(x - x_i) \leq v, i \in I_k; \\ g(x_j) + \xi_j^g(x - x_j) \leq 0; j \in J_k. \end{array} \right\}$$

The localization set $F_k(v_u)$ is a bounded polyhedral subset of the feasible region of $[MP_k]$ that contains any optimal solution $(x^*, f(x^*))$ to $[NDP]$. To see that, $f(x^*) \leq v_u^k$ as v_u^k is an upper bound on the optimal value. Furthermore, by the feasibility of x^* and the subgradient inequality $g(x_j) + \xi_j^g(x^* - x_j) \leq g(x^*) \leq 0$ for all $j \in J_k$ and by the fact that $[MP_k]$ is a relaxation of $[NDP]$, $f(x_i) + \xi_i^f(x^* - x_i) \leq v_k \leq f(x^*)$ for all $i \in I_k$.

Although every cutting plane method follows the above general scheme, different query point choices produce different algorithms. Prior to discussing a few of them, we would like to stress that some methods are explicitly based on the localization set, and require that this set be bounded. This might not hold in particular in the initialization phase, since no cut has is yet present, and no upper and lower bounds on the objective are known. Therefore, one has to introduce box constraints. This may not even suffice if the first generated cut is not an optimality one, because the localization set remains unbounded in the z variable. One must then proceed with a an auxiliary phase I problem.

3 Kelley's cutting plane method

This is the classical cutting plane method that was originally described by Kelley [27] and is present in the original Dantzig-Wolfe [8], [7] and Benders [4] decompositions. At iteration k, the method solves the relaxed master $[MP_k]$ and uses its solution x_k to generate further cuts.

As $[MP_k]$ is a relaxation of $[NDP]$, v_k gives a lower bound to $f(x^*)$ which is monotonically increasing. In addition, when x_k is feasible, $f(x_k)$ gives an upper bound that, unfortunately, is not monotonically decreasing. The difference between the updated bounds gives an estimate of the duality gap and can be used as a practical stopping criteria. The exact optimal solution of $[NDP]$ is detected when x_k is feasible and $f(x_k) = v_k$, which is equivalent to having $0 \in \partial f(x_k)$.

This optimal point strategy assumes that the relaxation is a good approximation of the original problem, but this is only true when a big number of cuts has been generated. The method is globally convergent [48], but in practice, it sometimes shows a slow pattern of convergence [36].

4 Centre of Gravity Method

This the method was first proposed by Levin [33] as the first cutting-plane method that generates query points at the centre of the localization set. Choosing the *centre of gravity* to generate query points seems to be the natural choice. In fact cutting through the centre of gravity of the a localization set of volume V and dimension n produces two sets, each with a volume of at least

$$\left[1 - (1 - \frac{1}{n+1})^n\right] V.$$

Therefore, after k iterations the volume of the localization set shrinks at a constant rate of $\left(\frac{1}{1-\exp(1)}\right)$. This rate of convergence is the best that can be got [47]. Unfortunately, finding a single centre of gravity could be as difficult as solving the original problem.

For some simple convex bodies such as ellipsoids, cubes or spheres, calculating the centre of gravity is relatively easy. This idea is the motivation behind the *largest inscribed sphere method* of Elzinga and Moore [12], *the Volumetric method* of Vaidya [44] and the *Analytic centre cutting plane method* (ACCPM) of Goffin and Vial [21], [22], [24].

5 Largest Inscribed Sphere Method

The query point of this method is chosen as the centre of the largest inscribed sphere in the localization set. Its calculation is based on works of Nemhauser and Widhelm [35] who showed that the minimization of the simple linear program

$$\begin{array}{ll} \min & \sigma \\ s.t. & a_i^T x + ||a_i^T||\sigma \le b_i \quad i \in I \end{array}$$

gives the radius σ and the centre x of the largest inscribed sphere in the bounded polyhedron $\{x : a_i^T x \le b_i, \quad i \in I\}$. The method is detailed in [12].

6 Volumetric Method

Vaidya [44] proposed the volumetric centre as a query point. It is the maximizer of the determinant of the Hessian matrix of the logarithmic barrier function. This choice is motivated by the observation that for every point of the localization set, an inscribed ellipsoid can be constructed. The point that gives the maximum-volume inscribed ellipsoid is the minimizer of

$$\frac{1}{2}\log\left(\det\left(\sum \frac{a_i a_i^T}{(b_i - a_i^T x)^2}\right)\right)$$

where a_i are the columns of the matrix defining the localization set.

7 Bundle Methods

Bundle methods were first proposed by Lemaréchal [31]. The method has developed over the years building upon the pioneering works of Kiwiel [28]and Lemaréchal [32].

The method adds a regularization term to the estimation of f and solves

$$\begin{array}{ll} \min & \left\{ \max\limits_{i \in I_k} f(x_i) + \xi_i^f (x - x_i) \right. \\ & \qquad \left. + \frac{1}{2\alpha_k} \|x - x_{k-1}\|^2 \right\} \\ \text{s.t.} & \max\limits_{j \in J_k} \{g(x_j) + \xi_j^g (x - x_j)\} \le 0 \end{array} \tag{4}$$

to get the next query point x_k. The main idea is based on the fact that (2) is a good approximation of f when the next iterate is not far from the previous ones. Problem (4) is equivalent to the quadratic program

$$\begin{array}{ll} \min & v + \frac{1}{2\alpha_k} \|x - x_{k-1}\|^2 \\ \text{s.t.} & f(x_i) + \xi_i^f (x - x_i) \le v, \quad i \in I_k, \\ & g(x_j) + \xi_j^g (x - x_j) \le 0, \quad j \in J_k. \end{array}$$

The nice feature of Bundle methods is that they limit the number of hyperplanes that are used to approximate f. The set of subgradients (the bundle) is updated at each iteration and kept moderately small. As a result, (4) is solved very quickly.

Techniques to estimate α_k are treated in [29] where three different values are proposed.

8 The Analytic Centre Cutting Plane Method (ACCPM)

To overcome the difficulty associated with the calculation of centres of polyhedrons and still use a central point strategy, ACCPM uses concepts form the interior point literature that have proven to be highly efficient.

The query point for ACCPM is the *analytic center* of the localization set $F_k(v_u)$. This new notion of centre was first introduced by Sonnevend[41], [42], [43] as the the unique pair $\left(x_a^k, v_a^k\right)$ that minimizes

$$\begin{aligned}&\log(v_u^k - v) + \sum_{i\in I_k} \log[v - f(x_i) + \xi_i^f (x - x_i)]\\ &+ \sum_{i\in J_k} \log[-g(x_i) + \xi_i^g (x - x_i)].\end{aligned}$$

This function is a *potential function* similar to the one used by Karmarkar[26] when presenting the first interior point algorithm. Its minimization is equivalent to the solution of a linear program by any interior point method. The *primal projective algorithm* is favored due to its ability to deal with dual infeasibilities. It is, in principle, a modified Newton method applied to the minimization of the potential function [9].

The Newton procedure for computing a central point $\left(x_a^k, v_a^k\right)$ of $F_k(v_u)$ identifies, upon termination, a dual feasible solution to $[MP_k]$. Evaluating the dual objective at this point, gives a lower bound to the optimal solution of $[NDP]$. In addition, if x_a^k is feasible to $[NDP]$, i.e. $g(x_a^k) \le 0$, then $f(x_a^k)$ is an upper bound. Unfortunately, x_a^k never coincides with the original optimizer x^*, as it can never be a vertex, so the stopping criteria can only be based on the difference between the upper a lower bounds. The algorithm would stop if that gap is sufficiently small.

The convergence analysis of the method is done in [20] and [37]. ACCPM is readily implemented [25] and has shown promising results in solving different large-scale problems [2], [3], [19], [34]. The method was modified to use weighted analytic centres as a query point [23], to add multiple cuts at once [17], to use a primal-dual interior point algorithm for the calculation of analytic centres [11] and to accommodate quadratic cuts [10].

9 Concluding Remarks

Cutting plane methods are an effective solution approach for nondifferentiable optimization. This has proven to be true when advanced methods such as Bundle and Analytic Centre methods were designed. Not only they make use of recent advances in optimization theory, they also resulted in efficient computer implementations.

Manipulating the set of cuts is an important enhancement factor to cutting plane methods. In the addition of cuts, introducing a whole set at once is more effective than introducing single cuts as it allows the fast accumulation of information about the problem. This is possible for certain classes of problems where disaggregation is possible. This is true for Dantzig-Wolfe and Benders decomposition with primal and dual block diagonal matrices respectively.

The second improvement is in the manipulation of number of cuts in the relaxed master problem. Keeping all cuts may lead to a better approximation but it also requires huge amounts of storage space and solution time. Thus well defined cut-dropping strategies will considerably improve the performance of the method.

A final remark concerns the use of heuristics with cutting planes. They can be used to initialize the method so that sufficient cutting planes are obtained to start the method, or take over the search for optimal solutions when the method stalls.

References

[1] D. S. Atkinson and P. M. Vaidya, "A cutting plane algorithm that uses analytic centers." *Mathematical Programming, Series B* **69** (1995), 1–43.

[2] O. Bahn, J. L. Goffin, J. P. Vial, and O. Du Merle. "Implementation and behavior of an interior point cutting plane algorithm for convex programming : An application to geometric programming." *Discrete Applied Mathematics***49** (1994), 3–23.

[3] O. Bahn, O. Du Merle J. L. Goffin and J. P. Vial. "A cutting plane method from analytic centers for stochastic programming." *Mathematical Programming, Series B,* **69** (1995) 45–73, eds. J.-L. Goffin and J.-P. Vial.

[4] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems", *Numerische Mathematik* **4** (1962), 238-252.

[5] W. Cheney and A.A. Goldstein "Newton's method for convex programming and Chebyshev approximation", *Numerische Mathematik* **1-5** (1959), 253-268.

[6] F. H. Clarke, "Optimization and Nonsmooth Analysis ", *Les publications CRM, Montreal* (1989).

[7] G. B. Dantzig and P. Wolfe, "The decomposition algorithm for linear programming", *Econometrica* **29** (4), (1961), 767-778.

[8] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs", *Operations Research*, **8**, (1960) 101-111.

[9] G. de Ghellinck and J. P. Vial. "A polynomial Newton method for linear programming" *Algorithmica.* **1**, (1986),425-453.

[10] M. Denault and J.-L. Goffin, "Variational Inequalities with Quadratic Cuts", GERAD Tech. Report G-98-69, 41 pp, December (1998).

[11] M. Denault and J.-L. Goffin, "On a Primal–Dual Analytic Center Cutting Plane Method for Variational Inequalities", GERAD Tech. Report G-97-56, 26 pp, September 1997, revised August (1998), accepted for Computational Optimization and Applications.

[12] J. Elzinga and T. G. Moore, "A central cutting plane algorithm for the convex programming problem", *Mathematical Programming* **8** (1975) 134–145.

[13] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, New York, 1968. Reprint : Volume 4 of *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, PA 19104–2688, USA, 1990.

[14] M. L. Fisher, "The Lagrangian relaxation method for solving integer programming problems", *Management Science* **27** (1981) 1-18.

[15] A. M. Geoffrion, "Lagrangean relaxation for integer programming, *Mathematical Programming Study*, **2** (1974) 82-114.

[16] A. M. Geoffrion, "Generalized Benders decomposition", *Journal of Optimization Theory and Applications* **10** (1972) 237-260.

[17] J.-L. Goffin and J.-P. Vial, "Multiple cuts in the analytic center cutting plane method", HEC/Logilab Technical Report 98.10, June 1998, Logilab, Department of Management Studies, University of Geneva : also GERAD Tech. Report G-98-26.

[18] J.-L. Goffin and J.-P. Vial, "Interior point methods for nondifferentiable optimization", Operations Research Proceedings 1997 P. Kishka, H.W. Lorenz, U. Derigs, W. Domschke, P. Kleinschmidt, R. Moehring, editors, Springer-Verlag, Berlin, (1998), 35–49.

[19] J.-L. Goffin, J. Gondzio, R. Sarkissian and J.-P. Vial, "Solving nonlinear multicommodity flow problem by the analytic centre cutting plane method", *Mathematical Programming, Series B* (1997), **76** 131-154.

[20] J-L. Goffin, Z.-Q. Luo, Y. Ye, "Complexity Analysis of an Interior Cutting Plane Method for Convex Feasibility Problems", *SIAM Journal on Optimization,* **6** (1996) 638-652.

[21] J.-L. Goffin, A. Haurie, and J.-P. Vial. "Decomposition and nondifferentiable optimization with the projective algorithm." *Management Science*, **38 (2)**, (1992), 284-302.

[22] J.-L. Goffin, A. Haurie, J.-P. Vial, and D. L. Zhu. "Using central prices in the decomposition of linear programs". *European Journal of Operational Research,* **64**, (1993), 393-409.

[23] J.-L. Goffin and J.-P. Vial. "On the computation of weighted analytic centers and dual ellipsoids with the projective algorithm." *Mathematical programming*, **60**,(1993) 81–92.

[24] J.-L. Goffin and J.-P. Vial. "Cutting planes and column generation techniques with the projective algorithm" *Journal of Optimization Theory and Applications*, **65**, (1990) 409–429.

[25] J. Gondzio, O. du Merle, R. Sarkissian and J. P. Vial "ACCPM-A library for convex optimization based on analytic centre cutting plane method" *European Journal of Operational Research*, **94**, (1994) 206-211.

[26] N. K. Karmarkar. "A new polynomial–time algorithm for linear programming" *Combinatorica*, **4**, (1984), 373–395.

[27] J. E. Kelley, "The cutting plane method for solving convex programs", *Journal of the SIAM* **8** (1960), 703-712.

[28] K. C. Kiwiel, "Methods of Descent for Nondifferentiable Optimization", lecture Notes in Mathematics 1133, Springer-Verlag, 1985.

[29] K. C. Kiwiel, "Proximity control in Bundle methods for convex nondifferentiable optimization", *Mathematical Programming* **46** (1990), 105-122.

[30] L. Lasdon, *Optimization Theory for Large Scale Systems* , MacMillan, (1970).

[31] C. Lemaréchal, "Bundle Methods in Nonsmooth Optimization", in Nonsmooth Optimization, Proceedings of the IIASA Workshop March 28 - April 8,1977, Lemaréchal, C., and Mifflin, R., eds. Pergamon Press, (1978).

[32] C. Lemaréchal, A. Nemrirovskii and Y. Nesterov, "New variants of bundle methods", in *Nonsmooth Optimization*, *Mathematical Programming* **69** (1995), 111–147.

[33] A. Y. Levin, " On an algorithm for the minimization of convex functions over convex functions", *Soviet Mathematical Doklady* **6** (1965), 286–290.

[34] O. du Merle, P. Hansen, B. Jaumard and N. Mladenovic, "An interior point algorithm for minimum sum of squares clustering" *GERAD technical Report G97-53* (1997).

[35] G. L. Nemhauser, W. B. Widhelm, " A modified linear program for columnar methods in mathematical programming", *Operations Research* **19** (1971),1051–1060.

[36] A. S. Nemirovskii and D. B. Yudin, "Problem complexity and method efficiency in optimization", John Wiley, Chichester (1983).

[37] Y. Nesterov "Cutting plane methods from analytic centers: efficiency estimates ", *Mathematical programming, Series B,* **69**,(1996), 149–176.

[38] C. Roos, T. Terlaky and J.-P. Vial *Interior Point Methods: Theory and Algorithms* springer Verlag 1997.

[39] N. Z. Shor, "Subgradient methods: A survey of Soviet research", *Nonsmooth optimization: Proceedings of the IIASA workshop March 28–April 8,1977* C. Lemaréchal and R. Mifflin eds. Pergamon Press 1978.

[40] N. Z. Shor, *Minimization Methods for Non-differentiable Functions* (in Russian), Naukova Dumka, Kiev, 1979 [English translation: Springer, Berlin, 1985].

[41] G. Sonnevend, "An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming", in *Lecture Notes in Control and Information Sciences 84* (Springer, New York, 1985) 866-876.

[42] G. Sonnevend. "New algorithms in convex programming based on a notion of *centre* (for systems of analytic inequalities) and on rational extrapolation. " In K. H. Hoffmann, J. B. Hiriat-Urruty, C. Lemaréchal, and J. Zowe, editors, *Trends in Mathematical Optimization : Proceedings of the 4th French–German Conference on Optimization in Irsee, West–Germany, April 1986*, volume 84 of *International Series of Numerical Mathematics*, pages 311–327. Birkhäuser Verlag, Basel, Switzerland, 1988.

[43] G. Sonnevend. "Applications of the notion of analytic center in approximation (estimation) problem." *Journal of Computational and Applied Mathematics*, 28:349–358, 1989.

[44] P. Vaidya. "A new algorithm for minimizing convex functions over convex sets" *Mathematical programming*, **73** (1996) 291–341.

[45] P. Wolfe. "A method of conjugate subgradients for minimizing nondifferentiable functions," *Mathematical programming study*, **3** (1975) 145–173.

[46] Y. Ye. "A potential reduction algorithm allowing column generation,"*SIAM Journal on Optimization*, **2** (1992) 7–20.

[47] Y. Ye. "Interior point algorithms: Theory and Analysis" *John Wiley & sons*, 1997.

[48] W. I. Zangwill "Nonlinear Programming: A unified approach" *Prentice Hall*, 1969.