NONSMOOTH OPTIMIZATION

C. Lemaréchal and F. Oustry
Inria Rhône-Alpes, Montbonnnot, Saint Ismier, France

J. Zowe
Sieglitzhoferstrasse 53, Erlangen, Germany.

Keywords: optimization, large-scale systems, SDP optimization, Lagrangian relaxation, duality, convexity, bundle method

Contents

Summary

Nonsmooth optimization deals with functions whose first derivatives are not continuous. Many such optimization problems come from Operations Research; they are often large-scale, involving many variables and/or constraints. These problems necessitate special algorithms. After a short introduction, we state the framework and mention the main class of applications: Lagrangian relaxation, or duality. Then, we explain the leading principles to design suitable algorithms, more precisely bundle methods. Finally, we give some illustrative examples.

1. Introduction

Optimization problems are often encountered, in which the objective f has discontinuous derivatives; they call for nonsmooth optimization. (For a discussion of methods for solving optimization problems with differentiable functions see *Nonlinear Programming*.) In fact, classical gradient methods fail in a nonsmooth context. First of all, there is usually no derivative at an optimum point, so that the standard optimality condition $\nabla f(x) = 0$ becomes meaningless. Besides, the information contained in gradients becomes less and less relevant when approaching a point of nondifferentiability.

A way out of this dilemma is to work at each iterate with some or all previously computed gradients. This simple idea of bundling the information is the basic ingredient of all so-called bundle methods, which form an important class of nonsmooth optimization methods.

The reader is warned against a common belief, that in nonsmooth (nondifferentiable) optimization, derivatives are not used – and not computed. Actually, and this might seem paradoxical, nonsmooth optimization makes a heavy use of gradients (or rather subgradients) which, another paradox, are usually very easy to compute.

We restrict ourselves to convex f, a situation in which things are more easily explained. For similar reasons, we consider only unconstrained problems. It should be mentioned, however, that a complete theory has been developed for bundle methods, which does not require convexity from f (although the simultaneous presence of nonsmoothness and nonconvexity results in substantially less efficient algorithms).

In section 2, we state the general problem and one of its main motivations: Lagrangian relaxation, or duality, a very useful methodology in various branches of applied mathematics. The algorithmic part makes up section 3, where we give a condensed introduction to bundle methods. Section 4, illustrates the use of nonsmooth optimization codes.

2. The General Problem and Its Motivation

We start this section by the general statement of a convex nonsmooth optimization problem. Then we present Lagrangian relaxation, which is by far the main source for such problems.

2.3 The Nonsmooth Problem

Throughout the following, f is a convex functional on the n-dimensional Euclidean space $\mathbb{R}^n$ and we study the problem

$$\text{minimize } f(x) \text{ on } \mathbb{R}^n \tag{1}$$

There are no constraints; but, in contrast with the standard situation, we do not require f to be smooth. The *subdifferential*

$$\partial f(x) = \{s \in \mathbb{R}^n \mid \langle s, z - x\rangle \le f(z) - f(x) \text{ for all } z \in \mathbb{R}^n\} \tag{2}$$

of the convex f at x will serve as substitute for the gradient. This $\partial f(x)$ is a non-empty, convex and compact set, which shrinks to the gradient $\nabla f(x)$ whenever f is differentiable at x. We make the general assumption:

$$\text{at every } x \text{, we know } f(x) \text{ and one (arbitrary) } s \in \partial f(x)\,. \tag{3}$$

This assumption is fairly natural and the next subsection will show that such an $s \in \partial f(x)$

(a so-called *subgradient*) can often be computed. In other words: we have a black box (sometimes called an *oracle*) which, given x, answers $f(x)$ and some $s \in \partial f(x)$. Thus, the situation is similar to that in ordinary smooth optimization, except that s will not vary continuously with x.

The optimality condition for the convex f to be minimal at x is well-known: $\partial f(x)$ must contain the 0-vector (the definition itself of $\partial f(x)$ shows that this condition is obviously necessary and sufficient). Yet, the poor knowledge about f, as specified by (3), makes it impossible to check the optimality condition directly: a mechanism is needed to somehow build the whole of ∂f, out of the single vectors given by the black box. This is what bundle methods are about.

2.4 Lagrangian Relaxation

Consider an abstract optimization problem

$$\max g(u), \text{ subject to } c_j(u) = 0,\ j = 1, \ldots, n. \tag{4}$$

Suppose that this problem would be “simple” if the constraints were not present. This is for example the case in *large-scale* optimization where each function is a sum of a big number of simpler functions; say

$$g(u) = \sum_{i=1}^{N} g_i(u_i) \text{ and } c_j(u) = \sum_{i=1}^{N} c_{ji}(u_i),\ j = 1, \ldots, n. \tag{5}$$

We see in this decomposable situation that the constraints link together the simpler variables u_i. An attractive idea is then to eliminate these constraints.

One way of doing this, without destroying completely the necessity for such constraints, is to introduce the Lagrangian function, which depends on u and on n real parameters x_i (one for each constraint). For the general problem (4), this Lagrangian is

$$L(u, x) := g(u) + \sum_{j=1}^{n} x_j c_j(u) = g(u) + \langle x, c(u) \rangle. \tag{6}$$

The idea is then to maximize $L(\cdot, x)$ on the whole space, for fixed x. When the problem is decomposable as above, we have

$$L(u, x) = \sum_{i=1}^{N} g_i(u_i) + \sum_{j=1}^{n} x_j \sum_{i=1}^{N} c_{ji}(u_i) = \sum_{i=1}^{N} L_i(u_i, x), \tag{7}$$

where we have set $L_i(u_i, x) := c_i(u_i) + \langle x, c_i(u_i) \rangle$. Here, maximizing the Lagrangian with respect to u reduces to the maximization of N independent functions L_i , each of which depends on u_i only. We see the decomposition effect of the Lagrangian. This technique is called Lagrangian relaxation.

Now comes the theory. Of course, the maximal value of the Lagrangian depends on x. Let us denote by f the resulting so-called dual function: $f(x) := \max_u L(u, x)$. Then the following properties hold:

i. $f(x) \leqslant g(u)$ for all $x \in \mathbb{R}^n$ and all u feasible in (4), this is called the *weak duality* relation;
ii. each $f(x)$ thus provides an upper bound of the optimal value in (4) and computing the best upper bound – i.e. minimizing f over $\mathbb{R}^n$ – is of importance;
iii. f turns out to be a convex function of x, and
iv. if u_x maximizes the Lagrangian at x (u_x solves $max_u\ L(u, x)$), then the vector $c(u_x) \in \mathbb{R}^n$ (the partial derivatives of L) is a subgradient of f at x.

We see that Lagrangian relaxation places us exactly in the framework described in the previous subsection.

3. Algorithms for Convex Optimization

There are two seminal algorithms for convex optimization: subgradients and cutting planes. The latter is already a form of bundle method. It has serious shortcomings, which can be cured, and this is the subject of the present section.

-
-
-

Bibliography

Cheney E. and Goldstein A. (1959). Newton's Method for Convex Programming and Tchebycheff Approximations. *Numerische Mathematik*, **1, pp.** 253-268. [In this paper the cutting plane algorithm has been defined.]

Geoffrion A.M. (1971). Duality in Nonlinear Programming: A Simplified Applications-Oriented Development. *SIAM Review*, **13**(1), pp. 1-37. [This paper describes Lagrangian relaxation seen from a practical point of view.]

Goffin J.L., Haurie A., and Vial J.Ph. (1992). Decomposition and Nondifferentiable Optimization With the Projective Algorithm. *Management Science*, **38**(2), pp. 284-302. [The algorithm of analytic centers (ACCPM) alluded to in the beginning of paragraph 3.3 appeared in this publication.]

Hiriart-Urruty J.B. and Lemaréchal C. (1993). *Convex Analysis and Minimization Algorithms*. Berlin: Springer-Verlag. [This monograph on convex analysis is more devoted to numerical optimization giving especially technical details on Lagrangian relaxation.]

Kelley J.E. (1960). The Cutting Plane Method for Solving Convex Programs. J. Soc. Indust. Appl. Math., **8, pp.** 703-712. [In this paper the cutting plane algorithm has been defined.]

Kočvara M. and Outrata, J.V. (1994) *A Numerical Approach to the Design of Masonry Structures*. System

Modeling and Optimization (ed. J. Henry and J.-P. Yvon), L.N. in Control and Information Sciences 197, pp. 195-205. Springer-Verlag. [The masonry example of 4.3. with its illustration is taken from this paper.]

Nesterov Yu.E. and Nemirovskii A.S. (1994). *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics 13, 405 pp. Philadelphia: SIAM. [This paper gives the theoretical development for interior point methods.]

Oustry F. (2000). A Second-Order Bundle Method to Minimize the Maximum Eigenvalue Function. *Mathematical Programming*, **89**(1), pp. 1-33. [In this paper an enriched bundle method is described.]

Polyak B.T. (1966). A General Method for Solving Extremum Problems. *Soviet Mathematics Doklady*, **8**, pp. 593-597. [This paper gives probably the best review of subgradient optimization.]

Reeves C.R. (1993). *Modern Heuristic Techniques for Combinatorial Problems*, 320 pp. New York: Blackwell Scientific Publications. [Chapter 6 of this monograph describes Lagrangian relaxation for combinatorial applications.]

Rockafellar R.T. (1970). *Convex Analysis*, 451 pp. Princeton University Press. [Nonsmooth optimization is closely related to convex analysis. This is a standard reference to convex analysis.]

Vandenberghe L. and Boyd S. (1996). Semidefinite Programming. *SIAM Review*, **38**(1), pp. 49-95. [This paper specializes interior point methods to SDP optimization.]

Biographical Sketches

Claude Lemaréchal graduated at the University of Toulouse in 1967 and obtained his PhD (Doctorat d'État) at the University of Paris IX. He spent most of his career at Inria, the French Institute for Numerical Mathematics and Computer Science; and stayed at the International Institute for Systems Analysis (IIASA) in Vienna for an international research project. He has published nearly 100 papers, and is recognized as a leading expert in nonsmooth optimization, especially from a numerical point of view; and he is one of the promoters of bundle methods. His current research activities are oriented toward spreading, in various fields such as combinatorics or non regular dynamics, the tools of convex analysis and nonsmooth optimization. He is Dantzig Prize winner1994.

Francois Oustry graduated at the University of Paris I, where he obtained his PhD in 1997. He paid several visits to Stanford University (Professor Boyd), Courant Institute New York (Professor Overton), CORE in Belgium (Professor Nesterov). He stayed at Inria as a Research Scientist, and then created his own firm: Raise Partner, devoted to risk analysis. His distinguished works in optimization problems involving eigenvalues (SDP optimization) establish fruitful links between differential geometry, convex analysis and nonsmooth optimization. In 1999, he received two Prizes for young researchers: the French ”Prix Jean-Claude Dodu”, and the international INFORMS Optimization Prize.

Jochem Zowe graduated from the University of Würzburg in 1969, where he obtained his PhD in 1971, and his habilitation in 1976. From 1971 to 1978, Jochem Zowe was Assistant Professor at the University of Würzburg, and from 1979 to 1995, he was Professor for Numerical Mathematics at the University of Bayreuth. From 1995 to 2001 he held a chair in Applied Mathematics at the University of Erlangen-Nürnberg. Jochem Zowe is a recognized researcher in the field of nonlinear optimization. He has published more than 60 papers and done numerous research projects and visits with colleagues in the USA, Canada, Chile, France and Israel. His latest research activities have been oriented towards the application of optimization theory and optimization methods for design problems in engineering.