Convex problems

A Descent Method with Relaxation Type Step

Ulf Brännlund*

Department of Mathematics,
Division of Optimization and Systems Theory,
KTH, Stockholm

Abstract

A new bundle method for minimizing a convex nondifferentiable function \( f : \mathbb{R}^n \to \mathbb{R} \) is presented. At each iteration a master problem is solved to get a search direction \( d \). This master problem is a quadratic programming problem of the type

\[
\min_d \frac{1}{2} d^T d \\
\text{s.t. } v_c \geq g_i^T d - \epsilon_i, \forall i \in I,
\]

where \( v_c \) is a parameter, which is an estimate the predicted decrease obtainable from the current iteration point and \( g_i \) are \( \epsilon_i \)-subgradients at the current iteration point.

It is shown that each sequence of \( \{x_k\} \) generated by the algorithm minimizes \( f \), i.e. \( f(x_k) \downarrow \inf\{f(x) | x \in \mathbb{R}^n\} \), and that \( \{x_k\} \) converges to a minimum point whenever \( f \) attains its infimum.

Some numerical experiments on some nondifferentiable test problems found in the literature are performed with satisfactory and encouraging results.

1 Introduction

This paper is concerned with a method to minimize a convex nonsmooth function \( f : \mathbb{R}^n \to \mathbb{R} \).

The subdifferential of \( f \) at \( x \) is defined by

\[
\partial f(x) = \{g \in \mathbb{R}^n | f(y) \geq f(x) + g^T (y-x), \forall y \in \mathbb{R}^n\}. \tag{1}
\]

An element \( g \) of the subdifferential \( \partial f(x) \) is called a subgradient of \( f \) at \( x \). It is assumed that an "oracle" (a black box), which given an \( x \) can deliver \( f(x) \) and one element of \( \partial f(x) \), is available.

For convex functions the fundamental property of a minimum point \( x^* \) is that

\[
0 \in \partial f(x^*), \text{ i.e.}
\]

\[
f(x^*) \leq f(y), \forall y \in \mathbb{R}^n \text{ if and only if } 0 \in \partial f(x^*). \tag{2}
\]

The first methods for solving convex nonsmooth optimization problems of the above type were the so called relaxation methods, see Polyak [Pol69] and the cutting

* Research supported by the Swedish Research Council for Engineering Sciences.
plane methods of Kelley [Kel60] and Cheney & Goldstein [ChG59]. The convergence of
the former methods depend on that at each iteration the new iteration point is closer
to the optimum than the previous. Convergence of the latter methods depend on that
at each iteration an approximation of the objective function is improving. Thus, in
neither case does convergence depend on that the objective function is monotonically
decreasing, and generally it is not.

Bundle methods, whose convergence do depend on a monotonic decrease of the
objective function, have received considerable attention ever since the seminal works
of Lemaréchal [Lem75] and Wolfe [Wol75]. Bundle methods are related more or less
closely to a number of well known methods in differentiable and nondifferentiable
optimization among them cutting plane methods, the method of conjugate gradients,
to the method of steepest descent, proximal point methods, trust-region methods
as well as to relaxation methods. The latter was pointed out by the author in two
previous papers contained in Brännlund [Brä93].

In this paper the connection between bundle methods and relaxation methods
is made explicit; in section 2 we present a descent algorithm in which the control
parameter is an estimate of the optimal function value just as in relaxation methods.
In section 3 the convergence of the method is established and in section 5 some
numerical experience is reported.

Bundle methods of this so called level type have been suggested previously by
Lemaréchal et al. [LNN91]. However, in contrast to theirs, our method is globally
convergent without any compactness assumptions and requires bounded storage. As
this work was completed we have learned that Kiwiel [Kiw93] also has derived a level
method with these properties. We refer to Brännlund [Brä93] for a discussion on
different bundle methods and a motivation of our approach.

2 A descent method of level type

We are dealing with an iterative method, in which we at each iteration \( x_k \), have a
"bundle" of information consisting of the previous iteration points, \( x_1, \ldots, x_k \), to-
gether with subgradients generated at these points, \( g_i \in \partial f(x_i) \).

We define the linearization error at \( x_k \), if \( f \) is linearized in \( x_i \), as

\[
\epsilon^k_i = \epsilon(x_k, x_i, g_i) = f(x_k) - (f(x_i) + g_i^T (x_k - x_i)),
\]

which is easily shown to be non-negative for a convex function. These linearization
errors can be updated recursively.

A vector \( g \) is said to be an \( \epsilon \)-subgradient at \( x \), \( g \in \partial_\epsilon f(x) \), if

\[
f(y) \geq f(x) + g^T (y - x) - \epsilon, \quad \forall y \in \mathbb{R}^n.
\]

A point \( z \) is said to be \( \epsilon \)-optimal if

\[
f(y) \geq f(x) - \epsilon \|y - x\| - \epsilon, \quad \forall y \in \mathbb{R}^n.
\]

Hence, if \( \|g\| \leq \epsilon \) and \( g \in \partial_\epsilon f(x) \) then \( x \) is \( \epsilon \)-optimal.

In order to get the search direction, we solve the problem

\[
\min_d \quad \frac{1}{2} d^T d
\]

s.t. \( \nu_i \geq g_i^T d - \epsilon_i, \quad \forall i \in I, \)

\( \epsilon \)