Open access • Journal Article • DOI:10.1080/10556780410001648112

A DC piecewise affine model and a bundling technique in nonconvex nonsmooth minimization — Source link

Antonio Fuduli, Manlio Gaudioso, Giovanni Giallombardo

Institutions: University of Calabria

Published on: 01 Feb 2004 - Optimization Methods & Software (Taylor & Francis Group)

Topics: Piecewise linear function, Piecewise, Trust region, Convexity and Convex function

A DC PIECEWISE AFFINE MODEL AND A BUNDLING TECHNIQUE IN NONCONVEX NONSMOOTH MINIMIZATION

A. FUDULI*, M. GAUDIOSO§, AND G. GIALLOMBARDO§

Abstract. We introduce an algorithm to minimize a function of several variables with no convexity nor smoothness assumptions.
The main peculiarity of our approach is the use of an the objective function model which is the difference of two piecewise affine convex functions.
Bundling and trust region concepts are embedded into the algorithm.
Convergence of the algorithm to a stationary point is proved.

Key words. Nonsmooth optimization, cutting planes, bundle methods, DC functions.

AMS subject classifications. 90C26, 65K05

1. Introduction. The methods currently available to find the unconstrained minima of nonconvex and not necessarily smooth functions appear in general to be an adaptation of methods originally designed for dealing with the corresponding convex problem.

Among such methods we mention here those due to Kiwiel, Makëla and Neittaanmäki, Schramm and Zowe [6, 7, 10].

They, according to the terminology used by Schramm and Zowe, combine the bundling idea, due to Lemarechal, and the trust region approach. Bundling is in turn derived from cutting plane approximation, where a piecewise affine approximation (the model) of the objective function is minimized in order to obtain an approximate solution to the original minimization problem. The quality of the model is improved at each iteration and this is the key argument to guarantee convergence.

The model is defined as the pointwise maximum of a set of affine functions. The original function, if convex, is interpolated by the model at a set of points, which is constituted by at least one point, normally the best one available, which we will refer to as to the stability center.

If the function to be minimized is nonconvex, the interpolation property of the model can be lost. Thus in [6, 7, 10], in order to keep on working with a model defined as the pointwise maximum of affine functions, some of the affine pieces are vertically translated, whenever necessary, in order to retain interpolation at least at the stability center.

*Dipartimento di Ingegneria dell'Innovazione, Università di Lecce, Via Monteroni, 73100 Lecce (LE), Italia. E-mail: antonio.fuduli@unile.it

§Dipartimento di Elettronica Informatica e Sistemistica, Università della Calabria, 87036 Rende (CS), Italia. E-mail: {gaudioso,giallo}@deis.unical.it

In [4] a different approach has been introduced, where the model takes into account explicitly the nonconvex behavior of the objective function.

Here we extend such approach by defining a piecewise affine type model which can be put in a DC form, i.e. it is the sum of a piecewise affine convex function and a piecewise affine concave function. Thus our approach benefits of some ideas coming from the theory of quasidifferentiable functions (see [3]).

In connection with our model, we adopt all the machinery coming from the bundle-trust philosophy.

The paper is organized as follows. In section 2 we describe our model and its main characteristics. The algorithm is presented in section 3 and the convergence properties are proved in section 4, while some conclusions are drawn in section 5. Some properties of the DC subproblem are introduced in the appendix.

The following notations are adopted throughout the paper. We denote by $\|\cdot\|$ the euclidean norm in $\mathbb{R}^n$, by $a^T b$ the inner product of the vectors a and b, and by e a vector of ones of appropriate dimension. The generalized gradient of a Lipschitz function $f : \mathbb{R}^n \mapsto \mathbb{R}$ at any point x is denoted by $\partial f(x)$.

2. The model. Consider the minimization problem

$$\min_{x \in \mathbb{R}^n} \ f(x) \ ,$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is not necessarily convex nor differentiable.

We assume that f is locally Lipschitz, i.e. it is Lipschitz on every bounded set, then it is differentiable almost everywhere. Under the above hypothesis, it is defined at each point x the generalized gradient [2] (or Clarke's gradient or subdifferential)

$$\partial f(x) = \text{conv}\{g \ | g \in \mathbb{R}^n, \nabla f(x_k) \to g, \ x_k \to x, \ x_k \notin \Omega_f\} \ ,$$

where Ω_f is the set (of zero measure) where f is not differentiable. An extension of the generalized gradient is the *Goldstein ϵ-subdifferential* $\partial_\epsilon^G f(x)$ defined as

$$\partial_\epsilon^G f(x) = \text{conv}\{\partial f(y) \ | \|y - x\| \leq \epsilon\} \ .$$

We assume also that we are able to calculate at each point x both the objective function value and a subgradient $g \in \partial f(x)$, i.e. an element of the generalized gradient and that, for any point x_0, the set

$$\mathcal{F}_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$$

is compact.

Now we describe the model adopted in our method, focusing on the differences with respect to the methods tailored on the convex case. We denote by x_j the current estimate of the minimum in an iterative procedure (coinciding with the stability

center) and by g_j any subgradient of f at x_j. The bundle of available information is the set of elements

$$(x_i, f(x_i), g_i, \alpha_i, a_i) \quad i \in I \ ,$$

where x_i , $i \in I$, are the points touched in the procedure, g_i is a subgradient of f at x_i, α_i is the linearization error between the actual value of the objective function at x_j and the linear expansion generated at x_i and evaluated at x_j, i.e.

$$\alpha_i \stackrel{\triangle}{=} f(x_j) - f(x_i) - g_i^T(x_j - x_i)$$

and

$$a_i \stackrel{\triangle}{=} \|x_j - x_i\| \ .$$

The classical cutting plane method [1, 5] minimizes at each iteration the cutting plane function $f_j(x)$ defined as

$$f_j(x) = \max_{i \in I} \left\{ f(x_i) + g_i^T(x - x_i) \right\} \ .$$

The minimization of $f_j(x)$ can be put in the form of a linear program as

$$\left\{ \begin{array}{ll} \min\limits_{\theta, x} & \theta \\ & \theta \geq f(x_i) + g_i^T(x - x_i) \quad i \in I \ , \end{array} \right. \tag{2.1}$$

which is equivalent to solve

$$\left\{ \begin{array}{ll} \min\limits_{v, d} & v \\ & v \geq g_i^T d - \alpha_i \quad i \in I \ , \end{array} \right. \tag{2.2}$$

where d is the "displacement" from x_j, i.e. $d \stackrel{\triangle}{=} x - x_j$.

It is worth noting that in the nonconvex case α_i may be negative, since the first order expansion at any point does not necessarily support from below the epigraph of the function.

Thus we partition the set I in two sets I_+ and I_- defined as follows

$$I_+ \stackrel{\triangle}{=} \{i | \alpha_i \geq 0\} \qquad I_- \stackrel{\triangle}{=} \{i | \alpha_i \leq 0\} \ . \tag{2.3}$$

We remark that the two index sets I_+ and I_- do not have empty intersection since, by definition, at least the index corresponding to the bundle element $(x_j, f(x_j), g_j, 0, 0)$ belongs to both I_+ and I_-.

The bundles defined by the index sets I_+ and I_- are characterized by points that somehow exhibit respectively a "convex behavior" and a "concave behavior" relatively to x_j.

The basic idea of our approach is to treat differently the two bundles in the construction of a piecewise affine model.

We define the following piecewise affine functions:

$$\Delta^+(d) \stackrel{\triangle}{=} \max_{i \in I_+} \left\{ g_i^T d - \alpha_i \right\}$$

and

$$\Delta^-(d) \stackrel{\triangle}{=} \min_{i \in I_-} \left\{ g_i^T d - \alpha_i \right\} \ .$$

We remark that $\Delta^+(d)$ is convex, while $\Delta^-(d)$ is concave. Moreover we have

$$\Delta^+(d) \geq \Delta^-(d) \quad \forall d \tag{2.4}$$

since $I_+ \cap I_- \neq \emptyset$.

Now, for any choice of the scalar $p \in (0,1)$, we can define the function

$$\Delta_p(d) \stackrel{\triangle}{=} p\Delta^+(d) + (1-p)\Delta^-(d) \ ,$$

which can be interpreted as an approximation to the difference function

$$h(d) \stackrel{\triangle}{=} f(x_j + d) - f(x_j)$$

obtained through a weighted average of the two approximations $\Delta^+(d)$ and $\Delta^-(d)$ that both interpolate $h(d)$ at $d = 0$.

Function $\Delta_p(d)$ is the model function that we adopt to find a descent step for f. It is a DC (Difference of two Convex) function and it is piecewise affine.

We introduce into our approach proximity control, aimed at defining implicitly a kind of trust region, by adding to the suitably weighted model function $\Delta_p(d)$ a quadratic penalty term. Thus we come out with the complete model function

$$f_{p\gamma}(d) \stackrel{\triangle}{=} \gamma \Delta_p(d) + \frac{1}{2}\|d\|^2 \ ,$$

where γ (the proximity control parameter) is any positive scalar. We emphasize in the notation the fact that the complete model depends, *ceteris paribus*, on the two scalar parameters p and γ.

The function $f_{p\gamma}(d)$ is quasidifferentiable (see [3]). A necessary and sufficient condition for $d_{p\gamma}$ to be a local minimum for $f_{p\gamma}(d)$ is derived in the appendix. We

note in passing that $f_{p\gamma}(d)$ is coercive and, in general, it may admit several local minima.

The following lemma provides a bound on the norm of $d_{p\gamma}$.

LEMMA 2.1. *For any $\gamma > 0$ it holds:*

$$\|d_{p\gamma}\| \leq 2\gamma \|g_{I_-}\| \ ,$$

where $\|g_{I_-}\| \stackrel{\triangle}{=} \max_{i \in I_-} \{\|g_i\|\}$.

Proof. From the definition of $\Delta_p(d)$, taking into account (2.4) and noting that $\alpha_i \leq 0$, $\forall i \in I_-$, we have

$$\Delta_p(d) \geq \Delta^-(d) \geq \min_{i \in I_-} g_i^T d \geq -\|g_{I_-}\| \|d\| \ .$$

Thus

$$f_{p\gamma}(d) \geq -\gamma \|g_{I_-}\| \|d\| + \frac{1}{2}\|d\|^2$$

and the thesis follows from the fact that $f_{p\gamma}(d_{p\gamma}) \leq 0$. □

In our method a significant role is played by the strictly convex program

$$\min_d \gamma \Delta^+(d) + \frac{1}{2}\|d\|^2 \ , \tag{2.5}$$

which, by introducing the scalar variable v, can be rewritten as a quadratic program of the form:

$$QP(\gamma) \begin{cases} \min\limits_{v,d} & \gamma v + \frac{1}{2}\|d\|^2 \\ & \\ & v \geq g_i^T d - \alpha_i \quad i \in I_+ \ . \end{cases}$$

The dual of $QP(\gamma)$ can be written in the form:

$$DP(\gamma) \begin{cases} \min\limits_{\lambda \geq 0} & \frac{1}{2}\|G_+\lambda\|^2 + \alpha_+^T \lambda \\ & \\ & e^T \lambda = \gamma \ , \end{cases}$$

where G_+ is the matrix whose columns are the vectors g_i, $i \in I_+$. Analogously, the scalars α_i, $i \in I_+$ are grouped in the vector α_+, while λ is the nonnegative multiplier vector of appropriate dimension.

The optimal primal solution $(d_\gamma, \Delta^+(d_\gamma))$ of (2.5) is related to the optimal dual solution λ_γ of $DP(\gamma)$ by the following formulae:

$$d_\gamma = -G_+\lambda_\gamma \tag{2.6a}$$

$$\Delta^+(d_\gamma) = -\frac{1}{\gamma}\left(\|d_\gamma\|^2 + \alpha_+^T\lambda_\gamma\right) \ . \tag{2.6b}$$

3. The algorithm. We describe now our algorithm, which is based on the iterative solution of problem

$$\min_{d\in\mathbb{R}^n} f_{p\gamma}(d).$$

As the function $f_{p\gamma}(d)$ is nonconvex, by "solving" we mean here to find a global optimal solution. See the appendix for a discussion on how to find such a solution.

In the sequel we will indicate the stability center x_j by y in order to emphasize its peculiar role with respect to the other available points.

Moreover, in the algorithm we refer to the "main iteration" as to the set of steps where the stability center remains unchanged. We can exit from the "main iteration" either because the whole algorithm terminates, due to the satisfaction of an approximate stationarity condition, or because the stability center is updated, due to the satisfaction of a sufficient decrease condition.

The initialization of the algorithm requires a starting point $x_0 \in \mathbb{R}^n$. The initial stability center y_0 is set equal to x_0. We have initially just one bundle element $(y_0, f(y_0), g(y_0), 0, 0)$, where $g(y_0) \in \partial f(y_0)$. The corresponding index is put in both I_+ and I_- which are consequently both a singleton. The following global parameters are to be set:

- the stationarity tolerance $\delta > 0$ and the proximity measure $\epsilon > 0$;
- the descent parameter $m \in (0, 1)$;
- the reduction parameter $r \in (0, 1)$ and the increase parameter $R > 1$;
- the balance parameter $p_0 \in (0, 1)$.

A short description of the algorithm is the following:

Algorithm Outline

1. Initialization.
2. Execute the "main iteration".
3. Update the bundle of information with respect to the new stability center and return to 2.

In the following description we do not index the "main iteration" for sake of notation simplicity. Thus the stability center y is to be intended as the current stability center. The following local parameters are set each time the "main iteration" is entered (they are subject to possible modifications during the execution):

- the safeguard parameters $0 < \gamma_{min} < \gamma_{max}$;
- the descent threshold parameter $\sigma > 0$, the bundle insertion threshold parameter $\phi > 0$ and the approximation parameter $\eta > 0$;
- the balance parameter $p = p_0$.

The following conditions on the parameters are imposed during the "main iteration":

$$\gamma_{min} < \frac{\epsilon}{2\|g_{I_-}\|} \tag{3.1}$$

$$\delta \geq \sqrt{\frac{(\sigma+\eta)}{\gamma_{min}}}\ . \tag{3.2}$$

In particular, we set initially $\gamma_{min} = \dfrac{r\epsilon}{2\|g_{I_-}\|}$, so that condition (3.1) is satisfied ($g_{I_-} = g(y_0)$ at the beginning of the first "main iteration") and we set also $\gamma_{max} = R\gamma_{min}$.

The remaining local parameters are set so that conditions (3.2) is satisfied; in particular we set σ satisfying the condition $\delta = \sqrt{\dfrac{2\sigma}{\gamma_{min}}}$ (i.e. $\sigma = \dfrac{\gamma_{min}\delta^2}{2}$) and then we set $\eta = \sigma$. Finally we set $\phi = \eta = \sigma$.

We remark that in general the "main iteration" maintains the (updated) bundle of information from previous iterations (the quantities α_i and a_i are in fact dependent on the stability center). Note also that the assignment of any bundle element index to I_+ or to I_- or to both depends also on the current stability center.

ALGORITHM 3.1 (Main Iteration).

0. *If* $\|g(y)\| \leq \delta$ *then* STOP *(stationarity achieved).*

1. *Set* $\gamma_{min} = \min\{\gamma_{min}, \dfrac{r\epsilon}{2\|g_{I_-}\|}\}$, *set* $\sigma = \dfrac{\gamma_{min}\delta^2}{2}$ *and* $\phi = \eta = \sigma$.

Select $\gamma \in (\gamma_{min}, \gamma_{max})$ *and calculate* $d_{p\gamma}$, *a global minimizer of* $f_{p\gamma}(d)$, *and* d_γ, *the minimizer of* $QP(\gamma)$.

2. *Calculate* $\Delta^+(d_{p\gamma})$.

If $\Delta^+(d_{p\gamma}) \leq -\sigma$ *go to* 3. *If* $\Delta^+(d_{p\gamma}) - \Delta^+(d_\gamma) > \eta$ *then set* $p = p + r(1-p)$ *and return to* 1. *Else go to* 5.

3. *Set* $\hat{x} := y + d_{p\gamma}$ *and calculate* $f(\hat{x})$. *If*

$$f(\hat{x}) \leq f(y) + m\Delta^+(d_{p\gamma}) \tag{3.3}$$

set the new stability center $y := \hat{x}$ *and* EXIT *from the main iteration.*

4. *Calculate* $\hat{g} \in \partial f(\hat{x})$ *and set*

$$\hat{\alpha} := f(y) - f(\hat{x}) + \hat{g}^T d_{p\gamma} \ .$$

If $\hat{\alpha} \leq -\phi$ *and* $\|d_{p\gamma}\| > \epsilon$ *then set* $\gamma := \gamma - r(\gamma - \gamma_{min})$ *and return to* 1.
If $\hat{\alpha} \geq \phi$ *then insert the element* $(\hat{x}, f(\hat{x}), \hat{g}, \hat{\alpha}, \|d_{p\gamma}\|)$ *into the bundle for an appropriate value of* $i \in I_+$ *and return to* 1.
If $\hat{\alpha} \geq 0$ *insert the element* $(\hat{x}, f(\hat{x}), \hat{g}, 0, \|d_{p\gamma}\|)$ *into the bundle twice, for two appropriate values of the indices one belonging to* I_+ *and the other to* I_- *and return to* 1.
Else find a scalar $t \in (0,1)$ *such that* $g(t) \in \partial f(y + td_{p\gamma})$ *satisfies the condition*

$$g(t)^T d_{p\gamma} \geq m\Delta^+(d_{p\gamma}), \tag{3.4}$$

insert the element $(y + td_{p\gamma}, f(y + td_{p\gamma}), g(t), 0, t\|d_{p\gamma}\|)$ *into the bundle twice, for two appropriate values of the indices one belonging to* I_+ *and the other to* I_- *and return to* 1.

5. *Set*

$$I_+ := I_+ \setminus \{i \in I_+ \mid a_i > \epsilon\}$$

and

$$I_- := I_- \setminus \{i \in I_- \mid a_i > \epsilon\} \ .$$

Calculate

$$\|g_*\| = \min_{g \in \mathrm{conv}\{g_i | i \in I_+ \cup I_-\}} \|g\|.$$

If $\|g_*\| \leq \delta$ *then* STOP *(stationarity achieved).*
Else set:

$$\gamma_{max} := \gamma_{max} - r(\gamma_{max} - \gamma_{min}) \tag{3.5}$$

and go to 1.

On exit from the main iteration, once the stability center has been updated, the bundle is updated too, according to the following procedure:

Algorithm 3.2 (Bundle update).

1. *Calculate the linearization errors* α_i *and the distances* a_i *with respect to the new stability center* $\forall i \in I_+ \cup I_-$.

2. *If* $|\alpha_i| \leq \phi$ *then insert the element* $(x_i, f(x_i), g_i, 0, a_i)$ *into the bundle twice, for two appropriate values of the indices, one belonging to* I_+ *and the other to* I_-.

3. *If* $\alpha_i < -\phi$ *then insert the element* $(x_i, f(x_i), g_i, \alpha_i, a_i)$ *into the bundle for an appropriate value of the index* $i \in I_-$.

4. *If* $\alpha_i > \phi$ *then insert the element* $(x_i, f(x_i), g_i, \alpha_i, a_i)$ *into the bundle for an appropriate value of the index* $i \in I_+$.

5. *Delete from the bundle all the elements such that:*

$$i \in I_- \quad and \quad x_i \notin \mathcal{F}_0. \tag{3.6}$$

Remark. The local parameters $\gamma_{min}, \sigma, \phi$ and η are subject to possible modification during the execution of the "main iteration", since γ_{min} depends on $\|g_{I_-}\|$.

The parameter γ_{min} is however bounded away from zero. In fact the bundle insertion rules at step 4 of the main iteration and the bundle deletion condition (3.6) ensure that $\|g_{I_-}\| \leq L_f$ where L_f is the Lipschitz constant of f on the set of points whose distance from $\mathcal{F}_0$ is not greater than ϵ. Thus throughout the algorithm $\gamma_{min} \geq \bar{\gamma} \stackrel{\triangle}{=} \dfrac{r\epsilon}{2L_f}$ and consequently $\sigma = \phi = \eta \geq \bar{\rho} \stackrel{\triangle}{=} \dfrac{\bar{\gamma}\delta^2}{2}$.

The value of the parameter ϕ used in the "bundle update" procedure is that one available on exit from the "main iteration".

4. Convergence. Throughout the section we make the following assumptions:

A1 f is locally Lipschitz;

A2 f is weakly semismooth (see [8, 10]).

In particular A2 is a technical assumption (see [9]) which ensures that at the step 4 the problem of finding the scalar t satisfying the condition (3.4) is well posed.

Before proving convergence, we state the following lemma which ensures that, by setting the parameter p sufficiently close to 1, the value assumed by function Δ^+ at the minimum of $f_{p\gamma}(d)$ is close to that assumed by the same function at the minimum of problem $QP(\gamma)$.

LEMMA 4.1. *Given any positive* γ*, for any positive scalar* η *there exists a positive threshold* $p_{max} < 1$ *such that* $\forall p \geq p_{max}$ *the following relation holds :*

$$\Delta^+(d_{p\gamma}) \leq \Delta^+(d_\gamma) + \eta,$$

where d_γ *is the optimal solution to problem (2.5).*

Proof. From the definition of $d_{p\gamma}$ the following inequality holds $\forall p \in (0, 1)$:

$$\gamma\{p\Delta^+(d_{p\gamma}) + (1-p)\Delta^-(d_{p\gamma})\} + \frac{1}{2}\|d_{p\gamma}\|^2 \leq \tag{4.1}$$

$$\gamma\{p\Delta^+(d_\gamma) + (1-p)\Delta^-(d_\gamma)\} + \frac{1}{2}\|d_\gamma\|^2.$$

Moreover convexity of $\Delta^+(d)$ implies $\forall p \in (0,1)$ and $\forall u \in \partial\Delta^+(d_\gamma)$:

$$\gamma p \Delta^+(d_{p\gamma}) + \frac{1}{2}\|d_{p\gamma}\|^2 \geq \tag{4.2}$$

$$\gamma p \Delta^+(d_\gamma) + \frac{1}{2}\|d_\gamma\|^2 + (\gamma p u + d_\gamma)^T (d_{p\gamma} - d_\gamma) + \frac{1}{2}\|d_{p\gamma} - d_\gamma\|^2 \ .$$

Now let $w(d) \stackrel{\triangle}{=} \gamma\Delta^+(d) + \frac{1}{2}\|d\|^2$ be the objective function of problem $QP(\gamma)$ and let $w'(d_\gamma, d_{p\gamma} - d_\gamma)$ be the directional derivative of w evaluated at d_γ along the direction $(d_{p\gamma} - d_\gamma)$. The optimality of d_γ implies that $w'(d_\gamma, d_{p\gamma} - d_\gamma) \geq 0$. Moreover, since

$$w'(d_\gamma, d_{p\gamma} - d_\gamma) = \max_{u \in \partial\Delta^+(d_\gamma)} (\gamma u + d_\gamma)^T (d_{p\gamma} - d_\gamma),$$

there exists a subgradient, say $u^* \in \partial\Delta^+(d_\gamma)$, such that

$$(\gamma u^* + d_\gamma)^T (d_{p\gamma} - d_\gamma) \geq 0 \ . \tag{4.3}$$

From (4.1), taking into account that (4.2) holds for $u = u^*$, we obtain

$$\gamma p \Delta^+(d_\gamma) + \frac{1}{2}\|d_\gamma\|^2 + (\gamma p u^* + d_\gamma)^T (d_{p\gamma} - d_\gamma) + \frac{1}{2}\|(d_{p\gamma} - d_\gamma)\|^2 + \gamma(1-p)\Delta^-(d_{p\gamma}) \leq$$

$$\gamma p \Delta^+(d_{p\gamma}) + \frac{1}{2}\|d_{p\gamma}\|^2 + \gamma(1-p)\Delta^-(d_{p\gamma}) \leq$$

$$\gamma p \Delta^+(d_\gamma) + \gamma(1-p)\Delta^-(d_\gamma) + \frac{1}{2}\|d_\gamma\|^2.$$

Thus we have

$$(\gamma p u^* + d_\gamma)^T (d_{p\gamma} - d_\gamma) + \frac{1}{2}\|(d_{p\gamma} - d_\gamma)\|^2 \leq \gamma(1-p)(\Delta^-(d_\gamma) - \Delta^-(d_{p\gamma})),$$

which can be rewritten as:

$$(\gamma u^* + d_\gamma)^T (d_{p\gamma} - d_\gamma) - (1-p)(\gamma u^* + d_\gamma)^T (d_{p\gamma} - d_\gamma) + (1-p) d_\gamma^T (d_{p\gamma} - d_\gamma) + \frac{1}{2}\|(d_{p\gamma} - d_\gamma)\|^2$$

$$\leq \gamma(1-p)(\Delta^-(d_\gamma) - \Delta^-(d_{p\gamma})).$$

Consequently, taking into account (4.3), we have

$$\|d_{p\gamma} - d_\gamma\|^2 \leq 2(1-p)\{(\gamma u^* + d_\gamma)^T (d_{p\gamma} - d_\gamma) + \gamma(\Delta^-(d_\gamma) - \Delta^-(d_{p\gamma})) - d_\gamma^T (d_{p\gamma} - d_\gamma)\} \leq$$

$$4(1-p)(\gamma L + D) \ ,$$

where L is the biggest between the Lipschitz constants of Δ^+ and Δ^- and D is the diameter of the set where the points $d_{p\gamma}$ and d_γ are located (boundedness of both $d_{p\gamma}$ and d_γ is ensured by boundedness of γ_{max}).

Finally, since:

$$|\Delta^+(d_{p\gamma}) - \Delta^+(d_\gamma)| \leq L\|d_{p\gamma} - d_\gamma\| \leq \sqrt{4(1-p)L(\gamma L + D)},$$

the thesis follows provided that $4(1-p)L(\gamma L + D) \leq \eta$, i.e. whenever $p \geq 1 - \dfrac{\eta^2}{4L(\gamma L + D)}$. □

Now we prove that the "main iteration" terminates.

LEMMA 4.2. *The* "main iteration" *terminates after a finite number of steps.*

Proof. We observe first that, as consequence of Lemma 4.1 and of the rule for updating p in case $\Delta^+(d_{p\gamma}) - \Delta^+(d_\gamma) > \eta$ at step 2, the algorithm cannot loop infinitely many times between step 1 and step 2.

Thus, to prove finiteness, we need to show that the algorithm cannot pass infinitely many times through step 3 nor through step 5, where respectively an exit test and a stopping test are executed.

We prove first that the algorithm cannot pass infinitely times through step 3.

Assume by contradiction that this is the case and that the descent condition (3.3) is never satisfied. Thus the step 4 is executed infinitely many times as well and let us index by k all the quantities referred to the k-th passage.

It is easy to verify that the insertion rules at step 4 imply that there exists an index $\bar{k}$ such that $\forall k \geq \bar{k}$ all the newly generated bundle elements are assigned an index belonging to I_+ (and possibly also an index belonging to I_-); moreover in the stored bundle element we set:

$$\hat{\alpha}_k = \max(0, \hat{\alpha}_k) \ .$$

In addition the insertion rules at step 4, together with the deletion rule at step 5 of the bundle update procedure, ensure also that $d^k_{p\gamma}$ is bounded in norm, and, consequently there exists a subsequence, say $\{d^k_{p\gamma}\}_{k\in\mathcal{K}}$, converging to a limit $\hat{d}$.

The corresponding subsequence $\{\Delta^+(d^k_{p\gamma})\}_{k\in\mathcal{K}}$ is bounded and hence admits in turn a subsequence converging for $k \in \mathcal{K}' \subseteq \mathcal{K}$, say, to $\hat{\Delta}^+ \leq -\hat{\sigma}$ where $\hat{\sigma} \stackrel{\triangle}{=} \lim_{k\to\infty} \sigma_k > 0$.

Now let s and t be two successive indices in $\mathcal{K}'$. We have

$$\Delta^+(d^t_{p\gamma}) \geq \hat{g}^{(s)^T} d^t_{p\gamma} - \max(0, \hat{\alpha}_s) \ , \tag{4.4}$$

and

$$m\Delta^+(d^s_{p\gamma}) \leq \hat{g}^{(s)^T} d^s_{p\gamma} - \max(0, \hat{\alpha}_s) \ ,$$

the latter being consequence of both the definition of $\hat{\alpha}_s$ and of (3.4). The above relations, passing to the limit, imply

$$(1-m)\hat{\Delta}^+ \geq 0$$

which contradicts $\hat{\Delta}^+ \leq -\hat{\sigma}$.

To complete the proof we need to show that the algorithm cannot pass infinitely many times through step 5.

We assume by contradiction that this is the case and we index by $k \in \mathcal{K}$ all the quantities referred to the k-th passage.

We observe that $\forall k$ we have:

$$\Delta^+(d^k_\gamma) \geq \Delta^+(d^k_{p\gamma}) - \eta_k > -(\sigma_k + \eta_k) \ . \tag{4.5}$$

Moreover every time the step 5 is executed γ_{max} is reduced according to (3.5). Thus, taking into account that γ^k_{min} is monotonically nonincreasing and bounded, γ^k_{max} becomes arbitrarily close to γ^k_{min}. Consequently from Lemma 2.1 and since $\gamma^k_{min} < \dfrac{\epsilon}{2\|g^k_{I_-}\|}$, we have asintotically

$$\|d^k_{p\gamma}\| \leq \epsilon \ ,$$

which in turn implies that the bundle elements are definitely assigned an index belonging to I_+ (and possibly also an index in I_-). Moreover in the stored bundle element we set:

$$\hat{\alpha}_k = \max(0, \hat{\alpha}_k) \ .$$

Note also that, as consequence of the bundle deletion criterion at step 5 of the "main iteration", it is $a_i \leq \epsilon$, $\forall i \in I_+$.

Now, from (4.5) and recalling (2.6) we have

$$\|d^k_\gamma\| \leq \sqrt{(\sigma_k + \eta_k)\gamma_k}$$

and

$$-\frac{1}{\gamma_k} d^k_\gamma \in \mathrm{conv}\{g_i \mid i \in I^{(k)}_+\} \ ,$$

which imply

$$\|g^k_*\| \leq \sqrt{\frac{(\sigma_k + \eta_k)}{\gamma_k}} \leq \sqrt{\frac{(\sigma_k + \eta_k)}{\gamma^k_{min}}}$$

that, taking into account $\delta \geq \sqrt{\frac{(\sigma_k + \eta_k)}{\gamma^k_{min}}}$, contradicts the fact that the algorithm does not stop.
□

Now we prove the overall finiteness of the algorithm.

THEOREM 4.3. *For any $\epsilon > 0$ and $\delta > 0$, the algorithm stops in a finite number of "main iterations" at a point y satisfying the approximate stationarity condition*

$$\|g_*\| \leq \delta \quad \text{with } g_* \in \partial^G_\epsilon f(y) \ . \tag{4.6}$$

Proof. We need to prove that the stopping condition tested both at step 0 and at step 5 is verified in a finite number of executions of the "main iteration".

Suppose by contradiction that this is not the case. From Lemma 4.2 it follows that the descent condition (3.3) at step 3 is satisfied infinitely many times. Let $y^{(k)}$ be the stability center at the k-th "main iteration" and let us index by $k \in \mathcal{K}$ all the quantities referred to the k-th passage.

From

$$f(y^{(k+1)}) \leq f(y^{(k)}) + m\Delta^+(d^{(k)}_{p\gamma})$$

we would have:

$$f(y^{(k+1)}) - f(y^{(0)}) \leq m \sum_{i=0}^{k} \Delta^+(d^{(i)}_{p\gamma}) \ ,$$

and, consequently, taking into account $\Delta^+(d^{(i)}_{p\gamma}) \leq -\sigma_k$ and σ_k is bounded away from zero, we would obtain:

$$\lim_{k\to\infty} f(y^{(k+1)}) - f(y^{(0)}) \leq -\infty,$$

which is impossible, since f is bounded from below by hypothesis. □

5. Conclusions. We have described a model algorithm for the minimization of nonconvex and nonsmooth functions based on the simultaneous construction of two piecewise affine approximations and we have proved that the algorithmic scheme converges to a stationary point.

However, to have a practically implementable algorithm, a number of issues are still to be treated.

We mention in particular the problem of finding an appropriate subgradient aggregation technique to cope with bounded storage availability, the problem of effective

selection of the parameter γ at step 1 of the "main iteration" and, finally, that of the efficient implementation of the line search at step 4.

Appendix. We discuss the problem of finding a global minimum of the problem

$$\min_d f_{p\gamma}(d) \stackrel{\triangle}{=} \min_d \gamma \Delta_p(d) + \frac{1}{2}\|d\|^2 \ ,$$

which, according to the definitions given in section 2, can be written as

$$\min_d \gamma[p \max_{i\in I_+} \left\{g_i^T d - \alpha_i\right\} + (1-p) \min_{i\in I_-} \left\{g_i^T d - \alpha_i\right\}] + \frac{1}{2}\|d\|^2. \tag{5.1}$$

The problem above consists of minimizing an objective function which is the sum of a strictly convex (nonsmooth) function and of a piecewise affine concave function.

We can define for any $k \in I_-$ the following strictly convex nonsmooth optimization problem $P^{(k)}$:

$$\min_d f^k_{p\gamma}(d) \stackrel{\triangle}{=} \min_d \gamma[p \max_{i\in I_+} \left\{g_i^T d - \alpha_i\right\} + (1-p) \left\{g_k^T d - \alpha_k\right\}] + \frac{1}{2}\|d\|^2.$$

It is obvious that

$$f_{p\gamma}(d) \leq f^k_{p\gamma}(d) \ \ \forall d \in \mathbb{R}^n \ \ and \ \ \forall k \in I_-$$

and that for any $d \in \mathbb{R}^n$ there exists $k \in I_-$ such that

$$f_{p\gamma}(d) = f^k_{p\gamma}(d).$$

Hence it follows that a global minimizer $d_{p\gamma}$ of $f_{p\gamma}(d)$ can be located by finding the "best" from among all the minimizers of the functions $f^k_{p\gamma}(d)$, $k \in I_-$. Finding the (global) minimizer of any function $f^k_{p\gamma}(d)$ requires in turn to solve a convex quadratic programming of the type usually solved in bundle type methods.

In conclusion, the global minimization of $f_{p\gamma}(d)$ requires solution to $|I_-|$ quadratic programming problems.

The (necessary and sufficient) condition for point $d^{(k)}_{p\gamma}$ to be optimal for problem $P^{(k)}$ is:

$$0 \in d^{(k)}_{p\gamma} + \gamma p \partial \Delta^+(d_{p\gamma}) + (1-p) g_k. \tag{5.2}$$

Thus, by letting $K_- \subseteq I_-$ be the subset of indices such that $d_{p\gamma} = d^{(k)}_{p\gamma}$, condition (5.2) becomes

$$0 \in d_{p\gamma} + \gamma p \partial \Delta^+(d_{p\gamma}) + (1-p) g_k \ \ \forall k \in K_-$$

which can be easily recognized to be the local optimality condition for quasidifferential functions (see [3]). It is worth noting that, due to the particular structure of function $f_{p\gamma}(d)$, such condition is both necessary and sufficient.

REFERENCES

[1] E. W. Cheney and A. A. Goldstein, *Newton's method for convex programming and Tchebycheff approximation*, Numerische Mathematik, 1 (1959), pp. 253–268.

[2] F. Clarke, *Optimization and nonsmooth analysis*, John Wiley and Sons, 1983.

[3] V. F. Demyanov and A. Rubinov, *Constructive nonsmooth analysis*, Verlag Peter Lang, 1995.

[4] A. Fuduli, M. Gaudioso, and G. Giallombardo, *Minimizing nonconvex nonsmooth functions via cutting planes and proximity controls*. Submitted for publication, 2002.

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

[6] K. C. Kiwiel, *Restricted step and Levenberg-Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization*, Siam Journal on Optimization, 6 (1996), pp. 227–249.

[7] M. Makëla and P. Neittaanmäki, *Nonsmooth optimization*, World Scientific, 1992.

[8] R. Mifflin, *An algorithm for constrained optimization with semismooth functions*, Mathematics of Operations Research, 2 (1977), pp. 191–207.

[9] G. Outrata, M. Kocvara, and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints. Theory, applications and numerical results*, Kluwer Academic Publishers, 1998.

[10] H. Schramm and J. Zowe, *A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis*, Siam Journal on Optimization, 1 (1992), pp. 121–152.