TECHNICAL REPORT
Mechanical Engineering Program- COPPE
Federal University of Rio de Janeiro
July 2009

A Feasible Directions Method for Nonsmooth Convex Optimization

Jose Herskovits[1] Wilhelm P. Freire, [2] Mario Tanaka Fo [3]

Abstract: We propose a new technique for minimization of convex functions not necessarily smooth. Our approach employs an equivalent constrained optimization problem and approximated linear programs obtained with cutting planes. At each iteration a search direction and a step length are computed. If the step length is considered "non serious", a cutting plane is added and a new search direction is computed. This procedure is repeated until a "serious" step is obtained. When this happens, the search direction is a feasible descent direction of the constrained equivalent problem. The search directions are computed with FDIPA, the Feasible Directions Interior Point Algorithm. We prove global convergence and solve several test problems very efficiently.

Keywords: Unconstrained convex optimization; Non-smooth optimization; Bundle methods; Interior point methods

1 Introduction

In this paper, we propose a new algorithm for solving the unconstrained optimization problem:

$$\left\{ \min_{x \in \mathbb{R}^n} f(x) \right. \tag{P}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a closed convex function, not necessarily smooth. Let $\partial f(x)$ be the subdifferential [3] of f at x. In what follows, it is assumed that one arbitrary subgradient $s \in \partial f(x)$ can be computed at any point $x \in \mathbb{R}^n$.

---

1. COPPE, Federal University of Rio de Janeiro , Caixa Postal 68503, 21945 970 Rio de Jaineiro, Brazil. E-mail: jose@optimize.ufrj.br
2. UFJF, Federal University of Juiz de Fora, Juiz de Fora, Brazil.
3. COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

A special feature of nonsmooth optimization is the fact that $\nabla f(x)$ can change discontinuously and is not necessarily small in the neighborhood of a local extreme of the objective function, see [11, 2]. For this reason, the usual smooth gradient based optimization methods cannot be employed.

Several methods have been proposed for solving (P), see [1, 2, 14, 19]. Cutting plane methods approximate the function with a set of tangent planes. At each iteration the approximated function is minimized and a new tangent plane is added. The classical reference is [13], where a trial point is computed by solving a linear programming problem. We mention [21, 5] for analytic center cutting plane methods and [24, 20] for logarithmic potential and volumetric barrier cutting plane methods. Bundle methods based on the stabilized cutting plane idea are numerically and theoretically well understood [2, 14, 19, 24].

Here we show that the techniques involved in FDIPA [6, 7, 8, 9], the Feasible Direction Interior Point Algorithm for constrained smooth optimization, can be successfully combined with bundle methods to obtain a nonsmooth solver with a simple structure that is easy to implement and without need of solving quadratic programming subproblems.

In this paper, the nonsmooth unconstrained problem (P) is reformulated as an equivalent constrained program (EP) with a linear objective function and one nonsmooth inequality constraint,

$$\begin{cases} \min\limits_{(x,z)\in\mathbb{R}^{n+1}} \quad z \\ \text{s.t.} \quad f(x) \leq z, \end{cases} \tag{EP}$$

where $z \in \mathbb{R}$ is an auxiliary variable. With the present approach, a decreasing sequence of feasible points $\{(x^k, z^k)\}$ converging to a minimum of $f(x)$ is obtained. That is, we have that $z^{k+1} < z^k$ and $z^k > f(x^k)$ for all k. To compute a feasible descent direction we employ a procedure that combines the cutting plane technique with the FDIPA, [8]. At each iteration, an auxiliary linear program is defined by the substitution of $f(x)$ by cutting planes. A feasible descent direction of the linear program is obtained employing FDIPA, and a step-length is computed. Then, a new iterate (x^{k+1}, z^{k+1}) is defined according to suitable rules. To determine a new iterate, the algorithm produces auxiliary points (y_i, w_i) and when a auxiliary point is an interior point of epi(f), we say that the step is “serious” and we take it as the new iterate. Otherwise, the iterate is not changed and we say that the step is “null”. A new cutting plane is then added and the procedure is repeated until a serious step is obtained. It will be proved that, when a serious step is obtained, the search direction given by FDIPA is also a feasible descent direction of (EP).

This paper is organized in six sections. In the next one we describe the FDIPA. In section 3 the main features of the new method are presented and global convergence of the algorithm is shown in section 4. In the subsequent section, numerical preliminary comparative results with two well known bundle methods show that our method is strong an efficient. The last section contains some concluding remarks.

2 The feasible direction interior point algorithm

In this section we describe the basic ideas of the feasible direction interior point algorithm. The FDIPA [8] is a numerical technique for smooth nonlinear optimization with equality and inequality constraints. In this paper, we consider the inequality constrained optimization problem

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x) \\ \text{s.t.} \;\; g(x) \leq 0, \end{cases} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. The FDIPA requires the following assumptions about Problem (1):

Assumptions

Assumption 1. Let $\Omega \equiv \{x \in \mathbb{R}^n / g(x) \leq 0\}$ be the feasible set. There exists a real number a such that the set $\Omega_a \equiv \{x \in \Omega;\; f(x) \leq a\}$ is compact and has an interior Ω_a^0.

Assumption 2. Each $x \in \Omega_a^0$ satisfies $g(x) < 0$.

Assumption 3. The functions f and g are continuously differentiable in Ω_a and their derivatives satisfy a Lipschitz condition.

Assumption 4. (Regularity Condition) A point $x \in \Omega_a$ is *regular* if the gradient vectors $\nabla g_i(x)$, for i such that $g_i(x) = 0$, are linearly independent. FDIPA requires regularity assumption at a local solution of (1).

Let us remind some well known concepts [16], widely employed in this paper.

Definitions

Definition 1. $d \in \mathbb{R}^n$ is a *descent direction* for a smooth function $\phi : \mathbb{R}^n \to \mathbb{R}$ if $d^T \nabla \phi < 0$.

Definition 2. $d \in \mathbb{R}^n$ is a *feasible direction* for the problem (1), at $x \in \Omega$, if for some $\theta > 0$ we have $x + td \in \Omega$ for all $t \in [0, \theta]$.

Definition 3. A vector field $d(x)$ defined on Ω is said to be a *uniformly feasible directions field* of the problem (1), if there exists a step length $\tau > 0$ such that $x + td(x) \in \Omega$ for all $t \in [0, \tau]$ and for all $x \in \Omega$.

It can be shown that d is a feasible direction if $d^T \nabla g_i(x) < 0$ for any i such that $g_i(x) = 0$. Definition 2.3 introduces a condition on the vector field $d(x)$, which is stronger than the simple feasibility of any element of $d(x)$. When $d(x)$ constitutes a uniformly feasible directions field, it supports a feasible segment $[x, x + \theta(x)d(x)]$, such that $\theta(x)$ is bounded below in Ω by $\tau > 0$.

Let x^* be a regular point of Problem (1). Karush-Kuhn-Tucker (KKT) first order necessary optimality conditions are expressed as follows: If x^* is a local minimum of

(1) then there exists $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) + \nabla g(x^*)\lambda^* = 0 \tag{2}$$
$$G(x^*)\lambda^* = 0 \tag{3}$$
$$\lambda^* \geq 0 \tag{4}$$
$$g(x^*) \leq 0, \tag{5}$$

where $G(x)$ is a diagonal matrix with $G_{ii}(x) \equiv g_i(x)$.
We say that x such that $g(x) \leq 0$ is a "Primal Feasible Point", and $\lambda \geq 0$ a "Dual Feasible Point". Given an initial feasible pair (x^0, λ^0), FDIPA finds KKT points by solving iteratively the nonlinear system of equations (2, 3) in (x, λ), in such a way that all the iterates are primal and dual feasible. Therefore, convergence to feasible points is obtained.
A Newton-like iteration to solve the nonlinear system of equations (2, 3) in (x, λ) can be stated as

$$\begin{bmatrix} S^k & \nabla g(x^k) \\ \Lambda^k \nabla g^T(x^k) & G(x^k) \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda_\alpha^{k+1} - \lambda^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) + \nabla g(x^k)\lambda^k \\ G(x^k)\lambda^k \end{bmatrix} \tag{6}$$

where (x^k, λ^k) is the starting point of the iteration and $(x^{k+1}, \lambda_\alpha^{k+1})$ is a new estimate, and Λ a diagonal matrix with $\Lambda_{ii} \equiv \lambda_i$.
In the case when $S^k \equiv \nabla^2 f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 g_i(x^k)$, (6) is a Newton iteration. However, S^k can be a quasi-Newton approximation or even the identity matrix. FDIPA requires S^k symmetric and positive definite. Calling $d_\alpha^k = x^{k+1} - x^k$, we obtain the following linear system in $(d_\alpha^k, \lambda_\alpha^{k+1})$:

$$S^k d_\alpha^k + \nabla g(x^k)\lambda_\alpha^{k+1} = -\nabla f(x^k) \tag{7}$$
$$\Lambda^k \nabla g^T(x^k) d_\alpha^k + G(x^k)\lambda_\alpha^{k+1} = 0. \tag{8}$$

It is easy to prove that d_α^k is a descent direction of the objective function, [8]. However, d_α^k cannot be employed as a search direction, since it is not necessarily a feasible direction. In effect, in the case when $g_l(x^k) = 0$ it follows from (8) that $\nabla g_l(x^k)^T d_\alpha^k = 0$. To obtain a feasible direction, the following perturbed linear system with unknowns d^k and $\bar{\lambda}^{k+1}$ is defined, by adding the negative vector $-\rho^k \lambda^k$ to the right side of (8), with $\rho^k > 0$,

$$S^k d^k + \nabla g(x^k)\bar{\lambda}^{k+1} = -\nabla f(x^k)$$
$$\Lambda^k \nabla g^T(x^k) d^k + G(x^k)\bar{\lambda}^{k+1} = -\rho^k \lambda^k.$$

The addition of a negative vector in the right hand side of (8) produces the effect of deflecting d_α^k into the feasible region, where the deflection is proportional to ρ^k. As the deflection of d_α^k grows with ρ^k and d_α^k is a descent direction of f, it is necessary to bound ρ^k, in a way to ensure that d^k is also a descent direction. Since $d_\alpha^{k^T} \nabla f(x^k) < 0$, we can get these bounds by imposing

$$d^{k^T} \nabla f(x^k) \leq \xi d_\alpha^{k^T} \nabla f(x^k), \tag{9}$$

with $\xi \in (0,1)$, which implies $d^{k^T} \nabla f(x^k) < 0$. Thus, d^k is a feasible descent direction. To obtain the upper bound on ρ^k, the following auxiliary linear system in $(d_\beta^k, \lambda_\beta^k)$ is solved:

$$\begin{aligned} &S^k d_\beta^k + \nabla g(x^k) \lambda_\beta^{k+1} = 0 \\ &\Lambda^k \nabla g^T(x^k) d_\beta^k + G(x^k) \lambda_\beta^{k+1} = -\lambda^k. \end{aligned}$$

Now defining $d^k = d_\alpha^k + \rho^k d_\beta^k$ and substituting in (9), it follows that d^k is a descent direction for any $\rho^k > 0$ in the case when $d_\beta^{k^T} \nabla f(x^k) \leq 0$. Otherwise, the following condition is required,

$$\rho^k \leq (\xi - 1) d_\alpha^{k^T} \nabla f(x^k) / d_\beta^{k^T} \nabla f(x^k).$$

In [8], ρ is defined as follows: If $d_\beta^{k^T} \nabla f(x^k) \leq 0$, then $\rho^k = \varphi \|d_\alpha^k\|^2$. If not, $\rho^k = \min[\varphi \|d_\alpha^k\|^2, (\xi - 1) d_\alpha^{k^T} \nabla f(x^k) / d_\beta^{k^T} \nabla f(x^k)]$.
A new feasible primal point with a lower objective value is obtained through an inexact line search along d^k, [8]. FDIPA has global convergence in the primal space for any way of updating S and λ, provided that S^{k+1} is positive definite and $\lambda^{k+1} > 0$ [8].
The following updating rule for λ can be employed:
Set, for $i = 1, ..., m$,

$$\lambda_i := \max\, [\lambda_{\alpha i};\ \epsilon \|d_\alpha\|^2],\ \epsilon > 0.$$

3 Description of the present technique for nonsmooth optimization

We employ ideas of the cutting planes method [13], to build piecewise linear approximations of the constraints of (EP). Let $g_i^k(x,z)$ be the current set of cutting planes such that

$$g_i^k(x,z) = f(y_i^k) + (s_i^k)^T (x - y_i^k) - z, \quad i = 0, 1, ..., \ell$$

where $y_\ell^k \in \mathbb{R}^n$ are auxiliary points, $s_i^k \in \partial f(y_i^k)$ are subgradients at those points and ℓ represents the number of current cutting planes. Let be,

$$\tilde{g}_\ell^k(x,z) \equiv [g_0^k(x,z), ..., g_\ell^k(x,z)]^T, \ \ \tilde{g}_\ell^k : \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathbb{R}^{\ell+1}$$

and the current auxiliary problem

$$\left\{ \begin{array}{ll} \displaystyle\min_{(x,z) \in \mathbb{R}^{n+1}} & \psi(x,z) = z \\ \text{s.t.} & \tilde{g}_\ell^k(x,z) \leq 0. \end{array} \right. \tag{AP_ℓ^k}$$

Instead of solving this problem, the present algorithm merely computes with FDIPA a search direction d_ℓ^k of (AP_ℓ^k). We note that d_ℓ^k can be computed even if (AP_ℓ^k) has not a finite minimum.
The largest feasible step is $t \equiv \max\{t \mid \tilde{g}_\ell^k((x^k, z^k) + td_\ell^k) \leq 0\}$. Since t is not always finite, it is taken

$$t_\ell^k := \min\{t_{max}/\mu, \quad t\}$$

where $\mu \in (0,1)$. Then,

$$(x_{\ell+1}^k, z_{\ell+1}^k) = (x^k, z^k) + t_\ell^k d_\ell^k \tag{11}$$

is feasible with respect to (AP_i^k). Next we compute the following auxiliary point

$$(y_{\ell+1}^k, w_{\ell+1}^k) = (x^k, z^k) + \mu t_\ell^k d_\ell^k, \tag{12}$$

where $\mu \in (0,1)$. If $(y_{\ell+1}^k, w_{\ell+1}^k)$ is feasible with respect to (EP), that is, if $w_{\ell+1}^k > f(y_{\ell+1}^k)$ we consider that the current set of cutting planes is a good local approximation of $f(x)$ in a neighborhood of x^k. Then, we say that the "step is serious" and set the new iterate $(x^{k+1}, z^{k+1}) = (y_{\ell+1}^k, w_{\ell+1}^k)$. Otherwise, a new cutting plane $g_{\ell+1}^k(x,z)$ is added to the approximated problem and the procedure repeated until a serious step is obtained. We are now in position to state our algorithm.

3.1 Algorithm - FD_NS

Parameters. $\xi, \mu \in (0,1)$, $\varphi > 0$, $t_{max} > 0$.
Data. x^0, $z^0 > f(x^0)$, $\lambda_0^0 \in \mathbb{R}^+$, $B^0 \in \mathbb{R}^{n+1} \times \mathbb{R}^{n+1}$ symmetric and positive definite. Set $y_0^0 = x^0$, $k = 0$ and $\ell = 0$.
Step 1) Compute $s_\ell^k \in \partial f(y_\ell^k)$. A new cutting plane at the current iterate (x^k, z^k) is defined by

$$g_\ell^k(x,z) = f(y_\ell^k) + (s_\ell^k)^T(x - y_\ell^k) - z, \quad g_\ell^k(x,z) \in \mathbb{R}.$$

Consider now

$$\nabla g_\ell^k(x,z) = \begin{bmatrix} (s_\ell^k)^T \\ \\ -1 \end{bmatrix}, \quad \nabla g_\ell^k(x,z) \in \mathbb{R}^{n+1}$$

define

$$\tilde{g}_\ell^k(x,z) = [g_0^k(x,z), ..., g_\ell^k(x,z)]^T, \quad \tilde{g}_\ell^k(x,z) \in \mathbb{R}^{\ell+1},$$

and

$$\nabla \tilde{g}_\ell^k(x,z) = [\nabla g_0^k(x,z), ..., \nabla g_\ell^k(x,z)]^T, \quad \nabla \tilde{g}_\ell^k(x,z) \in \mathbb{R}^{(n+1)\times(\ell+1)}.$$

Step 2) Calculation of a Feasible Descent Direction d_ℓ^k for (AP_ℓ^k)
i) Compute $d_{\alpha\ell}^k$ and $\lambda_{\alpha\ell}^k$, solving

$$B^k d_{\alpha\ell}^k + \nabla \tilde{g}_\ell^k(x^k, z^k)\tilde{\lambda}_{\alpha\ell}^k = -\nabla\psi(x,z) \tag{13}$$

$$\tilde{\Lambda}_\ell^k [\nabla \tilde{g}_\ell^k(x^k, z^k)]^T d_{\alpha\ell}^k + \tilde{G}_\ell^k(x^k, z^k)\tilde{\lambda}_{\alpha\ell}^k = 0. \tag{14}$$

Compute $d_{\beta\ell}^k$ and $\lambda_{\beta\ell}^k$, solving

$$B^k d_{\beta\ell}^k + \nabla\tilde{g}_\ell^k(x^k,z^k)\tilde{\lambda}_{\beta\ell}^k = 0 \tag{15}$$

$$\tilde{\Lambda}_\ell^k[\nabla\tilde{g}_\ell^k(x^k,z^k)]^T d_{\beta\ell}^k + \tilde{G}_l^k(x^k,z^k)\tilde{\lambda}_{\beta\ell}^k = -\tilde{\lambda}_\ell^k, \tag{16}$$

where $\tilde{\lambda}_{\alpha\ell}^k := (\lambda_{\alpha 0}^k, ..., \lambda_{\alpha\ell}^k)$, $\tilde{\lambda}_{\beta\ell}^k := (\lambda_{\beta 0}^k, ..., \lambda_{\beta\ell}^k)$, $\tilde{\lambda}_\ell^k := (\lambda_0^k, ..., \lambda_\ell^k)$, $\tilde{\Lambda}_\ell^k := \text{diag}(\lambda_0^k, ..., \lambda_\ell^k)$ and $\tilde{G}_\ell^k(x,z) := \text{diag}(g_0^k(x,z), ..., g_\ell^k(x,z))$.

ii) If $d_{\beta\ell}^{k^T}\nabla\psi(x,z) > 0$, set $\rho = \varphi\|d_{\alpha\ell}^k\|^2$.
Otherwise, set $\rho = \min\left\{\varphi\|d_{\alpha\ell}^k\|^2,\ (\xi-1)\dfrac{d_{\alpha\ell}^{k^T}\nabla\psi(x,z)}{d_{\beta\ell}^{k^T}\nabla\psi(x,z)}\right\}$.
iii) Compute the feasible descent direction $d_\ell^k = d_{\alpha\ell}^k + \rho d_{\beta\ell}^k$.
Step 3) Compute the step length

$$t_\ell^k = \min\left\{t_{max}/\mu,\ \max\{t \mid \tilde{g}_\ell^k((x^k,z^k) + t d_\ell^k) \le 0\}\right\}. \tag{17}$$

Step 4) Compute a new point
i) Set $(y_{\ell+1}^k, w_{\ell+1}^k) = (x^k,z^k) + \mu t_\ell^k d_\ell^k$.
ii) If $w_{\ell+1}^k \le f(y_{\ell+1}^k)$, we have a ***null step***. Then, define $\lambda_{\ell+1}^k$ and set $\ell := \ell+1$.
Otherwise, we have a ***serious step***. Then, call $d^k = d_\ell^k$, $d_\alpha^k = d_{\alpha\ell}^k$, $d_\beta^k = d_{\beta\ell}^k$ $\lambda_\alpha^k = \lambda_{\alpha\ell}^k$, $\lambda_\beta^k = \lambda_{\beta\ell}^k$ and $\ell^k = \ell$. Take $(x^{k+1}, z^{k+1}) = (y_{\ell+1}^k, w_{\ell+1}^k)$, define λ_0^{k+1}, B^{k+1} and set $k = k+1$, $\ell = 0$, $y_0^k = x^k$.
iii) Go to Step 1). □

4 Convergence analysis

In this section, we prove global convergence of the present algorithm. We first show that the search direction d_ℓ^k is a descent direction for z. Then, we prove that the number of null steps at each iteration is finite. That is; since $(x^k,z^k) \in \text{int}(\text{epi}\, f)$, after a finite number of subiterations, we obtain $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi}\, f)$. In consequence, the sequence $\{(x^k,z^k)\}_{k\in\mathbb{N}}$ is bounded and belongs the interior of the epigraph of f. Then, we show that any accumulation point of the sequence $\{(x^k,z^k)\}_{k\in\mathbb{N}}$ is a solution of the problem (P). For this, among other results, we have to show that d^k converges to zero when $k \to \infty$. This fact is employed to establish a stopping criterium for the present algorithm.
Finally, we show that for the accumulations points (x^*,z^*) of the sequence $\{(x^k,z^k)\}_{k\in\mathbb{N}}$, the optimality condition $0 \in \partial f(x^*)$ is satisfied.
In some cases indices will be omitted to simplify the notation. We introduce the following assumptions about B, λ and the set of cutting planes.

Assumption 1. There exist positive numbers σ_1 and σ_2 such that $\sigma_1\|v\|^2 \le v^T B v \le \sigma_2\|v\|^2$ for any $v \in \mathbb{R}^{n+1}$.

Assumption 2. There exist positive numbers λ^I, λ^S, such that $\lambda^I \le \lambda_i \le \lambda^S$, for $i = 0, 1, \ldots, \ell$.

We remark that the solutions d_α, λ_α, d_β, and λ_β of the linear systems (13), (14), and (15), (16) are unique. This fact is a consequence of a lemma proved in [22, 25] and stated as follows:

Lemma 4.1. *For any vector $(x, z) \in int(\text{epi}\, f)$ and any positive definite matrix $B \in \mathbb{R}^{(n+1)\times(n+1)}$, the matrix*

$$\begin{bmatrix} B & \nabla\tilde{g}(x,z) \\ \tilde{\Lambda}[\nabla\tilde{g}(x,z)]^T & \tilde{G}(x,z) \end{bmatrix},$$

is nonsingular. In addition we assume that the set of cutting planes is selected in such a way that that the previous matrix remains bounded bellow.

It follows that $d_\alpha, d_\beta, \lambda_\alpha$ and λ_β are bounded in epi f. Since ρ is bounded above we also have that $\bar{\lambda} = \lambda_\alpha + \rho\lambda_\beta$ is bounded.

Lemma 4.2. *The vector d_α satisfies $d_\alpha^T \nabla\psi(x,z) \le -d_\alpha^T B d_\alpha$.*

Proof. It follows from (13)

$$d_\alpha^T B d_\alpha + d_\alpha^T \nabla\tilde{g}(x,z)\lambda_\alpha = -d_\alpha^T \nabla\psi(x,z), \tag{18}$$

and from (14)

$$d_\alpha^T \nabla\tilde{g}(x,z) = -\lambda_\alpha^T \tilde{\Lambda}^{-1}\tilde{G}(x,z). \tag{19}$$

Replacing (19) in (18) we have $d_\alpha^T \nabla\psi(x,z) = -d_\alpha^T B d_\alpha + \lambda_\alpha^T \tilde{\Lambda}^{-1}\tilde{G}(x,z)\lambda_\alpha$. Since $\tilde{\Lambda}^{-1}\tilde{G}(x,z)$ is negative semidefinite, the result of the lemma is obtained. □

As a consequence, we also have that the search direction d_α is descent for the objective function the problem (AP_ℓ^k).

Proposition 4.3. *The direction $d = d_\alpha + \rho d_\beta$ is a descent direction for the objective function of the problem (AP_ℓ^k) in point x^k.*

Proof. Since $d = d_\alpha + \rho d_\beta$, we have $d^T\nabla\psi(x,z) = d_\alpha^T\nabla\psi(x,z) + \rho d_\beta^T \nabla\psi(x,z)$.

In the case when $d_\beta^T\nabla\psi(x,z) > 0$, we have $\rho \le (\xi - 1)\dfrac{d_\alpha^T\nabla\psi(x,z)}{d_\beta^T\nabla\psi(x,z)}$. Therefore, $d^T\nabla\psi(x,z) \le d_\alpha^T\nabla\psi(x,z) + (\xi-1)d_\alpha^T\nabla\psi(x,z) = \xi d_\alpha^T\nabla\psi(x,z) < 0$. On the other hand, when $d_\beta^T\nabla\psi(x,z) \le 0$, it follows from Lemma 4.2 that $d^T\nabla\psi(x,z) \le d_\alpha^T\nabla\psi(x,z) < 0$ for any $\rho > 0$. □

As a consequence of previous Lemma, we have that $z^{k+1} < z^k$ for all k.

Proposition 4.4. *The sequence $\{(x^k, z^k)\}_{k\in\mathbb{N}}$ generated by the present algorithm is bounded.*

Proof. Since f is a closed convex function, we have that the level sets of f are bounded. Then, the sequence $\{(x^k, z^k)\}_{k\in\mathbb{N}}$ is contained in the bounded set epi $f \cap \{(x,z) \in \mathbb{R}^{n+1} \mid f(x^k) \le z^0\}$. □

Lemma 4.5. *Let $X \subset \mathbb{R}^n$ be a convex set. Consider $x_0 \in intX$ and $\bar{x} \in X$. Let $\{\bar{x}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n - X$ be a sequence such that $\bar{x}^k \to \bar{x}$. Let $\{x^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ be a sequence defined by $x^k = x_0 + \mu(\bar{x}^k - x_0)$ with $\mu \in (0,1)$. Then there exist $k_0 \in \mathbb{N}$ such that $x^k \in intX, \ \forall\ k > k_0$.*

Proof. We have $x^k = x_0 + \mu(\bar{x}^k - x_0) \longrightarrow x_0 + \mu(\bar{x} - x_0) = x_\mu$. Since the segment $[x_0, \bar{x}] \subset X$ and $\mu \in (0,1)$ we have that $x_\mu \in \text{int}X$ and, in consequence there exist $\delta > 0$ such that $B(x_\mu, \delta) \subset \text{int}X$. Since $x^k \longrightarrow x_\mu$ there exist $k_0 \in \mathbb{N}$ such that $x^k \in B(x_\mu, \delta) \subset \text{int}X$, $\forall k > k_0$. □

Proposition 4.6. *Consider the sequence $\{(x_\ell^k, z_\ell^k)\}_{\ell\in\mathbb{N}}$ defined in (11) for k fixed. If and $(\tilde{x}^k, \tilde{z}^k)$ is an accumulation point of the sequence, then $\tilde{z}^k = f(\tilde{x}^k)$.*

Proof. By definition of the sequence $\{(x_\ell^k, z_\ell^k)\}_{\ell\in\mathbb{N}}$, we have always that $\tilde{z}^k \le f(\tilde{x}^k)$. Suppose now that $\tilde{z}^k < f(\tilde{x}^k)$ and consider a convergent sequence $\{(x_\ell^k, z_\ell^k)\}_{\ell\in\mathbb{N}'} \to (\tilde{x}^k, \tilde{z}^k)$ such that $\{s_\ell^k\}_{\ell\in\mathbb{N}'} \to \tilde{s}^k$, where $\mathbb{N}' \subset \mathbb{N}$. This sequence exists because $\{(x_\ell^k, z_\ell^k)\}_{\ell\in\mathbb{N}'}$ as well as $\{s_\ell^k\}_{\ell\in\mathbb{N}'}$ are in compact sets by Assumption 3. The corresponding cutting plane is represented by $f(x_\ell^k) + s_\ell^T(x - x_\ell^k) - z = 0$. Then, $z(\tilde{x}^k) = f(x_\ell^k) + s_\ell^T(\tilde{x}^k - x_\ell^k)$ is the vertical projection of $(\tilde{x}^k, \tilde{z}^k)$ on the cutting plane. Taking the limit for $\ell \to \infty$, we get $z(\tilde{x}^k) = f(\tilde{x}^k)$. Then, for $\ell \in \mathbb{N}' > L$ large enough, $(\tilde{x}^k, \tilde{z}^k)$ is under the ℓth cutting plane. Then, we arrived to a contradiction and $\tilde{z}^k = f(\tilde{x}^k)$. □

Proposition 4.7. *Let $(x^k, z^k) \in int(\text{epi}\, f)$. The next iterate $(x^{k+1}, z^{k+1}) \in int(\text{epi}\, f)$ is obtained after a finite number of subiterations.*

Proof. Our proof starts with the observation that in the step 4) of the algorithm we have that $(x^{k+1}, z^{k+1}) = (y_{\ell+1}^k, w_{\ell+1}^k)$ only if $w_{\ell+1}^k > f(y_{\ell+1}^k)$ (i.e., if we have a serious step), consequently, we have that $(x^{k+1}, z^{k+1}) \in \text{int}(\text{epi}\, f)$.
The sequence $\{(x_\ell^k, z_\ell^k)\}_{\ell\in\mathbb{N}}$ is bounded by construction and, by Proposition 4.6, it has an accumulation point $(\tilde{x}^k, \tilde{z}^k)$ such that $\tilde{z}^k = f(\tilde{x}^k)$. Considering now the sequence defined by (12),

$$(y_\ell^k, w_\ell^k) = (x^k, z^k) + \mu \left\| (x_\ell^k, z_\ell^k) - (x^k, z^k) \right\|, \quad \mu \in (0,1)$$

it follows from Lemma 4.5, that there exist $k_0 \in \mathbb{N}$ such that $(y_\ell^k, w_\ell^k) \in \text{int}(\text{epi}\, f)$, for $k > k_0$. But this is the condition for a serious step and the proof is complete. □

Lemma 4.8. *There exists $\tau > 0$ such that $\tilde{g}_\ell((x,z) + td) \le 0$, $\forall\ \ t \in [0, \tau]$ for any $(x,z) \in int(\text{epi}\, f)$ and any direction d given by the algorithm.*

Proof. Let us denote by b a vector such that $b_i = s_i^T x_i - f(x_i)$ for all $i = 0, 1, ..., \ell$. Then, $\tilde{g}_\ell((x,z) + td) = (\nabla \tilde{g}_\ell(x,z))^T(x,z) - b$, since

$$g_i(x,z) = f(y_i) + s_i^T(x - y_i) - z = [s_i^T \ \ -1](x,z) - b_i = (\nabla g_i(x,z))^T(x,z) - b_i$$

for all $i = 0, 1, ..., \ell$. The step length t is defined in the equation (17) in the Step 3) of the algorithm. Since the constraints of (AP) are linear, to satisfy the line search condition, the following inequalities must be true:

$$\begin{aligned} g_i((x,z)+t_id) &= (\nabla g_i((x,z)+t_id)^T((x,z)+t_id) - b_i \\ &= g_i(x,z) + t_i(\nabla g_i(x,z))^T d \leq 0, \quad \text{for all } i = 0,1,...,\ell. \end{aligned} \tag{20}$$

If $(\nabla g_i(x,z))^T d \leq 0$ the inequality is satisfy for all $t > 0$. Otherwise, it follows from iii) in the Step 2) that,

$$(\nabla g_i(x,z))^T d = (\nabla g_i(x,z))^T (d_\alpha + \rho d_\beta).$$

But

$$(\nabla g_i(x,z))^T d_\alpha = -g_i(x,z)\frac{\lambda_{\alpha i}}{\lambda_i}$$

and

$$(\nabla g_i(x,z))^T d_\beta = -1 - g_i(x,z)\frac{\lambda_{\beta i}}{\lambda_i}.$$

Then (20) is equivalent to

$$g_i(x,z)(1 - t_i\frac{\bar{\lambda}_i}{\lambda_i}) - \rho t_i \leq 0.$$

Since $\rho t_i > 0$, the last inequality will be satisfied when $t_i \leq \lambda_i/\bar{\lambda}_i$.
By construction, $\lambda > 0$ is bounded, $\bar{\lambda}$ is bounded from above. Thus, there exists $0 < \tau < t_{max}/\mu$ such that $\tau < \lambda_i/\bar{\lambda}_i$ for all $i = 0,1,...,\ell$. Therefore, for all $t \in [0,\tau]$ the line search condition $g_i((x,z)+td) \leq 0$ is satisfied for all $i = 0,1,...,\ell$. □

Proposition 4.9. *Let d_α^* be a accumulation point of the sequence $\{d_\alpha^k\}_{k\in\mathbb{N}}$. Then $d_\alpha^* = 0$.*

Proof. From Step 3) of the algorithm, we have $(x^{k+1}, z^{k+1}) = (x^k, z^k) + \mu t^k(d_x^k, d_z^k)$, thus

$$z^{k+1} = z^k + \mu t^k d_z^k. \tag{21}$$

The sequence $\{z^k\}_{k\in\mathbb{N}}$ is decreasing and bounded by Assumption 3. Let us denote by $z^* = \lim_{k\to\infty} z^k$ and $\mathbb{N}' \subset \mathbb{N}$ such that $\{t^k\}_{k\in\mathbb{N}'} \to t^*$. It follows from Lemma 4.8 that $t^* > 0$. When $k \to \infty$, $k \in \mathbb{N}'$ on (21) we have $z^* = z^* + \mu t^* d_z^*$, thus $d_z^* = 0$.
From Proposition 4.3 it follows that $0 = d_z^* \leq \xi d_\alpha^* \nabla\psi(x,z) = \xi d_{\alpha z}^* \leq 0$, thus $d_{\alpha z}^* = 0$.
Further, by Lemma 4.2, we have $0 = d_{\alpha z}^* = d_\alpha^* \nabla\psi(x,z) \leq -d_\alpha^* B d_\alpha^* \leq 0$, thus $d_\alpha^* = 0$, since B is positive definite. □

It follows from the previous result that $d^k = d_\alpha^k + \rho_k d_\beta^k \to 0$ when $k \to \infty$, $k \in \mathbb{N}'$, since $\rho_k \to 0$ if $d_\alpha^k \to 0$.

Proposition 4.10. *For any accumulation point (x^*, z^*) of the sequence $\{(x^k, z^k)\}_{k\in\mathbb{N}}$, we have $0 \in \partial f(x^*)$.*

Proof. Let be the following convex optimization problem,

$$\begin{cases} \min\limits_{(x,z)\in\mathbb{R}^{n+1}} & \phi(x,z) \\ \text{s.t.} & \tilde{g}_\ell^k(x,z) \le 0 \end{cases}$$

where $\phi(x,z) = \psi(x,z)+(d_\alpha^k)^t Bx$. A Karush-Kuhn-Tucker point (x_ϕ, z_ϕ) of the problem above satisfies,

$$\nabla\psi(x,z) + Bd_\alpha^k + \nabla\tilde{g}(x_\phi,z_\phi)\tilde{\lambda}_\phi = 0 \tag{22}$$

$$\tilde{G}(x_\phi,z_\phi)\tilde{\lambda}_\phi = 0 \tag{23}$$

$$\tilde{\lambda}_\phi \ge 0 \tag{24}$$

$$\tilde{g}_\ell^k(x_\phi,z_\phi) \le 0. \tag{25}$$

Now observe that the system (13), (14) in the step 2) of the algorithm, can be rewritten as

$$\nabla\psi(x,z) + B^k d_\alpha^k + \nabla\tilde{g}_\ell^k(x^k,z^k)\tilde{\lambda}_\alpha^k = 0 \tag{26}$$

$$\tilde{G}_\ell^k(x^k,z^k)\tilde{\lambda}_\alpha^k = \delta^k \tag{27}$$

where $\delta^k = -\tilde{\Lambda}_\ell^k[\nabla\tilde{g}_\ell^k(x^k,z^k)]^T d_\alpha^k$. When $d_\alpha^k \to 0$ we have that $\delta^k \to 0$ and then, given $\varepsilon_1 > 0$, there exists $K_1 > 0$ such that,

$$\left\|\tilde{\lambda}_\alpha^k - \tilde{\lambda}_\phi\right\| < \varepsilon_1 \quad \text{for} \quad k > K_1.$$

Then, as $\tilde{\lambda}_\phi \ge 0$ from (24), we deduce that $\tilde{\lambda}_\alpha^k \ge 0$ for k large enough.
Consider now $\mathbb{Y} := \{y_0^0, y_1^0, ..., y_{l^0}^0, y_0^1, y_1^1, ..., y_{l^1}^1, ..., y_0^k, y_1^k, ..., y_{\ell^k}^k, ...\}$, the sequence of all the points obtained by the sub-iterations of the algorithm. Since this sequence is bounded, we can find a subsequence $\bar{\mathbb{Y}} \subset \mathbb{Y}$ such that $\bar{\mathbb{Y}} \to x^*$.
We call $\bar{\mathbb{Y}}^k \equiv \bar{\mathbb{Y}} \cap \{y_0^k, y_1^k, ..., y_{\ell^k}^k\}$. It follows from (13) in the step 2) of the algorithm, that

$$\lim_{k\to\infty} \sum_{i=0}^{\ell^k} \lambda_{\alpha i}^k s_i^k = 0 \quad \text{and} \quad \lim_{k\to\infty} \sum_{i=0}^{\ell^k} \lambda_{\alpha i}^k = 1. \tag{28}$$

Let be $\mathbb{I}^k = \{i \mid y_i^k \in \bar{\mathbb{Y}}^k\}$. Then,

$$\lim_{k\to\infty} \sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k s_i^k = 0 \quad \text{and} \quad \lim_{k\to\infty} \sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k = 1. \tag{29}$$

Consider the auxiliary point y_i^k and the subgradient $s_i^k \in \partial f(y_i^k)$ such that i is the index of active constraint. By definition of the subdifferential [3], we can write

$$f(x) \ge f(y_i^k) + (s_i^k)^T(x - y_i^k) = f(x^k) + (s_i^k)^T(x - x^k) - \varepsilon_k^i,$$

where $\varepsilon_k^i = f(x^k) - f(y_i^k) - (s_i^k)^T(x^k - y_i^k)$. Thus, $s_i^k \in \partial_{\varepsilon_k^i} f(x^k)$, where $\partial_\varepsilon f(x)$ represents the ε-subdifferential of f, [3]. Then, we can write

$$\left(\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k\right) f(x) \geq \left(\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k\right) f(x^k) + (\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k s_i^k)^T(x - x^k) - \sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k \varepsilon_k^i,$$

and

$$f(x) \geq f(x^k) + (\sum_{i\in\mathbb{I}^k} \frac{\lambda_{\alpha i}^k}{\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k} s_i^k)^T(x - x^k) - \varepsilon_k,$$

where $\varepsilon_k = \frac{\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k \varepsilon_k^i}{\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k}$. Let be $s^k = (\sum_{i\in\mathbb{I}^k} \frac{\lambda_{\alpha i}^k}{\sum_{i\in\mathbb{I}^k} \lambda_{\alpha i}^k} s_i^k)$.
It follows from (29) that $s^k \in \partial_{\varepsilon_k} f(x^k)$ and also that $0 \in \partial f(x^*)$. □

With this result the proof of convergence is complete.

5 Numerical results

In this section, we give the numerical results obtained with the present algorithm employing a set of fixed default parameters, (FD_NS/DP), and with parameters selected looking for better results, (FD_NS/BR). We compare our results with the standard bundle method described in [18], and with the proximal bundle method, described in [10]. In the last case the results with two different values of the ε-subgradient, $\varepsilon_1 := 10^{-5}$ and $\varepsilon_2 := 10^{-2}$, are described. A collection of well known convex test problems, that can be found in [17] or in [19] is employed. The results are reported in Tables 1 and 2.

Table 1: Comparison with Standard Bundle Method

	Bundle method			FD_NS/BR			FD_NS/DP			
Problem	NI	NF	f	NI	NF	f	NI	NF	f	f^*
CB2	31	33	1.95222	17	18	1.95224	17	18	1.95256	1.95222
CB3	14	16	2.00000	15	16	2.00015	30	31	2.00016	2
DEM	17	19	-3.00000	15	16	-2.99162	31	32	-2.99861	-3
QL	13	15	7.20000	21	22	7.20002	21	22	7.20002	7.2
LQ	11	12	-1.41421	08	09	-1.41429	19	20	-1.41417	-1.41421
Mifflin1	66	68	-0.99999	20	21	-0.99997	22	23	-0.99991	-1
Rosen	43	45	-43.99999	44	45	-43.99994	48	49	-43.99996	-44
Shor	27	29	22.60016	41	42	22.60028	60	61	22.60016	22.60016
Maxquad	74	75	-0.84140	62	63	-0.84140	134	135	-0.84140	-0.84140
Maxq	150	151	0.16712e-06	157	158	1.29621e-8	244	245	3.32418e-8	0
Maxl	39	40	0.12440e-12	51	52	2.39888e-4	75	76	2.40691e-4	0
TR48	245	251	-638530.48	161	162	-638564.99	161	162	-638564.99	-638565
Goffin	52	53	0.11665e-11	64	65	5.88385e-5	77	78	2.88556e-4	0
Badguy	-	-	-	5	6	3.20e-4	5	6	3.20e-4	0

We call NI the number of iterations, NF the number of function evaluations, f^* the known optimal function value and f the computed one.

Table 2: Comparison with the Proximal Bundle Method

	PB (ε_1)			PB (ε_2)			FD_NS/BR			
Problem	NI	NF	f	NI	NI	f	NI	NF	f	f^*
CB2	23	32	1.95222	11	32	1.95270	17	18	1.95224	1.95222
CB3	22	33	2.00000	17	32	2.000022	15	16	2.00015	2
QL	40	70	7.20000	22	59	7.20009	22	23	7.20001	7.2
Mifflin1	30	59	-0.99999	27	58	-0.99999	20	21	-0.99997	-1
Rosen	52	98	-43.99999	32	68	-43.99972	44	45	-43.99994	-44
Shor	56	135	22.60016	41	128	22.60097	41	42	22.60028	22.60016
Maxq	319	329	0.00000	252	267	0.00000	170	171	4.95919e-8	0

The default parameters for the present method are: $B = 1/2^I$, where I is the iteration number, $\mu = 0.75$, $\varphi = 0.1$, $\xi = 0.7$, $\mathrm{t_{max}} = 1$. Furthermore we store up to $5n$ subgradients for all the test problems. The iterates stop when $\left\|d^k\right\| \leq 10^{-4}$.

6 Conclusions

In this paper, a new approach for unconstrained nonsmooth convex optimization was introduced. The present algorithm, that is very simple to code, does not require the solution of quadratic programming subproblems but just of two linear systems with the same matrix. Global convergence was proved and some numerical results were presented. This results compare favorably with well established techniques. A set of test problems was efficiently solved with the same values of parameters, indicating that our approach is strong and the corresponding code can be employed by nonexperts in mathematical programming.

7 Acknowledgments

The authors wish to thank Napsu Karmitsa (University of Turku, Finland) for her careful reading of the first version and for helpful comments. This work was financially supported by CNPQ and FAPERJ.

References

[1] Auslender, A. *Numerical methods for nondifferentiable convex optimization.* Mathematical Programming Study, (30)(1987), 102-126.

[2] Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C. *Optimization Numérique, Aspects Théoriques et Pratiques, Mathématiques & Applications, 27.* Springer-Verlag, Berlin, 1997.

[3] Clarke, F. H. *Optimization and Nonsmooth Analysis.* Wiley-Interscience, New York, 1983.

[4] Freire W.P. *Um algoritmo de direções viáveis para otimização convexa não diferenciável*, DSc. Tese, Universidade Federal do Rio de Janeiro, 2005.

[5] Goffin, J.L. and Vial, J.P. *Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method.* Optimization Methods and Software 17, (5)(2002), 808-8678.

[6] Herskovits, J. *Développement d'une Méthode Númerique pour l'Optimisation non-Linéaire.* DSc. Tese, Paris IX University, (in English), 1982.

[7] Herskovits, J. *A two-stage feasible directions algorithm for nonlinear constrained optimization.* Mathematical Programming, vol.36(1986), pp. 19-38.

[8] Herskovits, J. *Feasible direction interior-point technique for nonlinear optimization.* JOTA Journal of Optimization Theory and Applications, Londres, v. 99, n. 1, p. 121-146, 1998.

[9] Herskovits, J., Goulart, E., Aroztegui J.M., *Optimization of Large Scale Systems.* In: Jasbir Arora. (Org.). Optimization of Structural and Mechanical Systems. Singapura: World Scientific Publishing / Imperial College Press, v. , p. 35-57, 2007.

[10] Hintermüller, M. *A Proximal Bundle Method Based on Approximate Subgradients.* Journal of Optimization Theory and Applications 99, (1)(1998),121-146.

[11] Hiriart-Urruty, J.B., Lemaréchal, C., *Convex Analysis and Minimization Algorithms I: Fundamentals.* Springer-Verlag, Berlin, 1993.

[12] Hiriart-Urruty, J.B., Lemaréchal, C., *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods.* Springer-Verlag, Berlin, 1993.

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

[14] Kiwiel, K. C. *Methods of Descent for Nondifferentiable Optimization. Springer-Verlag, New York* 1985.

[15] Lemarechal, C. and Sagastizábal, C. *Variable metric bundle methods: from conceptual to implementable forms*, Mathematical Programming, vol.76(1997), pp.393-410.

[16] Luenberger, D. G. *Linear and Nonlinear Programming, 2nd. Edition*, Addison-*Wesley* 1984.

[17] Lukšan, L., and Vlček, J. *Test problems for nonsmooth unconstrained and linearly constrained optimization. Tech. Report 798, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague* 2000.

[18] Lukšan, L., and Vlček, J. *Variable metric methods for nonsmooth optimization. Tech. Report 837, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague* 2001.

[19] Mäkelä, M. M., and Neittaanmäki, P. *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control.* World Scientific Publishing Co., Singapore, 1992.

[20] Mitchell, J. E. *Polynomial interior point cutting plane methods.* Optimization Methods and Software 18, (5)(2003), 507-534.

[21] Nesterov, Y. Péton, O. and Vial, J.-P. *Homogeneous analytic center cutting plane methods with approximate centers.* Optimization Methods and Software 11, (12)(1999), 243-273.

[22] Panier, E. R., Tits, A. L., and Herskovits, J. N. *A QP-free, globally convergent, locally super-linearly convergent algorithm for inequality constrained optimization.* SIAM Journal on Control and Optimization 26, (4)(1988), 788-811.

[23] Rockafellar, R.T. *Convex Analysis.* Princeton University Press, Princeton, New Jersey, 1970.

[24] Schramm, H., Zowe, J. *A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, Convergence analysis, Numerical results.* SIAM Journal on Optimization, vol.2, N.1 (1992), pp.121-152.

[25] Tits, A. L., Wächter, A., Bakhtiari, S., Urban, T. J., and Lawrence, C. T. *A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties.* SIAM Journal on Optimization 14, (1)(2003), 173-199.