

## RANDOM PERTURBATION OF THE PROJECTED VARIABLE METRIC METHOD FOR NONSMOOTH NONCONVEX OPTIMIZATION PROBLEMS WITH LINEAR CONSTRAINTS

ABDELKRIM EL MOUATASIM <sup>\*,\*\*</sup>, RACHID ELLAIA <sup>\*\*</sup>, EDUARDO SOUZA DE CURSI <sup>\*\*\*</sup>

<sup>\*</sup> Department of Mathematics, Faculty of Science  
Jazan University, P.B. 2097, Jazan, Saudi Arabia  
e-mail: aelmouatasim@jazanu.edu.sa

<sup>\*\*</sup> Laboratory of Study and Research in Applied Mathematics, Mohammadia School of Engineers  
Mohammed V Agdal University, Ab Ibn sina, BP 765, Agdal, Rabat, Morocco  
e-mail: ellaia@emi.ac.ma

<sup>\*\*\*</sup> National Institute for Applied Sciences, Rouen  
Avenue de l'Université BP 8, Saint-Etienne du Rouvray, France  
e-mail: souza@insa-rouen.fr

We present a random perturbation of the projected variable metric method for solving linearly constrained nonsmooth (i.e., nondifferentiable) nonconvex optimization problems, and we establish the convergence to a global minimum for a locally Lipschitz continuous objective function which may be nondifferentiable on a countable set of points. Numerical results show the effectiveness of the proposed approach.

**Keywords:** global optimization, linear constraints, variable metric method, stochastic perturbation, nonsmooth optimization.

### 1. Introduction

Continuous nonconvex nonsmooth optimization problems involving linear restrictions arise in practical situations stemming from many fields such as optimal control (Kryazhinskii, 2001; Malanowski, 2004; Makela and Neittaanmaki, 1992), integer nonlinear programming problems (Kowalczuk, 2006; Zhang, 2009), minimax estimation (El Mouatasim and Al-Hossain, 2009; Petersen, 2006), and the clustering problem (Bagirov and Yearwood, 2006). A typical situation is the determination of a column vector  $\mathbf{x}^* \in E = \mathbb{R}^n$  such that

$$\mathbf{x}^* = \arg \min_S f, \quad S = \left\{ \mathbf{x} \in E \mid \mathbf{A}\mathbf{x} \leq \mathbf{b} \right\}, \quad (1)$$

where the function  $f : E \rightarrow \mathbb{R}$  does not satisfy convexity assumptions and may be nondifferentiable on a finite or countable subset of  $E$ —this is the case when, for instance,  $f$  is not assumed to be convex differentiable but only locally Lipschitz continuous.  $\mathbf{A}$  is an  $m \times n$  matrix,  $\mathbf{b}$  is an  $m \times 1$  matrix and  $S$  is assumed to be bounded:

there are two vectors  $\ell \in E$  and  $\mathbf{u} \in E$  such that

$$S \subset [\ell, \mathbf{u}] = \left\{ \mathbf{x} \in E \mid \ell \leq \mathbf{x} \leq \mathbf{u} \right\}. \quad (2)$$

The numerical solution to the model problem (1) is usually sought with descent methods, which start at an initial guess  $\mathbf{x}_0$  and generate a sequence of points  $\{\mathbf{x}_k\}_{k \geq 0} \subset E$ : at each iteration number  $k \geq 0$ , both a descent direction  $\mathbf{d}_k \in E$  and a step  $\omega_k \in \mathbb{R}$  ( $\omega_k \geq 0$ ) are determined in order to define

$$\mathbf{x}_0 \in S \text{ given, } \forall k \geq 0: \mathbf{x}_{k+1} = \mathbf{x}_k + \omega_k \mathbf{d}_k. \quad (3)$$

The descent direction is often determined by using the information furnished by the previous points  $\mathbf{x}_k, \mathbf{x}_{k-1}, \dots, \mathbf{x}_0$ . For instance, the classical steepest descent uses the information provided by the gradient  $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$  of the objective function at the point  $\mathbf{x}_k$  and sets  $\mathbf{d}_k = -\mathbf{g}_k$ . In variable metric methods, the determination of the descent direction usually involves the information provided by  $\mathbf{x}_k$  and  $\mathbf{x}_{k-1}$ . For instance, the Davidon–Fletcher–Powell approach (Davidon, 1991) uses  $\mathbf{d}_k = -\mathbf{B}_k \mathbf{g}_k$ ,

where  $\{\mathbf{B}_k\}_{k \geq 0}$  is a sequence of  $n \times n$  matrices such that  $\mathbf{B}_0 = \mathbf{I}$  (the  $n \times n$  identity matrix) and

$$\mathbf{B}_k = \mathbf{B}_{k-1} + \frac{(\mathbf{x}_k - \mathbf{x}_{k-1})(\mathbf{x}_k - \mathbf{x}_{k-1})^t}{(\mathbf{x}_k - \mathbf{x}_{k-1})^t(\mathbf{g}_k - \mathbf{g}_{k-1})} - \frac{\mathbf{B}_{k-1}(\mathbf{g}_k - \mathbf{g}_{k-1})(\mathbf{g}_k - \mathbf{g}_{k-1})^t \mathbf{B}_{k-1}}{(\mathbf{g}_k - \mathbf{g}_{k-1})^t \mathbf{B}_{k-1}(\mathbf{g}_k - \mathbf{g}_{k-1})}.$$

In the sequel, we consider descent vectors corresponding to a general variable metric method given by a function  $\mathbf{u}_k : E \times E \rightarrow E$ :

$$\mathbf{d}_k = \mathbf{u}_k(\mathbf{x}_k, \mathbf{x}_{k-1}). \tag{4}$$

The determination of the step  $\omega_k \geq 0$  often involves a one dimensional search and a previously established maximal step  $\bar{\omega}$ . For instance, the optimal step is

$$\omega_k = \arg \min_W f(\mathbf{x}_k + \omega \mathbf{d}_k),$$

$$W = \left\{ \omega \mid \mathbf{x}_k + \omega \mathbf{d}_k \in S, 0 \leq \omega \leq \bar{\omega} \right\}.$$

Consequently, the step is given by a function  $\omega : E \times E \rightarrow E$  such that

$$\omega_k = \omega(\mathbf{x}_k, \mathbf{d}_k), \quad 0 \leq \omega(\mathbf{x}_k, \mathbf{d}_k) \leq \bar{\omega}. \tag{5}$$

When solving the general problem stated in Eqn. (1), there are three essential difficulties with using the iterations defined in Eqns. (3)–(5). First, the determination of the descent direction  $\mathbf{d}_k$  usually involves the determination of the gradient  $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$  of the objective function at the point  $\mathbf{x}_k$ , which is not defined everywhere, since  $f$  is not anywhere differentiable. Second, the iterations must ensure that  $\{\mathbf{x}_k\}_{k \geq 0} \subset S$ , i.e., that the points generated remain feasible. Third, under the lack of both the convexity and the differentiability of  $f$ , the convergence to a global minimum  $\mathbf{x}^*$  is not ensured.

The first of these difficulties is usually settled in convex optimization by using subgradient information: whenever a subgradient may be defined, it carries information about the growth of the objective function. Variants of the subgradient approach are bundle or level methods. Both these variants try to obtain more information about the behavior of  $f$  by gathering the information provided by the subgradients obtained in the preceding iterations. This information is contained in the set of affine functions associated with these subgradients and the bundle which furnishes a local affine approximation of  $f$ . In convex situations, the descent direction can be determined by using the single information furnished by the bundle, which leads to cutting-plane methods (Kelley, 1960), or by solving a quadratic direction finding problem (Makela and Neittaanmaki, 1992). The convergence of subgradient or bundle methods may be established for convex situations

(Hiriart-Urruty and Lemaréchal, 1993). In the case of bundle methods with a limited number of stored subgradients, the convergence can be guaranteed by using a subgradient aggregation strategy (Kiwiel, 1985), which accumulates information from the previous iterations (Lemaréchal et al., 1981; Schramm and Zowe, 1992). For a nonconvex  $f$ , subgradients are not in general anywhere defined. Alternative methods of construction of a local affine approximation of  $f$  must be supplied in order to get the adequate information about the local growth of  $f$ . For instance, we may introduce other generalized gradient definitions, such as Clarke’s generalized gradients, or simply the gradient of an affine lower estimate. The standard gradient or an  $\varepsilon$ -subgradient may be used, whenever one of these quantities is defined (see Section 2).

The second difficulty is usually settled by projection, whenever an operator of projection onto  $S$  is available. This is just the case of the problem (1). There are usually two possibilities for the introduction of the projection operator according to its use in order to determine feasible points or feasible directions. For instance, one approach consists in introducing a projection operator  $\mathbf{proj}_S : E \rightarrow S$  and determining the descent direction and the step as follows:

$$\begin{aligned} \mathbf{t}_{k+1} &= \mathbf{x}_k + \eta_k \mathbf{v}_k, \\ \mathbf{d}_k &= \mathbf{proj}_S(\mathbf{t}_{k+1}) - \mathbf{x}_k, \\ \omega_k &= 1, \end{aligned}$$

where  $\mathbf{v}_k$  and  $\eta_k$  are a descent direction and a step, respectively. Both  $\mathbf{v}_k$  and  $\eta_k$  are generated by a standard method which does not take the restrictions, i.e.,  $S$ , into account (such as, for instance, the standard gradient descent method). The point  $\mathbf{t}_{k+1}$  is called a trial point and we have  $\mathbf{x}_{k+1} = \mathbf{proj}_S(\mathbf{t}_{k+1})$ . In this approach, the projection operator is used to get a feasible point  $\mathbf{x}_{k+1}$  from the eventually infeasible trial point  $\mathbf{t}_{k+1}$ . For instance, this is the case of bundle or level methods involving proximal projection.

Another approach consists in using

$$\mathbf{d}_k = \mathbf{proj}_{T(S, \mathbf{x}_k)}(\mathbf{v}_k),$$

$$0 \leq \omega_k \leq \omega_{\max} = \max \left\{ \omega \mid \mathbf{A}(\mathbf{x}_k + \omega \mathbf{d}_k) \leq \mathbf{b} \right\},$$

where  $T(S, \mathbf{x}_k)$  is the tangent cone to  $S$  at  $\mathbf{x}_k$ ,  $\mathbf{proj}_{T(S, \mathbf{x}_k)} : E \rightarrow T(S, \mathbf{x}_k)$  is the orthogonal projection onto  $T(S, \mathbf{x}_k)$ ,  $\mathbf{v}_k$  is generated by a standard method which does not take the restrictions into account. In this method, the descent direction  $\mathbf{d}_k$  is projected to get a descent direction containing feasible points. This is the case of the popular projected subgradient method (Correa and Lemaréchal, 1993; Kiwiel, 1985; Larsson et al., 1996), which is used in this work. For linearly constrained problems, an interesting variant is offered by  $\varepsilon$ -active set

methods, which have the reputation of avoiding zigzag (Panier, 1987), and generalized pattern search methods (Bogani *et al.*, 2009).

The third difficulty yields that, as previously observed, a sophisticated approach may become necessary in order to get information about the local growth of the objective function. Moreover, the convergence of the sequence  $\{\mathbf{x}_k\}_{k \geq 0}$  to a point of global minimum  $\mathbf{x}^*$  is not ensured under the lack of convexity: we introduce a controlled random search based on stochastic perturbations of the descent method (3) (Dorea, 1990; El Mouatasim *et al.*, 2006; Pogu and Souza de Cursi, 1994; Souza de Cursi *et al.*, 2003). In this approach,  $\{\mathbf{x}_k\}_{k \geq 0}$ ,  $\{\mathbf{d}_k\}_{k \geq 0}$ ,  $\{\omega_k\}_{k \geq 0}$  become random vectors  $\{\mathbf{X}_k\}_{k \geq 0}$ ,  $\{\mathbf{D}_k\}_{k \geq 0}$ ,  $\{\Omega_k\}_{k \geq 0}$  and the descent iterations are modified as follows:

$$\mathbf{X}_0 = \mathbf{x}_0 \in S \text{ given,} \quad (6)$$

$$\forall k \geq 0: \mathbf{X}_{k+1} = \mathbf{X}_k + \Omega_k \mathbf{D}_k + \mathbf{P}_k, \quad (7)$$

$$\mathbf{D}_k = \mathbf{u}_k(\mathbf{X}_k, \mathbf{X}_{k-1}), \quad (8)$$

$$\Omega_k = \omega(\mathbf{X}_k, \mathbf{D}_k), \quad 0 \leq \omega(\mathbf{X}_k, \mathbf{D}_k) \leq \bar{\omega}, \quad (9)$$

where  $\mathbf{P}_k$  is a suitable random vector the stochastic perturbation. A convenient choice of  $\{\mathbf{P}_k\}_{k \geq 0}$  ensures the convergence of this sequence to  $\mathbf{x}^*$  (see Section 4).

In the sequel, we consider the projected variable metric method applied to the problem (1). After introducing the notation (Section 2), the method is introduced in Section 3. In Section 4, we introduce the stochastic perturbations and we establish the convergence results. The results of numerical experiments are given in Section 5.

## 2. Notation and assumptions

As previously introduced,  $E = \mathbb{R}^n$  is the standard  $n$ -dimensional Euclidean space formed by  $n$ -tuples of real numbers. The elements of  $E$  are denoted using bold lowercase: for instance,  $\mathbf{x} = (x_1, \dots, x_n)^t$ , where the symbol  $t$  denotes the transpose. The usual inner product in  $E$  is denoted by  $(\cdot, \cdot)$ , and the associated Euclidean norm is denoted by  $\|\cdot\|$ :

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t \mathbf{y} = \sum_{i=1}^n x_i y_i,$$

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})} = \sqrt{\mathbf{x}^t \mathbf{x}}.$$

We denote by  $\|\cdot\|$  the matrix norm induced by this norm: if  $\mathbf{C} = (\mathbf{C}_{ij})$ ,  $(1 \leq i \leq m, 1 \leq j \leq n)$  is a  $m \times n$  matrix ( $0 < m < n$ ) formed by real numbers, we have  $\|\mathbf{C}\mathbf{x}\| \leq \|\mathbf{C}\| \|\mathbf{x}\|$  and

$$\|\mathbf{C}\| = \sup \left\{ \|\mathbf{C}\mathbf{x}\| : \|\mathbf{x}\| = 1 \right\}.$$

Let us introduce vectors  $\mathbf{b} = (b_1, b_2, \dots, b_m)^t \in \mathbb{R}^m$ ,  $\boldsymbol{\ell} = (\ell_1, \ell_2, \dots, \ell_n)^t \in E$ ,  $\mathbf{u} = (u_1, u_2, \dots, u_n)^t \in E$  and a real  $m \times n$  matrix  $\mathbf{A} = (A_{ij})$  ( $1 \leq i \leq m$ ,  $1 \leq j \leq n$ ). We have  $\mathbf{A} \equiv [\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_m]^t$ ,

$$\mathbf{A}_i = (A_{i1}, A_{i2}, \dots, A_{in})^t \in E, \quad i = 1, \dots, m.$$

No loss of generality is implied if we assume that

$$\|\mathbf{A}_i\| = 1, \quad i = 1, \dots, m. \quad (10)$$

The feasible set is  $S = \{\mathbf{x} \in E \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$ , i.e.,

$$S = \left\{ \mathbf{x} \in E \mid \sum_{j=1}^n A_{ij} x_j - b_i \leq 0, \quad i = 1, 2, \dots, m \right\}. \quad (11)$$

We assume that

$$S \subset [\boldsymbol{\ell}, \mathbf{u}] = \left\{ \mathbf{x} \in E \mid \ell_i \leq x_i \leq u_i, \quad 1 \leq i \leq n \right\}. \quad (12)$$

Hence  $S$  is a bounded closed convex subset of  $E$ . For any  $\mathbf{x}_1, \mathbf{x}_2 \in S$ , and every  $\theta \in (0, 1)$  we have

$$\begin{aligned} \mathbf{A}(\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2) &= \theta \mathbf{A}\mathbf{x}_1 + (1 - \theta) \mathbf{A}\mathbf{x}_2 \\ &\leq \theta \mathbf{b} + (1 - \theta) \mathbf{b} = \mathbf{b}. \end{aligned}$$

On the other hand,

$$\begin{aligned} \|\mathbf{x}_1 - \mathbf{x}_2\| &\leq L_{12} = \|\boldsymbol{\ell} - \mathbf{u}\|, \\ \|\mathbf{x}_1\| &\leq L = \max\{\|\boldsymbol{\ell}\|, \|\mathbf{u}\|\}. \end{aligned} \quad (13)$$

We recall that the *tangent cone* to  $S$  at a point  $\mathbf{x}$  is the set  $T(S, \mathbf{x}) \subset E$  defined by

$$\mathbf{d} \in T(S, \mathbf{x}) \iff \exists \{(\mathbf{h}_n, \lambda_n)\}_{n > 0} \subset E \times \mathbb{R}_+^*,$$

$$\lambda_n \rightarrow 0, \quad \mathbf{h}_n \rightarrow \mathbf{d}, \quad \mathbf{x} + \lambda_n \mathbf{h}_n \in S.$$

This property is exploited in the sequel.

Practical determination of  $T(S, \mathbf{x})$  is performed by using *active constraints*. Let  $\mathbf{x} \in S$ . The  $i$ -th constraint is *active* at  $\mathbf{x}$  if and only if  $\mathbf{A}_i^t \mathbf{x} - b_i = 0$ . The set of active constraints  $I_{ac}(\mathbf{x})$  and the number of active constraints  $m_{ac}(\mathbf{x})$  at  $\mathbf{x}$  are, respectively,

$$I_{ac}(\mathbf{x}) = \left\{ i : 1 \leq i \leq m, \mathbf{A}_i^t \mathbf{x} - b_i = 0 \right\},$$

$$m_{ac}(\mathbf{x}) = \text{card}(I_{ac}(\mathbf{x})).$$

We set  $\mathbf{A}_N(\mathbf{x}) = [\mathbf{A}_i : i \in I_{ac}(\mathbf{x})]^t$ .  $\mathbf{A}_N(\mathbf{x})$  is the  $m_{ac}(\mathbf{x}) \times n$  submatrix of  $\mathbf{A}$  formed by the lines corresponding to the active constraints at the point  $\mathbf{x}$ . In the particular situation where  $I_{ac}(\mathbf{x}) = \emptyset$ , we have  $m_{ac}(\mathbf{x}) = 0$ ,  $T(S, \mathbf{x}) = E$  and we take  $\mathbf{A}_N(\mathbf{x}) = \mathbf{0} = (0, \dots, 0)$ . In the sequel, we shall use the following properties of  $T(S, \mathbf{x})$ .

**Proposition 1.** Let  $\mathbf{x} \in S$ . We have

$$\begin{aligned} \forall \mathbf{x} \in S : T(S, \mathbf{x}) &= \left\{ \mathbf{d} \in E \mid \mathbf{A}_i^t \mathbf{d} \leq 0, i \in I_{ac}(\mathbf{x}) \right\} \\ &= \left\{ \mathbf{d} \in E \mid \mathbf{A}_N(\mathbf{x}) \mathbf{d} \leq \mathbf{0} \right\}. \end{aligned}$$

Moreover

$$\forall \mathbf{x} \in S : S \subset \{ \mathbf{x} \} + T(S, \mathbf{x}),$$

and the orthogonal projection from  $E$  onto  $T(S, \mathbf{x})$ ,  $\mathbf{proj}(\mathbf{x}, \cdot) : E \rightarrow T(S, \mathbf{x})$  satisfies

$$\forall \mathbf{x} \in S : S \subset \{ \mathbf{x} \} + \text{Im}(\mathbf{proj}(\mathbf{x}, \cdot)), \quad (14)$$

$$\forall \mathbf{x} \in S : \|\mathbf{proj}(\mathbf{x}, \mathbf{w})\| \leq \|\mathbf{w}\|, \quad \forall \mathbf{w} \in E. \quad (15)$$

*Proof.* The result is immediate for  $m_{ac}(\mathbf{x}) = 0$  ( $I_{ac}(\mathbf{x}) = \emptyset$ ), since  $T(S, \mathbf{x}) = E$ ,  $\mathbf{proj}_{T(S, \mathbf{x})}(\mathbf{x}, \mathbf{w}) = \mathbf{w}$ .

Assume that  $m_{ac}(\mathbf{x}) > 0$  and let  $\mathbf{b}_N(\mathbf{x}) = [b_i : i \in I_{ac}(\mathbf{x})]^t$ . Analogously to  $\mathbf{A}_N(\mathbf{x})$ ,  $\mathbf{b}_N(\mathbf{x})$  is formed by the lines of  $\mathbf{b}$  corresponding to the indexes in  $I_{ac}(\mathbf{x})$ .

We denote by  $I_{ac}^c(\mathbf{x})$  the complement of  $I_{ac}(\mathbf{x})$ ,

$$I_{ac}^c(\mathbf{x}) = \left\{ i : 1 \leq i \leq m, \mathbf{A}_i^t \mathbf{x} - b_i < 0 \right\}.$$

Let

$$\eta(\mathbf{x}) = \min \left\{ b_i - \mathbf{A}_i^t \mathbf{x} : i \in I_{ac}^c(\mathbf{x}) \right\} > 0.$$

We assume that  $\mathbf{d} \in T(S, \mathbf{x})$  and wish to show that  $\mathbf{A}_N(\mathbf{x}) \mathbf{d} \leq \mathbf{0}$ . For any sequence  $\{(\mathbf{h}_n, \lambda_n)\}_{n>0} \subset E \times \mathbb{R}_+^*$  such that  $\lambda_n \rightarrow 0$  and  $\mathbf{h}_n \rightarrow \mathbf{d}$ , we have  $\lambda_n \mathbf{A} \mathbf{h}_n \rightarrow \mathbf{0}$ . Thus, there exists an index  $n_0$  such that

$$\begin{aligned} n \geq n_0 &\implies \|\lambda_n \mathbf{A} \mathbf{h}_n\| < \eta(\mathbf{x}) \\ &\implies \mathbf{A}_i^t (\mathbf{x} + \lambda_n \mathbf{h}_n) - b_i \leq 0, \quad \forall i \in I_{ac}^c(\mathbf{x}). \end{aligned}$$

In addition,

$$\mathbf{x} + \lambda_n \mathbf{h}_n \in S \implies$$

$$\mathbf{A}_N(\mathbf{x}) \mathbf{h}_n = \frac{\mathbf{A}_N(\mathbf{x}) (\mathbf{x} + \lambda_n \mathbf{h}_n) - \mathbf{b}_N}{\lambda_n} \leq \mathbf{0}.$$

Passing to the limit in this inequality, we obtain the claim  $\mathbf{A}_N(\mathbf{x}) \mathbf{d} \leq \mathbf{0}$ .

Now we assume that  $\mathbf{A}_N(\mathbf{x}) \mathbf{d} \leq \mathbf{0}$  and wish to show that  $\mathbf{d} \in T(S, \mathbf{x})$ . Let  $\lambda_n = 1/n$  and  $\mathbf{h}_n = \mathbf{d}$ . We have

$$n \geq \frac{\|\mathbf{A} \mathbf{d}\|}{\eta(\mathbf{x})} \implies$$

$$\mathbf{A}_i^t (\mathbf{x} + \lambda_n \mathbf{h}_n) - b_i \leq 0, \quad \forall i \in I_{ac}^c(\mathbf{x}).$$

In addition,

$$\mathbf{A}_N(\mathbf{x}) (\mathbf{x} + \lambda_n \mathbf{h}_n) - \mathbf{b}_N = \mathbf{A}_N(\mathbf{x}) \mathbf{d} \leq \mathbf{0}.$$

Thus,  $\mathbf{x} + \lambda_n \mathbf{h}_n \in S$  and we obtain the claim

$$\mathbf{d} \in T(S, \mathbf{x}).$$

In this way, the first assertion of the proposition is established. For the second one, let  $\mathbf{y} \in S$  and  $\mathbf{d} = \mathbf{y} - \mathbf{x}$ . Then  $\lambda_n = 1/n > 0$ ,  $\mathbf{h}_n = \mathbf{d}$  and  $\mathbf{x} + \lambda_n \mathbf{h}_n = (1 - 1/n)\mathbf{x} + (1/n)\mathbf{y} \in S$ . Thus,  $\mathbf{d} \in T(S, \mathbf{x})$  and we have

$$S - \{ \mathbf{x} \} = \left\{ \mathbf{d} = \mathbf{y} - \mathbf{x} \mid \mathbf{y} \in S \right\} \subset T(S, \mathbf{x}).$$

Hence,

$$S \subset \{ \mathbf{x} \} + T(S, \mathbf{x}).$$

Since  $T(S, \mathbf{x}) = \text{Im}(\mathbf{proj}(\mathbf{x}, \cdot))$ , we have  $S \subset \{ \mathbf{x} \} + \text{Im}(\mathbf{proj}(\mathbf{x}, \cdot))$ . The inequality  $\|\mathbf{proj}(\mathbf{x}, \mathbf{w})\| \leq \|\mathbf{w}\|$  results from the standard properties of orthogonal projections. ■

For a given element  $\mathbf{v} \in E$ , we have  $\mathbf{proj}(\mathbf{x}, \mathbf{v}) = \mathbf{\Pi}_T(\mathbf{x}, \mathbf{v}) \mathbf{v}$ , where  $\mathbf{\Pi}_T(\mathbf{x}, \mathbf{v})$  is an  $n \times n$  matrix, determined as follows. Let

$$I_+(\mathbf{x}, \mathbf{v}) = \left\{ i \in I_{ac}(\mathbf{x}) : \mathbf{A}_i^t \mathbf{v} > 0 \right\},$$

$$m_+(\mathbf{x}, \mathbf{v}) = \text{card}(I_+(\mathbf{x})).$$

If  $I_+(\mathbf{x}, \mathbf{v}) = \emptyset$ , we set  $\mathbf{\Pi}_T(\mathbf{x}, \mathbf{v}) = \mathbf{Id}$ , the  $n \times n$  identity matrix. If  $I_+(\mathbf{x}, \mathbf{v}) \neq \emptyset$ , we set  $\mathbf{A}_+(\mathbf{x}, \mathbf{v}) = [\mathbf{A}_i : i \in I_+(\mathbf{x}, \mathbf{v})]^t$ . There is no loss of generality in assuming that

$$\text{rank}(\mathbf{A}_+(\mathbf{x}, \mathbf{v})) = m_+(\mathbf{x}, \mathbf{v}). \quad (16)$$

Otherwise, we extract from  $\mathbf{A}_+(\mathbf{x}, \mathbf{v})$  a maximal rank submatrix and the associated lines. Then  $\mathbf{\Pi}_T(\mathbf{x}, \mathbf{v})$  is the matrix associated with the operator  $\mathbf{proj}_{\mathbf{x}}(\mathbf{v}) = \mathbf{v} - \mathbf{\Pi}_+(\mathbf{x}, \mathbf{v}) \mathbf{v}$ , where  $\mathbf{\Pi}_+(\mathbf{x}, \mathbf{v})$  corresponds to the orthogonal projection onto the subspace spanned by the vectors forming  $\mathbf{A}_+$  :

$$\begin{aligned} N_+(\mathbf{x}) &= \text{span} [\mathbf{A}_i^t : i \in I_+(\mathbf{x})] \\ &= \left\{ \mathbf{d} \in E \mid \mathbf{d} = \sum_{i \in I_+(\mathbf{x})} \lambda_i \mathbf{A}_i \right\}. \end{aligned}$$

We have (Luenberger, 1973)

$$\mathbf{\Pi}_T(\mathbf{x}, \mathbf{v}) = \mathbf{Id} - \mathbf{\Pi}_+(\mathbf{x}, \mathbf{v}),$$

$$\mathbf{\Pi}_+(\mathbf{x}, \mathbf{v}) = \mathbf{A}_+^t (\mathbf{A}_+ \mathbf{A}_+^t)^{-1} \mathbf{A}_+.$$

We shall also use the following properties of the step.

**Proposition 2.** Let  $\mathbf{x} \in S$  and  $\mathbf{d} \in T(S, \mathbf{x})$ . The maximal allowable step in the direction  $\mathbf{d}$  at the point  $\mathbf{x} \in S$  is

$$\omega_{\max}(\mathbf{x}, \mathbf{d}) = \max \left\{ \omega \mid \mathbf{A}(\mathbf{x} + \omega \mathbf{d}) \leq \mathbf{b} \right\}.$$

We have

$$\forall \mathbf{x} \in S \text{ and } \mathbf{d} \in T(S, \mathbf{x}) : \omega_{\max}(\mathbf{x}, \mathbf{d}) > 0.$$

Moreover, for any  $\mathbf{x} \in S$  there is an  $\varepsilon > 0$  such that

$$\mathbf{d} \in T(S, \mathbf{y}), \quad \forall \mathbf{y} \in \mathbf{x} + B_\varepsilon$$

and

$$\min_{\mathbf{y} \in \mathbf{x} + B_\varepsilon} \omega_{\max}(\mathbf{y}, \mathbf{d}) > 0,$$

and for any  $\mathbf{d} \in T(S, \mathbf{x})$  there is an  $\varepsilon > 0$  such that

$$(\mathbf{d} + B_\varepsilon) \cap T(S, \mathbf{x}) \neq \emptyset$$

and

$$\min_{\mathbf{t} \in (\mathbf{d} + B_\varepsilon) \cap T(S, \mathbf{x})} \omega_{\max}(\mathbf{x}, \mathbf{t}) > 0,$$

where  $B_\varepsilon = \{\mathbf{u} \in E \mid \|\mathbf{u}\| \leq \varepsilon\}$  is the ball with center  $\mathbf{0}$  and radius  $\varepsilon$ .

*Proof.* We have

$$\omega_{\max}(\mathbf{x}, \mathbf{d}) = \min_{1 \leq i \leq m} \{\omega_i\},$$

$$\omega_i = \begin{cases} b_i - \frac{\sum_{j=1}^n A_{ij}x_j}{\sum_{j=1}^n A_{ij}d_j} & \text{if } \sum_{j=1}^n A_{ij}d_j > 0, \\ +\infty & \text{otherwise} \end{cases}$$

and

$$b_i - \sum_{j=1}^n A_{ij}x_j > 0.$$

Thus,  $\omega_{\max}(\mathbf{x}, \mathbf{d}) > 0$ .

Let  $\mathbf{d} \in T(S, \mathbf{x})$ . Assume that for each  $n > 0$  there exists  $\mathbf{y}_n$  such that  $\|\mathbf{y}_n\| \leq 1/n$  and

$$\mathbf{d} \notin T(S, \mathbf{x}_n), \quad \mathbf{x}_n = \mathbf{x} + \mathbf{y}_n.$$

Thus, there exists  $i(n)$  such that

$$\mathbf{A}_{i(n)}^t \mathbf{x}_n - b_{i(n)} = 0, \quad \mathbf{A}_{i(n)}^t \mathbf{d} > \mathbf{0}.$$

Let

$$r(k) = \max \{i(n) : n \geq k\},$$

$$n(k) = \min \{n : n \geq k \text{ and } i(n) = r(k)\}.$$

By construction,  $\{r(k)\}_{k>0} \subset \{1, \dots, m\}$  is decreasing and bounded below. Thus,  $r(k) \rightarrow r$  for  $k \rightarrow \infty$ . Since  $\{1, \dots, m\}$  is discrete, there is a  $k_0$  such that  $k \geq k_0 \implies r(k) = r$ . We have

$$k \geq k_0 \implies \mathbf{A}_r^t \mathbf{x}_{n(k)} - b_r = 0 \text{ and } \mathbf{A}_r^t \mathbf{d} > \mathbf{0}.$$

Passing to the limit as  $k \rightarrow \infty$ , we have, since  $\mathbf{y}_n \rightarrow \mathbf{0}$ ,

$$\mathbf{A}_r^t \mathbf{x} - b_r = 0 \text{ and } \mathbf{A}_r^t \mathbf{d} > \mathbf{0}.$$

Thus,  $\mathbf{d} \notin T(S, \mathbf{x})$  and we have a contradiction. Hence, there is an  $n > 0$  such that

$$\|\mathbf{y}\| \leq \frac{1}{n} \implies \mathbf{d} \in T(S, \mathbf{x} + \mathbf{y}).$$

Let  $\mathbf{d} \in T(S, \mathbf{x})$ . Assume that

$$\forall \varepsilon > 0 : \min_{\mathbf{y} \in \mathbf{x} + B_\varepsilon} \omega_{\max}(\mathbf{x}, \mathbf{d}) = 0.$$

Then for any  $n > 0$  there is a  $\mathbf{y}_n$  such that

$$\|\mathbf{y}_n\| \leq \frac{1}{n}, \quad \omega_{\max}(\mathbf{x}_n, \mathbf{d}) \leq \frac{1}{n}, \quad \mathbf{x}_n = \mathbf{x} + \mathbf{y}_n.$$

Thus, there exists  $i(n)$  such that

$$\frac{b_{i(n)} - \mathbf{A}_{i(n)}^t \mathbf{x}_n}{\mathbf{A}_{i(n)}^t \mathbf{d}} \leq \frac{1}{n},$$

$$b_{i(n)} - \mathbf{A}_{i(n)}^t \mathbf{x}_n > 0, \quad \mathbf{A}_{i(n)}^t \mathbf{d} > \mathbf{0}.$$

Let

$$r(k) = \max \{i(n) : n \geq k\},$$

$$n(k) = \min \{n : n \geq k \text{ and } i(n) = r(k)\}.$$

Analogously to the above argument, there exists  $k_0$  such that  $k \geq k_0 \implies r(k) = r$ , and we have

$$k \geq k_0 \implies b_r - \mathbf{A}_r^t \mathbf{x}_{n(k)} \leq \frac{1}{n} \mathbf{A}_r^t \mathbf{d},$$

$$b_r - \mathbf{A}_r^t \mathbf{x}_{n(k)} > 0, \quad \mathbf{A}_r^t \mathbf{d} > \mathbf{0}.$$

By taking the limit for  $k \rightarrow \infty$ , we have, since  $\mathbf{y}_n \rightarrow \mathbf{0}$ ,

$$b_r - \mathbf{A}_r^t \mathbf{x} \leq 0, \quad b_r - \mathbf{A}_r^t \mathbf{x} > 0, \quad \mathbf{A}_r^t \mathbf{d} > \mathbf{0}.$$

and we obtain a contradiction.

Let  $\mathbf{d} \in T(S, \mathbf{x})$ . We have  $(\mathbf{d} + B_\varepsilon) \cap T(S, \mathbf{x}) \neq \emptyset$  for any  $\varepsilon > 0$ , since  $\mathbf{d} \in (\mathbf{d} + B_\varepsilon) \cap T(S, \mathbf{x})$ . Assume that

$$\forall \varepsilon > 0 : \min_{\mathbf{t} \in (\mathbf{d} + B_\varepsilon) \cap T(S, \mathbf{x})} \omega_{\max}(\mathbf{x}, \mathbf{t}) = 0.$$

Then for any  $n > 0$  there is a  $\mathbf{t}_n$  such that

$$\|\mathbf{t}_n\| \leq \frac{1}{n}, \quad \omega_{\max}(\mathbf{x}, \mathbf{d}_n) \leq \frac{1}{n}, \quad \mathbf{d}_n = \mathbf{d} + \mathbf{t}_n.$$

Thus, there exists  $i(n)$  such that

$$\frac{b_{i(n)} - \mathbf{A}_{i(n)}^t \mathbf{x}}{\mathbf{A}_{i(n)}^t \mathbf{d}_n} \leq \frac{1}{n},$$

$$b_{i(n)} - \mathbf{A}_{i(n)}^t \mathbf{x} > 0, \quad \mathbf{A}_{i(n)}^t \mathbf{d}_n > \mathbf{0}.$$

Let

$$r(k) = \max \{i(n) : n \geq k\},$$



$$n(k) = \min \{ n : n \geq k \text{ and } i(n) = r(k) \}.$$

Analogously to the demonstration above, there exists  $k_0$  such that  $k \geq k_0 \implies r(k) = r$ , and we have

$$k \geq k_0 \implies b_r - \mathbf{A}_r^t \mathbf{x} \leq \frac{1}{n} \mathbf{A}_r^t \mathbf{d}_{n(k)},$$

$$b_r - \mathbf{A}_r^t \mathbf{x} > 0, \quad \mathbf{A}_r^t \mathbf{d}_{n(k)} > \mathbf{0}.$$

Passing to the limit as  $k \rightarrow \infty$ , we have, since  $\mathbf{t}_n \rightarrow \mathbf{0}$ ,

$$b_r - \mathbf{A}_r^t \mathbf{x} \leq 0, \quad b_r - \mathbf{A}_r^t \mathbf{x} > 0, \quad \mathbf{A}_r^t \mathbf{d} \geq \mathbf{0},$$

and we get a contradiction. ■

As mentioned above, the objective function  $f : E \rightarrow \mathbb{R}$  is assumed to be locally Lipschitz continuous: it may have a countable number of points of nondifferentiability. Moreover,  $f$  is not assumed to be convex. Since  $S$  is closed and bounded, and  $f$  is continuous, there exists  $\theta^* \in \mathbb{R}$  such that

$$\min_S f = \theta^* \in \mathbb{R}. \tag{17}$$

Let  $\theta > \theta^*$ . We denote by  $S_\theta$  the set

$$S_\theta = \{ \mathbf{x} \in S \mid \theta^* \leq f(\mathbf{x}) < \theta \}.$$

In the sequel, we consider

$$\theta_{\max} = \max \{ \theta \mid S_\theta \cap S \neq \emptyset \}.$$

The continuity of  $f$  implies that

$$\begin{aligned} \theta^* < \theta < \theta_{\max} \\ \implies \text{meas}(S_\theta) > 0 \text{ and } \text{meas}(S - S_\theta) > 0. \end{aligned} \tag{18}$$

### 3. Projected variable metric method

The class of *variable metric* methods was originally introduced by Davidon (1991) along with Fletcher and Powell (1963) in an attempt to get information about the curvature of the objective function by using a variable symmetric positive definite  $n \times n$  matrix  $\mathbf{B}_k$  and

$$\mathbf{v}_k = \arg \min \{ \mathbf{g}_k^t \mathbf{v} : \mathbf{v}^t \mathbf{B}_k \mathbf{v} = 1 \}.$$

The properties of  $\mathbf{B}_k$  show that  $\mathbf{v}^t \mathbf{B}_k \mathbf{v}$  is a norm:  $\mathbf{v}_k$  is the element of the generalized circle  $C_k = \{ \mathbf{v} : \mathbf{v}^t \mathbf{B}_k \mathbf{v} = 1 \}$  having the most negative Euclidean projection on the direction of  $\mathbf{g}_k$ . This method is known as the DFP descent method. Several variants may be found in the literature, such as the BFGS descent method (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) and other quasi-Newton methods.

As mentioned above, the determination of the descent direction  $\mathbf{v}_k$  usually involves the gradient  $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$  of the objective function at the point  $\mathbf{x}_k$ , which may be not defined due to the lack of regularity of  $f$  (Peng

and Heying, 2009; Uryasev, 1991). In addition, the objective function is not assumed to be convex and its subdifferential may be empty.

These considerations provide a simple way to extend descent methods based on the gradient to the nonsmooth situation under consideration: if the objective function  $f$  is differentiable at  $\mathbf{x}_k$ , the descent direction  $\mathbf{d}_k$  is determined by using the standard gradient  $\mathbf{g}_k = \nabla f(\mathbf{x}_k)$ . Otherwise, we consider a local affine underestimate or overestimate  $\gamma_k(\mathbf{y}) = (\mathbf{p}_k, \mathbf{y} - \mathbf{x}_k) + f(\mathbf{x}_k)$ , and we use  $\mathbf{g}_k = \mathbf{p}_k$  for the determination of the descent direction (for more, see El Mouatasim et al., 2006). In practice,  $\gamma_k$  may be numerically approximated by using the values of  $f$  or  $\nabla f$  at points close to  $\mathbf{x}_k$ . This approach is particularly suitable for the situation under consideration, since  $f$  is differentiable almost everywhere (i.e., except for a set having zero Lebesgue measure (Makela and Neittaanmaki, 1992)).

### 4. Stochastic perturbation

As previously observed, the lack of convexity yields that the convergence to a global minimum cannot be ensured. In order to solve this difficulty, the original sequence generated by the iterations,  $\{\mathbf{x}_k\}_{k \geq 0}$ , is replaced by a sequence of random variables  $\{\mathbf{X}_k\}_{k \geq 0}$  defined by Eqns. (7)–(9).

In previous works, an analogous strategy has been applied to smooth unconstrained (Pogu and Souza de Cursi, 1994) or smooth constrained situations (El Mouatasim et al., 2006; Souza de Cursi et al., 2003), involving iterations of the form  $\mathbf{X}_{k+1} = \mathbf{Q}_k(\mathbf{X}_k) + \mathbf{P}_k$ , which corresponds to a Markov chain with the memory length equal to one, since only the last result intervenes. In the situation under consideration, the iterate number  $k + 1$  depends on the whole preceding history (see Step 7 of the algorithm). This corresponds to a particular kind of the Markov chain, where the variable is not  $\mathbf{X}_k$  but the whole history  $\mathbf{X}_{\leq k}$ . Thus, the preceding theoretical results do not apply immediately and must be modified in order to match the situation under consideration.

In this section, we establish the convergence results concerning the general iterations given by

$$\forall k \geq 1 : \mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{h}_k(\mathbf{X}_{\leq k}) + \mathbf{P}_k, \tag{19}$$

where  $\mathbf{X}_0 = \mathbf{x}_0 \in S$  and  $\mathbf{X}_1 = \mathbf{x}_1 \in S$  are given. It is assumed that  $\mathbf{h}_k(\cdot)$  is bounded on  $S^{k+1}$ , i.e., there exists a real number  $\Lambda \geq 0$  such that

$$\forall \mathbf{x}_{\leq k} \in S^{k+1} : \|\mathbf{h}_k(\mathbf{x}_{\leq k})\| \leq \Lambda. \tag{20}$$

The algorithm corresponds to

$$\forall k \geq 1 : \mathbf{h}_k(\mathbf{x}_{\leq k}) = \omega_k \mathbf{proj}(\mathbf{x}_k, \mathbf{s}_k(\mathbf{x}_{\leq k})).$$

Equations (5) and (15) show that this definition satisfies the inequality (20). Nevertheless, the mathematical results apply to a larger context: for instance,  $\mathbf{h}_k(\mathbf{x}_{\leq k}) = \mathbf{0}$  also satisfies (20); in this case, the algorithm becomes a purely stochastic search. Analogously, these assumptions take into account situations where  $\mathbf{h}_k(\mathbf{x}_{\leq k})$  is not always a descent direction, but remains bounded. If  $\mathbf{h}_k(\mathbf{x}_{\leq k})$  is not a descent direction, the stochastic perturbation drives the process and yields a descent at each iteration. Here  $\mathbf{h}_k(\mathbf{x}_{\leq k})$  is expected to drive the iterations in the neighbourhood of a minimum, in order to accelerate the convergence compared with a pure random search.

The proof of the results follows the lines of El Mouatasim *et al.* (2006), Pogu and Souza de Cursi (1994) as well as Souza de Cursi *et al.* (2003). It must be noticed that smoothness arguments are not directly involved in the probabilistic results of convergence established in the sequel (but they are involved in the definition of the deterministic term  $\mathbf{h}_k(\mathbf{x}_{\leq k})$ ). The convergence of the iterations is a consequence of the following fundamental theorem.

**Theorem 1.** *Let  $\{\mathbf{X}_k\}_{k \geq 0} \subset S$  be a sequence of random variables defined by Eqn. (19), where  $\mathbf{h}_k(\mathbf{x}_{\leq k})$  satisfies the inequality (20). Assume that  $\mathbf{P}_k$  is the restriction to  $S$  of a random variable  $\mathbf{T}_k$  taking its values on the whole space  $E$ , such that its density  $\phi_k$  satisfies the conditions*

$$\forall k \geq 0 : \phi_k(\mathbf{p}) \geq \psi_k(\|\mathbf{p}\|) > 0,$$

$$\forall M \geq 0 : \sum_{k=0}^{+\infty} \psi_k(M) = +\infty,$$

where  $\psi_k : \mathbb{R} \rightarrow \mathbb{R}$  is a decreasing function.

Let

$$U_k = \min \{f(\mathbf{X}_i) : 1 \leq i \leq k\}.$$

Then there exists  $U \geq \theta^*$  such that  $U_k \rightarrow U$  as  $k \rightarrow +\infty$  and  $U = \theta^*$  almost surely.

A simple way for the generation of perturbations  $\mathbf{P}_k$  satisfying these assumptions consists in considering an  $n$ -sample  $\mathbf{Z}$  from  $N(0, 1)$  (i.e.,  $\mathbf{Z}$  is an  $n$ -dimensional vector, independent of  $\mathbf{X}_k$ , formed by independent variables of the same law  $N(0, 1)$ ) and a decreasing sequence  $\{\xi_k\}_{k \geq 0}$  of strictly positive real numbers converging to zero. We set  $\mathbf{T}_k = \xi_k \mathbf{Z}$ , and  $\mathbf{P}_k$  is the restriction of  $\mathbf{T}_k$  to the values such that  $\mathbf{X}_k + \mathbf{h}_k(\mathbf{X}_{\leq k}) + \mathbf{T}_k \in S$ . We have

$$P(\mathbf{T}_k < \mathbf{p}) = P\left(\mathbf{Z} < \frac{\mathbf{p}}{\xi_k}\right)$$

and

$$\begin{aligned} \phi_k(\mathbf{p}) &= \frac{1}{(\xi_k)^n} \rho_k\left(\frac{\mathbf{p}}{\xi_k}\right) \\ &= \frac{1}{(\xi_k)^n} \psi\left(\frac{\|\mathbf{p}\|}{\xi_k}\right) = \psi_k(\|\mathbf{p}\|), \end{aligned}$$

where  $n = \dim(E)$ . In practice, the generation of the restriction of  $\mathbf{T}_k$  may lead to the rejection of a large number of the points generated. Thus, we shall use

$$\begin{aligned} \mathbf{P}_k &= \omega_k \xi_k \mathbf{Z}_k, \\ \mathbf{Z}_k &= \text{proj}(\mathbf{X}_k, \mathbf{Z}) = \mathbf{\Pi}_k \mathbf{Z}, \quad \mathbf{\Pi}_k = \mathbf{\Pi}_T(\mathbf{X}_k), \end{aligned} \quad (21)$$

where  $\omega_k$  is the step associated with the direction  $\mathbf{d}_k + \xi_k \mathbf{Z}_k$ . Since  $\mathbf{Z}$  is an  $n$ -sample from  $N(0, 1)$  and  $\text{proj}(\mathbf{X}_k, \cdot)$  is an orthogonal projection operator, and the components of  $\mathbf{Z}_k$  in any orthonormal basis form a sample from  $N(0, 1)$  (Bouleau, 1986; Souza de Cursi, 1991). In addition, Proposition 1 shows that  $\mathbf{X}_{k+1}$  spans  $S$ . This approach generates only admissible points.

Theorem 1 is a consequence of the following result.

**Proposition 3.** *Let  $\{U_n\}_{n \geq 0}$  be a decreasing sequence, lower bounded by  $\theta^*$ . Then there exists  $U$  such that*

$$U_n \rightarrow U \quad \text{as } n \rightarrow +\infty.$$

Assume that, in addition, for any  $\theta \in ]\theta^*, \theta_{\max}[$ , there is a sequence of strictly positive real numbers  $\{c_k(\theta)\}_{k \geq 0}$  such that for every  $k \geq 0$  we have

$$\begin{aligned} P(U_{k+1} < \theta \mid U_k \geq \theta) &\geq c_k(\theta) > 0, \\ \sum_{k=0}^{+\infty} c_k(\theta) &= +\infty. \end{aligned}$$

Then  $U = \theta^*$  almost surely.

*Proof.* See, for instance, the results of Pinter (1996) or Pogu and Souza de Cursi (1994). ■

*Proof of Theorem 1.* Let us introduce

$$\begin{aligned} \mathcal{S}_k &= \left\{ \mathbf{z} \in E \mid \exists (\mathbf{x}_{\leq k}) \in S^{k+1} \right. \\ &\quad \left. \text{such that } \mathbf{x}_k + \mathbf{h}_k(\mathbf{x}_{\leq k}) + \mathbf{z} \in S \right\}. \end{aligned}$$

Since  $S$  is bounded and  $|\mathbf{h}_k(\mathbf{x}_{\leq k})| \leq \Lambda$ ,  $\mathcal{S}_k$  is bounded. Thus, there is a real number  $\Gamma > 0$  such that  $\|\mathbf{z}\| \leq \Gamma$ ,  $\forall \mathbf{z} \in \mathcal{S}_k$ . In addition, the assumption (18) shows that  $\text{meas}(\mathcal{S}_k) > 0$ .

Let  $\mathbf{z} \in \mathcal{S}_k$ , and let  $\Phi_k$  denote the cumulative function of  $\mathbf{P}_k$  and  $H_k = \mathbf{h}_k(\mathbf{X}_{\leq k})$ . We have

$$\begin{aligned} P(\mathbf{X}_{k+1} < \mathbf{z} \mid \mathbf{X}_{\leq k} = \mathbf{x}_{\leq k}) &= P(\mathbf{X}_k + H_k + \mathbf{P}_k < \mathbf{z} \mid \mathbf{X}_{\leq k} = \mathbf{x}_{\leq k}) \\ &= P(\mathbf{P}_k < \mathbf{z} - \mathbf{x}_k - \mathbf{h}_k(\mathbf{x}_{\leq k})). \end{aligned}$$

Thus, the conditional cumulative function of  $\mathbf{X}_{k+1}$  is

$$F_{k+1}(\mathbf{z} | \mathbf{X}_{\leq k} = \mathbf{x}_{\leq k}) = \Phi_k(\mathbf{z} - \mathbf{x}_k - \mathbf{h}_k(\mathbf{x}_{\leq k})).$$

and the associated density of probability  $f_{k+1}$  is

$$f_{k+1}(\mathbf{z} | \mathbf{X}_{\leq k} = \mathbf{x}_{\leq k}) = \phi_k(\mathbf{z} - \mathbf{x}_k - \mathbf{h}_k(\mathbf{x}_{\leq k})).$$

Hence, we have

$$f_{k+1}(\mathbf{z} | \mathbf{X}_{\leq k} = \mathbf{x}_{\leq k}) \geq \psi_k(\|\mathbf{z} - \mathbf{x}_k - \mathbf{h}_k(\mathbf{x}_{\leq k})\|).$$

Since

$$\|\mathbf{z} - \mathbf{x}_k - \mathbf{h}_k(\mathbf{x}_{\leq k})\| \leq \|\mathbf{z} - \mathbf{x}_k\| + \|\mathbf{h}_k(\mathbf{x}_{\leq k})\| \leq L_{12} + \Lambda,$$

where  $L_{12} = \|\ell - \mathbf{u}\|$  and  $\psi_k$ , is decreasing,

$$\psi_k(\|\mathbf{z} - \mathbf{x}_k - \mathbf{h}_k(\mathbf{x}, \mathbf{y})\|) \geq \psi_k(L_{12} + \Lambda),$$

and we have

$$f_{k+1}(\mathbf{z} | \mathbf{X}_{\leq k} = \mathbf{x}_{\leq k}) \geq \psi_k(M) > 0, \quad M = L_{12} + \Lambda > 0. \quad (22)$$

Let  $k \geq 2$ . Since  $\{\mathbf{X}_k\}_{k \geq 0} \subset S$ , we have

$$P(\mathbf{X}_{\leq k-1} \in S^k) = 1.$$

Thus,

$$P(\mathbf{X}_k \notin S_\theta) = P(\mathbf{X}_k \notin S_\theta, \mathbf{X}_{\leq k-1} \in S^k).$$

Moreover,

$$P(\mathbf{X}_k \notin S_\theta | \mathbf{X}_{\leq k-1} \in S^k) = \frac{P(\mathbf{X}_k \notin S_\theta, \mathbf{X}_{\leq k-1} \in S^k)}{P(\mathbf{X}_{\leq k-1} \in S^k)} = P(\mathbf{X}_k \notin S_\theta).$$

Hence

$$P(\mathbf{X}_k \notin S_\theta) = \int_{S^k} P(\mathbf{X}_{\leq k-1} \in d\mathbf{x}_{\leq k-1}) \times \int_{S-S_\theta} f_k(\mathbf{z} | \mathbf{X}_{\leq k-1} \in S^k) d\mathbf{z}.$$

Thus, from Eqn. (22),

$$P(\mathbf{X}_k \notin S_\theta) \geq \int_{S^k} P(\mathbf{X}_{\leq k-1} \in d\mathbf{x}_{\leq k-1}) \times \int_{S-S_\theta} \psi_{k-1}(M) d\mathbf{z}$$

and

$$P(\mathbf{X}_k \notin S_\theta) \geq \text{meas}(S - S_\theta) \psi_{k-1}(M) \times \int_{S^k} P(\mathbf{X}_{\leq k-1} \in d\mathbf{x}_{\leq k-1}).$$

We have

$$\int_{S^k} P(\mathbf{X}_{\leq k-1} \in d\mathbf{x}) = P(\mathbf{X}_{\leq k-1} \in S^k) = 1.$$

Thus,

$$P(\mathbf{X}_k \notin S_\theta) \geq \text{meas}(S - S_\theta) \psi_{k-1}(M) > 0. \quad (23)$$

We have

$$P(\mathbf{X}_{k+1} \in S_\theta, \mathbf{X}_k \notin S_\theta) = P(\mathbf{X}_{k+1} \in S_\theta, \mathbf{X}_k \notin S_\theta, \mathbf{X}_{\leq k-1} \in S^k).$$

Thus,

$$P(\mathbf{X}_{k+1} \in S_\theta, \mathbf{X}_k \notin S_\theta) = \int_{(S-S_\theta) \times S^k} P(\mathbf{X}_k \in d\mathbf{x}, \mathbf{X}_{\leq k-1} \in d\mathbf{y}_{\leq k-1}) \times \int_{S_\theta} f_{k+1}(\mathbf{z} | \mathbf{X}_k = \mathbf{x}, \mathbf{X}_{\leq k-1} = \mathbf{y}_{\leq k-1}) d\mathbf{z}.$$

From Eqn. (22),

$$P(\mathbf{X}_{k+1} \in S_\theta, \mathbf{X}_k \notin S_\theta) \geq \text{meas}(S_\theta) \psi_k(M) \times \int_{(S-S_\theta) \times S} P(\mathbf{X}_k \in d\mathbf{x}, \mathbf{X}_{k-1} \in d\mathbf{y}),$$

that is to say,

$$P(\mathbf{X}_{k+1} \in S_\theta, \mathbf{X}_k \notin S_\theta) \geq \text{meas}(S_\theta) \psi_k(L_{12}) P(\mathbf{X}_k \notin S_\theta).$$

Thus, from Eqn. (23),

$$P(\mathbf{X}_{k+1} \in S_\theta | \mathbf{X}_k \notin S_\theta) = \frac{P(\mathbf{X}_{k+1} \in S_\theta, \mathbf{X}_k \notin S_\theta)}{P(\mathbf{X}_k \notin S_\theta)} \geq \text{meas}(S_\theta) \psi_k(M). \quad (24)$$

By construction, the sequence  $\{U_n\}_{n \geq 0}$  is decreasing and bounded from below by  $\theta^*$ . Thus, there exists  $U \geq \theta^*$  such that  $U_k \rightarrow U$  as  $k \rightarrow +\infty$ . Moreover,

$$P(U_{k+1} < \theta | U_k \geq \theta) = P(\mathbf{X}_{k+1} \in S_\theta | \mathbf{X}_k \notin S_\theta) \geq c_k(\theta),$$

where

$$c_k(\theta) = \text{meas}(S_\theta) \psi_k(M).$$

The result follows from Proposition 3. ■

### Random perturbation of the projected variable metric algorithm.

Step 0. Parameter:  $b_{\text{step}} = 0.1$ . Data:  $\mathbf{x}_0 = \mathbf{X}_0 \in S$ .

Step 1. Initialization. Set  $k = 0$ ,  $\mathbf{B}_0 = \mathbf{I}$ .



Step 2. Generalized gradient calculation:  $\mathbf{g} \in \partial f(\mathbf{X}_k)$ .

Step 3. Generalized gradient normalization:  $\mathbf{g}_k = \mathbf{g}/\|\mathbf{g}\|$ .

Step 4. Set  $\mathbf{g}_b = \mathbf{B}_k^t \mathbf{g}_k$ ,  $\mathbf{g}_m = \mathbf{B}_k \mathbf{g}_b$ .

Step 5. Direction calculation:

$$\mathbf{v}_k = \mathbf{s}_k(\mathbf{X}_{\leq k}) = \begin{cases} -\frac{\mathbf{g}_m}{\|\mathbf{g}_b\|} & \text{if } \mathbf{g}_m \neq \mathbf{0}, \\ \mathbf{0} & \text{if } \mathbf{g}_m = \mathbf{0}, \end{cases}$$

where  $\mathbf{X}_{\leq k} = (\mathbf{X}_k, \mathbf{X}_{k-1}, \dots, \mathbf{X}_0)$ .

Step 6. Calculation of the optimal step  $\Omega_k$ .

Step 7. Set

$$\mathbf{X}_{k+1} = \mathbf{X}_k + \Omega_k \text{proj}(\mathbf{X}_k, \mathbf{s}_k(\mathbf{X}_{\leq k})) + \mathbf{P}_k.$$

Step 8. Set  $\mathbf{g} \in \partial f(\mathbf{X}_{k+1})$ .

Step 9. Generalized gradient normalization:

$$\mathbf{g}_{k+1} = \mathbf{g}/\|\mathbf{g}\|.$$

Step 10. Set

$$\mathbf{B}_{k+1} = \mathbf{B}_k + b_{\text{step}}(\mathbf{g}_k \mathbf{g}_{k+1}^t \mathbf{B}_k + \mathbf{g}_{k+1} \mathbf{g}_k^t \mathbf{B}_k).$$

Step 11. Set  $k = k + 1$ .

Step 12. Go to Step 4.

The step  $\Omega_k$  has to be determined by an independent rule. Classical choices are, for instance, the fixed step, Wolfe's rule or the optimal step. In our calculations, we shall use the optimal step approach.

### 5. Numerical experiments

In this section, we describe practical implementation of random perturbations and we present the results of some numerical experiments which illustrate the numerical behavior of the method.

At the iteration number  $k \geq 0$ , we have that  $\mathbf{X}_{\leq k}$  is known and  $\mathbf{X}_{k+1}$  has to be determined. From the numerical standpoint, we consider finite samples of  $\mathbf{P}_k$ . Let  $k_{\text{sto}}$  be a nonnegative integer and  $\mathcal{P}_k = (\mathbf{P}_k^1, \dots, \mathbf{P}_k^{k_{\text{sto}}})$  from  $\mathbf{P}_k$  be a sample formed by  $k_{\text{sto}}$  variates from  $\mathbf{P}_k$ . By setting  $\mathbf{P}_0 = \mathbf{0}$ , Eqn. (19) furnishes  $k_{\text{sto}} + 1$  values from  $\mathbf{X}_{k+1}$ , denoted by  $\mathcal{X}_{k+1} = (\mathbf{X}_{k+1}^0, \mathbf{X}_{k+1}^1, \dots, \mathbf{X}_{k+1}^{k_{\text{sto}}})$ . Then, we estimate  $U_{k+1} \approx \min \{f(\mathbf{X}) : \mathbf{X} \in \{\mathbf{X}_k\} \cup \mathcal{X}_{k+1}\}$  and  $\mathbf{X}_{k+1} = \arg \min \{f(\mathbf{X}) : \mathbf{X} \in \{\mathbf{X}_k\} \cup \mathcal{X}_{k+1}\}$ .

In our experiments, the perturbation is generated according to Eqn. (21). The Gaussian variates are obtained from calls to standard generators. We use

$$\xi_k = \sqrt{\frac{a}{\log(k+2)}},$$

where  $a > 0$ .

According to Section 3, the descent direction is generated by using generalized gradients of the objective function. If the objective function is differentiable at  $\mathbf{X}_K$ , the gradient is used. Otherwise, we consider local affine underestimate or overestimate and the descent direction is random convex combination of these elements. For instance, if subgradients are available at a nondifferentiability point, then the descent direction is a random convex combination of elements of the subdifferential.

We introduce a maximum iteration number  $k_{\text{max}}$ : the iterations are stopped when  $k = k_{\text{max}}$ . We denote by  $f_{\text{opt}}$  and  $\mathbf{x}_{\text{opt}}$ , the estimations of the optimal value  $f$  and  $\mathbf{x}^*$  furnished by the method.  $f_{\text{mean}}$  and  $\mathbf{x}_{\text{mean}}$  are their mean values estimated from 100 independent runs. We denote by  $Vf_{\text{mean}}$  and  $\sigma f_{\text{mean}}$  the variance and standard deviation of  $f_{\text{mean}}$ , which are estimated from the results of the runs.

Our approach was programmed using Visual Fortran 6.1. As far as the experiments were concerned, they were performed on a workstation running an HP Intel (R) M processor (1.30 GHz, 224 MB RAM). The case  $k_{\text{sto}} = 0$  corresponds to unperturbed descent (deterministic) method.

### Results.

- Case 1:  $\bar{\omega} = 500$ ,  $k_{\text{max}} = 100$ ,  $k_{\text{sto}} = 500$  and  $a = 1$ .
- Case 2:  $\bar{\omega} = 500$ ,  $k_{\text{max}} = 500$ ,  $k_{\text{sto}} = 500$  and  $a = 1$ .

In Tables 1 and 2, we show the observed effect of the variation in a single parameter value while the others remain with their original value. Tables 1 and 2 contain the minimal value objective  $f(\mathbf{x}^*) = -0.38966$ ,  $-0.33036$ ,  $-1.8596$ ,  $128$  and  $-1964$  of the problems Mad 1, Mad 2, Pentagon, Wong 3 and HS114.

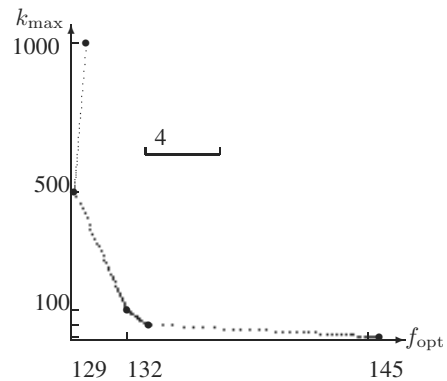


Fig. 1. Evaluation of the objective function with the iteration number for the Wong 3 problem.

Table 3. Results of 100 tests.

Example	1	2	3	4	5
$f_{\text{mean}}$	-0.38959	-0.330357	-1.85957	132	-1931
$vf_{\text{mean}}$	$5.55 \times 10^{-17}$	$4.02 \times 10^{-16}$	$2.96 \times 10^{-8}$	$9.09 \times 10^{-11}$	$2.33 \times 10^{-9}$
$\sigma f_{\text{mean}}$	$7.45 \times 10^{-9}$	$2.01 \times 10^{-8}$	$1.72 \times 10^{-4}$	$9.54 \times 10^{-6}$	$4.83 \times 10^{-9}$
$\mathbf{x}_{\text{mean}}$	-0.4002	-0.8929	-0.9776	-0.1916	2.15 2.30 7.87 5.02
	0.9002	0.1786	0.3799	-0.7785	1.17 1.87 1.34 9.87
			0.3519	0.9146	8.39 8.67 2.63 1.09
					6.09 14.21 0.96 0.62
					1.27 2.06 1.33 2.03
				1733.36 11989.23	
				39.49 3047.92	
				1985.10 89.96	
				95.00 8.39	
				1.20 154.44	

Table 1. Results of optimal values in Case 1.

Example	1	2	3	4	5	
$\bar{\omega}$	1	-0.38966	-0.33036	-1.8596	132	-1560
	10	-0.38966	-0.33036	-1.8596	130	-1930
	100	-0.38966	-0.33036	-1.8596	132	-1931
	500	-0.38959	-0.33036	-1.8596	132	-1931
	1000	-0.38954	-0.33036	-1.8596	132	-1931
$k_{\text{max}}$	10	-0.38959	-0.33036	-1.8590	146	-1348
	50	-0.38959	-0.33036	-1.8596	133	-1644
	100	-0.3896	-0.33036	-1.8596	132	-1931
	500	-0.38965	-0.33036	-1.8596	130	-1964
	1000	-0.38965	-0.33036	-1.8596	129	-1964
$k_{\text{sto}}$	0	-0.11088	-0.28878	-1.4153	248	-1210
	10	-0.38867	-0.33029	-1.8595	143	-1367
	50	-0.38945	-0.33036	-1.8596	138	-1413
	100	-0.38945	-0.330357	-1.8596	137	-1500
	500	-0.38959	-0.33036	-1.8596	132	-1931
	1000	-0.38959	-0.33036	-1.8596	131	-1920
$a$	0.1	-0.38965	-0.33036	-1.8596	131	-1758
	1	-0.38959	-0.33036	-1.8596	132	-1931
	10	-0.38965	-0.33035	-1.8596	133	-1668
	100	-0.38648	-0.33009	-1.8596	134	-1726

Table 2. Results of optimal values in Case 2.

Example	1	2	3	4	5	
$\bar{\omega}$	1	-0.38966	-0.33036	-1.8596	128	-1958
	10	-0.38966	-0.33036	-1.8596	128	-1961
	100	-0.38966	-0.33036	-1.8596	130	-1964
	500	-0.38965	-0.33036	-1.8596	130	-1964
	1000	-0.38966	-0.33036	-1.8596	130	-1964
$k_{\text{max}}$	10	-0.38959	-0.33036	-1.8590	146	-1348
	50	-0.38959	-0.33036	-1.8596	133	-1644
	100	-0.38959	-0.33036	-1.8596	132	-1931
	500	-0.38965	-0.33036	-1.8596	130	-1964
	1000	-0.38965	-0.33036	-1.8596	129	-1964
$k_{\text{sto}}$	0	-0.11088	-0.28878	-1.4153	248	-1210
	10	-0.38951	-0.33035	-1.8594	138	-1459
	50	-0.38965	-0.33036	-1.8596	135	-1744
	100	-0.38965	-0.33036	-1.8596	135	-1932
	500	-0.38965	-0.33036	-1.8596	130	-1964
	1000	-0.38966	-0.33036	-1.8596	130	-1941
$a$	0.1	-0.38965	-0.33036	-1.85956	130	-1953
	1	-0.38965	-0.33036	-1.8596	130	-1964
	10	-0.38964	-0.33036	-1.8596	131	-1960
	100	-0.38951	-0.33036	-1.8596	131	-1959

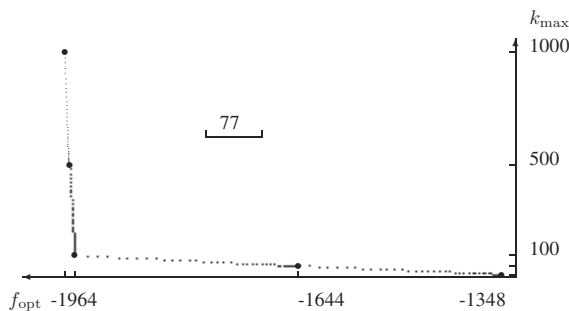


Fig. 2. Evaluation of the objective function with the iteration number for the HS114 problem.

In order to get some information about the robustness, we have studied the behavior of the method when

using different samples of random vectors. We observe that the results are stable, with small variance.

### 6. Concluding remarks

We have presented a stochastic modification of the projected variable metric method for nonsmooth optimization involving introduction of a stochastic perturbation. This approach leads to a stochastic descent method where the deterministic sequence generated by Clarke's generalized gradient is replaced with a sequence of random variables.

Numerical experiments show the effectiveness of the method. The use of stochastic perturbations improves the results furnished by Clarke's generalized gradient, and the robustness has been analyzed through the use of independent runs employed to estimate the resulting variable as

shown in Table 3. Thus, the robustness is increased. The main classical difficulty in the practical use of stochastic perturbations is connected with the tuning of the parameters. We have analyzed the practical effect of variations of  $a$ ,  $\bar{\omega}$ ,  $k_{\max}$  and  $k_{\text{sto}}$ . As shown in Tables 1 and 2, the more influenced parameters are  $k_{\max}$  and  $k_{\text{sto}}$ . We observe that, for a small number of iterations  $k_{\max}$ , a large number of perturbations  $k_{\text{sto}}$  is needed in order to obtain the best results, while a small  $k_{\text{sto}}$  requests a large  $k_{\max}$ . The parameters  $a$  and  $\bar{\omega}$  are less influential: the values of about 1 to 10 produced good results.

## References

- Bagirov, A. and Yearwood, J. (2006). A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems, *European Journal of Operational Research* **170**(2): 578–596.
- Bogani, C., Gasparo, M.G. and Papini, A. (2009). Generalized pattern search methods for a class of nonsmooth optimization problems with structure, *Journal of Computational and Applied Mathematics* **229**(1): 283–293.
- Bouveau, N. (1986). *Variables Aléatoires et Simulation*, Hermann Editions, Paris.
- Broyden, C. (1970). The convergence of a class of double-rank minimization algorithms, *Journal Institute of Mathematics and Its Applications* **6**(1): 76–90.
- Correa, R. and Lemaréchal, C. (1993). Convergence of some algorithms for convex minimization, *Mathematical Programming* **62**(2): 261–275.
- Davidon, W. (1991). Variable metric method for minimization, *SIAM Journal on Optimization* **1**(1): 1–17.
- Dorea, C. (1990). Stopping rules for a random optimization method, *SIAM Journal on Control and Optimization* **28**(4): 841–850.
- El Mouatasim, A. and Al-Hossain, A. (2009). Reduced gradient method for minimax estimation of a bounded Poisson mean, *Journal of Statistics: Advances in Theory and Applications* **2**(2): 183–197.
- El Mouatasim, A., Ellaia, R. and Souza de Cursi, J. (2006). Random perturbation of variable metric method for unconstrained nonsmooth nonconvex optimization, *International Journal of Applied Mathematics and Computer Science* **16**(4): 463–474.
- Fletcher, R. (1970). A new approach to variable metric algorithms, *Computer Journal* **13**(3): 317–322.
- Fletcher, R. and Powell, M. (1963). A rapidly convergent descent method for minimization, *Computer Journal* **6**(2): 163–168.
- Goldfarb, D. (1970). A family of variable metric methods derived by variational means, *Mathematics of Computation* **24**(109): 23–26.
- Hiriart-Urruty, J.-B. and Lemaréchal, C. (1993). *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*, Grundlehren der mathematischen Wissenschaften, Vol. 306, Springer-Verlag, Berlin.
- Kelley, J. (1960). The cutting plane method for solving convex programs, *Journal of the Society for Industrial and Applied Mathematics* **8**(4): 703–712.
- Kiwiel, K. (1985). *Method of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, Vol. 1133, Springer-Verlag, Berlin.
- Kowalczyk, Z., and Oliski K.O. (2006). Suboptimal fault tolerant control design with the use of discrete optimization, *International Journal of Applied Mathematics and Computer Science* **18**(4), DOI: 10.2478/v10006-008-0049-0.
- Kryazhinskii, A. (2001). Optimization problems with convex epigraphs. application to optimal control, *International Journal of Applied Mathematics and Computer Science* **11**(4): 773–801.
- Larsson, T., Patrcksson, M. and Stromberg, A. (1996). Conditional subgradient optimization-theory and applications, *European Journal of Operational Research* **88**(2): 382–403.
- Lemaréchal, C., Strodiot, J. and Bihain, A. (1981). On a bundle algorithm for nonsmooth optimization, in O. Mangasarian, R. Meyer and S. Robinson (Eds.), *Nonlinear Programming*, Vol. 4, Academic Press, New York, NY/London, pp. 245–282.
- Luenberger, D. (1973). *Introduction to Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, London.
- Lukšan, L. and Vlček, J. (2000). Test problems for nonsmooth unconstrained and linearly constraint optimization, *Technical Report TR-798*, Institute of Computer Sciences, Academy of Sciences of the Czech Republic, Prague.
- Makela, M. and Neittaanmaki, P. (1992). *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control*, World Scientific Publishing Co., London.
- Malanowski, K. (2004). Convergence of the Lagrange-Newton method for optimal control problems, *International Journal of Applied Mathematics and Computer Science* **14**(4): 531–540.
- Panier, E. (1987). An active set method for solving linearly constrained nonsmooth optimization problems, *Mathematical Programming* **37**(3): 269–292.
- Peng, Y. and Heying, Q. (2009). A filter-variable-metric method for nonsmooth convex constrained optimization, *Applied Mathematics and Computation* **208**(1): 119–128.
- Petersen, I. (2006). Minimax LQG control, *International Journal of Applied Mathematics and Computer Science* **16**(3): 309–323.
- Pinter, J. (1996). *Global Optimization in Action*, Kluwer, Dordrecht.
- Pogu, M. and Souza de Cursi, J. (1994). Global optimization by random perturbation of the gradient method with a fixed parameter, *Journal of Global Optimization* **5**(2): 159–180.
- Schramm, H. and Zowe, J. (1992). A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results, *SIAM Journal of Optimization* **2**: 121–152.

Shanno, D. (1970). Conditioning of quasi-Newton methods for function minimization, *Mathematics of Computation* **24**(111): 647–657.

Souza de Cursi, J. (1991). *Introduction aux probabilités*, Ecole Centrale Nantes, Nantes.

Souza de Cursi, J., Ellaia, R. and Bouhadi, M. (2003). Global optimization under nonlinear restrictions by using stochastic perturbations of the projected gradient, in C.A. Floudas and P.M. Pardalos (eds.), *Frontiers in Global Optimization*, Vol. 1, Dordrecht, Springer, pp.: 541–562.

Uryasev, S. (1991). New variable metric algorithms for nondifferentiable optimization problems, *Journal of Optimization Theory and Applications* **71**(2): 359–388.

Zhang, G. (2009). A note on: A continuous approach to nonlinear integer programming, *Applied Mathematics and Computation* **215**(6): 2388–2389.



**Abdelkrim El Mouatasim**, born in 1973, received a Ph.D. degree in applied mathematics and scientific computation in 2007 from Mohammedia Engineering School in Rabat, Morocco. Currently he works for the Department of Mathematics at the Faculty of Sciences, Jazan University, Saudi Arabia. His research interests include global optimization, nonsmooth optimization mathematical modelling, stochastic perturbation, water distribution system and combinatorial optimization. In 2001 he was decorated by King Mohammed VI.



**Rachid Ellaia** received a Ph.D. degree in applied mathematics and optimization in 1992 from Mohammed V University. Currently he is a full professor at the Mohammedia School of Engineers in Rabat, Morocco. His main research interests include areas such as global optimization, nonsmooth optimization, mathematical modelling, stochastic methods, convex analysis and financial mathematics. In 2009 he was a visiting professor at the National Institute for Applied Sciences, Rouen, and presently he is a visiting professor at the University of Nice, Sophia Antipolis, France.

**Eduardo Souza de Cursi**, born in 1957, graduate in physics: 1978, M.Sc.: 1980, Ph.D.: 1992. Presently: a professor at the National Institute for Applied Sciences, Rouen, the head of the Laboratory of Mechanics of Rouen, dean for European and international affairs and a delegate for European and international affairs of the group National Institute for Applied Sciences. He has experience in applied mathematics and mechanics, namely, theoretical and numerical aspects, stochastic methods and convex analysis.

### Appendix A

**Example 1.** (Lukšan and Vlček, 2000) Mad 1 problem.

$$\begin{cases} \min F(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})\}, \\ \text{subject to } h_1(\mathbf{x}) = x_1 + x_2 - 0.5 \geq 0, \end{cases}$$

where

$$\begin{aligned} f_1(\mathbf{x}) &= x_1^2 + x_2^2 + x_1x_2 - 1, \\ f_2(\mathbf{x}) &= \sin x_1, \\ f_3(\mathbf{x}) &= -\cos x_2. \end{aligned}$$

The starting point is  $\mathbf{x}_0 = (1, 2)$ .

**Example 2.** (Lukšan and Vlček, 2000) Mad 2 problem.

$$\begin{cases} \min F(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})\}, \\ \text{subject to } h_1(\mathbf{x}) = -3x_1 + x_2 - 2.5 \geq 0, \end{cases}$$

where  $f_i(\mathbf{x})$ ,  $i = 1, 2, 3$ , as in Example 1. The starting point is  $\mathbf{x}_0 = (-2, -1)$ .

**Example 3.** (Lukšan and Vlček, 2000) Pentagon problem.

$$\begin{cases} \min F(\mathbf{x}) = \max\{f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})\}, \\ \text{subject to } h_i(\mathbf{x}) = x_i \cos \frac{2\pi j}{5} + x_{i+1} \sin \frac{2\pi j}{5} \leq 1, \\ i = 1, 3, 5, \quad j = 0, 1, 2, 3, 4, \end{cases}$$

where

$$\begin{aligned} f_1(\mathbf{x}) &= -\sqrt{(x_1 - x_3)^2 + (x_2 - x_4)^2}, \\ f_2(\mathbf{x}) &= -\sqrt{(x_3 - x_5)^2 + (x_4 - x_6)^2}, \\ f_3(\mathbf{x}) &= -\sqrt{(x_5 - x_1)^2 + (x_6 - x_2)^2}. \end{aligned}$$

The starting point is  $\mathbf{x}_0 = (-1, 0, 0, -1, 0, 1)$ .

**Example 4.** (Lukšan and Vlček, 2000) Wong 3 problem.

$$\begin{cases} \min F(\mathbf{x}) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 \\ \quad + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 \\ \quad + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 \\ \quad + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + (x_{11} - 9)^2 \\ \quad + 10(x_{12} - 1)^2 + 5(x_{13} - 7)^2 \\ \quad + 4(x_{14} - 14)^2 + 27(x_{15} - 1)^2 + x_{16}^4 \\ \quad + (x_{17} - 2)^2 + 13(x_{18} - 2)^2 + (x_{19} - 3)^2 \\ \quad + x_{20}^2 + 95, \end{cases}$$

subject to

$$\begin{aligned} \psi_i(\mathbf{x}) &\leq 0, \quad i = 1, \dots, 13, \\ h_1(\mathbf{x}) &= 4x_1 + 5x_2 - 3x_7 + 9x_8 \leq 105, \\ h_2(\mathbf{x}) &= 10x_1 - 8x_2 - 17x_7 + 2x_8 \leq 0, \\ h_3(\mathbf{x}) &= -8x_1 + 2x_2 + 5x_9 - 2x_{10} \leq 12, \\ h_4(\mathbf{x}) &= x_1 + x_2 + 4x_{11} - 21x_{12} \leq 0, \end{aligned}$$

where

$$\begin{aligned} \psi_1(\mathbf{x}) &= 3(x_1 - 2)^2 + 4(x_2 - 3)^2 + 2x_3^2 - 7x_4 - 120, \\ \psi_2(\mathbf{x}) &= 5x_1^2 + 8x_2 + (x_3 - 6)^2 - 2x_4 - 40, \\ \psi_3(\mathbf{x}) &= 0.5(x_1 - 8)^2 + 2(x_2 - 4)^2 + 3x_5^2 - x_6 - 30, \\ \psi_4(\mathbf{x}) &= x_1^2 + 2(x_2 - 2)^2 - 2x_1x_2 + 14x_5 - 6x_6, \\ \psi_5(\mathbf{x}) &= -3x_1 + 6x_2 + 12(x_9 - 8)^2 - 7x_{10}, \\ \psi_6(\mathbf{x}) &= x_1^2 + 5x_{11} - 8x_{12} - 28, \\ \psi_7(\mathbf{x}) &= 4x_1 + 9x_2 + 5x_{13}^2 - 9x_{14} - 87, \\ \psi_8(\mathbf{x}) &= 3x_1 + 4x_2 + 3(x_{13} - 6)^2 - 14x_{14} - 10, \\ \psi_9(\mathbf{x}) &= 14x_1^2 + 35x_{15} - 79x_{16} - 92, \\ \psi_{10}(\mathbf{x}) &= 15x_2^2 + 11x_{15} - 61x_{16} - 54, \\ \psi_{11}(\mathbf{x}) &= 5x_1^2 + 2x_2 + 9x_{17}^4 - x_{18} - 68, \\ \psi_{12}(\mathbf{x}) &= x_1^2 - x_9 + 19x_{19} - 20x_{20} + 19, \\ \psi_{13}(\mathbf{x}) &= 7x_1^2 + 5x_2^2 + x_{19}^2 - 30x_{20}. \end{aligned}$$

In order to apply the algorithm, the nonlinear constraints are penalized, i.e.,  $F$  is replaced with

$$F_\lambda(\mathbf{x}) = F(\mathbf{x}) + \lambda \sum_{i=1}^{13} \psi_i^+(\mathbf{x}),$$

where

$$\psi_i^+(\mathbf{x}) = \max\{0, \psi_i(\mathbf{x})\}.$$

$F_\lambda$  satisfies the general assumptions. The numerical experiment uses  $\lambda = 10$ . The starting point is

$$\mathbf{x}_0 = (2, 3, 5, 5, 1, 2, 7, 3, 6, 10, 2, 2, 6, 15, 1, 2, 1, 2, 1, 3).$$

**Example 5.** (Lukšan and Vlček, 2000) HS114 problem.

$$\left\{ \begin{array}{l} \min F(\mathbf{x}) = 5.04x_1 + 0.035x_2 + 10x_3 + 3.36x_5 \\ \quad - 0.063x_4x_7, \\ \text{subject to} \\ \psi_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, 4, \\ \eta_j(\mathbf{x}) = 0, \quad j = 1, 2, \\ h_1(\mathbf{x}) = 0.222x_{10} + bx_9 \leq 35.82, \\ h_2(\mathbf{x}) = 0.222x_{10} + \frac{1}{b}x_9 \geq 35.82, \\ h_3(\mathbf{x}) = 3x_7 - ax_{10} \geq 133, \\ h_4(\mathbf{x}) = 3x_7 - \frac{1}{a}x_{10} \leq 133, \\ h_5(\mathbf{x}) = 1.22x_4 - x_1 - x_5 = 0, \end{array} \right.$$

where

$$\begin{aligned} \psi_1(\mathbf{x}) &= 1.12x_1 + 0.13167x_1x_8 - 0.00667x_1x_8^2 \\ &\quad - \frac{1}{a}x_4, \\ \psi_2(\mathbf{x}) &= -(1.12x_1 + 0.13167x_1x_8 - 0.00667x_1x_8^2 \\ &\quad - ax_4), \\ \psi_3(\mathbf{x}) &= 1.098x_8 - 0.038x_8^2 + 0.325x_6 \\ &\quad - \frac{1}{a}x_7 + 57.425, \\ \psi_4(\mathbf{x}) &= -(1.098x_8 - 0.038x_8^2 + 0.325x_6 \\ &\quad - ax_7 + 57.425), \\ \eta_1(\mathbf{x}) &= \frac{98000x_3}{x_4x_9 + 1000x_3} - x_6, \\ \eta_2(\mathbf{x}) &= \frac{x_2 + x_5}{x_1} - x_8, \end{aligned}$$

$a = 0.99$ ,  $b = 0.90$ , and sample constraints are

$$\begin{aligned} 10^{-5} &\leq x_1 \leq 2000, & 85 &\leq x_6 \leq 93, \\ 10^{-5} &\leq x_2 \leq 16000, & 90 &\leq x_7 \leq 95, \\ 10^{-5} &\leq x_3 \leq 120, & 3 &\leq x_8 \leq 12, \\ 10^{-5} &\leq x_4 \leq 5000, & 1.2 &\leq x_9 \leq 4, \\ 10^{-5} &\leq x_5 \leq 2000, & 145 &\leq x_{10} \leq 162, \end{aligned}$$

The starting point is

$$\mathbf{x}_0 = (1745, 12000, 110, 3048, 1974, 89.2, 92.8, 8, 3.6, 145).$$

As in the previous example, the nonlinear constraints are penalized. In addition, the restriction  $h_5$  is rewritten as  $h_5 \leq 0$  and  $-h_5 \leq 0$ . The restriction  $-h_5 \leq 0$  is treated as an ordinary affine one, while the other one is penalized. Thus we minimize

$$F_\lambda = F + \lambda \left[ \sum_{i=1}^4 \psi_i^+ + h_5^+ + |\eta_1| + |\eta_2| \right].$$

We use  $\lambda = 1000$ .

Received: 9 February 2010  
Revised: 6 November 2010