ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

**G. Liuzzi, A. Risi**

# A DECOMPOSITION ALGORITHM FOR UNCONSTRAINED OPTIMIZATION PROBLEMS WITH PARTIAL DERIVATIVE INFORMATION

**R. 09-22 2009**

**G. Liuzzi** – Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Consiglio Nazionale delle Ricerche, Viale Manzoni 30 - 00185 Rome, Italy, liuzzi@iasi.cnr.it.

**A. Risi** – Clinical & Therapy Development Consultant, Medtronic Italia, Via Lucrezio Caro 63 - 00193 Rome, Italy, arnaldo.risi@medtronic.it.

ISSN: 1128–3378

Collana dei Rapporti
Istituto di Analisi dei Sistemi ed Informatica “Antonio Ruberti”
Consiglio Nazionale delle Ricerche

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.cnr.it
URL: http://www.iasi.cnr.it

## Abstract

In this paper we consider the problem of minimizing a nonlinear function using partial derivative knowledge. Namely, the objective function is such that its derivatives with respect to a prespecified block of variables cannot be computed. To solve the problem we propose a block decomposition method that takes advantage of both derivative-free and derivative-based iterations to account for the features of the objective function. Under quite standard assumptions, we manage to prove global convergence of the method to stationary points of the problem.

**Keywords**. Unconstrained optimization, block decomposition method, derivative-free iteration

## 1. Introduction

We consider the following unconstrained minimization problem

$$\min_{x \in \Re^\ell} f(x) \tag{1}$$

where $f$ is a continuously differentiable function. We suppose that the vector $x \in \Re^\ell$ is partitioned into the component vectors $y \in \Re^n$ and $z \in \Re^m$ with $\ell = n + m$, and we denote $x = (y, z)$. Furthermore, we assume that first derivatives of $f$ with respect to the $y$ variables can be neither explicitly computed nor approximated.
Many real world problems in engineering such as parameter estimation of nonlinear models or optimal design of complex physical systems, can be stated as Problem (1). Moreover, statistical inference of diffusion processes based on the use of a likelihood function leads to problems of the form (1) (see e.g., [8],[9]).
As argued in [4], the use of block decomposition techniques has some strong motivations. For instance, when some variables are fixed, we can obtain subproblems of a special structure which can be fruitfully exploited by means of suitable optimization methods. For a thorough discussion on the relevance and importance of block decomposition methods we refer the reader to [4] and the references therein. In the present context, when the $y$ variables are held fixed, one can take advantage of some derivative-free methods [11, 6, 5, 2] for minimization of $f$ with respect to the $x$ variables. Conversely, when we held fixed the $x$ variables, the minimization of $f$ with respect to $y$ can be carried out by means of some efficient gradient-based method (see e.g. [10, 1, 7, 3]). More in particular, in the paper we propose the use of derivative-free iteration map of the pattern search type [11] for minimization with respect to the $x$ variables and of derivative-based Armijo-type iteration map [1] for minimization with respect to the $y$ variables. The convergence analysis of the proposed method will thus be developed in this particular setting.
Throughout the paper, we denote by $p_i$, $i = 1, \dots, r$, the directions used by the derivative-free minimization algorithm, whereas, $d \in R^m$ denotes the direction used by the Armijo-type linesearch procedure. Given a vector $v \in \Re^n$, by diag(v) we denote the $n \times n$ diagonal matrix with the components of vector $v$ on the main diagonal.
Finally, we require the following assumption to hold true throughout the paper.

**Assumption 1.** *Given an initial point $x^0 = (y^0, z^0) \in \Re^\ell$, the level set*

$$\mathcal{L}_0 = \{x : f(x) \le f(x^0)\}$$

*is compact.*

This is a basic assumption that is needed to guarantee the existence of solutions for Problem (1).
The paper is organized as follows. In Section 2 we an illustrative example that is useful to better understand the aim of the paper. In Section 3 we present the derivative-free iteration map and give its main theoretical properties. In Section 4 we introduce the derivative-based iteration map and recall the relevant properties. Section 5 is devoted to the definition of the algorithmic scheme along with its convergence analysis. Finally, we report in the appendix some useful technical results.

## 2. An illustrative example

Let us consider a physical system with state dynamics defined by the following system of Ordinary Differential Equations (ODE)

$$dv(t) = h(v(t); x)dt, \qquad v(0) = v_0 \tag{2}$$

where $h(\cdot; x) : \Re^m \to \Re^m$ is a vector of functions depending on the unknown parameters $x \in \Re^n$. Assume that the entire state variable $v$ is observable and suppose that we are given $N$ measurements

$u_1, \ldots, u_N \in \Re^m$ of $v(t)$ at time instants $t_1, \ldots, t_N$. Further, assume that the measures are affected by measurement noise, that is,

$$u_i = v_i + \varepsilon_i, \qquad i = 1, \ldots, N,$$

with $v_i = v(t_i) \in \Re^m$, and $\varepsilon_i \in \Re^m$, $i = 1, \ldots, N$, being $N$ realizations of a vector of random variables $\Sigma \in \Re^m$ whose components are identically, independent and normally distributed with zero mean and unknown standard deviations $y \in \Re^m_+$.

Now consider the problem of estimating the unknown parameters $y \in \Re^m$ and $x \in \Re^n$. Typically, the parameter estimation problem is formulated by using a likelihood function. More precisely, for any $x$, let $v_i(x)$, $i = 1, \ldots, N$ be the output of system (2) at times $t_1, \ldots, t_N$. Assuming that the measurement errors are stochastically independent from each other and that they are independent from the dynamic process, we can define the following likelihood function [8, 9].

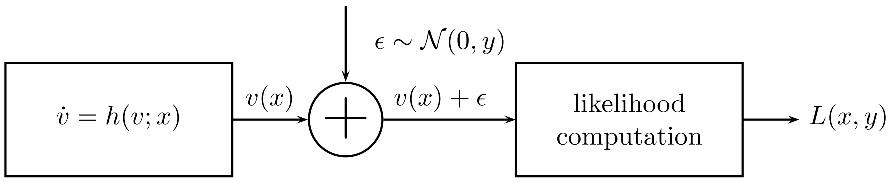


Figure 1: Computation schema of the likelihood function $L(x, y)$. In the figure, $\dot{v}$ stands for $dv/dt$; $\varepsilon \sim \mathcal{N}(0, y)$ means that the error $\varepsilon$ is statistically distributed according to a normal probability density function with zero mean and $y$ standard deviation.

$$L(x, y) = \prod_{i=1}^{N} \frac{1}{\|2\pi y\|} e^{-\dfrac{(v_i(x) - u_i)^\top \mathrm{diag}(y)^{-2}(v_i(x) - u_i)}{2}} \tag{3}$$

Then, the best values for the vector of unknown parameters $x$ and $y$ are the maximizers of function $L(x, y)$. We remark that, in practical situations, the system of ODE (2) cannot be solved analytically but only by means of some numerical integration method. For this reason, the analytical expression of the function $v(x)$ is not known so that, there is no derivative knowledge of the function $L(\cdot, y)$ with respect to $x$. On the contrary, for every fixed $x$, $L(x, \cdot)$, as a function of $y$ alone, has a perfectly known expression (which is given by (3) where $v_i(x)$ are constant).

Furthermore, we note that instead of maximizing function $L(x, y)$, it is possible to maximize the function $\log L(x, y)$ which is equal to

$$\log L(x, y) = -\log \|2\pi y\| - \sum_{i=1}^{N} \frac{(v_i(x) - u_i)^\top \mathrm{diag}(y)^{-2}(v_i(x) - u_i)}{2}$$

and this amounts to solving the problem

$$\min_{x,y} \log \|2\pi y\| + \sum_{i=1}^{N} \frac{(v_i(x) - u_i)^\top \mathrm{diag}(y)^{-2}(v_i(x) - u_i)}{2}.$$

## 3. A Derivative-free mapping

In this section we introduce the derivative-free iteration map, used to update the $y$ variables of Problem (1), and we briefly recall its main theoretical properties. Let

$$D = \{p_1, \ldots, p_r\}$$

be the set of directions that will be used by the derivative-free iteration map. Let us assume that, for $i = 1, \ldots, r$, $\|p_i\| = 1$ and that the set $D$ positively span $\Re^n$, that is, for any vector $v \in \Re^n$, $r$ non-negative scalars $\beta_1, \ldots, \beta_r$ exist such that

$$v = \sum_{i=1}^{r} \beta_i p_i.$$

Given a sequence $\{x^k\} \subset \Re^{n+m}$ and set $D$, we define, for each $k$, a derivative-free iteration map according to a pattern search derivative-free procedure with sufficient decrease (see, e.g. [11, 5]).

**Iteration Map** $DF(x^k, \Delta^k)$

INPUT: $x^k = (y^k, z^k) \in \Re^{n+m}$, $\Delta^k > 0$.

DATA: $D = \{p_1, \ldots, p_r\}$, $\gamma_1 > 0$, $\theta \in (0, 1)$.

STEP 1: If $f(y^k + \Delta^k p_i, z^k) \leq f(y^k, z^k) - \gamma_1 (\Delta^k)^2$ for some $p_i \in D$, then
Set $w^{k+1} = y^k + \Delta^k p_i$, $\Delta^{k+1} = \Delta^k$.
Otherwise Set $w^{k+1} = y^k$, $\Delta^{k+1} = \theta \Delta^k$.

OUTPUT: $w^{k+1}$, $\Delta^{k+1}$.

The *Iteration Map* DF is a procedure that, given a point $x^k = (y^k, z^k)$ and a tentative step $\Delta^k > 0$, returns

- either $w^{k+1} \neq y^k$ and $\Delta^{k+1} = \Delta^k$ such that $f(w^{k+1}, z^k) \leq f(y^k, z^k) - \gamma_1 (\Delta^{k+1})^2$;
- or $w^{k+1} = y^k$ and $\Delta^{k+1} = \theta \Delta^k < \Delta^k$.

In the next proposition, whose proof can be found in reference [5] and in the appendix for the sake of completeness, we recall the main theoretical properties of the iteration map DF that are of interest to prove convergence of the overall decomposition method to stationary points of $f$. Note that $\{x^k\}$ is a given sequence that may not depend on the iteration map $DF$, in the sense that $y^{k+1}$ is not necessarily equal to $w^{k+1}$.

**Proposition 3.1 (See [11, 5])** *Let $\{x^k\}$ be a sequence of points in $\Re^{n+m}$ and let $\{w^k\}$ and $\{\Delta^k\}$ be the sequences produced by the iteration*

$$(w^{k+1}, \Delta^{k+1}) = DF(x^k, \Delta^k).$$

*If $\lim_{k \to \infty} \left[ f(x^k) - f(w^{k+1}, z^k) \right] = 0$ and $\{x^k\}_H$ converges to $\bar{x}$ with $H = \{k : \Delta^{k+1} < \Delta^k\}$, then*

$$\lim_{k \to \infty} \Delta^k = 0, \tag{4}$$

$$\nabla_y f(\bar{w}) = 0. \tag{5}$$

**Proof**. The proof of the proposition is quite standard and is reported in the appendix for the sake of completeness. □

## 4. A Derivative-based linesearch mapping

In this section we introduce the derivative-based linesearch mapping, used to update the $z$ variables of Problem (1), and we recall its main theoretical properties. In order to do this, let us define $d^k \in \Re^m$ as follows

$$d^k = -H^k \nabla_z f(x^k). \tag{6}$$

where $\{x^k\}$ is a predefined sequence of points in $\Re^{n+m}$ and $\{H^k\} \in \Re^{m \times m}$ is a sequence of positive definite matrices, and that there exist numbers $0 < r \leq R$ such that it results, for all $k$, $0 < r \leq \lambda_m(H^k) \leq$

$\lambda_M(H^k) \leq R$, where $\lambda_m(H^k)$ and $\lambda_M(H^k)$ denote, respectively, the smallest and largest eigenvalues of $H^k$.
Note that $d^k$ is, by definition, a *gradient-related* direction [1], that is, it satisfies

$$\begin{aligned} \nabla_z f(x^k)^\top d^k &\leq -r\|\nabla_z f(x^k)\|^2; \\ \|d^k\| &\leq R\|\nabla_z f(x^k)\|. \end{aligned}$$

In the following we report an Armijo-type linesearch mapping that, given $x^k$ and $d^k$ computes a steplength $\beta^k$ and point $z^k + \beta^k d^k$.

**Iteration Map** $LS(x^k, d^k)$

INPUT: $x^k = (y^k, z^k) \in \Re^{n+m}$, $d^k \in \Re^m$.

DATA: $\gamma_2 \in (0,1)$, $\Lambda > 0$, $\delta_2 \in (0,1)$.

STEP 1: Compute $\beta^k = \max\{\delta_2^j \Lambda : j = 0, 1, \ldots\}$ such that

$$f(y^k, z^k + \beta^k d^k) \leq f(x^k) + \gamma_2 \beta^k \nabla_z f(x^k)^\top d^k. \tag{7}$$

OUTPUT: $z^k + \beta^k d^k$, $\beta^k$.

In the following proposition, whose proof can be found in reference [1] and in the appendix for the sake of completeness, we recall some basic properties of the Armijo-type Iteration Map. Basically, it shows that the line search procedure is well-defined and that the *Iteration Map LS* is able to force $\nabla f_z(x)$ to zero. Also in this case, we note that $\{x^k\}$ is a given sequence that may not depend on the iteration map $LS$, in the sense that $z^{k+1}$ is not necessarily equal to $z^k + \beta^k d^k$.

**Proposition 4.1 (See [1])** *Let $\{x^k\}$ be a sequence of points in $\Re^{n+m}$ and let $\{d^k\}$ be a sequence of directions defined as in 6. Let $\beta^k$ be computed by means of the Iteration Map LS when $\nabla_z f(x^k) \neq 0$ and set $\beta^k = 0$ whenever $\nabla_z f(x^k) = 0$. Then:*

*(i) there exists a finite integer $j$ such that $\beta^k = \delta_2^j \Delta^k$ satisfies the acceptability condition at Step 1.*

*(ii) if $\lim_{k\to\infty} \left[f(x^k) - f(y^k, z^k + \beta^k d^k)\right] = 0$ and $\{x^k\}$ converges to $\bar{x}$ then we have*

$$\nabla_z f(\bar{x}) = 0.$$

**Proof**. The proof of the proposition is quite standard and is reported in the appendix for the sake of completeness. □

## 5. A two-block decomposition algorithm

In this section we propose a block decomposition algorithm for the solution of Problem (1) in case of partial derivative knowledge. The algorithm is based on the use of the Iteration Maps $DF$ and $LS$ for the inexact minimization with respect to the $y$ and $z$ variables, respectively.

**Algorithm 1.**

**Data:** $x^0 = (y^0, z^0) \in \Re^{n+m}$, $\Delta^0, \gamma_1, \gamma_2 > 0$ and $\tau_1 \geq 1/\gamma_1$, a sequence of gradient-related directions $\{d^k\}$.

**Step 0:** Set $k = 0$.

**Step 1:** *Derivative-free line search w.r.t. $y$*

(a) Compute $\tilde{y}^{k+1}$, $\Delta^{k+1}$ by means of $DF(x^k, \Delta^k)$.

(b) Choose $y^{k+1}$ such that

$$f(y^{k+1}, z^k) \leq f(\tilde{y}^{k+1}, z^k);$$
$$\|y^{k+1} - y^k\|^2 \leq \tau_1(f(y^k, z^k) - f(y^{k+1}, z^k)).$$

**Step 2:** *Derivative-based line search w.r.t. z*

(a) Compute $\beta^k$ by means of $LS((y^{k+1}, z^k), d^k)$ (with $\beta^k = 0$ if $\nabla_z f(y^{k+1}, z^k) = 0$) and set $\tilde{z}^{k+1} = z^k + \beta^k d^k$.

(b) Choose $z^{k+1}$ such that

$$f(y^{k+1}, z^{k+1}) \leq f(y^{k+1}, \tilde{z}^{k+1}).$$

**Step 3:** Set $x^{k+1} = (y^{k+1}, z^{k+1})$, $k = k + 1$ and go to Step 1.

**Remark 1.** *We note that the conditions at Steps 1(b) and 2(b) can be satisfied by choosing*

$$\begin{aligned} y^{k+1} &= \tilde{y}^{k+1}, \\ z^{k+1} &= \tilde{z}^{k+1}. \end{aligned}$$

*Indeed, since $\tau_1 \geq \gamma_1$, it can be verified that the acceptance rule of the Iteration Maps DF and LS ensure that conditions at Steps 1(b) and 2(b) are verified.*

**Theorem 5.1.** *Let $\{x^k\}$ be the sequence generated by Algorithm 1. Then,*

*(i) $\{x^k\} \subset \mathcal{L}_0$;*

*(ii) $\{x^k\}$ admits limit points and one of these is a stationary point of $f$.*

**Proof**. By the instructions of Algorithm 1 and by the definitions of the mappings $DF$ and $LS$, it follows that $\{x^k\}$ is such that

$$f(x^{k+1}) \leq f(y^{k+1}, \tilde{z}^{k+1}) \leq f(y^{k+1}, z^k) \leq f(\tilde{y}^{k+1}, z^k) \leq f(x^k), \tag{8}$$

so that point (i) is proved. Now, let us define

$$H = \{k : \Delta^{k+1} < \Delta^k\}.$$

Since, $\{x^k\}_H$ belongs to $\mathcal{L}_0$ which, by Assumption 1, is compact, it follows that $\{x^k\}_H$ admits limit points. Let $\bar{x}$ be any limit point of $\{x^k\}_H$, that is

$$\lim_{k\to\infty, k\in\bar{H}} x^k = \bar{x},$$

with $\bar{H} \subseteq H$. The continuity of $f$ and the convergence of $\{x^k\}_{\bar{H}}$ imply that the sequence $\{f(x^k)\}$ has a convergent subsequence. As $\{f(x^k)\}$ is non-increasing, this, in turn, implies that $\{f(x^k)\}$ is bounded from below and has a limit. Therefore, recalling (8) we have that

$$\lim_{k\to\infty} f(y^k, z^k) - f(\tilde{y}^{k+1}, z^k) = 0, \tag{9a}$$
$$\lim_{k\to\infty} f(y^{k+1}, z^k) - f(y^{k+1}, \tilde{z}^{k+1}) = 0, \tag{9b}$$
$$\lim_{k\to\infty} f(y^k, z^k) - f(y^{k+1}, z^k) = 0, \tag{9c}$$
$$\lim_{k\to\infty} f(y^{k+1}, \tilde{z}^{k+1}) - f(y^{k+1}, z^k) = 0. \tag{9d}$$

Hence, by (9a) and Propositions 3.1, where we identify $w^{k+1}$ with $\tilde{y}^{k+1}$, we have that

$$\nabla_y f(\bar{x}) = 0. \tag{10}$$

Now, we show that a subset $\tilde{H} \subseteq \bar{H}$ exists such that

$$\lim_{k\to\infty, k\in\tilde{H}} \|y^{k+1} - y^k\| = 0. \tag{11}$$

By the instructions at Step 1(b) of the algorithm, we know that the following condition is satisfied

$$\|y^{k+1} - y^k\|^2 \leq \tau_1(f(y^k, z^k) - f(y^{k+1}, z^k)) \tag{12}$$

Now, we have that (11) holds by (9c) with $\tilde{H} = \bar{H}$. Thus, we obtain that

$$\lim_{k\to\infty, k\in\tilde{H}} y^{k+1} = \bar{y}.$$

Hence, we have that $\{x^k\}_{\tilde{H}}$ and $\{(y^{k+1}, z^k)\}_{\tilde{H}}$ are both convergent and have the same limit $\bar{x}$. Therefore, by (9d) and Proposition 4.1, where we identify $\{x^k\}$ with $\{(y^{k+1}, z^k)\}_{\tilde{H}}$, we have that

$$\nabla_z f(\bar{x}) = 0 \tag{13}$$

which, along with (10), concludes the proof. □

## 6. Concluding remarks

In the paper we considered the problem of minimizing the continuously differentiable function $f(x)$ where $x = (y, z)$ is such, that the derivative of $f$ with respect to $y$ cannot be neither computed nor approximated explicitly. For this kind of problem, we presented a block-decomposition method that takes advantage of both derivative-free and derivative-based minimization methods. The proposed method updates the current iterate by first changing the $y$ variables according to a derivative-free pattern search iteration map (see Section 3), and then updates the $z$ variables according to a derivative-based Armijo-type iteration map (see Section 4). This is a choice that we made to be able to prove convergence of the method and, at the same time, to keep clear the exposition of the paper. However, we remark that there is no conceptual difficulty in substituting the two iteration maps with any other derivative-free and, respectively, derivative-based mappings, as long as they are able to satisfy analogous properties to those stated in Propositions 3.1 and 4.1. Indeed, these results are necessary to be able to prove convergence of the overall method to stationary points of Problem 1. In particular, we could use any derivative-free linesearch mapping (see for instance [6]) to update the $y$ variables; As concerns the derivative-based mapping, it could be substituted by any safe-guarded linesearch procedure that allows extrapolation steps (see for instance [4]).

## Acknowledgements

The authors are indebted and wish to thank Prof. Marco Sciandrone for stimulating discussions on the topic of the paper, for useful suggestions and comments and for reading earlier versions of the paper.

## Appendix

## A. Technical results

**Proof of Proposition 3.1**. First we prove (4). By the definition of the Iteration Map $DF$, the sequence $\{\Delta^k\}$ is monotonically non-increasing, that is, $0 < \Delta^{k+1} \leq \Delta^k$. Hence, the sequence $\{\Delta^k\}$ is converging to a limit $M \geq 0$. Let us assume, by contradiction, that $M > 0$. If this were the case, then an index $\bar{k}$ should exist such that

$$\Delta^{k+1} = \Delta^k = M, \qquad \text{for all } k \geq \bar{k}.$$

By the instructions at Step 1 of the Iteration Map $DF$, this would imply that

$$\gamma_1 M^2 \leq f(x^k) - f(w^{k+1}, z^k), \qquad \text{for all } k \geq \bar{k},$$

which is a contradiction with the assumption that $\lim_{k\to\infty} f(x^k) - f(w^{k+1}, z^k) = 0$.
Now we prove (5). To this aim, note that for every index $k \in H$, it results

$$f(y^k + \Delta^k p_i, z^k) > f(y^k, z^k) - \gamma_1(\Delta^k)^2, \quad \forall\ p_i \in D,$$

which, in turn, implies that

$$\frac{f(y^k + \Delta^k p_i, z^k) - f(y^k, z^k)}{\Delta^k} > -\gamma_1\Delta^k, \quad \forall\ p_i \in D.$$

Taking the limit for $k \to \infty$, $k \in H$, in the above formula and recalling (4), we obtain

$$\nabla_y f(\bar{x})^\top p_i \geq 0, \quad \forall\ p_i \in D,$$

which, recalling that, by definition, the set $D$ positively spans $\Re^n$, implies that $\nabla_y f(\bar{x}) = 0$. □

**Proof of Proposition 4.1**. In order to prove point (i), let us assume, by contradiction, that $\nabla_z f(x^k) \neq 0$ and that condition (7) is violated for every $j \geq 0$, so that

$$\frac{f(y^k, z^k + \delta_2^j\Delta^k d^k) - f(x^k)}{\delta_2^j\Delta^k} > \gamma_2\nabla_z f(x^k)^\top d^k.$$

Then, taking the limit for $j \to \infty$, we would obtain

$$\gamma_2 \geq 1$$

thus contradicting the assumption $\gamma_2 \in (0, 1)$.
Now we prove point (ii). Since $\beta^k$ satisfies condition (7) and $d^k$, by definition, is such that

$$\frac{|\nabla_z f(x^k)^\top d^k|}{\|d^k\|} \geq \frac{r}{R^2}\|d^k\|, \tag{14}$$

we can write

$$f(x^k) - f(y^k, z^k + \beta^k d^k) \geq \gamma_2\beta^k|\nabla_z f(x^k)^\top d^k| \geq \gamma_2\frac{r}{R^2}\beta^k\|d^k\|^2.$$

Now, by assumption we know that $\lim_{k\to\infty} f(x^k) - f(y^k, z^k + \beta^k d^k) = 0$ so that

$$\lim_{k\to\infty} \beta^k|\nabla_z f(x^k)^\top d^k| = 0, \quad \lim_{k\to\infty} \beta^k\|d^k\|^2 = 0. \tag{15}$$

Let us first show that $\lim_{k\to\infty} \beta^k\|d^k\| = 0$. We proceed by contradiction and assume that a subset $K \subseteq \{0, 1, \ldots\}$ exists such that

$$\lim_{k\to\infty, k\in K} \beta^k\|d^k\| = \eta > 0. \tag{16}$$

This and (15) imply that

$$\lim_{k\to\infty, k\in K} \|d^k\| = 0, \tag{17}$$

which, again by (16), would imply that $\lim_{k\to\infty, k\in K} \beta^k = +\infty$, contradicting the fact that, by the definition of Iteration Map $LS$, $\beta^k \leq \Lambda$. Now we prove that

$$\lim_{k\to\infty} |\nabla_z f(x^k)^\top d^k| = 0.$$

We proceed again by contradiction and assume that a subset $K \subseteq \{0, 1, \ldots\}$ exists such that

$$\lim_{k\to\infty, k\in K} |\nabla_z f(x^k)^\top d^k| = \eta > 0. \tag{18}$$

This and (15) imply that

$$\lim_{k\to\infty, k\in K} \beta^k = 0. \tag{19}$$

Hence, for $k \in K$ and sufficiently large, it results $\beta^k < \Lambda$. Thus, by the instruction of the iteration map $LS$, we can write

$$f\left(y^k, z^k + \frac{\beta^k}{\delta_2} d^k\right) - f(x^k) > \gamma_2 \frac{\beta^k}{\delta_2} \nabla_z f(x^k)^\top d^k. \tag{20}$$

By the Mean Value Theorem, we have that

$$f\left(y^k, z^k + \frac{\beta^k}{\delta_2} d^k\right) = f(x^k) + \frac{\beta^k}{\delta_2} \nabla_z f(y^k, \xi^k)^\top d^k,$$

where $\xi^k = z^k + \theta^k \frac{\beta^k}{\delta_2} d^k$ and $\theta^k \in (0,1)$. Substituting the above equation into (20), we obtain

$$\nabla_z f(y^k, \xi^k)^\top d^k > \gamma_2 \nabla_z f(x^k)^\top d^k.$$

Taking the limit for $k \to \infty$ and $k \in K$ and considering that $\lim_{k\to\infty} \beta^k \|d^k\| = 0$ so that $(y^k, \xi^k) \to \bar{x}$, we get

$$\eta \leq \gamma_2 \eta$$

which yields $\gamma_2 \geq 1$ thus contradicting the assumption $\gamma_2 \in (0,1)$. Hence we have proved that

$$\lim_{k\to\infty} |\nabla_z f(x^k)^\top d^k| = 0. \tag{21}$$

Now, by the assumption on the search direction $d^k$ we know that

$$|\nabla_z f(x^k)^\top d^k| \geq r \|\nabla_z f(x^k)\|.$$

Taking the limit in the above relation, considering (21) and recalling that $\{x^k\}$ converges to $\bar{x}$, we obtain

$$\nabla_z f(\bar{x}) = 0$$

thus concluding the proof. □

## References

[1] D.P. Bertsekas, *Nonlinear Programming.* Athena Scientific, Massachusetts 1999

[2] K. Scheinberg A.R. Conn and L.N. Vicente. *Introduction to Derivative-free Optimization.* MPS-SIAM Series in Optimization, Philadelphia, (2009)

[3] J.E. Dennis Jr. and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* SIAM, Classics in Applied Mathematics 16, Philadelphia, (1996)

[4] L. Grippo and M. Sciandrone, *Globally convergent block-coordinate techniques for unconstrained optimization*, Optimization Methods and Software, 10 (1999), pp. 587–637.

[5] T.G. Kolda, R.M. Lewis and V. Torczon, *Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods*, SIAM Review, 45 (2003), pp. 385–482.

[6] S. Lucidi and M. Sciandrone, *On the global convergence of derivative free methods for unconstrained optimization*, SIAM J. on Optimization, 13 (2002), pp. 97–116.

[7] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer, Series in Operations Research, Berlin, Heidelberg, (1999)

[8] Pedersen, A. R. (1995). *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*, Scand. J. Statist., 22(1), 55–71.

[9] Pedersen, A. R. (2001). *Likelihood inference by monte carlo methods for incompletely discretely observed diffusion processes*, Technical report, Department of Biostatistics, University of Aarhus, Denmark.

[10] E. Polak. *Optimization, Algorithms and Consistent approximations.* Springer, Series in Applied Mathematical Sciences, Berlin, Heidelberg, (1997)

[11] V. Torczon, *On the convergence of pattern search algorithms*, SIAM J. on Optimization, 7 (1997), pp. 1–25.