Optim Lett (2012) 6:437–450
DOI 10.1007/s11590-010-0270-2

ORIGINAL PAPER

A decomposition algorithm for unconstrained optimization problems with partial derivative information

G. Liuzzi · A. Risi

Received: 23 November 2009 / Accepted: 9 December 2010 / Published online: 25 December 2010

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 pre-specified 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 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

G. Liuzzi (✉)
Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti" (IASI)-CNR,
Viale Manzoni 30, 00185 Rome, Italy
e-mail: giampaolo.liuzzi@iasi.cnr.it

A. Risi
Clinical & Therapy Development Consultant, Medtronic Italia,
Via Lucrezio Caro 63, 00193 Rome, Italy
e-mail: arnaldo.risi@medtronic.com

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 non-linear models or optimal design of complex physical systems, can be stated as Problem (1). In particular, we refer to multidisciplinary design optimization (MDO) problems [1,6], to penalty methods for generalized Nash equilibrium (GNE) problems [4] and to statistical inference (SI) problems of diffusion processes [12]. MDO is concerned with systematic approaches to the design optimization of complex, coupled engineering systems, where "multidisciplinary" refers to the different aspects that must be included in the design problem. The design process is extremely complex because engineering systems are often governed by the considerations of all the contributing disciplines. It often happens that partial derivatives with respect to variables related to some particular disciplines are unavailable or untrustworthy. In the context of GNE problems, let us refer to the specific case in which some of the players participating in the game cannot compute partial derivatives of their respective objective and/or constraint functions. In this case, a penalty function approach would yield to a problem with an objective function with partial derivative knowledge. Finally, as it will be further discussed in the following section, SI of stochastic processes based on the use of a likelihood function leads to problems of the form (1) (see e.g., [11,12]).

As argued in [5], the use of block decomposition techniques for this kind of problems 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 [5] and the references therein. In the present context, when the z variables are held fixed, one can take advantage of some derivative-free methods [7,8,14,15] for minimization of f with respect to the y variables. Conversely, when we hold fixed the y variables, the minimization of f with respect to z can be carried out by means of some efficient gradient-based method (see e.g. [2,3,9,13]). More in particular, in the paper we propose the use of a derivative-free iteration map of the pattern search type [15] for minimization with respect to the y variables and of a derivative-based Armijo-type iteration map [2] for minimization with respect to the z variables. The convergence analysis of the proposed method will thus be developed in this particular setting.

Throughout the paper, we denote by $p_i \in \Re^n$, $i = 1, \dots, r$, 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 $\text{diag}(v)$ we denote the $n \times n$ diagonal matrix with the components of vector v on the main diagonal. Given a sequence of points $\{x^k\} \subset \Re^n$ and an index set H, $\{x^k\}_H$ denotes the subsequence of iterates with indices in H, namely $\{x^k\}_H = \{x^k : k \in H\}$. According to [10], a function $\sigma : \Re^+ \to \Re^+$ that is continuous, non-decreasing and such that $\sigma(t) > 0$ for $t > 0$ and

$$\lim_{k\to\infty} \sigma(t_k) = 0 \quad \textit{implies} \quad \lim_{k\to\infty} t_k = 0$$

will be denoted *forcing function*. In the paper, $\|\cdot\|$ will denote the Euclidean norm.

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 guarantees existence of solutions for Problem (1), provided that the objective function f is at least continuous.

The paper is organized as follows. In Sect. 2 we present an illustrative example that is useful to better understand the aim of the paper. In Sect. 3 we present the derivative-free iteration map and give its main theoretical properties. In Sect. 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 depending on the unknown parameters $y \in \Re^n$ and defined by the following system of ordinary differential equations (ODE)

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

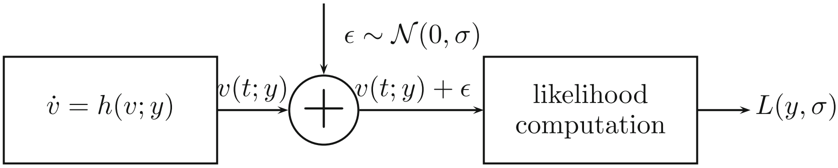
where $h(\cdot; y) : \Re^m \to \Re^m$ is a vector of functions depending on the unknown parameters $y \in \Re^n$. Assume that the entire state vector v is observable and suppose that we are given N measurements $u_1, \dots, u_N \in \Re^m$ of $v(t; y)$ at time instants $t_1, \dots, t_N$. Further, assume that the measures are affected by measurement noise, that is,

$$u_i = v(t_i; y) + \varepsilon_i, \quad i = 1, \dots, N,$$

with $\varepsilon_i \in \Re^m$, $i = 1, \dots, N$, 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 σ_i, $i = 1, \dots, m$.

Now consider the problem of estimating the unknown parameters $\sigma \in \Re^m_+$ and $y \in \Re^n$. Typically, the parameter estimation problem is formulated by using a likelihood function. More precisely, for any y, let $v_i(y) = v(t_i; y) \in \Re^m$, $i = 1, \dots, N$, be the output of system (2) at times $t_1, \dots, t_N$. 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 [11,12]

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



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

Then, the best values for the vector of unknown parameters y and σ are the maximizers of function $L(y,\sigma)$. 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 functions $v_i(y)$ are not known so that, there is no derivative knowledge of the function $L(\cdot,\sigma)$ with respect to y. On the contrary, for every fixed y, $L(y,\cdot)$, as a function of σ alone, has a perfectly known expression (which is given by (3) where $v_i(y)$ are constant).

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

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

and this amounts to solving the problem

$$\min_{y\in\Re^n,\sigma\in\Re^m_+} \tilde{f}(y,\sigma) = N \log 2\pi \|\sigma\| + \sum_{i=1}^{N} \frac{(v_i(y)-u_i)^\top \mathrm{diag}(\sigma)^{-2}(v_i(y)-u_i)}{2}. \tag{4}$$

We note that function $\tilde{f}(y,\sigma)$ is not continuously differentiable on the whole space $\Re^{n+m}$. Thus, $\tilde{f}(y,\sigma)$ does not satisfy the continuous differentiability assumption stated in the Sect. 1. However, we note that by posing $\sigma_i = z_i^2$, $i = 1,\dots,m$, Problem (4) could be approximated (see, e.g. [4]) by the following smooth problem

$$\begin{aligned}\min_{y,z} f(y,z) = N \log \left(2\pi (z^\top z + \varepsilon)^{1/2}\right) \\ + \sum_{i=1}^{N} \frac{(v_i(y)-u_i)^\top (\mathrm{diag}(z)^{-4} + \delta I)(v_i(y)-u_i)}{2}.\end{aligned}$$

where $\varepsilon > 0$ and $\delta > 0$ are two small positive numbers.

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 the 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. [7,15]).

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

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

DATA: $D = \{p_1, \ldots, p_r\}$, $\theta \in (0, 1)$, $\phi \geq 1$ and a forcing function $\sigma(t)$ such that $\lim_{t \to 0^+} \frac{\sigma(t)}{t} = 0$.

STEP 1: If $f(y^k + \Delta^k p_i, z^k) \leq f(y^k, z^k) - \sigma(\Delta^k)$ for some $p_i \in D$, then

Set $w^{k+1} = y^k + \Delta^k p_i$, $\Delta^{k+1} = \phi\Delta^k$.

Otherwise Set $w^{k+1} = y^k$, $\Delta^{k+1} = \theta\Delta^k$.

OUTPUT: w^{k+1}, Δ^{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 $\Delta^{k+1} = \phi\Delta^k \geq \Delta^k$ such that $f(w^{k+1}, z^k) \leq f(y^k, z^k) - \sigma(\Delta^k)$;
- or $w^{k+1} = y^k$ and $\Delta^{k+1} = \theta\Delta^k < \Delta^k$.

It is worth noting that the stepsize Δ^k is reduced only when it is not able to yield a sufficient decrease in the objective function along any of the directions in D.

In the next proposition, whose proof can be obtained reasoning as in, e.g., references [7,8,15] and is reported 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 2 *Let $\{x^k\}$ be a given sequence of points in $\Re^{n+m}$ and let $\{w^k\}$ and $\{\Delta^k\}$ be the sequences produced by the iteration map DF, i.e., for all $k \geq 0$ we have*

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

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,

$$\lim_{k\to\infty} \Delta^k = 0, \tag{5}$$

$$\nabla_y f(\bar{x}) = 0. \tag{6}$$

Proof The proof of the proposition is 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, we will use a search direction $d^k \in \Re^m$ and assume that the sequence $\{d^k\}$ satisfies the following assumption [2].

Assumption 3 Given sequences $\{x^k\} \subset \Re^{n+m}$ and $\{d^k\} \subset \Re^m$, two scalars $r > 0$ and $R > 0$ exist such that, for all k,

$$\nabla_z f(x^k)^\top d^k \leq -r\|\nabla_z f(x^k)\|^2; \tag{7a}$$

$$\|d^k\| \leq R\|\nabla_z f(x^k)\|. \tag{7b}$$

It can be easily verified that a sequence of directions $\{d^k\}$ satisfying Assumption 3 is *gradient-related* to $\{x^k\}$ as defined in [2]. Note that Assumption 3 can be satisfied, for instance, by defining d^k as

$$d^k = -G^k \nabla_z f(x^k). \tag{8}$$

where $\{x^k\}$ is a predefined sequence of points in $\Re^{n+m}$ and $\{G^k\} \in \Re^{m\times m}$ is a sequence of positive definite matrices such that, for all k, $0 < r \leq \lambda_m(G^k) \leq \lambda_M(G^k) \leq R$, where $\lambda_m(G^k)$ and $\lambda_M(G^k)$ denote, respectively, the smallest and largest eigenvalues of G^k and r and R are numbers satisfying $0 < r \leq R$.

In the following we report an Armijo-type linesearch mapping that, given x^k and d^k computes a steplength β^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: $\gamma \in (0,1)$, $\Lambda > 0$, $\delta \in (0,1)$.

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

$$f(y^k, z^k + \beta^k d^k) \le f(x^k) + \gamma \beta^k \nabla_z f(x^k)^\top d^k. \tag{9}$$

OUTPUT: $z^k + \beta^k d^k$, β^k.

In the following proposition, whose proof can be found in reference [2] 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 *Let* $\{x^k\} = \{(y^k, z^k)\}$ *be a sequence of points in* $\Re^{n+m}$ *and let* $\{d^k\}$ *be a sequence of directions satisfying Assumption* 3. *Let* β^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* $\beta^k = \delta^j \Delta^k$ *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*

$$\lim_{k\to\infty} \beta^k \|d^k\| = 0, \quad \textit{and} \quad \nabla_z f(\bar{x}) = 0.$$

Proof The proof of the proposition is 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:** $x^0 = (y^0, z^0) \in \Re^{n+m}$, $\Delta^0 > 0$, $\gamma \in (0, 1)$, $c \ge 1$, a forcing function $\sigma(t)$ such that $\lim_{t\to 0^+} \sigma(t)/t = 0$ and a sequence of directions $\{d^k\}$ satisfying Assumption 3.

Step 0: Set $k = 0$.

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

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

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

$$f(y^{k+1}, z^k) \le f(\tilde{y}^{k+1}, z^k);$$

$$\|y^{k+1} - y^k\| \le c \max\left\{\Delta^k,\ f(y^k, z^k) - f(y^{k+1}, z^k)\right\}.$$

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

(a) Compute β^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

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

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

Remark 5 We note that the conditions at Steps 1(b) and 2(b) can be trivially satisfied by choosing

$$y^{k+1} = \tilde{y}^{k+1},$$
$$z^{k+1} = \tilde{z}^{k+1},$$

and by considering that $\|\tilde{y}^{k+1} - y^k\| \le \Delta^k$ and $c \ge 1$.

It is worth noting that Steps 1(b) and 2(b) allow the user to choose any point that yields as much decrease in the objective function as that obtained by the derivative-free and, respectively, the derivative-based iterations. Hence, the points produced by the two iteration maps can be regarded as reference points that impose minimal descent conditions which are necessary to guarantee convergence of the overall method.

However, particular care must be taken when choosing the new point y^{k+1} so has to guarantee that the distance $\|y^{k+1} - y^k\|$ converges to zero and that the choice $y^{k+1} = \tilde{y}^{k+1}$ is acceptable. In this respect, the second condition at Step 1(b) has a twofold significance. On the one hand it implies that the distance $\|y^{k+1} - y^k\|$ is bounded to the maximum between the previous stepsize and the difference $f(y^k, z^k) - f(y^{k+1}, z^k)$, which allows to prove that $\|y^{k+1} - y^k\| \to 0$ and, ultimately, that $\{y^{k+1}\}$ and $\{y^k\}$ have the same limit point $\bar{y}$. On the other hand, it guarantees that the choice $y^{k+1} = \tilde{y}^{k+1}$ is acceptable. Finally, we stress that the second condition at Step 1(b) of the algorithm is not particularly strong both for the presence of the max function and of the multiplicative constant c which can be chosen so that $c \gg 1$.

Now we state the main convergence theorem of the paper.

Theorem 6 *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

$$f(x^{k+1}) \leq f(y^{k+1}, \tilde{z}^{k+1}) \leq f(y^{k+1}, z^k) \leq f(\tilde{y}^{k+1}, z^k) \leq f(x^k), \tag{10}$$

so that point (i) is proved. By virtue of point (i) and relation (10), we can conclude that the non-increasing sequence $\{f(x^k)\}$ is bounded from below and has a limit. Therefore, recalling (10), we have that

$$\lim_{k\to\infty} f(y^k, z^k) - f(\tilde{y}^{k+1}, z^k) = 0, \tag{11a}$$

$$\lim_{k\to\infty} f(y^k, z^k) - f(y^{k+1}, z^k) = 0, \tag{11b}$$

$$\lim_{k\to\infty} f(y^{k+1}, z^k) - f(y^{k+1}, \tilde{z}^{k+1}) = 0. \tag{11c}$$

By (11a) and Proposition 2, where we identify w^{k+1} with $\tilde{y}^{k+1}$, we know that

$$\lim_{k\to\infty} \Delta^k = 0. \tag{12}$$

Let us define

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

and note that, in order for (12) to hold true, H must have infinitely many elements. In fact, if this was not the case, a constant $M > 0$ and an index $\bar{k}$ would exist such that $\Delta^k \geq M$, for all $k \geq \bar{k}$, thus contradicting (12).

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 $\{x^k\}$, that is

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

with $K \subseteq H$. Thus, by (11a) and Proposition 2, where we identify w^{k+1} with $\tilde{y}^{k+1}$, we can conclude that

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

Now, we show that

$$\lim_{k\to\infty} \|y^{k+1} - y^k\| = 0. \tag{14}$$

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

$$\|y^{k+1} - y^k\| \leq c \max\left\{\Delta^k, f(y^k, z^k) - f(y^{k+1}, z^k)\right\}.$$

Taking the limit for $k \to \infty$ in the above relation and considering (11b) and (12), we obtain that (14) holds. Thus, $\lim_{k\to\infty} y^{k+1} = \bar{y}$ and, in particular, for the subset of indices K, we can write

$$\lim_{k\to\infty, k\in K} y^{k+1} = \bar{y}. \tag{15}$$

Hence, sequences $\{x^k\}_K$ and $\{(y^{k+1}, z^k)\}_K$ have the same limit $\bar{x}$.

Finally, by (11c), Proposition 4 (where we identify $\{x^k\}$ with $\{(y^{k+1}, z^k)\}_K$) and (15), we have

$$\nabla_z f(\bar{x}) = 0 \tag{16}$$

which, along with (13), concludes the proof. $\square$

6 Concluding remarks

In the paper we considered the problem of minimizing a 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 Sect. 3), and then updates the z variables according to a derivative-based Armijo-type iteration map (see Sect. 4).

The order in which the two block of variables are considered in Algorithm 1 is not important to prove convergence of the method. However, if we were to change the order of Steps 1 and 2, we would have to guarantee that the new point z^{k+1} is chosen in such a way that

$$\begin{aligned} &f(y^k, z^{k+1}) \le f(y^k, \tilde{z}^{k+1}); \\ &\|z^{k+1} - z^k\| \le c \max\left\{\beta^k \|d^k\|, f(y^k, z^k) - f(y^k, z^{k+1})\right\}, \end{aligned} \tag{17}$$

whereas, y^{k+1} is such that

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

We note that condition (17), which would replace condition at Step 1(b) of Algorithm 1, is basically needed to guarantee that $\|z^{k+1} - z^k\| \to 0$, which can be shown by analogous reasoning as in the proof of Theorem 6 and by considering that, by Proposition 4, $\beta^k \|d^k\| \to 0$.

Moreover, 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 2 and 4. These results are essential to prove convergence of the overall method to stationary points of Problem 1. For instance, in place of the derivative-free iteration map proposed in the paper, a derivative-free linesearch mapping (see [8]) to update the y variables could be used. Actually, what is really important to prove convergence is that a subsequence of iterates exists where the stepsize of the iteration map DF converges to zero. Finally, as concerns the derivative-based mapping, it could be substituted by any safe-guarded linesearch procedure that also allows extrapolation steps (see for instance [5]).

Acknowledgments The authors are indebted and wish to thank Prof. Marco Sciandrone for stimulating discussions on the topic of the paper and for reading earlier versions of the paper. We are also thankful to two anonymous Referees whose many useful suggestions and comments helped us to significantly improve the paper.

Appendix A: Technical results

Proof of Proposition 2 First we prove (5). To this aim we can split the iteration sequence $\{k\}$ into two parts, H and $\bar{H}$, namely,

$$\begin{aligned} H &= \{k : \Delta^{k+1} = \theta \Delta^k\}, \\ \bar{H} &= \{k : \Delta^{k+1} = \phi \Delta^k\}. \end{aligned}$$

In particular, if $k \in \bar{H}$, we have

$$f(w^{k+1}, z^k) \leq f(x^k) - \sigma(\Delta^k); \tag{18}$$

if $k \in H$, we have

$$f(y^k + \Delta^k p_i, z^k) > f(x^k) - \sigma(\Delta^k), \quad \text{for all } p_i \in D. \tag{19}$$

If $\bar{H}$ is an infinite subset, by (18), recalling that $\lim_{k\to\infty}\left[f(x^k) - f(w^{k+1}, z^k)\right] = 0$, the properties of the forcing function $\sigma(t)$, we obtain

$$\lim_{k\to\infty, k\in\bar{H}} \Delta^k = 0. \tag{20}$$

Now, for each $k \in H$ let m_k be the biggest index such that $m_k < k$ and $m_k \in \bar{H}$ (we assume $m_k = 0$ if $\bar{H}$ is empty). Then, we have

$$\Delta^{k+1} = \theta^{k-m_k} \Delta^{m_k}. \tag{21}$$

As $k \to \infty$ and $k \in H$, either $m_k \to \infty$ (if $\bar{H}$ is an infinite subset) or $k - m_k \to \infty$ (if $\bar{H}$ is finite). Therefore, (21) together with (20) or the fact that $\theta \in (0, 1)$ yields

$$\lim_{k\to\infty, k\in H} \Delta^k = 0. \tag{22}$$

Thus, by using (20) and (22), we can write

$$\lim_{k\to\infty} \Delta^k = 0. \tag{23}$$

Now we prove (6). Note that, in order for (5) to hold true, H must have infinitely many elements. In fact, if this was not the case, a constant $M > 0$ and an index $\bar{k}$ would exist such that $\Delta^k \geq M$, for all $k \geq \bar{k}$, thus contradicting (5). For every index $k \in H$, it results

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

which, in turn, implies that

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

Taking the limit for $k \to \infty$, $k \in H$, in the above formula, recalling (5), the property of the forcing function $\sigma(t)$ and that $\lim_{t\to 0^+} \sigma(t)/t = 0$, 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 $\nabla_y f(\bar{x}) = 0$. $\square$

Proof of Proposition 4 In order to prove point (i), let us assume, by contradiction, that $\nabla_z f(x^k) \neq 0$ and that condition (9) is violated for every $j \geq 0$, so that

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

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

$$\gamma \geq 1$$

thus contradicting the assumption $\gamma \in (0, 1)$.

Now we prove point (ii). Since β^k satisfies condition (9) and d^k, by definition, is such that

$$\frac{|\nabla_z f(x^k)^\top d^k|}{\|d^k\|} \geq \frac{r}{R^2}\|d^k\|, \tag{24}$$

we can write

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

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

$$\lim_{k\to\infty} \beta^k |\nabla_z f(x^k)^\top d^k| = 0, \quad \lim_{k\to\infty} \beta^k \|d^k\|^2 = 0. \tag{25}$$

Let us first show that $\lim_{k\to\infty} \beta^k \|d^k\| = 0$. We proceed by contradiction and assume that a subset $\bar{K} \subseteq \{1, 2, \ldots\}$ exists such that

$$\lim_{k\to\infty, k\in\bar{K}} \beta^k \|d^k\| = \eta > 0. \tag{26}$$

This and (25) imply that

$$\lim_{k\to\infty, k\in\bar{K}} \|d^k\| = 0, \tag{27}$$

which, again by (26), would imply that $\lim_{k\to\infty, k\in\bar{K}} \beta^k = +\infty$, contradicting the fact that, by the definition of Iteration Map LS, $\beta^k \le \Lambda$. 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 $\bar{K} \subseteq \{1, 2, \ldots\}$ exists such that

$$\lim_{k\to\infty, k\in\bar{K}} |\nabla_z f(x^k)^\top d^k| = \eta > 0. \tag{28}$$

This and (25) imply that

$$\lim_{k\to\infty, k\in\bar{K}} \beta^k = 0. \tag{29}$$

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

$$f\left(y^k, z^k + \frac{\beta^k}{\delta} d^k\right) - f(x^k) > \gamma \frac{\beta^k}{\delta} \nabla_z f(x^k)^\top d^k. \tag{30}$$

By the mean value theorem, we have that

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

where $\xi^k = z^k + \theta^k \frac{\beta^k}{\delta} d^k$ and $\theta^k \in (0, 1)$. Substituting the above equation into (30), we obtain

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

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

$$\eta \leq \gamma\eta$$

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

$$\lim_{k\to\infty} |\nabla_z f(x^k)^\top d^k| = 0. \tag{31}$$

Now, by Assumption 3 on the search direction d^k, we know that, for all k,

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

Taking the limit in the above relation, considering (31) and recalling that $\{x^k\}$ converges to $\bar{x}$, we obtain

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

thus concluding the proof. □

References

1. Arroyo, S.F., Cramer, E.J., Dennis, J.E. Jr., Frank, P.D.: Comparing problem formulations for coupled sets of components. Optim. Eng. **10**, 557–573 (2009)
2. Bertsekas, D.P.: Nonlinear programming. Athena Scientific, Massachusetts (1999)
3. Dennis, J.E. Jr., Schnabel, R.B.: Numerical methods for unconstrained optimization and nonlinear equations volume 16 of classics in applied mathematics. SIAM, Philadelphia (1996)
4. Facchinei, F., Kanzow, C.: Penalty methods for the solution of generalized nash equilibrium problems. SIAM J. Optim. **20**, 2228–2253 (2010)
5. Grippo, L., Sciandrone, M.: Globally convergent block-coordinate techniques for unconstrained optimization. Optim. Methods Softw. **10**, 587–637 (1999)
6. Haftka, R.T., Sobieszczanski-Sobieski, J.: Multidisciplinary aerospace design optimization: survey of recent developments. Struct. Multidiscip. Optim. **14**, 1–23 (1997)
7. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: new perspectives on some classical and modern methods. SIAM Rev. **45**, 385–482 (2003)
8. Lucidi, S., Sciandrone, M.: On the global convergence of derivative free methods for unconstrained optimization. SIAM J. Optim. **13**, 97–116 (2002)
9. Nocedal, J., Wright, S.J.: Numerical optimization. Series in operations research. Springer, Berlin (1999)
10. Ortega, J.M., Rheinboldt, W.C.: Iterative solution of nonlinear equations in several variables. Academic Press, New York (1970)
11. Pedersen, A.R.: A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Stat. **22**, 55–71 (1995)
12. Pedersen, A.R.: Likelihood inference by monte carlo methods for incompletely discretely observed diffusion processes. Technical report, Department of Biostatistics, University of Aarhus, Denmark (2001)
13. Polak, E.: Optimization, algorithms and consistent approximations. Series in applied mathematical sciences. Springer, Berlin (1997)
14. Scheinberg, K., Conn, A.R., Vicente, L.N.: Introduction to derivative-free optimization. Series in optimization. MPS-SIAM, Philadelphia (2009)
15. Torczon, V.: On the convergence of pattern search algorithms. SIAM J. Optim. **7**, 1–25 (1997)