Noname manuscript No.
(will be inserted by the editor)

A global hybrid derivative-free method for high-dimensional systems of nonlinear equations

Rodolfo G. Begiato · Ana L. Custódio · Márcia Ap. Gomes-Ruggiero

the date of receipt and acceptance should be inserted later

Abstract This work concerns the numerical solution of high-dimensional systems of nonlinear equations, when derivatives are not available for use, but assuming that all functions defining the problem are continuously differentiable. A hybrid approach is taken, based on a derivative-free iterative method, organized in two phases.

The first phase is defined by derivative-free versions of a fixed-point method that employs spectral parameters to define the steplength along the residual direction. The second phase consists on a matrix-free inexact Newton method that employs the Generalized Minimal Residual algorithm to solve the linear system that computes the search direction. This second phase will only take place if the first one fails to find a better point after a predefined number of reductions in the step size. In all stages, the criterion to accept a new point considers a nonmonotone decrease condition upon a merit function.

Convergence results are established and the numerical performance is assessed through experiments in a set of problems collected from the literature. Both the theoretical and the experimental analysis support the feasibility of the proposed hybrid strategy.

R. G. Begiato
DAMAT, UTFPR, 80230-901, Curitiba, PR, Brazil, E-mail: begiato@utfpr.edu.br

A. L. Custódio
Department of Mathematics, FCT-UNL-CMA, Campus de Caparica, 2829-516 Caparica, Portugal E-mail: alcustodio@fct.unl.pt
Support for this author was provided by Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) under the projects PTDC/MAT-APL/28400/2017 and UID/MAT/00297/2019.

M. A. Gomes-Ruggiero
Departamento de Matemática Aplicada, IMECC, UNICAMP, 13083-970, Campinas, SP, Brazil E-mail: marcia@ime.unicamp.br
Partially supported by PRONEX-Optimization; CNPq (Grants 306220/2009-1 and 304032/2010-7); FAPESP (Grants 2013/05475-7 and 2013/07375-0) and CAPES.

Keywords Nonlinear systems of equations · derivative-free optimization methods · spectral residual · inexact Newton · nonmonotone line search

1 Introduction

In this work, we propose a method for solving the system of nonlinear equations

$$\begin{aligned} &F(x) = 0, \\ &F : \mathbb{R}^n \to \mathbb{R}^n, \end{aligned} \tag{1}$$

where F is a continuously differentiable function, but derivatives are not available for use, neither could be approximated by numerical techniques. An analytical expression does not need to be available for F, which could be evaluated by numerical simulation. Thus, the algorithm proposed is derivative-free, only requiring zero-order information regarding function evaluation, although differentiability assumptions are considered when deriving theoretical results.

To solve problem (1), we use an iterative method, where each iterate takes the form

$$x_{k+1} = x_k + \lambda_k d_k, \tag{2}$$

with d_k a search direction and $\lambda_k \in (0, 1]$ a step length.

When derivatives are available for use, Newton method is a classical approach for solving (1). In this case, the search direction d_k is computed as the solution of the linear system:

$$J(x_k)d = -F(x_k), \tag{3}$$

where $J(x_k)$ denotes the Jacobian matrix of F at x_k.

Nevertheless, in the presence of a large number of variables, computing an exact solution of (3) could be unpractical, which motivates the use of inexact Newton methods [6]. In this case, the search direction should satisfy:

$$\|J(x_k)d_k + F(x_k)\| \leq \eta_k \|F(x_k)\|, \tag{4}$$

where $\eta_k \in [0, 1)$ is called a forcing term. Krylov methods [13] are a classical approach to compute d_k satisfying (4), allowing derivative-free versions [12, 13].

Fixed-point iterations have also been considered for solving (1) [5,4,11]. In this case, a multiple of the residual vector is used as search direction, $d_k = \alpha_k F(x_k)$, $\alpha_k \in \mathbb{R}$, avoiding derivative calculations and the use of matrices.

Under suitable assumptions, theoretical results can be derived establishing local convergence of the previous methods [13]. For global convergence, a merit function needs to be considered, traducing the solution of the nonlinear system into a minimization problem, and a globalization procedure needs to be adopted.

Merit functions are usually defined as $f : \mathbb{R}^n \to \mathbb{R}_0^+$, with $f(x) = \|F(x)\|_2$ or $f(x) = \|F(x)\|_2^2$. Regarding the globalization procedure, typical approaches lie on a line search with an Armijo type condition [1] to accept new points.

However, in some cases, this requirement can lead to a large number of function evaluations. More flexible criteria can be adopted, overcoming this difficulty without jeopardizing the global convergence properties. Grippo, Lampariello and Lucidi [9] proved the convergence of inexact Newton methods under a nonmonotone acceptance criterion. In Li and Fukushima [14] and La Cruz, Martínez and Raydan [4], nonmonotone acceptance criteria which do not require derivatives were proposed.

We consider a hybrid two-step approach to solve high-dimensional systems of nonlinear equations:

1. `step 1`: use of a fixed point method, where the computation of the step length for the residual direction is based on a *spectral approach* [2,5];
2. `step 2`: use of an inexact Newton method, where a matrix-free version of GMRES [16] is used to solve the inner linear system.

The two steps are applied sequentially. However, the second step is only applied in case of failure of the previous one. We will consider $\|\cdot\| \equiv \|\cdot\|_2$ in the definition of the merit function and use a nonmonotone globalization strategy based on La Cruz, Martínez and Raydan [4].

The paper is organized as follows. In Section 2, we start by recalling the structure and the basic properties of a spectral residual method and a FDGM-RES iteration, motivating the hybrid two-phase algorithm. The algorithmic structure of the new method is formalized in Section 3 and the convergence is established in Section 4. Numerical experiments on some test problems are reported in Section 5, comparing the hybrid approach with pure methods. Finally, Section 6 is dedicated to some concluding remarks.

2 Derivative-free methods and line search techniques

The Spectral Residual Method (SANE), introduced by La Cruz and Raydan [5], uses as search directions

$$d_k = (1/\alpha_k)F(x_k) \text{ and } d_k = -(1/\alpha_k)F(x_k), \tag{5}$$

with α_k a *spectral scaling parameter*.

This is a fixed point method where the parameter α_k is computed by adjusting the Barzilai-Borwein [2] procedure for computing step sizes to the solution of systems of nonlinear equations. The vector $d_k = -(1/\alpha_k)F(x_k)$ is not necessarily a descent direction for the merit function, what justifies the systematic evaluation of both directions (5).

Being a quasi-Newton method, the computation of the spectral parameter requires that a Jacobian approximation $B_k = \alpha_k I$ satisfies the secant equation $B_k(x_k - x_{k-1}) = F(x_k) - F(x_{k-1})$. Thus:

$$\alpha_k = \frac{s_k^\top y_k}{s_k^\top s_k} = \frac{(x_k - x_{k-1})^\top (F(x_k) - F(x_{k-1}))}{(x_k - x_{k-1})^\top (x_k - x_{k-1})}. \tag{6}$$

In practical applications, scaling gradients or residual directions using spectral parameters conducted to good algorithmic performances, although the corresponding theoretical justification is not completely understood. Thus, flexible globalization strategies are required, allowing the algorithms to accept the use of the spectral parameter as a step size, without any attempt of backtracking. Rigid strategies to accept a new point, for instance by imposing an Armijo-type condition, can conduct to a performance similar to the one of the steepest descent method.

In [5] a nonmonotone line search was considered to guarantee global convergence of the method. The following condition:

$$f(x_k + \lambda_k d_k) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \gamma \lambda_k \nabla f(x_k)^\top d_k, \tag{7}$$

where f is a merit function, $\gamma \in (0,1)$ and $M \in \mathbb{N}$, proposed by Grippo, Lampariello and Lucidi [9], was used as acceptance criterion for a new point, generating a sequence $\{x_k\}$ such that $\{f(x_k)\}$ is not necessarily decreasing.

La Cruz, Martínez and Raydan [4] proposed a derivative-free version of SANE. The new algorithm, named DFSANE, preserves the use of the residual direction and the spectral stepsize, but introduces a new acceptance criterion for new points, which does not use derivatives. This criteria blends (7) with the strategy proposed by Li and Fukushima [14].

In [14], a new point is accepted if

$$\|F(x_k + \lambda_k d_k)\| \leq (1 + \zeta_k)\|F(x_k)\| - \gamma \lambda_k^2 \|d_k\|^2, \tag{8}$$

with $\zeta_k > 0$ for all k, $\sum_k \zeta_k = \zeta < \infty$ and $\gamma \in (0,1)$. The approach followed by La Cruz, Martínez and Raydan [4] accepts new points that satisfy

$$f(x_k + \lambda_k d_k) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \zeta_k - \gamma \lambda_k^2 f(x_k) \tag{9}$$

where $f : \mathbb{R}^n \to \mathbb{R}_0^+$ is a merit function, $M \in \mathbb{N}$, $\gamma \in (0,1)$, $\zeta_k > 0$ for all $k \in \mathbb{N}$ and $\sum_k \zeta_k = \zeta < \infty$.

Algorithm 1 details an iteration of DFSANE.

Algorithm 1 DFSANE

Input parameters: $x_k \in \mathbb{R}^n$; $0 < \tau_{\min} < \tau_{\max} < 1$; $NBL_{\max}, M \in \mathbb{N}$; $\gamma \in (0,1)$, $\zeta_k > 0$ and $0 < \alpha_{\min} < \alpha_{\max}$.

1. Choose α_k such that $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$. Set $d = -(1/\alpha_k)F(x_k)$, $\lambda_+ = \lambda_- = 1$ and $NBL = 0$.
2. If $NBL = NBL_{\max}$, set $flag = 0$ and terminate.
3. If $f(x_k + \lambda_+ d) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \zeta_k - \gamma \lambda_+^2 f(x_k)$, define $d_k = d$, $\lambda_k = \lambda_+$, $flag = 1$ and terminate.
4. if $f(x_k - \lambda_- d) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \zeta_k - \gamma \lambda_-^2 f(x_k)$, define $d_k = -d$, $\lambda_k = \lambda_-$, $flag = 1$ and terminate.

5. Choose $\lambda_+ \in [\tau_{\min}\lambda_+, \tau_{\max}\lambda_+]$ and $\lambda_- \in [\tau_{\min}\lambda_-, \tau_{\max}\lambda_-]$, set $NBL = NBL + 1$ and go to Step 2.

Each iteration explores both directions (5), in a nonmonotone line search framework, using a backtracking strategy, until condition (9) is satisfied by one of the directions. A maximum number (NBL_{max}) of backtracks is allowed. The logical variable $flag$ does not play any active role in this algorithmic description. It is only defined to facilitate the presentation of the proposed hybrid approach, in Section 3. In fact, variable $flag$ set equal to 1 means that the nonmonotone line search procedure was successful.

Convergence results were established by assuming continuity of the partial derivatives of $F : \mathbb{R}^n \to \mathbb{R}^n$ and considering $f : \mathbb{R}^n \to \mathbb{R}_0^+$ with $f(x) = \|F(x)\|$ or $f(x) = \|F(x)\|^2$ as merit function [4].

Grippo and Sciandrone [12] proposed two different approaches to address the solution of a system of nonlinear equations. The first is an inexact Newton method, combining a nonmonotone watchdog phase [3] with a nonmonotone line search.

To compute d_k satisfying (4), a *matrix free* version of the classical Newton-GMRES [13] method was employed. In the classical GMRES method, the Jacobian matrix is required to solve the linear system (3). However, this matrix is only used in matrix-vector products. In the derivative-free case, these products are approximated by:

$$J(x_k)w \approx \frac{F(x_k + \sigma w) - F(x_k)}{\sigma}, \quad \sigma \in \mathbb{R} \setminus \{0\},\ w \in \mathbb{R}^n. \tag{10}$$

The new matrix free method is named FDGMRES and the corresponding inexact Newton method is known as Newton-FDGMRES [13].

In case of failure of the inexact Newton method, a second approach [12] adds a coordinate search phase to the first algorithm. This change allowed to weaken the assumptions required for establishing convergence and, in practical tests, improved the numerical robustness of the method. The coordinate search step consists in evaluating the objective function in the set $\{x_k \pm \lambda_k e_i : i = 1, 2, \ldots, n\}$, where e_i denotes the i-th column of the identity matrix.

As in [4], a nonmonotone line search is used as globalization strategy, but with different conditions associated to each type of steps. In the watchdog step the acceptance condition is

$$f(x_{k+1}) \leq \rho \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}),$$

whereas in the line search step it is considered

$$f(x_k + \lambda_k d_k) \leq (1 - \gamma\lambda_k) \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}), \tag{11}$$

with ρ, $\gamma \in (0, 1)$. Finally, in the coordinate search step it is used the condition

$$f(x_k + \lambda_k d_k) \leq (1 - \gamma\lambda_k^2) \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}).$$

3 A two-phase hybrid algorithm for systems of nonlinear equations

The practical numerical behavior of the spectral residual method indicates that it rarely succeeds in finding a better point when a high number of reductions in the stepsize needs to be performed. Therefore, we propose to impose a maximum limit to this number of reductions, replacing the spectral direction with a new search direction when this limit is reached. In the case, as in [12], we will use the inexact Newton direction. To accept new points, we consider the nonmonotone condition proposed by La Cruz, Martínez and Raydan (9), since it is more flexible than condition (11), proposed by Grippo and Sciandrone [12]. In fact,

$$(1-\gamma\lambda_k)\max_{0\leq j\leq \min\{k,M-1\}} f(x_{k-j}) < \max_{0\leq j\leq \min\{k,M-1\}} f(x_{k-j}) + \zeta_k - \gamma\lambda_k^2 f(x_k),$$

since $\lambda_k \in (0,1]$ and $\zeta_k > 0$.

Algorithm 2 describes the procedure adopted when using the inexact Newton direction.

Algorithm 2 Inexact Newton

Input parameters: $x_k \in \mathbb{R}^n$; $\mu, \gamma, \sigma, \theta_1, \theta_2, \theta_3 \in (0,1)$; $a \in (0,1]$; $\eta_k \in (0,1)$; $\zeta_k > 0$ and $0 < \xi_{\min} < \xi_{\max} < 1$.

1. Set $t = 0$ and $\eta = \eta_k$.
2. Compute d_k satisfying (4), using σ in the matrix-vector products (10). Set $\lambda = 1$.
3. If $f(x_k+\lambda d_k) \leq \max_{0\leq j\leq \min\{k,M-1\}} f(x_{k-j})+\zeta_k-\gamma\lambda^2 f(x_k)$, go to Step 8.
4. Set $\alpha = a, t = t+1$ and $i = 0$.
5. While $f(x_k+\alpha d_k) > \max_{0\leq j\leq \min\{k,M-1\}} f(x_{k-j}) + \zeta_k - \gamma\alpha^2 f(x_k)$, do:
   (a) If $\alpha < \mu a$, then $\lambda = 0$ and go to Step 7.
   (b) Choose $\xi \in [\xi_{\min}, \xi_{\max}]$ and set $\alpha = \alpha\,\xi$, $i = i+1$.
6. Set $\lambda = \alpha$ and go to Step 8.
7. Set $\sigma = \theta_1\sigma$, $\eta = \theta_2\eta$, $\eta_k = \eta$, $\mu = \theta_3\mu$ and go to Step 2.
8. Set $\tilde{\sigma}_k = \sigma$, $\tilde{\eta}_k = \eta$ and $\lambda_k = \lambda$.

As in [13], the FDGMRES method is used to define d_k in Step 2. The algorithm then tries to define an adequate step that allows to satisfy the nonmonotone condition (9). For that, again a backtracking strategy is adopted, allowing a minimum value for the stepsize (namely μa). When this value is reached, the algorithm reduces σ and η, which will be used for recomputing a more precise inexact Newton direction.

In a practical implementation of the algorithm, counters t and i, and sequences $\{\tilde{\sigma}_k\}, \{\tilde{\eta}_k\}$ are not required. They are only defined to facilitate the presentation of the convergence analysis.

To establish that Algorithm 2 is well defined, first we need to show that a direction satisfying (4) can be computed at Step 2. Moreover, we need to ensure that the line search ends with a nonzero value for λ, meaning that Algorithm 2 does not cycle between Step 2 and Step 7. Both results depend on the following assumption:

Hypothesis 1 *Function J is Lipschitz continuous on a convex set $\Omega \subseteq \mathbb{R}^n$, with Lipschitz constant $L_J > 0$. Function J is also nonsingular on Ω and satisfies $\|J^{-1}(y)\| \leq m_J$ for all $y \in \Omega$, with $m_J > 0$.*

Proposition 1, stated in [12], follows from Proposition 6.2.1 in [13]. It ensures that it is possible to compute a direction satisfying the inexact Newton condition (4) using the FDGMRES method proposed in [13], which is a GMRES matrix free method.

Proposition 1 *Let $x_k \in \mathbb{R}^n$ be a point such that $F(x_k) \neq 0$. Assume that F satisfies Hypothesis 1 for a convex set Ω_k, such that $x_k \in \Omega_k$ with $L_J = L_k$ and $m_J = c_k$. Let*

$$\hat{\sigma}_k = \frac{1}{2n^{1/2}L_k c_k} \tag{12}$$

and

$$C_k = 4n^{1/2}L_k c_k. \tag{13}$$

Then, for each $\sigma \in (0, \hat{\sigma}_k]$ and for each $\eta_k \in (0,1)$, procedure FDGMRES determines a direction d_k satisfying

$$\|J(x_k)d_k + F(x_k)\| \leq (\eta_k + C_k\sigma)\|F(x_k)\|. \tag{14}$$

By adjusting Lema 8.2.1 in [13], Grippo and Sciandrone [12] established Proposition 2, which guarantees that Hypothesis 1 is sufficient to ensure that Algorithm 2 does not cycle between Step 2 and Step 7.

Proposition 2 *Let $x \in \mathbb{R}^n$ be a point such that $F(x) \neq 0$ and satisfies Hypothesis 1 for some set $\Omega = \{y \in \mathbb{R}^n \mid \|x - y\| \leq r\}$, with $r > 0$. Let $d \in \mathbb{R}^n$ be a vector satisfying the Inexact Newton Condition:*

$$\|J(x)d + F(x)\| \leq \eta\|F(x)\| \tag{15}$$

with $0 \leq \eta \leq \bar{\eta} < (1 - \gamma)$ and $\gamma \in (0,1)$. Then, we have

$$\|F(x + \lambda d)\| \leq (1 - \gamma\lambda)\|F(x)\| \tag{16}$$

with $\lambda \in [0, \bar{\lambda}(x)]$, where

$$\bar{\lambda}(x) = \min\left(\frac{r}{m_J(1+\bar{\eta})\|F(x)\|}, \frac{2(1-\gamma-\bar{\eta})}{(1+\bar{\eta})^2 m_J^2 L_J \|F(x)\|}\right). \tag{17}$$

Proposition 2 holds a result similar to condition (9), used for accepting new points. In fact, using condition (16), since we consider the merit function $f(x) = \|F(x)\|^2$, and since λ is in the interval $[0,1]$, we have:

$$\begin{aligned} \|F(x_k + \lambda d)\|^2 &\leq (1-\lambda\gamma)^2\|F(x_k)\|^2 \leq (1-\lambda\gamma)\|F(x_k)\|^2 \\ &\leq \|F(x_k)\|^2 - \lambda^2\gamma\|F(x_k)\|^2 \\ &< \max_{0\leq j\leq \min\{k,M-1\}} \|F(x_{k-j})\|^2 + \zeta_k - \gamma\lambda^2\|F(x_k)\|^2. \end{aligned} \tag{18}$$

Proposition 3 adapts Proposition 3.1 in [12], allowing to ensure that the backtracking scheme, defined in Step 5 of Algorithm 2 is well defined.

Proposition 3 *Let $\mu \in (0,1)$ and $a \in (0,1]$ be fixed. Step 5 (Line Search) of Algorithm 2 determines, in a finite number of iterations, a scalar $\lambda \in [0,a]$ such that:*

$$f(x_k + \lambda d_k) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \zeta_k - \gamma \lambda^2 f(x_k). \tag{19}$$

Additionally, at least one of the following conditions holds:

1. $\lambda = 0$ *and*

$$\left\|F(x_k + \delta d_k)\right\|^2 > \max_{0 \leq j \leq \min\{k, M-1\}} \|F(x_{k-j})\|^2 + \zeta_k - \gamma \delta^2 \left\|F(x_k)\right\|^2 \geq \tag{20}$$

$$\geq (1 - \gamma \delta^2) \left\|F(x_k)\right\|^2 + \zeta_k, \text{ with } \delta < \mu a;$$

   or
2. $\lambda \geq \xi_{\min} \mu a$.

Proof Since λ is initially set equal to a and reduced, in each iteration of Step 5, by a factor $\xi \leq \xi_{\max} < 1$, Step 5 of Algorithm 2 ends on Step 5a with $\lambda = 0$ (thus trivially satisfying condition (19)) or it finds a nonzero value λ such that condition (lmr2) holds.

If Step 5 of Algorithm 2 ends on Step 5a we have $\lambda = 0$ and condition (20) is satisfied. Else, either the initial stepsize is accepted (in this case $\lambda = a$) or a new stepsize λ is computed such that $\lambda/\xi \geq \mu a$. In both cases, $\lambda \geq \xi_{\min} \mu a$.

Algorithm 3 **H2P** corresponds to the proposed two-phase hybrid procedure. Figure 1 presents a schematic description of it.

Algorithm 3 H2P Input parameters: $x_0 \in \mathbb{R}^n$; $NBL_{\max}, M \in \mathbb{N}$; $\gamma, \mu, \sigma, \theta_1, \theta_2, \theta_3 \in (0,1)$; $0 < \tau_{\min} < \tau_{\max} < 1$; $0 < \xi_{\min} < \xi_{\max} < 1$; $0 < \alpha_{\min} < \alpha_{\max}$; $\{\zeta_k\}$; $\{\eta_k\}$ and $a \in (0,1]$.

1. Set $k = 0$.
2. Compute $d_k, \lambda_k, flag$ using Algorithm 1 (DFSANE).
3. If $flag = 0$, compute d_k, λ_k using Algorithm 2 (Inexact Newton).
4. Set $x_{k+1} = x_k + \lambda_k d_k$ and $k = k + 1$.
5. If $F(x_{k+1}) = 0$, terminate. Else go to Step 2.

4 Convergence analysis of the hybrid two-phase method

In this section, we analyze the convergence of Algorithm 3. Similarly to the approach of [12], we establish the convergence of a subsequence generated by the algorithm to a critical point of the merit function or the existence of a critical point of the merit function generated by the algorithm. For that, we define $W_k = \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j})$ and the sequence $\{\nu(k)\}$ such that $f(x_{\nu(k)}) = W_k$.

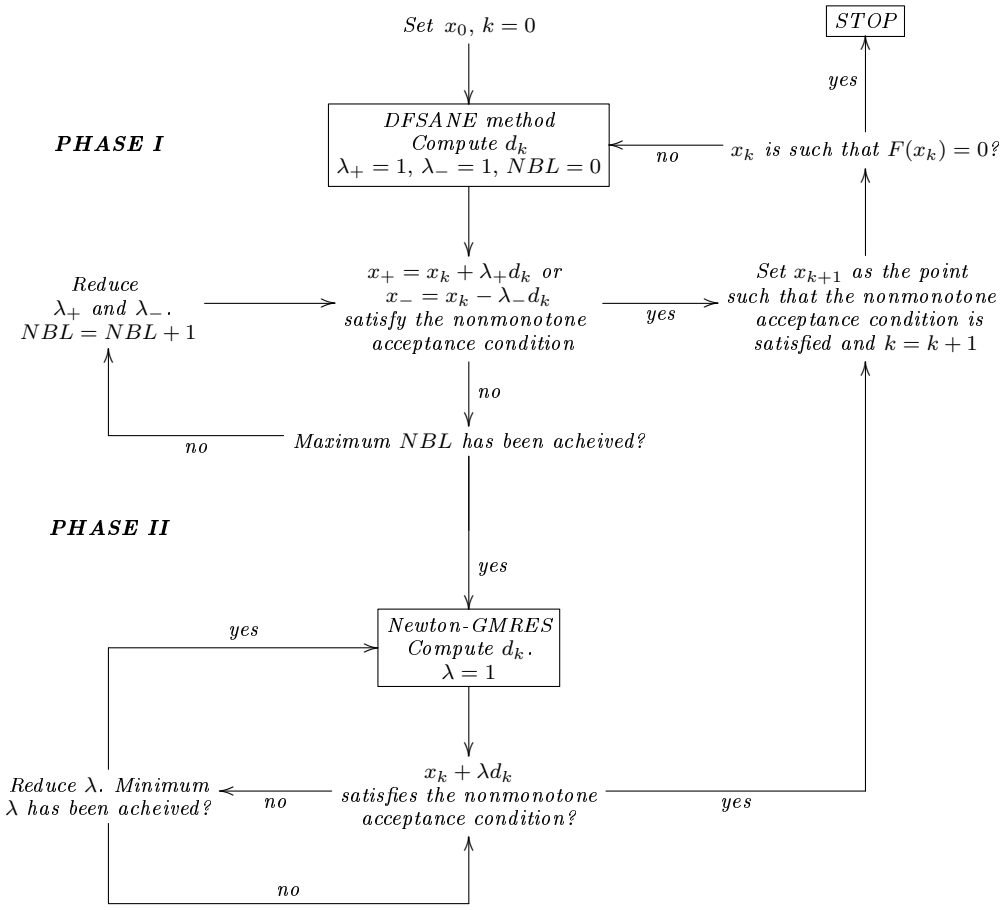

Fig. 1 Flow diagram corresponding to Algorithm 3 (`H2P`).

Proposition 4 *Let $\{x_k\} \subset \mathbb{R}^n$ be a sequence such that*

$$f(x_{k+1}) \leq W_k + \zeta_k, \tag{21}$$

with $\zeta_k > 0$ for all k and $\sum_{i=0}^{\infty} \zeta_i = \zeta < \infty$.

1. *Then for all k, $x_k \in \bar{\mathcal{L}}_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0) + \zeta\}$.*
2. *Moreover, assuming the existence of $\bar{k} \in \mathbb{N}$ such that*

$$f(x_{k+1}) \leq W_k \text{ for all } k \geq \bar{k}, \tag{22}$$

the sequence $\{W_k\}_{k>\bar{k}}$ is monotonically non increasing.

Proof Since $\min\{k+1, M-1\} \leq \min\{k, M-1\}+1$, by definition of $\nu(k)$, we have:

$$\begin{aligned} W_{k+1} = f(x_{\nu(k+1)}) &= \max_{0\leq j\leq \min\{k+1,M-1\}} f(x_{k+1-j}) \\ &\leq \max_{0\leq j\leq \min\{k,M-1\}+1} f(x_{k+1-j}) \\ &= \max\{\max_{1\leq j\leq \min\{k,M-1\}+1} f(x_{k+1-j}), f(x_{k+1})\} \\ &= \max\{f(x_{\nu(k)}), f(x_{k+1})\}. \end{aligned} \tag{23}$$

By hypothesis, $f(x_{k+1}) \leq W_k + \zeta_k$ and $f(x_{\nu(k)}) = W_k \leq W_k + \zeta_k$. So, we conclude that $W_{k+1} \leq W_k + \zeta_k$ for all k, and an inductive argument allow us to state:

$$W_{k+1} = f(x_{\nu(k+1)}) \leq f(x_{\nu(0)}) + \sum_{i=0}^{k} \zeta_i \leq f(x_0) + \zeta. \tag{24}$$

Since $f(x_{k+1}) \leq W_{k+1}$, we have $x_k \in \bar{\mathcal{L}}_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0) + \zeta\}$ for all k, and part 1. is established.

Part 2. is a direct consequence of the assumption regarding the existence of $\bar{k}$ sufficiently large such that $f(x_{k+1}) \leq f(x_{\nu(k)})$ for all $k \geq \bar{k}$ and of inequality (23).

Lemma 1 is an auxiliary result to prove Proposition 5 which, in turn, is necessary for the desired convergence result.

Lemma 1 *Consider* $f : \mathbb{R}^n \to \mathbb{R}_0^+$, $M \in \mathbb{N}$, $W_k = \max_{0\leq j\leq \min\{k,M-1\}} f(x_{k-j})$, $\varrho \in \mathbb{R}$ *with* $\varrho > 0$ *and* $\{\zeta_k\}$ *a sequence in* $\mathbb{R}^+$ *such that* $\sum_{i=0}^{\infty} \zeta_i = \zeta < \infty$. *Then*

$$\lim_{k\to\infty} [\varrho f(x_k) - \zeta_k] = 0 \Rightarrow \lim_{k\to\infty} W_k = 0. \tag{25}$$

Proof Since

$$\lim_{k\to\infty} [\varrho f(x_k) - \zeta_k] = 0, \tag{26}$$

and by assumption $\lim_{k\to\infty} -\zeta_k = 0$, then $\lim_{k\to\infty} \varrho f(x_k) = 0$. Consequently,

$$\lim_{k\to\infty} f(x_k) = 0. \tag{27}$$

This result allows us to state that $\lim_{k\to\infty} W_k = 0$.

Indeed, by (27), for all $\varepsilon > 0$, there exists $\bar{k} \in \mathbb{N}$ such that, for all $k > \bar{k}$, $f(x_k) < \varepsilon$. Thus, for all $k > \bar{k} + M$ we have $W_k = f(x_{\nu(k)})$ with $\nu(k) > \bar{k}$. Therefore, $W_k < \varepsilon$, which conducts to the desired result.

We are now in conditions of establishing Proposition 5.

Proposition 5 *Let* $f : \mathbb{R}^n \to \mathbb{R}_0^+$ *be a function and* $\{x_k\} \subset \mathbb{R}^n$ *a sequence such that:*

$$f(x_{k+1}) \leq W_k + \zeta_k - \varpi f(x_k), \tag{28}$$

where $\varpi > 0$, $\zeta_k > 0$ *for all* k *and* $\sum_{i=0}^{\infty} \zeta_i = \zeta < \infty$. *Moreover, assume that for each* $\varrho > 0$ *there exists* $\bar{k} \in \mathbb{N}$ *such that for all* $k > \bar{k}$ *inequality*

$$\zeta_k - \varrho f(x_k) < 0, \quad \textit{holds.} \tag{29}$$

Then $x_k \in \bar{\mathcal{L}}_0$ for all k and, furthermore,

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} W_k = 0. \tag{30}$$

Proof Condition (28) implies that inequality (21) is satisfied for all k. Thus, Proposition 4 ensures that $x_k \in \bar{\mathcal{L}}_0$ for all k. Since condition (29) holds for $k > \bar{k}$, again using Proposition 4 we have that the non negative sequence $\{W_k\}_{k>\bar{k}}$ is monotonically non increasing. Thus, there exists a limit $W_* \geq 0$ for this sequence, when $k \to \infty$.

Reasoning by contradiction, we assume that $W_* \neq 0$. Thus, there exists $k_1 \in \mathbb{N}$ such that, for all $k > k_1$, we have $W_k > t_1 > 0$. By Lemma 1, there exists $k_2 \in \mathbb{N}$ such that $|-\zeta_k + \varpi f(x_k)| > t_2 > 0$ for all $k > k_2$.

Without loss of generality, we can assume that $k_2 \geq \bar{k}$. By hypothesis, we have that

$$f(x_{k+1}) \leq W_k - t_2,\ k > k_2 \geq \bar{k}. \tag{31}$$

Taking $k \geq k_2 + M + 1$, we have $\nu(k) - 1 > k - M - 1 \geq k_2$. Consequently, using definition of $\nu(k)$ and equation (31), it follows:

$$f(x_{\nu(k)}) \leq f(x_{\nu(\nu(k)-1)}) - t_2. \tag{32}$$

Taking limits in both sides and being that $f(x_{\nu(k)}) \to W_*$, we have $t_2 \leq 0$, which contradicts our assumption.

The following theorem establishes that Algorithm 3 (H2P) is well defined and states the corresponding convergence.

Theorem 1 *Let $\{x_k\}$ be the sequence of iterates generated by Algorithm 3 (H2P). Consider the sequence $\{\zeta_k\}$, required for accepting new points, defined as $\zeta_k > 0$ for all k and $\sum_{k=0}^{\infty}\zeta_k = \zeta < \infty$ and define $\bar{\mathcal{L}}_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0) + \zeta\}$. Suppose that there exists $r > 0$ such that, for all $x \in \bar{\mathcal{L}}_0$, the closed ball $\bar{B}(x, r)$ is contained in an open convex set Ω where Hypothesis 1 is satisfied. For each $\varrho > 0$, assume that there exists $\bar{k} \in \mathbb{N}$ such that for all $k > \bar{k}$*

$$\zeta_k - \varrho f(x_k) < 0. \tag{33}$$

Then, Algorithm 3 (H2P) ends at some point x_k satisfying $F(x_k) = 0$, or it is well defined and generates a sequence $\{x_k\}$ such that

$$\lim_{k\to\infty} F(x_k) = 0. \tag{34}$$

Proof Consider $F(x_k) \neq 0, \forall k$. Let us start by showing that Algorithm 3 is well defined, meaning that Algorithm 3 will generate a stepsize parameter $\lambda_k > 0$ at each iteration.

Reasoning by contradiction, suppose not. Algorithm 1 (Step 2 on H2P Algorithm) indicates that $\lambda_k \geq \tau_{\min}^{NBL_{\max}} > 0$ in every iteration. Therefore, if $\lambda_k = 0$ then it should occur at Step 3 which corresponds to the inexact

Newton algorithm. In Algorithm 2, the line search occurs between Step 2 and Step 7, where the values σ, η and μ are reduced. At iteration k, let $\{t_l\}$ be the sequence used to count at Step 4 the current number of calls of the line search. We have $t_l \to \infty$ and sequences $\{\sigma_l\}$, $\{\eta_l\}$ and $\{\mu_l\}$ converge to zero. Otherwise, Algorithm 3 would have computed $\lambda_k > 0$.

Proposition 4 guarantees that $x_k \in \bar{\mathcal{L}}_0$ for all $k \in \mathbb{N}$. Consider L_J and m_J, the constants of Hypothesis 1 associated with the convex set Ω. Define $\hat{\sigma} = \frac{1}{2n^{1/2}L_J m_J}$ and $C = 4n^{1/2}L_J m_J$. For l sufficiently large $\eta_l \in (0,1)$ and $0 < \sigma_l \leq \hat{\sigma}$. Thus, Proposition 1 guarantees that procedure FDGMRES computes a direction d_l such that

$$\big\|J(x_k)d_l + F(x_k)\big\| \leq (\eta_l + C\sigma_l)\big\|F(x_k)\big\| \leq \bar{\eta}\big\|F(x_k)\big\| < (1-\gamma)\big\|F(x_k)\big\|, \quad (35)$$

with $\eta_l + C\sigma_l \leq \bar{\eta} < (1-\gamma)$.

By applying Proposition 2 we can conclude that

$$\|F(x_k + \lambda d_l)\| \leq (1 - \gamma\lambda)\|F(x_k)\|,$$

for $\lambda \in [0, \bar{\lambda}(x_k)]$ and $\bar{\lambda}(x_k)$ defined as in (17) with $F(x) = F(x_k)$.

We note that $a \in (0,1]$ and $\xi \in [\xi_{\min}, \xi_{\max}]$, resulting $\xi^i_{\min} a \leq \alpha_i \leq \xi^i_{\max}$. Therefore, it is possible choose i_* that satisfies

$$i_* \geq \max\left\{0, \frac{\log(\bar{\lambda}(x_k))}{\log(\xi_{\max})}\right\} \quad (36)$$

ensuring $0 < \alpha(i_*) \leq \bar{\lambda}(x_k)$. Since $\mu_l \to 0$, for l sufficiently large we have $\mu_l a \leq \xi^{i_*}_{\min} a \leq \alpha(i_*)$. Thus,

$$\|F(x_k + \alpha(i_*)d_l)\| \leq (1 - \gamma\alpha(i_*))\|F(x_k)\|$$

and, since $\alpha(i_*) \in (0,1]$, the acceptance condition

$$f(x_k + \alpha(i_*)d_l) \leq W_k + \zeta_k - \gamma\alpha(i_*)^2 f(x_k)$$

will be satisfied. Iteration k of Algorithm 2 ends, with a positive value for λ_k, ensuring that Algorithm 3 is well defined.

We will now prove that the sequence $\{\lambda_k\}$ is lower bounded by a constant $\varpi > 0$. Once more, reasoning by contradiction, let us suppose that there is $K \subseteq \mathbb{N}$ such that

$$\lim_{k \in K,\, k \to \infty} \lambda_k = 0. \quad (37)$$

In Algorithm 1 DFSANE, for all k we have $\lambda_k \geq \tau^{NBL_{\max}}_{\min} > 0$. Thus, for $k \in K$ sufficiently large, λ_k is generated by the inexact Newton algorithm. The line search procedure of Algorithm 2 (between Step 2 and Step 7) then implies that $\tilde{\sigma}_k \to 0$ and $\tilde{\eta}_k \to 0$, when $k \in K, k \to \infty$.

Again, let L_J and m_J be the constants of Hypothesis 1 associated with the convex set Ω. Define $\hat{\sigma} = \frac{1}{2n^{1/2}L_J m_J}$ and $C = 4n^{1/2}L_J m_J$. For $k \in K$

sufficiently large $\tilde{\eta}_k \in (0,1)$ and $0 < \tilde{\sigma}_k \leq \hat{\sigma}$. Thus, Proposition 1 guarantees that procedure FDGMRES computes a direction d_k such that

$$\left\|J(x_k)d_k+F(x_k)\right\| \leq (\tilde{\eta}_k+C\tilde{\sigma}_k)\left\|F(x_k)\right\| \leq \bar{\eta}\left\|F(x_k)\right\| < (1-\gamma)\left\|F(x_k)\right\| \quad (38)$$

with $\tilde{\eta}_k + C\tilde{\sigma}_k \leq \bar{\eta} < (1-\gamma)$.

Proposition 2 then establishes that

$$\|F(x_k + \lambda d_k)\| \leq (1 - \gamma\lambda)\|F(x_k)\|,$$

for $\lambda \in [0, \bar{\lambda}(x_k)]$ and $\bar{\lambda}(x_k)$ defined as in (17), with $F(x) = F(x_k)$. For $\lambda \in [0,1]$, the previous condition implies

$$f(x_k + \lambda d_k) \leq W_k + \zeta_k - \gamma\lambda^2 f(x_k).$$

Since $x_k \in \bar{\mathcal{L}}_0$, we have $f(x_k) \leq f(x_0) + \zeta$ and $\left\|F(x_k)\right\|^2 \leq \left\|F(x_0)\right\|^2 + \zeta$. By setting $b = \sqrt{2\max\{\left\|F(x_0)\right\|^2, \zeta\}}$ we can conclude that $\left\|F(x_k)\right\| \leq b$ for all $k \in K$. It is now possible to define a lower bound for $\bar{\lambda}(x_k)$, considering

$$0 < \omega = \min\left(\frac{r}{m_J(1+\bar{\eta})b}, \frac{2(1-\gamma-\bar{\eta})}{(1+\bar{\eta})^2 m_J^2 L_J b}\right) \leq \bar{\lambda}(x_k), \text{ for all } k \in K. \quad (39)$$

Since $\lambda_k \to 0$ for $k \in K$, for $k \in K$ sufficiently large $\lambda_k < a \leq 1$. In this case, λ_k was computed in the line search procedure of Step 5, and satisfies $\lambda_k \geq \xi_{\min}\bar{\lambda}(x_k)$, since $\lambda_k \neq 0$.

Thus, for all $k \in K$ we have

$$\lambda_k \geq \min\{a, \xi_{\min}\bar{\lambda}(x_k)\} \geq \min\{a, \xi_{\min}\omega\}, \quad (40)$$

which contradicts limit (37). Thus, $\{\lambda_k\}$ is lower bounded by some $\varpi > 0$.

Consequently, for all k, we have

$$\left\|F(x_k + \lambda_k d_k)\right\|^2 \leq W_k + \zeta_k - \gamma\varpi^2\left\|F(x_k)\right\|^2. \quad (41)$$

For k sufficiently large, condition (33) holds. Thus, the assumptions of Proposition 5 are satisfied for $f(x_k) = \left\|F(x_k)\right\|^2$, resulting

$$\lim_{k\to\infty} \left\|F(x_k)\right\| = 0. \quad (42)$$

Before ending this section, we would like to point out that condition (33) can be easily satisfied, for example, by defining $\zeta_k = \frac{\min\{f(x_0), f(x_k)\}}{(k+1)^{1.1}}$. Note that, in this case, we also have $\zeta_k > 0$ for all k and $\sum_{k=0}^{\infty}\zeta_k = \zeta < \infty$.

5 Numerical experiments

This numerical section intends to evaluate the contribution of the hybrid version to improve the practical performance of algorithms. For achieving this goal, we compared implementations of the Hybrid Algorithm 3 (`H2P`), the original version of `DFSANE` [4] and a derivative-free version of an inexact Newton method, based on Algorithm NM1 proposed in [12] (without considering the watchdog procedure). All codes were implemented in Matlab 7.0 and tested on a Intel(R) Core I3-2100 personal computer, with 3.10 GHz and 4 Gb RAM.

In the hybrid algorithm, whenever possible, we consider the original settings proposed in [4] and [12], for DFSANE and the inexact Newton method, respectively. Thus, the stepsize in DFSANE is reduced using a quadratic interpolation procedure, considering the identity matrix as an approximation to the Jacobian. As example, for computing a new value λ_+ on Step 5 of Algorithm 1, define

$$\begin{array}{l} \varphi : [0, \lambda_+] \to \mathbb{R} \\ \varphi(\lambda) = f(x_k - \lambda(1/\alpha_k)F(x_k)). \end{array} \tag{43}$$

The minimizer λ_{new} of (43) is computed and used to define the new stepsize λ_+, considering the safeguards

$$\lambda_+ = \begin{cases} \tau_{\min}\lambda_+, & \text{if } \lambda_{new} < \tau_{\min}\lambda_+ \\ \tau_{\max}\lambda_+, & \text{if } \lambda_{new} > \tau_{\max}\lambda_+ \\ \lambda_{new}, & \text{otherwise.} \end{cases} \tag{44}$$

The safeguards were set to $\tau_{\min} = 0.1$ and $\tau_{\max} = 0.5$.

Furthermore, we considered $\alpha_0 = 1$, $\gamma = 10^{-4}$, $M = 7$ and, to define α_k, we use equation (6), if $\alpha_k \in [10^{-10}, 10^{10}]$. Otherwise, a new α_k is computed, depending on $F(x_k)$, as:

$$\alpha_k = \begin{cases} 1, & \text{if } \left\|F(x_k)\right\| > 1, \\ \left\|F(x_k)\right\|, & \text{if } 10^{-5} \leq \left\|F(x_k)\right\| \leq 1, \\ 10^{-5}, & \text{if } \left\|F(x_k)\right\| < 10^{-5}. \end{cases} \tag{45}$$

Regarding the settings of [4], the only modification is the definition of $\zeta_k = \frac{\min\{f(x_0), f(x_k)\}}{(k+1)^{1.1}}$, to satisfy the Hypothesis of Theorem 1.

In the inexact Newton method, following [13], we use the GMRES(m). This strategy restarts the GMRES method after one cycle of m iterations, where m is a predefined integer. At the end of each cycle, the last direction d_m (if not satisfactory) is used as initialization for the new cycle of m iterations. This strategy intends to reduce the memory requirements and the computational cost associated with the increase in the number of iterations of GMRES, which is responsible for its inefficiency when used for solving larger problems. By adopting GMRES(m), we do not guarantee the conditions required to establish the theoretical properties of the method. Nevertheless, this is a common practice for other authors [12] due to the good practical performance.

We set the maximum number of GMRES iterations $m = 30$ and the maximum number of GMRES cycles $nc_{\max} = 30$. Differently from Algorithm NM1 [12], where a sequence of constant forcing terms is used, we adopt the sequence proposed by Eisenstat and Walker [8] both in our Hybrid Algorithm and in the implementation of the inexact Newton method (based on `NM1`) :

$$\eta_k = \gamma \left(\frac{\|F(x_k)\|}{\|F(x_{k-1})\|} \right)^{\alpha}, \tag{46}$$

with $\gamma = 1$ and $\alpha = 0.5(1 + \sqrt{5})$ (in this last case we use a safeguard $[10^{-6}, 10^{-2}]$).

As in [4], we consider the stopping criterion:

$$\frac{\|F(x_k)\|}{\sqrt{n}} \leq \varepsilon_a + \frac{\varepsilon_r \|F(x_0)\|}{\sqrt{n}}, \tag{47}$$

where $\varepsilon_a = 10^{-5}$ and $\varepsilon_r = 10^{-4}$.

Moreover, all algorithms will end with a failure ($flag = 0$) in the following conditions:

FII If the number of inner GMRES(m) iterations equals or exceeds $nc_{\max} = 30$ cycles of $m = 30$ iterations;
FST If the stepsize is equal or lower than 10^{-12};
FFE If the total number of function evaluations equals or exceeds 10000;
FOU *Overflow* or *underflow* cases.

If condition (47) is satisfied, then the algorithms end with a success (denoted by **S**, when reporting the numerical results).

Two versions of the Hybrid Algorithm 3 (H2P) were implemented. In the first, named as `H2P1`, the line search procedure is not performed in the DFSANE method ($NBL_{\max} = 0$). In the second version, named as `H2P6`, the stepsize accepts a maximum of five reductions ($NBL_{\max} = 5$). DFSANE implementation is named as `DFSANE` and the derivative-free implementation of the inexact Newton method is named as `NI`.

All algorithms were tested in two sets of high-dimension problems, one collected from La Cruz and Raydan [5] (problems 1-20) and another collected from Section 4 of Lukšan and Vlček [15] (problems 1-21).

Problems collected from [5] were solved for dimensions $n = 100$, 500, 1000, 2000 and 5000, except problems 4, 7 and 18, for which we have considered $n = 99$, 498, 999, 1998 and 4998.

For each problem, 10 initializations were uniformly randomly generated in a neighborhood of the initial points proposed in [5]. Let $x_0 = (x_1, x_2, \ldots, x_n)^\top$ be the proposed initial point for a given problem. The i^{th}-component of the new initialization is uniformly randomly generated in the interval

$$[x_i + \min\{-5, -5|x_i|\}, x_i + \max\{5, 5|x_i|\}].$$

Additionally, 10 other initializations were randomly generated for each problem, considering a normal distribution. The i^{th}-component was generated using x_i as mean and $\max\{5, 5|x_i|\}$ as standard deviation.

The numerical results in this first set of problems are reported in Table 1 and in the performance profiles [7] of Figure 2. The two left plots are more adequate to analyze efficiency, since $\tau \in [1, 2]$. The right graphs are more suitable to evaluate robustness, since $\tau \in [1, 10]$.

	FII	**FST** ($\lambda < 10^{-12}$)	**FFE** (max.10000)	**FOU**	**S**
H2P1	54.5%	0%	0%	6.4%	39.1%
DFSANE	—	1.5%	44.7%	3.1%	50.8%
H2P6	39.9%	0.4%	3.1%	3.4%	53.2%
NI	57.2%	0%	0%	3.6%	39.2%

Table 1 Performance on the first test set - Problems 1-20 in [5], considering random initializations.

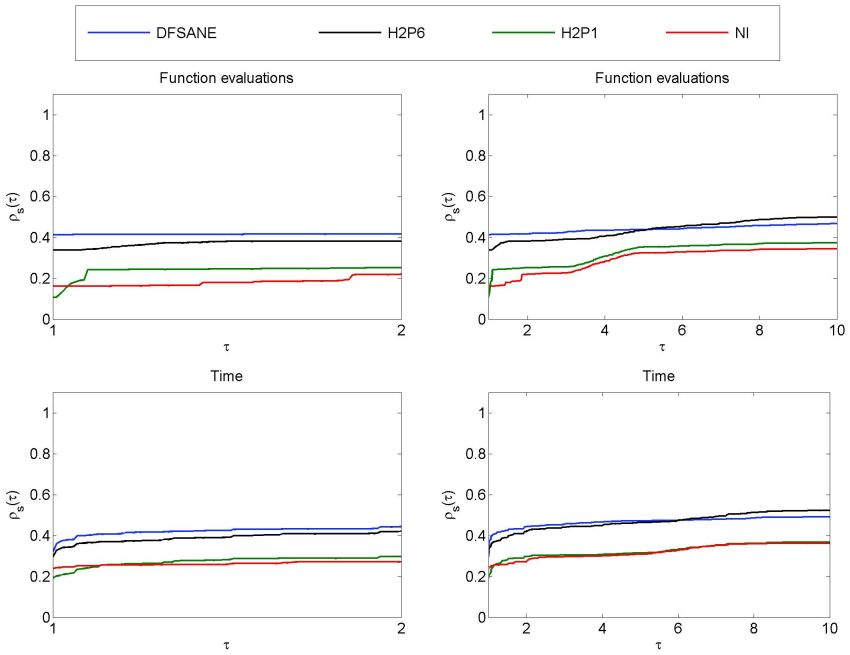

Fig. 2 Performance profiles - Problems 1-20 in [5], considering random initializations.

DFSANE method is characterized by carrying out many low cost iterations, when compared with inexact Newton methods. Additionally, in the class of problems suited for derivative-free optimization, function evaluation is mainly

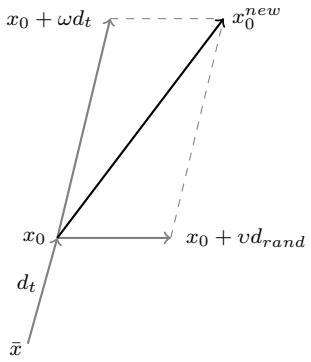

Fig. 3 Computation of the new initialization for Problems 1-21 from [15].

responsible for the computational cost. So, we considered that the number of iterations is not an adequate indicator of efficiency, using the number of function evaluations as the main performance measure. Computational time was used as a secondary criterion.

Problems 1-21 from [15] were initially tested for the initialization suggested and a dimension $n = 5000$, except Problem 5 for which $n = 4999$. Since some of these initial points were close to the problem solution, we decided to generate a new set of initializations, increasing the distance from the problem solution. In this new test we have only included problems for which at least one of the solvers tested had succeeded with the initialization proposed in [15]. As result, problems 1, 2, 5 and 10 were excluded.

Twelve new initializations were generated for each selected problem. With this purpose, we defined the vector $d_t = x_0 - \bar{x}$, where x_0 is the initial point reported in [15] and $\bar{x}$ is an approximation to the problem solution computed with the initial numerical test procedure. Additionally, we generated a random direction d_{rand} such that the cosine of the angle between d_t and d_{rand} is lower than 0.95. As we can see in Figure 3, the new initialization is defined as $x_0^{new} = x_0 + \omega d_t + \upsilon d_{rand}$, where the parameter ω takes the values 1, 20 and 200 and υ is set equal to 0, 1, 20 and 200, in a total of 12 combinations.

Table 2 and the performance profiles of Figure 4 correspond to the results in this new set of problems.

	FII	**FST** ($\lambda < 10^{-12}$)	**FFE** (max.10000)	**FOU**	**S**
H2P1	35.01%	1.11%	2.22%	4.44%	57.22%
DFSANE	0.00%	1.67%	41.67%	3.33%	53.33%
H2P6	29.44%	1.11%	3.33%	4.44%	61.67%
NI	34.44%	1.11%	1.11%	3.33%	60.01%

Table 2 Performance on the second test set - Selected problems from [15], considering random initializations.

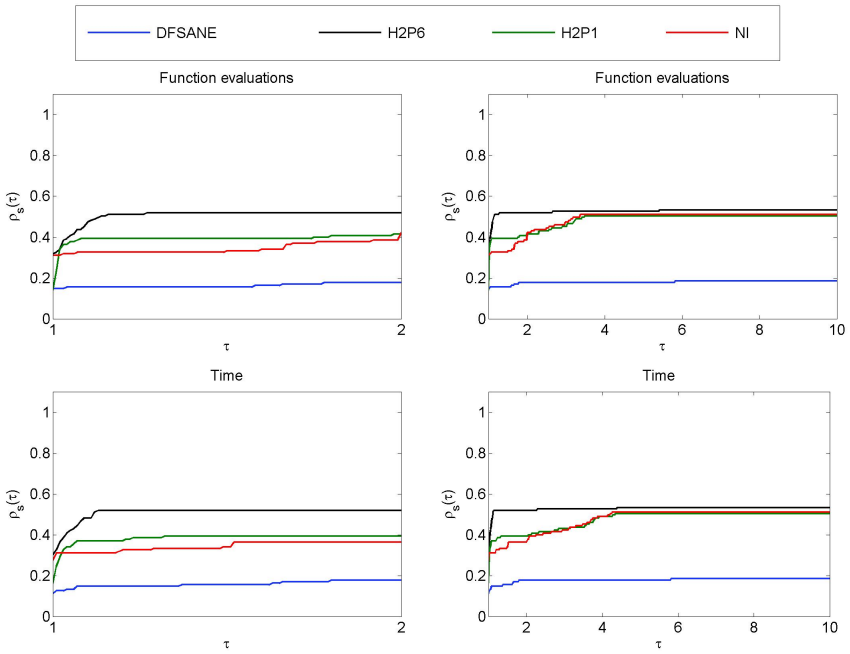

Fig. 4 Performance profiles - Selected problems from [15], considering random initializations.

In the two test sets, the hybrid algorithm, particularly the version `H2P6`, presented a good performance both in terms of efficiency and robustness. In the problems collected from [5], algorithm `H2P6` was the most robust and the second more efficient. For this test set, `DFSANE` presented the best performance in terms of efficiency. In this test set, algorithms which tend to often use spectral directions presented a superior performance, which may indicate that some characteristics of the problems favor algorithms enhanced with this type of directions.

In the second test set, where problems were selected from [15], algorithm `H2P6` was the most robust and the most efficient. Contrary to the previous results, `DFSANE` presented the worst performance. Although this fact indicates some unsuitability of spectral methods to solve this test set of problems, the good performance of `H2P6` reinforces the advantage of hybrid algorithms.

Due to the variability on the type of directions considered in each step, it was expected that the use of a hybrid strategy could bring advantages in what respects to robustness, when compared with pure methods. However, the additional good performance obtained in terms of efficiency validates the benefit of using hybrid algorithms for solving high-dimensional nonlinear systems of equations.

6 Final remarks

In this work, we proposed a hybrid approach to address the resolution of high-dimensional systems of nonlinear equations, in the situation where derivatives are not available for use. A two-steps Algorithm 3 (H2P), combining the Spectral Residual Method ([5, 4]) and the inexact Newton method was developed and analyzed.

Theorem 1 is the main theoretical contribution of this work, ensuring that the sequence of iterates generated by Algorithm 3 (H2P) determines at least one accumulation point that is a solution of the nonlinear system.

We highlight that, theoretical convergence only depends on the last step. In this sense, the initial step in H2P is designed only to facilitate the practical understanding of the algorithm. The theoretical results extend to methods in which the initial steps do not exist or differ from those proposed in this work. For example, the results presented for algorithm H2P can be extended to pure inexact Newton methods, considering as acceptance criteria (9), which is more flexible than the one used in [9].

From the numerical point of view, according to our test sets, Algorithm 3 is competitive when compared with a pure version of an inexact Newton method and DFSANE.

Acknowledgements We are thankful to Sandra Augusta Santos, from University of Campinas, Brazil, for suggestions and valuable discussions on the subject, which greatly helped us to improve the paper. We also acknowledge the suggestions of two anonymous referees, which greatly contributed for the paper organization.

References

1. Armijo, L.: ,Minimization of functions having Lipschitz continuous first partial derivatives. Pacific Journal of Mathematics 16, 1-3 (1966).
2. Barzilai, J. , Borwein, J. M.: Two-point step size gradient methods, IMA J. Numerical Analysis 8, 141-148 (1988).
3. Chamberlain, R. M., Powell, M. J. D., Lemarechal, C., Pedersen, H. C. .: The watchdog technique for forcing convergence in algorithms for constrained optimization. In: Buckley A.G., Goffin J.L. (eds.) Algorithms for Constrained Minimization of Smooth Nonlinear Functions. Mathematical Programming Studies, 16, pp. 1-17, Springer Berlin Heidelberg, Berlin, Heidelberg, (1982).
4. La Cruz, W., Martínez, J. M., Raydan, M.: Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. Mathematics of Computation 75, 1429-1448 (2006).
5. La Cruz, W., Raydan, M.: Nonmonotone spectral methods for large-scale nonlinear systems. Optimization Methods and Software 18, 583-599 (2003).
6. Dembo, R. S., Eisenstat, S. C., Steihaug, T.: Inexact Newton methods. SIAM J. Numerical Analysis. 19, 400-408 (1982).
7. Dolan, E. D., Moré, J. J.: Benchmarking optimization software with performance profiles. Mathematical Programming 91, 201-213 (2002).
8. Eisenstat, S. C., Walker, H. F.: Choosing the forcing terms in an inexact-Newton method. SIAM J. Scientific Computing 17, 16-32 (1996).

9. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton's method. SIAM J. Numerical Analysis, 23, 707-716 (1986).
10. Grippo L., Sciandrone, M.: Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. Computational Optimization and Applications 23, 143-169 (2002).
11. Grippo, L., Sciandrone, M.: Nonmonotone derivative-free methods for nonlinear equations. Computational Optimization and Applications 37, 297-328 (2007).
12. Grippo, L., Sciandrone, M.: Nonmonotone globalization of the finite-difference Newton-GMRES method for nonlinear equations. Optimization Methods and Software 25, 971-999 (2010).
13. Kelley, C. T.: Iterative Methods for Linear and Nonlinear Equations. SIAM, Philadelphia (1995).
14. Li, D.H., Fukushima, M.: A derivative-free line search and global convergence of Broyden-like method for nonlinear equations. Optimization Methods and Software 13, 181-201 (2000).
15. Lukšan, L., Vlček, J.: Sparse and partially separable test problems for unconstrained and equality constrained optimization. Report V-767, ICS AS CR, Prague (1998).
16. Saad, Y., Schultz, M. H.: GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM J. Scientific and Statistical Computing 7, 856-869 (1986).