COMPUTATIONAL & APPLIED MATHEMATICS

Volume 30, N. 1, pp. 217–245, 2011
Copyright © 2011 SBMAC
ISSN 0101-8205
www.scielo.br/cam

Two derivative-free methods for solving underdetermined nonlinear systems of equations

N. ECHEBEST[1], M.L. SCHUVERDT[2*] and R.P. VIGNAU[1]
[1]Department of Mathematics, FCE, University of La Plata
CP 172, 1900 La Plata Bs. As., Argentina
[2]CONICET, Department of Mathematics, FCE, University of La Plata
CP 172, 1900 La Plata Bs. As., Argentina
E-mails: schuverd@mate.unlp.edu.ar / opti@mate.unlp.edu.ar / vignau@mate.unlp.edu.ar

Abstract. In this paper, two different approaches to solve underdetermined nonlinear system of equations are proposed. In one of them, the derivative-free method defined by La Cruz, Martínez and Raydan for solving square nonlinear systems is modified and extended to cope with the underdetermined case. The other approach is a Quasi-Newton method that uses the Broyden update formula and the globalized line search that combines the strategy of Grippo, Lampariello and Lucidi with the Li and Fukushima one. Global convergence results for both methods are proved and numerical experiments are presented.

Mathematical subject classification: Primary: 65H10; Secondary: 90C57.

Key words: underdetermined nonlinear systems, Quasi-Newton method, derivative-free line search, spectral step length, global convergence.

1 Introduction

We consider the problem of determining a solution $x^* \in \mathbb{R}^n$ that verifies the nonlinear system of equations

$$F(x) = 0 \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function and $m \leq n$, making special emphasis when $m < n$. We are interested in systems for which

#CAM-323/10. Received: 15/VIII/10. Accepted: 05/I/11.

the Jacobian matrix of F, denoted by $J(x)$, is not available or requires a prohibitive amount of storage. This situation is common when functional values come from experimental measures, for example: from physics, chemistry or economics.

Such kind of problems also appears as the feasible set of general nonlinear programming problems. Our main motivation is the application of the proposed methods as subalgorithms for finding feasible points in derivative-free nonlinear programming algorithms such as, for example two-phases algorithms [3, 8, 16, 26], feasible methods [1, 12, 18, 19, 20, 22, 23, 24, 25] and inexact restoration methods [15, 17].

The resolution of square nonlinear systems without using derivatives has been addressed using the spectral residual approach in [11] and the Broyden Quasi Newton approach in [13]. Some ideas of those papers are incorporated in the present work. In [11], the authors defined the derivative-free spectral algorithm for nonlinear equations called DF-SANE. Furthermore, global convergence was proved by using a derivative-free line search that combines the nonmonotone strategy of Grippo, Lampariello and Lucidi [9] with the Li and Fukushima scheme [13].

In [13], a Quasi Newton method based on the Broyden update formula and on a nonmonotone derivative-free line search was defined for the square case. Also in [4] an inexact Quasi-Newton method was introduced with a similar line search technique to the one introduced in [13] and using $B_k = J(x_k)$ periodically. Under appropriate hypotheses, global and superlinear convergence were proved in both papers.

In the present paper we define two approaches for the nonsquare system of equations based on the ideas appearing in [11] and [13]. The proposed algorithms use a generalization of the derivative-free nonmonotone line search defined in [11]. The search direction in [11] is computed using the residual vector $F(x_k)$ and the spectral coefficient [2]. For the underdetermined system, the current point is computed by considering a fixed number of L_m directions that are the solution of some appropriate square systems using a spectral coefficient under the idea explained in [11]. The direction in the second algorithm is computed as an approximate solution of a linear system using the nonsquare

Broyden update formula for matrices. It is well established in the literature that, for the square case, this is the most used secant approximation to the Jacobian and it works very well locally [6]. As we mentioned before, the Broyden update formula has been previously used by Li and Fukushima and combined with a derivative-free line search for the square case. In [13] the current direction is the solution to the linear system

$$B_k d + F(x_k) = 0 \tag{2}$$

and the update matrix B_{k+1} is defined as

$$B_{k+1} = B_k + \beta_k \frac{(y_k - B_k s_k)s_k^T}{\|s_k\|^2} \tag{3}$$

where $y_k = F(x_k) - F(x_{k-1})$, $s_k = x_k - x_{k-1}$ and the parameter β_k is chosen such that $|\beta_k - 1| \leq \bar{\beta} < 1$ for which B_{k+1} is nonsingular. In this paper, when there is a solution of the linear system (2) we use such solution as a search direction and, when there is none, we solve the linear system approximately as proposed in [14]. Consequently, we avoid the necessity of choosing the parameter β_k.

Under appropriate hypotheses, global convergence of the sequence generated by both methods will be proved. The global convergence result obtained for the algorithm based on the spectral ideas extends the convergence result in [11]. For the Quasi Newton method that uses the Broyden update formula, we obtain convergence using a Dennis-Moré type condition. We show that this condition can be dropped out for a particular derivative-free line search. Thus, both algorithms can be viewed as extensions of the well known methods defined in [11] and [13] for square nonlinear systems.

We consider the usual continuously differentiable merit function $f\colon \mathbb{R}^n \to \mathbb{R}$, which consists in a measure of the residual $F(x)$ at x, $f(x) = \frac{1}{2}\|F(x)\|^2$.

The iterative algorithms generate a sequence $\{x_k\}$, for $k = 1, 2, \ldots$, starting from a given initial point x_0. A subsequence of $\{x_k\}$ will be indicated by $\{x_k\}_{k\in K}$ where K is some infinite index set.

This paper is organized as follows. In Section 2 we present the algorithm that performs the derivative-free line search and we establish there some of its properties. In Section 3 we define DF-SAUNE Algorithm, a modification of

DF-SANE Algorithm of [11] for handling efficiently problem (1) and we prove the global convergence results. In Section 4 we define the Quasi Newton method using the Broyden update formula and the derivative-free line search introduced in Section 2. We analyze the conditions under which it is possible to obtain global convergence. Both algorithms are tested and the numerical results are presented in Section 5. Finally, some conclusions are drawn in Section 6.

Notation.

- $\|.\|$ denotes the Euclidean norm.
- Given a matrix $B \in \mathbb{R}^{m\times n}$, $\mathcal{N}(B)$ denotes the null space of the matrix B.
- For $i = 1, \ldots, n$; e_i is the canonical vector in $\mathbb{R}^n$.
- I_n denotes the identity matrix in $\mathbb{R}^{n\times n}$.
- $g(x) = J(x)^T F(x) = \nabla(\frac{1}{2}\|F(x)\|^2)$.

2 The nonmonotone line search without derivatives

In this section we shall be concerned with the nonmonotone derivative-free line search that will be used in the methods defined in the following sections. As we mentioned before, the strategy is based on the line search proposed in [11]. Given the current iterate x_k and a search direction d_k, the algorithm looks for a steplength α_k such that

$$f(x_k + \alpha_k d_k) \leq \max_{0\leq j\leq M-1} f(x_{k-j}) + \eta_k - \gamma\alpha_k^2 f(x_k) \tag{4}$$

where M is a nonnegative integer, $0 < \gamma < 1$ and $\displaystyle\sum_{k=0}^{\infty} \eta_k = \eta < \infty$.

This procedure combines the well known nonmonotone line search strategy for unconstrained optimization introduced by Grippo, Lampariello and Lucidi [9]:

$$f(x_k + \alpha_k d_k) \leq \max_{0\leq j\leq M-1} f(x_{k-j}) + \gamma\alpha_k \nabla f(x_k)^T d_k, \tag{5}$$

with the Li-Fukushima derivative-free line search scheme [13]:

$$\|F(x_k + \alpha_k d_k)\| \leq (1 + \eta_k)\|F(x_k)\| - \gamma\alpha_k^2\|d_k\|^2. \tag{6}$$

The combined strategy (4) produces a robust nonmonotone derivative-free line search that takes into account the advantages of both schemes. The line search (4) is strongly based on the fact that the search direction comes from the residual vector. For details we refer to [11]. In such paper this strategy was implemented and tested using an extensive set of numerical experiments which showed its competitiveness for square systems of nonlinear equations.

In this paper, given a general search direction d_k, based on (5) and (6), we consider the following line search condition:

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq M-1} f(x_{k-j}) + \eta_k - \gamma \alpha_k^2 \|d_k\|^2, \tag{7}$$

where $\|d_k\|^2$ takes the place of $f(x_k)$ in (4), obtaining a more general strategy. For completeness, we establish here the implemented process.

Algorithm 1. *Nonmonotone line search*

Given $d \in \mathbb{R}^n$, $0 < \tau_{\min} < \tau_{\max} < 1$, $0 < \gamma < 1$, $M \in \mathbb{N}$, $\{\eta_k\}$ *such that* $\eta_k > 0$ *and* $\sum_{k=0}^{\infty} \eta_k = \eta < \infty$

Step 1: *Compute* $\overline{f}_k = \max\{f(x_k), \ldots, f(x_{\max\{0,k-M+1\}})\}$
$\alpha_+ = \alpha_- = 1$

Step 2: *If* $f(x_k + \alpha_+ d) \leq \overline{f}_k + \eta_k - \gamma \alpha_+^2 \|d\|^2$,
define $d_k = d$, $\alpha_k = \alpha_+$, $x_{k+1} = x_k + \alpha_k d_k$
else if $f(x_k - \alpha_- d) \leq \overline{f}_k + \eta_k - \gamma \alpha_-^2 \|d\|^2$,
define $d_k = -d$, $\alpha_k = \alpha_-$, $x_{k+1} = x_k + \alpha_k d_k$
else
choose $\alpha_{+new} \in [\tau_{\min}\alpha_+, \tau_{\max}\alpha_+]$, $\alpha_{-new} \in [\tau_{\min}\alpha_-, \tau_{\max}\alpha_-]$
$\alpha_+ = \alpha_{+new}$, $\alpha_- = \alpha_{-new}$ *and go to step 2*

Proposition 2.1. *Algorithm* 1 *is well defined.*

Proof. See Proposition 1 of [11]. □

The new algorithms for solving (1) will follow the next procedure.

Algorithm 2. *General Algorithm*

Given x_0, $F(x_0)$, $M \in \mathbb{N}$, $0 < \tau_{\min} < \tau_{\max} < 1$, $0 \leq \epsilon < 1$, $0 < \gamma < 1$, $\{\eta_k\}$ *such that* $\eta_k > 0$ *and* $\sum_{k=0}^{\infty} \eta_k = \eta < \infty$.

Set $k \leftarrow 0$.

Step 1: *If* $\|F(x_k)\| \leq \epsilon \max\{1, \|F(x_0)\|\}$ *stop.*

Step 2: *Compute a search direction* d_k.

Step 3: *Find* α_k *and* $x_{k+1} = x_k + \alpha_k d_k$ *using Algorithm 1.*

Update $k \leftarrow k + 1$ *and go to Step 1.*

By considering the procedure above it is possible to obtain results (see below) that will be used in the next sections for obtaining the convergence results. The proofs of Propositions 2.2 and 2.3 below follow from Propositions 2 and 3 in [11] updated for the line search condition (7). We establish them here for completeness.

Proposition 2.2. *For all* $k \in \mathbb{N}$ *consider*

$$U_k = \max\{f(x_{(k-1)M+1}), \ldots, f(x_{kM})\}$$

and define $\nu(k) \in \{(k-1)M + 1, \ldots, kM\}$ *the index for which* $f(x_{\nu(k)}) = U_k$. *Then for all* $k = 1, 2, \ldots$

$$f(x_{\nu(k+1)}) \leq f(x_{\nu(k)}) + \eta$$

where $\eta = \sum_{i=0}^{\infty} \eta_i$.

Proof. We have that

$$\begin{aligned} f(x_{kM+1}) &\leq \max\{f(x_{(k-1)M+1}), \ldots, f(x_{kM})\} + \eta_{kM} - \gamma \alpha_{kM}^2 \|d_{kM}\|^2 \\ &= U_k + \eta_{kM} - \gamma \alpha_{kM}^2 \|d_{kM}\|^2 \leq U_k + \eta_{kM} \end{aligned}$$

then

$$\begin{aligned} f(x_{kM+2}) &\leq \max\{f(x_{(k-1)M+2}), \ldots, f(x_{kM+1})\} + \eta_{kM+1} - \gamma \alpha_{kM+1}^2 \|d_{kM+1}\|^2 \\ &\leq \max\{U_k, f(x_{kM+1})\} + \eta_{kM+1} \leq U_k + \eta_{kM} + \eta_{kM+1}. \end{aligned}$$

Thus, by an induction argument we obtain:

$$f(x_{kM+l}) \leq U_k + \sum_{j=0}^{l-1} \eta_{kM+j} - \gamma \alpha_{kM+l-1}^2 \|d_{kM+l-1}\|^2,$$

for $l = 1, 2, \ldots$.

Since $\nu(k+1) \in \{kM+1, \ldots, kM+M\}$

$$U_{k+1} = f(x_{\nu(k+1)}) \leq U_k + \sum_{j=0}^{M-1} \eta_{kM+j} - \gamma \alpha_{\nu(k+1)-1}^2 \|d_{\nu(k+1)-1}\|^2.$$

Thus, for all $k = 1, 2, \ldots$ we have that

$$f(x_{\nu(k+1)}) \leq f(x_{\nu(k)}) + \sum_{j=0}^{M-1} \eta_{kM+j} - \gamma \alpha_{\nu(k+1)-1}^2 \|d_{\nu(k+1)-1}\|^2 \leq f(x_{\nu(k)}) + \eta$$

as we wanted to prove. □

Proposition 2.3.

$$\lim_{k\to\infty} \alpha_{\nu(k)-1}^2 \|d_{\nu(k)-1}\|^2 = 0.$$

Proof. For all $k = 1, 2, \ldots$ we have

$$f(x_{\nu(k+1)}) \leq f(x_{\nu(k)}) + \sum_{j=0}^{M-1} \eta_{kM+j} - \gamma \alpha_{\nu(k+1)-1}^2 \|d_{\nu(k+1)-1}\|^2.$$

Writing the last inequality for $k = 1, 2, \ldots, L$ and adding these L inequalities we obtain

$$f(x_{\nu(L+1)}) \leq f(x_{\nu(1)}) + \sum_{j=M}^{(L+1)M-1} \eta_j - \gamma \sum_{j=1}^{L} \alpha_{\nu(j+1)-1}^2 \|d_{\nu(j+1)-1}\|^2.$$

Therefore

$$\gamma \sum_{j=1}^{L} \alpha_{\nu(j+1)-1}^2 \|d_{\nu(j+1)-1}\|^2 \leq f(x_{\nu(1)}) + \sum_{j=M}^{(L+1)M-1} \eta_j - f(x_{\nu(L+1)})$$
$$\leq f(x_{\nu(1)}) + \eta.$$

Thus, the series $\sum_{j=1}^{\infty} \alpha^2_{\nu(j+1)-1} \|d_{\nu(j+1)-1}\|^2$ is convergent and

$$\lim_{k\to\infty} \alpha^2_{\nu(k)-1} \|d_{\nu(k)-1}\|^2 = 0$$

as claimed. □

Proposition 2.4. *The sequence* $\{x_k\}$ *generated by the General Algorithm is contained in*

$$\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_{\nu(1)}) + \eta\}.$$

Proof. From Proposition 2.2 we have that, for all $k \geq 1$,

$$f(x_{\nu(k+1)}) \leq f(x_{\nu(1)}) + \eta.$$

Therefore, $f(x_{k+1}) \leq f(x_{\nu(k+1)}) \leq f(x_{\nu(1)}) + \eta$ as we wanted to prove. □

The results obtained up to here depend strongly on the line search technique without taking into account the way in which the direction d_k in the step 2 of the Algorithm 2 was computed.

From now on we will consider the set

$$K = \{\nu(1) - 1, \nu(2) - 1, \nu(3) - 1, \ldots\} \tag{8}$$

and from Proposition 2.3 we have that

$$\lim_{k\in K} \alpha_k^2 \|d_k\|^2 = 0. \tag{9}$$

Observe that from the proof of Propositions 2.3 and 2.4 we obtain the following result:

Proposition 2.5. *If we take* $M = 1$ *in Algorithm* 1 *then*

- *the sequence* $\{x_k\}$ *generated by the General Algorithm is contained in* $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0) + \eta\}$.
- *the series* $\sum_{k=1}^{\infty} \alpha_k^2 \|d_k\|^2$ *is convergent.*

3 Derivative-Free Spectral Algorithm for solving Underdetermined Nonlinear Equations (DF-SAUNE)

In this section, we develop a derivative-free method based on the algorithm DF-SANE [11] updated for the underdetermined case. DF-SANE is a derivative-free method for solving square nonlinear systems that uses the $n-$dimensional residual vector as a search direction together with a spectral step length and a globalization strategy that produces a nonmonotone process. The spectral coefficient is the inverse of an approximation of the Rayleigh quotient with respect to a secant approximation of the Jacobian:

$$\sigma_k = \frac{\langle s_k, s_k\rangle}{\langle y_k, s_k\rangle}, \tag{10}$$

where $y_k = F(x_k) - F(x_{k-1})$, $s_k = x_k - x_{k-1}$, see [2, 11].

The iterative process in DF-SANE can be viewed as a Quasi Newton method considering the matrix $B_k = \frac{1}{\sigma_k} I_n$ together with the iteration $x_{k+1} = x_k - \alpha_k B_k^{-1} F(x_k) = x_k - \alpha_k \sigma_k F(x_k)$ where α_k is computed using the derivative-free nonmonotone strategy (4).

In order to solve the underdetermined case we propose to combine the idea of the augmented Jacobian algorithm, see for example [29], with the spectral residual vector explained above. In order to do that we consider L_m as the ceil number of $\frac{n}{m}$, that is,

$$L_m = \begin{cases} \dfrac{n}{m} & \text{if } \dfrac{n}{m} \in \mathbb{N} \\ \left[\dfrac{n}{m}\right] + 1 & \text{if not.} \end{cases}$$

For each $j = 0, \ldots, L_m - 1$ we define the matrices $E_j \in \mathbb{R}^{m\times n}$ as follows. If $L_m = \frac{n}{m}$ then E_j is the matrix whose rows are the m canonical vectors $e_{jm+1}, e_{jm+2}, \ldots, e_{jm+m}$ in $\mathbb{R}^n$. If $L_m = [\frac{n}{m}] + 1$ then for $j = 0, \ldots, L_m - 2$, E_j is the same matrix defined above and E_{L_m-1} is the matrix whose rows are the m canonical vectors $e_{(L_m-1)m+1}, \ldots, e_n, e_1, \ldots, e_{L_m m-n}$ in $\mathbb{R}^n$.

Given x_k, for $j = 0, \ldots, L_m - 1$ we consider the direction $d_{k+j} \in \mathbb{R}^n$ as the solution to the square linear system

$$\frac{1}{\sigma_{k+j}} \begin{pmatrix} E_j \\ V_j \end{pmatrix} d_{k+j} = \begin{pmatrix} -F(x_{k+j}) \\ 0 \end{pmatrix} \tag{11}$$

where E_j was defined above, $V_j \in \mathbb{R}^{(n-m)\times n}$ is the matrix whose rows are the canonical vectors in $\mathbb{R}^n$ that span $\mathcal{N}(E_j)$ and σ_{k+j} is a spectral coefficient. Thus we can observe that $d_{k+j} = -\sigma_{k+j}\ E_j^T\ F(x_{k+j})$ and

$$\|d_{k+j}\|^2 = |\ \sigma_{k+j}\ |^2\ \|F(x_{k+j})\|^2. \tag{12}$$

Once the direction d_{k+j} is obtained, the line search is performed using Algorithm 1.

Observe that the system (11) is a square system that resembles the one used in DF-SANE for obtaining its current direction.

Given an arbitrary initial point $x_0 \in \mathbb{R}^n$, the algorithm that allows us to obtain the next iterate is given below:

Algorithm 3. *DF-SAUNE*

Given x_0, $F(x_0)$, $0 < \gamma < 1$, $0 < \tau_{\min} < \tau_{\max} < 1$, $\sigma_0 = 1$,
$0 < \sigma_{\min} < \sigma_{\max} < \infty$, $0 \leq \epsilon < 1$.
Set $k \leftarrow 0$.

Main Step: *Given* x_k, $F(x_k)$, σ_k.

(1) *If* $\|F(x_k)\| \leq\ \epsilon \max\{1, \|F(x_0)\|\}$, *stop.*

(2) *For* $j = 0 : L_m - 1$

- *Compute* $d_{k+j} = -\sigma_{k+j}\ E_j^T\ F(x_{k+j})$
- *Find* α_j *and* $x_{k+j+1} = x_{k+j} + \alpha_j d_{k+j}$ *using the Algorithm 1.*
- *If* $\|F(x_{k+j+1})\| \leq\ \epsilon \max\{1, \|F(x_0)\|\}$, *stop.*
- *Compute* $s = x_{k+j+1} - x_{k+j}$,
  $y = F(x_{k+j+1}) - F(x_{k+j})$ *and* $\sigma = \frac{\langle E_j s, E_j s\rangle}{\langle y, E_j s\rangle}$.
- *If* $|\sigma| \in [\sigma_{\min}, \sigma_{\max}]$ *define* $\sigma_{k+j+1} = \sigma$. *If not, choose* σ_{k+j+1} *such that* $|\sigma_{k+j+1}| \in [\sigma_{\min}, \sigma_{\max}]$

End

(3) *Update* $k \leftarrow k + L_m$ *and repeat the main step.*

End

By (9) we have that

$$\lim_{k\in K} \alpha_k^2 \|d_k\|^2 = 0.$$

Thus, using (12) and considering that each $|\sigma_k| \in [\sigma_{\min}, \sigma_{\max}]$ we obtain that

$$\lim_{k\in K} \alpha_k^2 f(x_k) = 0. \tag{13}$$

In the following Theorem we prove the main convergence result associated to DF-SAUNE Algorithm. The proof follows the idea of the Theorem 1 in [11] updated for the Algorithm 3.

Theorem 3.1. *Assume that $\{x_k\}_{k\in\mathbb{N}}$ is the sequence generated by Algorithm 3. Then, for every limit point x^* of $\{x_k\}_{k\in K}$ there exists an index $j_0 \in \{0, \ldots, L_m - 1\}$ such that*

$$\langle J(x^*)^T F(x^*), E_{j_0}^T F(x^*)\rangle = 0 \tag{14}$$

Proof. Let x^* be a limit point of $\{x_k\}_{k\in K}$. Thus, there is an infinite index set $K_1 \subset K$ such that

$$\lim_{k\in K_1} x_k = x^*.$$

Then, by (13),

$$\lim_{k\in K_1} \alpha_k^2 f(x_k) = 0. \tag{15}$$

We have two possibilities:

(1) The sequence $\{\alpha_k\}_{k\in K_1}$ does not tend to zero;

(2) The sequence $\{\alpha_k\}_{k\in K_1}$ tends to zero.

In the first case there exists an infinite sequence of indices $K_2 \subset K_1$ such that $\alpha_k \geq c > 0$ for all $k \in K_2$. Then, by (15),

$$\lim_{k\in K_2} f(x_k) = 0.$$

Since f is continuous this implies that $f(x^*) = 0$.

Suppose that case 2 happens. Once the direction d_k was computed, in the step 2 of the Algorithm 1, one tests the inequality

$$f(x_k + \alpha_+ d_k) \leq \bar{f}_k + \eta_k - 2\gamma\alpha_+^2\sigma_k^2 f(x_k). \tag{16}$$

If this inequality does not hold, one tests the inequality

$$f(x_k - \alpha_- d_k) \leq \bar{f}_k + \eta_k - 2\gamma\alpha_-^2\sigma_k^2 f(x_k). \tag{17}$$

The first trial points at (16)–(17) are $\alpha+ = \alpha_- = 1$. Since $\lim_{k\in K_1} \alpha_k = 0$, there exists $k_0 \in K_1$ such that $\alpha_k < 1$ for all $k \in K_1, k \geq k_0$. Therefore, for those iterations k, there are steps α_k^+ and α_k^- that do not satisfy (16) and (17) respectively for which

$$\lim_{k\in K_1} \alpha_k^+ = \lim_{k\in K_1} \alpha_k^- = 0.$$

So, for all $k \in K_1, k \geq k_0$, we have that

$$f(x_k + \alpha_k^+ d_k) > \bar{f}_k + \eta_k - 2\gamma(\alpha_k^+)^2\sigma_k^2 f(x_k), \tag{18}$$

$$f(x_k - \alpha_k^- d_k) > \bar{f}_k + \eta_k - 2\gamma(\alpha_k^-)^2\sigma_k^2 f(x_k). \tag{19}$$

Since $|\sigma_k| \in [\sigma_{\min}, \sigma_{\max}]$, we have that (18) implies

$$f(x_k + \alpha_k^+ d_k) > \bar{f}_k + \eta_k - \overline{\gamma}(\alpha_k^+)^2 f(x_k) \tag{20}$$

and (19) implies

$$f(x_k - \alpha_k^- d_k) > \bar{f}_k + \eta_k - \overline{\gamma}(\alpha_k^-)^2 f(x_k) \tag{21}$$

where $\overline{\gamma} = 2\gamma\sigma_{\max}^2$.

The inequality (20) implies that

$$f(x_k + \alpha_k^+ d_k) > f(x_k) - \overline{\gamma}(\alpha_k^+)^2 f(x_k).$$

So,

$$f(x_k + \alpha_k^+ d_k) - f(x_k) > -\overline{\gamma}(\alpha_k^+)^2 f(x_k).$$

By Proposition 2.4, $\{f(x_k)\}$ is a sequence bounded above by a constant $C > 0$. Thus,

$$f(x_k + \alpha_k^+ d_k) - f(x_k) \geq -\overline{\gamma} C(\alpha_k^+)^2 \tag{22}$$

which implies that

$$\|F(x_k + \alpha_k^+ d_k)\|^2 - \|F(x_k)\|^2 \geq -\overline{\gamma} C(\alpha_k^+)^2.$$

So,

$$\frac{\|F(x_k+\alpha_k^+ d_k)\|^2 - \|F(x_k)\|^2}{\alpha_k^+} \geq -\overline{\gamma} C \alpha_k^+.$$

By the Mean Value Theorem, there exists $\xi_k \in [0,1]$ such that

$$\langle g(x_k+\xi_k\alpha_k^+ d_k), d_k\rangle \geq -\overline{\gamma} C \alpha_k^+.$$

Since L_m is finite there exists $j_0 \in \{0,\ldots,L_m-1\}$, $K_2 \subset K_1$ such that, for all $k \in K_2$

$$d_k = -\sigma_k \begin{pmatrix} E_{j_0}^T & V_{j_0}^T \end{pmatrix} \begin{pmatrix} F(x_k) \\ 0 \end{pmatrix}.$$

Thus, we have that

$$-\sigma_k \left\langle g(x_k+\xi_k\alpha_k^+ d_k), \begin{pmatrix} E_{j_0}^T & V_{j_0}^T \end{pmatrix} \begin{pmatrix} F(x_k) \\ 0 \end{pmatrix} \right\rangle \geq -\overline{\gamma} C \alpha_k^+.$$

Now, if $\sigma_k > 0$ for infinitely many indices $k \in K_2$ the last inequality implies that, for $k \in K_2$, $k \geq k_0$

$$\left\langle g(x_k+\xi_k\alpha_k^+ d_k), \begin{pmatrix} E_{j_0}^T & V_{j_0}^T \end{pmatrix} \begin{pmatrix} F(x_k) \\ 0 \end{pmatrix} \right\rangle \leq \frac{\overline{\gamma} C \alpha_k^+}{\sigma_k} \leq \frac{\overline{\gamma} C \alpha_k^+}{\sigma_{\min}}. \quad (23)$$

Using (21) and proceeding in the same way, we obtain that, for $k \in K_2$, $k \geq k_0$,

$$\left\langle g(x_k-\xi_k^{'}\alpha_k^- d_k), \begin{pmatrix} E_{j_0}^T & V_{j_0}^T \end{pmatrix} \begin{pmatrix} F(x_k) \\ 0 \end{pmatrix} \right\rangle \geq -\frac{C\overline{\gamma}\alpha_k^-}{\sigma_k} \geq -\frac{C\overline{\gamma}\alpha_k^-}{\sigma_{\min}}. \quad (24)$$

for some $\xi_k^{'} \in [0,1]$.

Since $\{\sigma_k\}$ and $\{f(x_k)\}$ are bounded we have that $\{\|d_k\|\}$ is bounded.

Then, using that $\alpha_k^+ \to 0, \alpha_k^- \to 0$, and taking limits in (23) and (24), we obtain that

$$\left\langle g(x^*), \begin{pmatrix} E_{j_0}^T & V_{j_0}^T \end{pmatrix} \begin{pmatrix} F(x^*) \\ 0 \end{pmatrix} \right\rangle = 0.$$

If $\sigma_k < 0$ for infinitely many indices, proceeding in an analogous way, we deduce the same equation as we wanted to prove. □

Remark 3.2. We have proved that there exists an index $j_0 \in \{0, \ldots, L_m - 1\}$ such that the gradient of $\|F(x)\|^2$ at x^* is orthogonal to the residual $E_{j_0}^T F(x^*)$. Note that when $m = n$, the result of Theorem 3.1 coincides with the convergence result obtained in [11]. Thus, we can say that DF-SAUNE Algorithm is an extension of DF-SANE.

Corollary 3.3. *Assume that x^* is a limit point of the sequence $\{x_k\}_{k\in K}$ generated by Algorithm* 3 *and suppose that for all $j \in \{0, \ldots, L_m - 1\}$ we have that $\langle J(x^*)E_j^T v, v\rangle \neq 0$ for all $v \in \mathbb{R}^n,\ v \neq 0$. Then $F(x^*) = 0$.*

4 Derivative-Free Quasi Newton method using the Broyden update formula (DF-QNB)

In this section we will define a Quasi Newton method based on the Broyden update formula for the matrices with the derivative-free line search defined in Algorithm 1 and we will establish the convergence results.

We will define the search direction as an approximate solution to the linear system $B_k d = F(x_k)$ and we will use the nonsquare rank one Broyden update formula:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)s_k^T}{s_k^T\ s_k} \tag{25}$$

where $s_k = x_{k+1} - x_k$ and $y_k = F(x_{k+1}) - F(x_k)$.

As previously stated, the aim is to solve the linear system accurately when it is possible and in an approximate way in any other case, as considered in [14].

Algorithm 4. *DF-QNB*

Given x_0, $B_0 \in \mathbb{R}^{m\times n}$, $F(x_0)$, $0 < \gamma < 1$, $0 \leq \theta_0 < 1$, $\Delta > 0$, $0 < \tau_{\min} < \tau_{\max} < 1$, $0 \leq \epsilon < 1$.
Set $k \leftarrow 0$.

Step 1: *If $\|F(x_k)\| \leq \epsilon \max\{1, \|F(x_0)\|\}$ stop.*

Step 2: *Computing the direction d_k*

Step 2.1: *Find d such that*

$$B_k d + F(x_k) = 0 \quad \text{and} \quad \|d\| \leq \Delta. \tag{26}$$

If such direction d is found, define $d_k = d$, $\theta_{k+1} = \theta_k$ and go to Step 3.

Step 2.2: *Find a solution d of* $\min_{d \in \mathbb{R}^n} \|B_k d + F(x_k)\|$.

If d satisfies

$$\|B_k d + F(x_k)\| \leq \theta_k \|F(x_k)\| \quad \text{and} \quad \|d\| \leq \Delta \tag{27}$$

define $d_k = d$, $\theta_{k+1} = \theta_k$ and go to Step 3.
Else, set $d_k = 0$, $x_{k+1} = x_k$, $\theta_{k+1} = \frac{\theta_k + 1}{2}$ and go to Step 5.

Step 3: *Find α_k and $x_{k+1} = x_k + \alpha_k d_k$ using Algorithm 1.*

Step 4: *Update B_{k+1} using* (25).

Step 5: *Update $k \leftarrow k + 1$ and go to Step 1.*

In [14], the subproblems in Step 2 were defined using two different matrices.

Remark 4.1. When $m = n$ the Algorithm 4 is similar to the Algorithm defined in [13], in the sense that both of them use the Broyden update formula. In [13] the authors solve the linear system (2) by using the update (3). We acknowledge that, in large scale problems, it is difficult to solve linear systems, even when they do have a solution. Having that in mind we have tried to find an approximate solution in the sense of (27) for those cases.

The following theorems establish the necessary hypotheses for obtaining the convergence results.

Theorem 4.2. *Assume that Algorithm* 4 *generates an infinite sequence* $\{x_k\}$. *Suppose that*

$$\lim_{k \to \infty} \langle (B_k - J(x_k)) d_k, F(x_k) \rangle = 0 \tag{28}$$

and there exists $k_0 \in \mathbb{N}$ such that, for all $k \geq k_0$, $\theta_k = \overline{\theta} < 1$. Then, every limit point of $\{x_k\}_{k \in K}$ is a solution of (1), *where K is given by* (8).

Proof. Let x^* be a limit point of $\{x_k\}_{k\in K}$, then there exists $K_1 \subset K$ such that $\lim_{k\in K_1} x_k = x^*$.

We know that $\lim_{k\in K} \alpha_k^2 \|d_k\|^2 = 0$, so $\lim_{k\in K_1} \alpha_k^2 \|d_k\|^2 = 0$.

We will consider two cases: *a)* $\lim_{k\in K_1} \alpha_k \neq 0$ and *b)* $\lim_{k\in K_1} \alpha_k = 0$.

Let suppose the first case. Then, we have that

$$\lim_{k\in K_1} \|d_k\| = 0. \tag{29}$$

(1) First we assume that in the process of the Algorithm 4, the direction d_k was calculated finitely times solving the linear system $B_k d + F(x_k) = 0$.

Then, there exists $k_1 \in K_1$ such that $\forall k \geq k_1, k \in K_1$, d_k verifies the formula $\|B_k d + F(x_k)\| \leq \theta_k \|F(x_k)\|$. Thus, for all $k \in K_1$, $k \geq \max\{k_0, k_1\}$ we have that

$$\langle B_k d_k, F(x_k)\rangle \leq \frac{\overline{\theta}^2 - 1}{2} \|F(x_k)\|^2.$$

This implies that

$$\langle (B_k - J(x_k))d_k, F(x_k)\rangle + \langle J(x_k)d_k, F(x_k)\rangle \leq \frac{\overline{\theta}^2 - 1}{2} \|F(x_k)\|^2.$$

Taking limits for $k \in K_1, k \geq \max\{k_0, k_1\}$ in the last expression, and using the continuity of F and J and (28)–(29), we obtain that

$$0 \leq \frac{\overline{\theta}^2 - 1}{2} \|F(x^*)\|^2.$$

Since $\overline{\theta} < 1$, we have that $\|F(x^*)\| = 0$ as we wanted to prove.

(2) Let assume now that the direction d_k was obtained infinitely many times solving the linear system $B_k d + F(x_k) = 0$.

Then, there exists $K_2 \subset K_1$ such that for all $k \in K_2$ we have that $B_k d_k + F(x_k) = 0$.

Thus, for all $k \in K_2$,

$$\langle (B_k - J(x_k))d_k, F(x_k)\rangle + \langle J(x_k)d_k, F(x_k)\rangle = -\|F(x_k)\|^2.$$

Taking limits for $k \in K_2, k \geq k_0$ and using the continuity of F and J and (28)–(29), we obtain that

$$\|F(x^*)\| = 0.$$

Let suppose the case b). Since the sequence $\{d_k\}_{k\in\mathbb{N}}$ is bounded (and then it is bounded in K_1), there exists $K_2 \subset K_1$ and $\bar{d} \in \mathbb{R}^n$ such that $\lim_{k\in K_2} d_k = \overline{d}$.

In the line search, to choose the step α_k the algorithm DF-QNB tests the following inequalities

$$f(x_k + \alpha_+ d_k) \leq \overline{f}_k + \eta_k - \gamma\alpha_+^2\|d_k\|^2 \tag{30}$$

$$f(x_k - \alpha_- d_k) \leq \overline{f}_k + \eta_k - \gamma\alpha_-^2\|d_k\|^2. \tag{31}$$

The initial values of α_+ and α_- are 1. Since $\lim_{k\in K_2} \alpha_k = 0$, there exists $\overline{k} \in K_2$ such that $\alpha_k < 1$ for all $k \in K_2, k \geq \overline{k}$. Thus, for those iterations k there exist steps α_k^+ and α_k^- that do not satisfy (30) and (31) and $\lim_{k\in K_2} \alpha_k^+ = \lim_{k\in K_2} \alpha_k^- = 0$. So we have that

$$f(x_k + \alpha_k^+ d_k) > \overline{f}_k + \eta_k - \gamma(\alpha_k^+)^2\|d_k\|^2 \tag{32}$$

$$f(x_k - \alpha_k^- d_k) > \overline{f}_k + \eta_k - \gamma(\alpha_k^-)^2\|d_k\|^2. \tag{33}$$

Considering (32), the following inequality holds

$$f(x_k + \alpha_k^+ d_k) > \overline{f}_k + \eta_k - \gamma(\alpha_k^+)^2\|d_k\|^2 > f(x_k) - \gamma(\alpha_k^+)^2\|d_k\|^2.$$

Since $\|d_k\| \leq \Delta$ we obtain that

$$\frac{f(x_k + \alpha_k^+ d_k) - f(x_k)}{\alpha_k^+} > -\gamma\alpha_k^+\Delta^2.$$

By the Mean Value Theorem there exists $\xi_k^+ \in [0, 1]$ such that

$$\langle g(x_k + \xi_k^+\alpha_k^+ d_k), d_k\rangle > -\gamma\alpha_k^+\Delta^2. \tag{34}$$

Considering (33), we have that

$$f(x_k - \alpha_k^- d_k) > \overline{f}_k + \eta_k - \gamma(\alpha_k^-)^2\|d_k\|^2 > f(x_k) - \gamma(\alpha_k^-)^2\|d_k\|^2.$$

Since $\|d_k\| \leq \Delta$ we obtain that

$$\frac{f(x_k - \alpha_k^- d_k) - f(x_k)}{\alpha_k^-} > -\gamma \alpha_k^- \Delta^2.$$

By the Mean Value Theorem there exists $\xi_k^- \in [0, 1]$ such that

$$\langle g(x_k - \xi_k^- \alpha_k^- d_k), d_k \rangle < \gamma \alpha_k^- \Delta^2. \tag{35}$$

Taking limits in (34) and (35) when $k \to \infty$, $k \in K_2$, we obtain that

$$\langle g(x^*), \overline{d} \rangle = \langle J(x^*)^T F(x^*), \overline{d} \rangle = 0.$$

Thus,

$$\langle J(x^*)\, \overline{d}, F(x^*) \rangle = 0. \tag{36}$$

Now, as we did before, we have to consider two cases: (i) the direction d_k was calculated finitely many times solving the linear system $B_k d + F(x_k) = 0$; and (ii) d_k was obtained infinitely many times solving the linear system $B_k d + F(x_k) = 0$.

Proceeding in analogous way as we did when $\lim_{k \in K_1} \alpha_k \neq 0$ and using (28) and (36) we obtain that

$$\|F(x^*)\| = 0$$

as we wanted to prove. □

Observe that, according to Proposition 2.4, we have that $\|F(x_k)\|$ is bounded. Thus, if

$$\lim_{k \to \infty} \frac{\|(B_k - J(x_k))d_k\|}{\|d_k\|} = 0$$

we obtain the hypothesis (28). The last condition is known as a necessary and sufficient condition for obtaining q-superlinear convergence of classical Quasi Newton methods [6].

If we can not guarantee that $\theta_k = \bar{\theta} < 1$ for $k \geq k_0$ then we can prove the following result which is similar to one obtained in [14].

Theorem 4.3. *Suppose that, in Algorithm* 4, θ_k *is increased infinitely many times and define*

$$K_* = \{k \in \mathbb{N} \mid \theta_{k+1} > \theta_k\}.$$

Assume that

$$\lim_{k\in K_*} \|B_k - J(x_k)\| = 0. \tag{37}$$

Then, every limit point x^* of the sequence $\{x_k\}_{k\in K_*}$ is a solution of (1) *or it is a global minimizer of $\|F(x^*) + J(x^*)(x - x^*)\|$.*

Proof. If $F(x^*) = 0$ we are done. Let us assume that $\|F(x^*)\| > 0$. Suppose that x^* is not a global minimizer of $\|F(x^*) + J(x^*)(x - x^*)\|$, therefore, there exists d such that $\|d\| \leq \frac{\Delta}{2}$ and

$$\|F(x^*) + J(x^*)d\| < \|F(x^*)\|$$

thus

$$\frac{\|F(x^*) + J(x^*)d\|}{\|F(x^*)\|} \leq r < 1.$$

By (37) and the continuity of F and J, we have that

$$\frac{\|F(x_k) + B_k(x^* - x_k + d)\|}{\|F(x_k)\|} \leq \frac{r+1}{2} \tag{38}$$

for all k large enough $k \in K_*$. But, since $\|d\| \leq \frac{\Delta}{2}$ and $\lim_{k\in K_*} x_k = x^*$ we have that, for all large enough $k \in K_*$, $\|x^* - x_k + d\| \leq \Delta$. So, (38) contradicts the fact that: $\theta_k \to 1$ and a direction verifying (27) can be obtained. □

A point x^* that is a global minimizer of the function $\|F(x^*) + J(x^*)(x - x^*)\|$ can be viewed as the solution of the linear least squares problem

$$\min_{x\in\mathbb{R}^n} \|A(x - x^*) - b\| \tag{39}$$

where $A = J(x^*)$ and $b = -F(x^*)$. The linear function is the affine model of the function F around x^*. We can not expect, in general, to find x^* such that $F(x^*) = 0$ since the problem could not have a solution. Likewise, we can not expect to find x^* such that $F(x^*) + J(x^*)(x - x^*) = 0$ since this is an underdetermined linear system of equations and $J(x^*)$ could not have full rank. Because of that, it seems reasonable to find a global minimizer of (39) when the problem has no solutions.

4.1 Analysis of the case $M = 1$

In this subsection we analyze the case in which the derivative-free line search used in Algorithm DF-QNB considers $M = 1$. By the presence of η_k the line search is still nonmonotone but it imposes an almost monotone behavior of the merit function when x_k is close to a solution. Thus, $M = 1$ is plausible considering we are working with a Quasi Newton method. $M = 1$, as pointed out in [11], could be inconvenient for the spectral residual method because it performs highly nonmonotone even in a neighborhood of an isolated solution.

Next, we will demonstrate that, in this case, our algorithm verifies the assumption (28).

Firstly, under this case, using Proposition 2.5 it can be observed that

$$\sum_{k=1}^{\infty} \|s_k\|^2 = \sum_{k=1}^{\infty} \alpha_k^2 \|d_k\|^2 < \infty. \tag{40}$$

Secondly, it will be convenient to define the matrix

$$A_{k+1} = \int_0^1 J(x_k + ts_k)dt.$$

Thus, $A_{k+1}s_k = y_k$ and

$$B_{k+1} = B_k + \frac{(A_{k+1} - B_k)s_k s_k^T}{\|s_k\|^2}.$$

Finally, we can use the result that appears in Lemma 2.6 of [13] and that we present here for completeness.

Lemma 4.4 (Lemma 2.6, [13]). *Let us suppose that the set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0) + \eta\}$ is bounded and that $J(x)$ is Lipschitz continuous in Ω. If* (40) *is verified then*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \rho_i^2 = 0$$

where

$$\rho_k = \frac{\|(A_{k+1} - B_k)d_k\|}{\|d_k\|}.$$

In particular, there is a subsequence of $\{\rho_k\}$ tending to zero.

Thus, we can prove the following convergence result.

Theorem 4.5. *Assume that Algorithm* 4 *generates an infinite sequence* $\{x_k\}$*, that* $M = 1$ *in the line search and that the hypotheses of Lemma* 4.4 *hold. Then, if there exists* k_0 *such that* $\theta_k = \bar{\theta} < 1$ *for all* $k \geq k_0$*, there is a limit point* x^* *of* $\{x_k\}_{k\in\mathbb{N}}$ *that is a solution of* (1).

Proof. By Proposition 2.5 we have that (40) holds. Thus, by Lemma 4.4, there is a subset $\tilde{K} \subset \mathbb{N}$ such that

$$\lim_{k\in\tilde{K}} \frac{\|(A_{k+1} - B_k)d_k\|}{\|d_k\|} = 0. \tag{41}$$

Let x^* be a limit point of the subsequence $\{x_k\}_{k\in\tilde{K}}$. Note that

$$| \langle (B_k - J(x_k))d_k, F(x_k)\rangle | \leq \|(B_k - J(x_k))d_k\| \, \|F(x_k)\|$$

$$\leq (\|(A_{k+1} - B_k)d_k\| + \|(J(x_k) - A_{k+1})d_k\|)\|F(x_k)\|.$$

By Proposition 2.5 we have that $\|F(x_k)\|$ is bounded. Thus, taking limit when $k \in \tilde{K}, k \to \infty$ and using that $\{d_k\}$ is bounded and (41), we obtain that $\|(A_{k+1} - B_k)d_k\| \to 0$. Also, by definition of A_{k+1} we obtain that $\|(J(x_k) - A_{k+1})d_k\| \to 0$. Thus, we prove that (28) happens for $k \in \tilde{K}$ and the proof follows from Theorem 4.2. □

Observe that this particular line search improves the results of Theorem 4.2.

5 Numerical experiments

In this section we present some computational results obtained with a Fortran 77 implementation of DF-SAUNE and DF-QNB algorithms. All experiments were run on a personal computer with INTEL(R) Core (TM) 2 Duo CPU E8400 at 3.00 GHz and 3.23 GB of RAM.

As it is usual in derivative-free optimization articles we are interested in the number of function evaluations for both codes. We also report for medium size problems the CPU time obtained for both algorithms and we include a comparison with NEWUOA algorithm developed by M.J.D. Powell for unconstrained optimization [21]. In these experiments, NEWUOA algorithm solves the least squares problem: $\min \|F(x)\|^2$.

5.1 Test problems

The problems used for these numerical experiments, of the form $F(x) = 0$, $F : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, were a set of problems defined by feasible sets of nonlinear programming problems in Hock and Schittkowski [10], where the number of variables ranges from 2 to 10, and the number of equations from 1 to 4. Also, conceiving tests for larger dimension problems, we tested some problems described in [5]. Some of the test problems analyzed here have been previously examined in [7] for problems with bound constraints and using derivatives in order to achieve similar results to those pursued here; that is, to solve underdetermined nonlinear systems.

In Table 1 we show the data of the problems. In column 1 we show the number of the problem, in column 2 the source of the problem and in the last columns the number of equations (m) and variables (n).

Initial points were the same as in the cited references.

Remark 5.1. The case (a) in Problems 2 and 4 of [5] corresponds to the use of the initial point $x_0(1 : n) = 2$. The case (b) in Problem 2 and 4 of [5] corresponds to the use of $x_0(1 : n) = 150$ as initial point.

5.2 Implementation

In the implementations of DF-SAUNE and DF-QNB algorithms:

1. The parameters for Algorithm 1 were:

   $M = 2$, $\tau_{\min} = 0.1$, $\tau_{\max} = 0.5$, $\gamma = 10^{-4}$, $\eta_0 = 1$,

   - $\forall\, k \in \mathbb{N}, k \geq 1 \;\; \eta_k = \dfrac{\|F(x_0)\|}{2^k}$ for small size problems.
   - $\forall\, k \in \mathbb{N}, k \geq 1, \;\; \eta_k = \dfrac{\|F(x_0)\|}{(1+k)^2}$ for medium size problems.

2. The parameters for DF-SAUNE Algorithm were:

   $\epsilon = 10^{-6}$, $\sigma_0 = 1$, $\sigma_{\min} = 10^{-10}$, $\sigma_{\max} = 10^{10}$.

3. The parameters for DF-QNB Algorithm were:

   $\epsilon = 10^{-6}$, $\theta_0 = 0.95$, $\Delta = 10^{12}$.

Problem	Source	m	n
1	Problem 6 of [10]	1	2
2	Problem 7 of [10]	1	2
3	Problem 8 of [10]	2	2
4	Problem 26 of [10]	1	3
5	Problem 27 of [10]	1	3
6	Problem 39 of [10]	2	4
7	Problem 40 of [10]	3	4
8	Problem 42 of [10]	2	4
9	Problem 46 of [10]	2	5
10	Problem 47 of [10]	3	5
11	Problem 48 of [10]	2	5
12	Problem 53 of [10]	3	5
13	Problem 56 of [10]	4	7
14	Problem 61 of [10]	2	3
15	Problem 63 of [10]	2	3
16	Problem 77 of [10]	2	5
17	Problem 78 of [10]	3	5
18	Problem 79 of [10]	3	5
19	Problem 81 of [10]	3	5
20	Problem 111 of [10]	3	10
21	Problem 2(a) of [5]	150	300
22	Problem 2(b) of [5]	150	300
23	Problem 4(a) of [5]	150	300
24	Problem 4(b) of [5]	150	300
25	Problem ARWHEAD of [21]	149	150
26	Problem ARWHEAD of [21]	299	300

Table 1 – Data of the problems.

4. In DF-QNB, the first B_0 matrix was computed by finite differences as an aproximation to the Jacobian matrix in x_0.
5. In DF-QNB, we have used ACCIM algorithm described in [27] to find a solution to $B_k d = -F(x_k)$. For solving the least squares problem $\min_{\|d\| \leq \Delta} \|B_k d + F(x_k)\|^2$ we have used BVLS algorithm described in [28].
6. The stopping condition for the two new algorithms was: $\|F(x_k)\| \leq 10^{-6} \max\{1, \|F(x_0)\|\}$.

7. The maximum number of function evaluations allowed was:

   - $MAX_{FE} = 5000$, for small size problems,
   - $MAX_{FE} = 15000$, when $n = 150$,
   - $MAX_{FE} = 30000$, when $n = 300$.

   For DF-QNB method we have also added the required evaluations to calculate the initial matrix B_0.

The implementation of NEWUOA is the original version of M.J.D. Powell [21] with its stopping criterion, that is, the algorithm stops when the trust-region radius is lower than a tolerance $\rho_{end} = 10^{-6}$. As previously mentioned, NEWUOA algorithm solved the least squares problem min $\|F(x)\|^2$ in our trials.

5.3 Numerical results

In Table 2 we show the results obtained by DF-SAUNE (DF-S) and DF-QNB (DF-B) algorithms taking into account the final value $\|F(x_{end})\|$ and the number function evaluations. The results correspond to the stopping criterion satisfaction or to internal conditions that do not allow further improvement.

Table 2 also shows the number of problems (column 1), the number of iterations (Iter, column 2), the number of function evaluations (Feval, column 3), and the final functional values obtained for both codes ($\|F(x_{end})\|$, column 4).

These results illustrate DF-QNB effectiveness, although DF-SAUNE has also a satisfactory behavior. We can see in 15 out of the 20 test problems DF-QNB used less function evaluations than DF-SAUNE. It is also worth mentioning that when DF-QNB requires more function evaluations than DF-SAUNE such difference becomes particulary significant. These problems are too small to consider useful showing CPU time readings.

In problem 5, we have seen that, in many interations the norm of the solution d_k of the linear system $B_k d + F(x_k) = 0$ is bigger than Δ and, in those iterations, DF-QNB has to solve subproblem (27). Because of that, the decrease of the merit function $\|F(x)\|$ is very slow and the algorithm requires many functional evaluations.

In problem 19, DF-SAUNE, stopped without obtaining a satisfactory decrease in the residual when reaching the maximum number of function evaluations

Problem	Iter		Feval		$\|F(x_{end})\|$	
	DF-S	DF-B	DF-S	DF-B	DF-S	DF-B
1	83	4	85	7	2.555333D-08	1.460382D-08
2	44	62	70	844	5.616128D-08	8.074138D-06
3	52	10	54	19	1.852329D-07	1.029118D-06
4	49	61	85	929	1.487057D-07	1.188074D-07
5	1	251	2	4815	0.0D+00	1.341756D-06
6	57	30	105	76	5.707520D-08	3.311382D-08
7	122	5	327	10	8.140108D-09	7.144642D-09
8	56	15	137	20	9.481413D-09	8.774760D-07
9	143	19	448	25	9.368079D-08	8.287063D-06
10	80	8	133	14	1.483448D-08	6.553991D-07
11	1	1	4	7	0.0D+00	5.264796D-06
12	1	1	3	7	0.0D+00	1.026234D-09
13	92	5	163	13	5.098483D-08	3.366883D-07
14	101	10	206	22	3.382254D-08	1.863196D-09
15	67	7	134	16	7.382254D-08	2.947364D-07
16	54	10	125	16	3.911728D-07	2.765711D-07
17	318	6	1176	12	2.243582D-06	1.839195D-07
18	68	7	169	13	6.362777D-08	1.753428D-06
19	696	8	5000	14	1.137512D-01	1.641720D-08
20	132	12	461	29	1.200486D-08	4.219128D-09

Table 2 – Small size problems.

allowed. We believe that the bad perfomance of DF-SAUNE it is related to the strategy used to define the matrices E_j when $\frac{n}{m} \notin \mathbb{N}$. We think that it is an interesting future work to study a better strategy to complete the last matrix E_j in that specific case.

In Table 3 we show the number of iterations, number of function evaluations and the CPU time in seconds required by DF-SAUNE, DF-QNB and NEWUOA algorithms running the medium size problems 21, 22, 23, 24, 25 and 26. In that table we indicate the number of equations (m) and variables (n) of the problems.

Firstly, an overall review of the numerical results shows that final residual values are similar for all tested methods.

Secondly, we observe DF-SAUNE performed a more significant amount of function evaluations in the last two problems. It should be highlighted that CPU

Problem	Method	Iter	Feval	$\|F(x_{end})\|$	CPU
21	DF-SAUNE	105	123	1.096064D-06	0.01
	DF-QNB	2	303	4.544647D-11	0.05
	NEWUOA	3768	5765	4.681655D-10	245.12
22	DF-SAUNE	440	566	1.656464D-08	0.04
	DF-QNB	1	302	2.002649D-05	0.05
	NEWUOA	4010	6043	4.284213D-10	270.10
23	DF-SAUNE	147	185	5.481636D-06	0.01
	DF-QNB	128	649	4.283559D-06	2.43
	NEWUOA	15472	30000	8.203095D-07	1286.79
24	DF-SAUNE	94	122	9.933035D-07	0.01
	DF-QNB	394	731	7.489404D-05	9.67
	NEWUOA	18874	29205	1.334399D-07	1248.00
25	DF-SAUNE	3476	9848	2.696620D-07	0.34
	DF-QNB	13	164	3.628420D-05	4.23
	NEWUOA	7533	15000	1.092628D-09	124.15
26	DF-SAUNE	3001	8664	1.393098D-06	0.33
	DF-QNB	13	314	5.139963D-05	52.20
	NEWUOA	15072	30000	1.453638D-08	1237.54

Table 3 – Medium size problems.

time for DF-SAUNE was always substantially shorter than the one for DF-QNB. The reason is that DF-QNB algorithm has to solve a linear system of equations or a least squares problem in every iteration.

Finally we ran the well known NEWUOA solver in order to measure the number of function evaluations of our algorithms. Although NEWUOA was designed to solve unconstrained optimization problems, we can conclude that our methods performs satisfactorily.

6 Conclusions

Many practical optimization methods require specific algorithms for obtaining feasible points at every iteration. Thus, our aim in this paper was to define algorithms capable of dealing with the feasible set defined by nonsquare system of equations. We present two derivative-free algorithms that exploit this structure: one of them is based on the spectral residual approach and the other

on the Broyden Quasi Newton method. The algorithms can be viewed as generalizations of the algorithms defined in [11] and [13], the last one combined with [14].

From a theoretical point of view we were able to obtain some convergence results. Under usual assumptions on the Jacobian matrix we establish global convergence of the scheme that uses the spectral residual idea. This convergence result can be seen as the underdetermined counterpart of the result presented in [11] for the square case.

For the Broyden Quasi Newton method we obtain global convergence under a Dennis Moré type condition. We have shown that this condition can be dropped out for a particular line search. It remains a challenge to reduce the restrictive hypotheses required in Theorem 4.3 of Section 4.

Numerical experiments suggest that the algorithms behave as expected. We consider both approaches are promising even though we also believe that it is necessary to test a more challenging set of problems in order to decide which is more suitable. Furthermore, such decision could depend on the requirements of each user.

Since many nonlinear programming problems have also box constraints, future research will consider the extension of this type of algorithms to solve underdetermined nonlinear systems with bound constraints.

Acknowledgements. The authors are indebted to two anonymous referees and to the Guest Editor for their careful reading of this paper.

REFERENCES

[1] J. Abadie, J. Carpentier, *Generalization of the Wolfe Reduced-Gradient Method to the Case of Nonlinear Constraints*. Optimization, Edited by R. Fletcher, Academic Press, New York (1968), 37–47.

[2] J. Barzilai and J.M. Borwein, *Two-point stepsize gradient methods*. IMA J. of Numerical Analysis, **8** (1988), 141–148.

[3] R.H. Bielschowsky and F.A.M. Gomes, *Dynamic control of infeasibility in equality constrained optimization*. SIAM J. on Opt., **19** (2008), 1299-1325.

[4] E.G. Birgin, N. Krejić and J.M. Martínez, *Globally convergent inexact quasi-Newton methods for solving nonlinear systems*. Numerical Algorithms, **32** (2003), 249–260.

[5] N. Dan, N. Yamashita and M. Fukushima, *Convergence properties of inexact Levenberg-Marquardt method under local error bound conditions*. Opt. Methods and Software, **17** (2002), 605–626.

[6] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, 1983.

[7] J.B. Francisco, N. Krejić and J.M. Martínez, *An interior point method for solving box-constrained underdetermined nonlinear systems*. J. of Comp. and Applied Math., **177** (2005), 67–88.

[8] N.I.M. Gould and Ph. L. Toint, *Nonlinear programming without a penalty function or a filter*. Math. Prog., **122** (2010), 155–196.

[9] L. Grippo, F. Lampariello and S. Lucidi, *A nonmonotone line search technique for Newton's method*. SIAM J. on Numerical Analysis, **23** (1986), 707–716.

[10] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*. Springer Series Lectures Notes in Economics Mathematical Systems, (1981).

[11] W. La Cruz, J.M. Martínez and M. Raydan, *Spectral residual method without gradient information for solving large-scale nonlinear systems of equations*. Mathematics of Computation, **75** (2006), 1429–1448.

[12] L.S. Lasdon, *Reduced Gradient Methods*. Nonlinear Optimization 1981, edited by M.J.D. Powell, Academic Press, New York (1982), 235–242.

[13] D.H. Li and M. Fukushima, *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*. Opt. Meth. and Software, **13** (2000), 181–201.

[14] J.M. Martínez, *Quasi-Inexact Newton methods with global convergence for solving constrained nonlinear systems*. Nonlinear Analysis, **30** (1997), 1–7.

[15] J.M. Martínez, *Inexact Restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming*. J. of Opt. Theory and Applications, **111** (2001), 39–58.

[16] J.M. Martínez, *Two-Phase Model Algorithm with Global Convergence for Nonlinear Programming*. J. of Opt. Theory and Applications, **96** (1998), 397–436.

[17] J.M. Martínez and E.A. Pilotta, *Inexact Restoration algorithms for constrained optimization*. J. of Opt. Theory and Applications, **104** (2000), 135–163.

[18] A. Miele, H.Y. Huang and J.C. Heideman, *Sequential Gradient-Restoration Algorithm for the Minimization of Constrained Functions, Ordinary and Conjugate Gradient Version*. J. of Opt. Theory and Applications, **4** (1969), 213–246.

[19] A. Miele, A.V. Levy and E.E. Cragg, *Modifications and Extensions of the Conjugate-Gradient Restoration Algorithm for Mathematical Programming Problems*. J. of Opt. Theory and Applications, **7** (1971), 450–472.

[20] A. Miele, E.M. Sims and V.K. Basapur, *Sequential Gradient-Restoration Algorithm for Mathematical Programming Problems with Inequality Constraints, Part 1, Theory*. Rice University, Aero-Astronautics Report No. 168 (1983).

[21] M.J.D. Powell, *The NEWUOA software for unconstrained optimization without derivatives*. Nonconvex Opt. and its Applications, **83** (2006), 255–297.

[22] M. Rom and M. Avriel, *Properties of the Sequential Gradient-Restoration Algorithm (SGRA), Part 1: Introduction and Comparison with Related Methods*. J. of Opt. Theory and Applications, **62** (1989), 77–98.

[23] M. Rom and M. Avriel *Properties of the Sequential Gradient-Restoration Algorithm (SGRA), Part 2: Convergence Analysis*. J. of Opt. Theory and Applications, **62** (1989), 99–126.

[24] J.B. Rosen, *The Gradient Projection Method for Nonlinear Programming, Part 1, Linear Constraints*. SIAM J. on Applied Math., **8** (1960), 181–217.

[25] J.B. Rosen, *The Gradient Projection Method for Nonlinear Programming, Part 2, Nonlinear Constraints*. SIAM J. on Applied Math., **9** (1961), 514–532.

[26] J.B. Rosen, *Two-Phase Algorithm for Nonlinear Constraint Problems*. Nonlinear Prog. 3, Edited by O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press, London and New York (1978), 97–124.

[27] H.D. Scolnik, N. Echebest, M.T. Guardarucci and M.C. Vacchino, *Incomplete Oblique Projections for Solving Large Inconsistent Linear Systems*. Math. Prog. B, **111** (2008), 273–300.

[28] P.B. Stark and R.L. Parker, *Bounded Variable Least Squares: An Algorithm and Applications*. J. Computational Statistics, **10** (1995), 129–141.

[29] H.F. Walker and L.T. Watson, *Least-Change secant update methods for underdetermined systems*. SIAM J. on Numerical Analysis, **27** (1990), 1227–1262.