Gauss-Newton methods with approximate projections for solving constrained nonlinear least squares problems

Max L. N. Gonçalves [*] Tiago C. Menezes[†]

January 9, 2020

Abstract

This paper is concerned with algorithms for solving constrained nonlinear least squares problems. We first propose a local Gauss-Newton method with approximate projections for solving the aforementioned problems and study, by using a general majorant condition, its convergence results, including results on its rate. By combining the latter method and a nonmonotone line search strategy, we then propose a global algorithm and analyze its convergence results. Finally, some preliminary numerical experiments are reported in order to illustrate the advantages of the new schemes.

Keywords: Constrained nonlinear least squares problems; approximate projection; Gauss-Newton methods; nonmonotone line search; global convergence.

1 Introduction

This paper addresses the numerical solution of the constrained nonlinear least squares problem

$$\min_{x\in C} G(x) := \frac{1}{2}\|F(x)\|^2, \tag{1}$$

where $\mathbb{X}$ and $\mathbb{Y}$ are real or complex Hilbert spaces, $\Omega \subseteq \mathbb{X}$ an open set containing the nonempty convex closed set C and $F:\Omega\to\mathbb{Y}$ is a continuously differentiable nonlinear function such that F' has a closed image in Ω. The constraint set C in (1) may naturally arise in order to exclude solutions of the model with no physical meaning, or it may be considered artificially due to some knowledge about the problem itself (see, e.g., [1, 2, 4, 24] for some applications which can be reformulated as a constrained nonlinear least squares problem).

*IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil. (E-mails: maxlng@ufg.br and tiagmenezes1@gmail.com). The work of these authors was supported in part by CAPES, FAPEG/CNPq/PRONEM-201710267000532, and CNPq Grants 302666/2017-6 and 408123/2018-4.

In the unconstrained case (i.e., $C = \Omega$), the Gauss-Newton method and its variations are the most efficient methods known to solve (1). If $F'(x)$ is injective and has closed image for all $x \in \Omega$, the standard Gauss-Newton method generates a sequence $\{x_k\}$ as follows: Given an initial point $x_0 \in \Omega$, define

$$x_{k+1} = x_k + S_k, \qquad F'(x_k)^* F'(x_k) S_k = -F'(x_k)^* F(x_k), \qquad k = 0, 1, \ldots,$$

where A^* denotes the adjoint of the operator A. Works dealing with the convergence of unconstrained Gauss-Newton methods include, for example, [9, 10, 14, 20].

On the other hand, when $F'(x)$ is injective and has closed image for all $x \in \Omega$, in order to solve the harder (constrained) case, a proximal Gauss-Newton method for solving a more general class of constrained nonlinear least squares problems was proposed in [26]. The proximal Gauss-Newton method specified for (1) generates a sequence $\{x_k\}$ defined as

$$x_{k+1} = P_C^{H_k}(x_k - [F'(x_k)^* F'(x_k)]^{-1} F'(x_k)^* F(x_k)), \qquad k = 0, 1, \ldots, \tag{2}$$

where $P_C^{H_k}$ is the projection operator with respect to the metric defined by the operator $H_k := F'(x_k)^* F'(x_k)$, i.e., for every $y \in \mathbb{X}$,

$$P_C^{H_k}(y) = \operatorname{argmin}_{x \in C} \frac{1}{2} \|x - y\|_{H_k}^2. \tag{3}$$

Under the assumption that F' is Lipschitz continuous, local convergence results of the proximal Gauss-Newton method were established in [26]. Moreover, some numerical experiments were presented showing the effectiveness of the method. A regularized version and some local convergence results under a more general condition on F' of the proximal Gauss-Newton method were also studied in [13] and [3], respectively. It is worth pointing out that different algorithms have been proposed and studied in the literature for solving (1). Strategies based on sequential quadratic programming, quasi-Newton and trust-region methods have been used; see, for instance, [21, 22, 25].

Depending on the application, the exact projection in (3) can be extremely difficult to obtain. Consequently, the first goal of this paper is to propose an extension of the algorithm in (2) in which inexact projections can be admitted. Toward this goal, we introduce a concept of approximate projection which, if necessary, can be efficiently computed by an iterative method; see Definition 1 and the remarks after it. Hence, the first method to be proposed here basically consists of computing the unconstrained Gauss-Newton step, and then an approximate projection of it, with respect to the metric defined by $H_k = F'(x_k)^* F'(x_k)$ onto C. From the theoretical viewpoint, we provide an estimate of the convergence radius, for which well-definedness and convergence of the method are ensured. Furthermore, results on its convergence rates are also established. Our analysis is done by using a majorant condition, which allows us to study convergence results of Newton and Gauss-Newton methods in a unified way; see, for example, [9, 10, 14]. Thus, our local analysis covers two large families of nonlinear functions, namely, one satisfying a Lipschitz condition and another one satisfying a Smale condition, which includes a substantial class of analytic functions.

Another issue in [26] is that no globalization strategy was considered. However, as it is well-known, strategies of globalization become, in general, more robustness methods. Therefore, the second goal of this paper is to propose a global version of our first method. Our globalization technique is based on the efficient nonmonotone line search in [18]. As has been reported by many authors, the nonmonotone strategy has been shown more efficient due to the fact that enforcing monotonicity of the function values may make the method converge slower. Under mild assumptions, we prove that any accumulation point of the sequence generated by our global method is a stationary point of (1).

In order to illustrate the robustness and effectiveness of the new schemes, we report some numerical experiments on a set of box- and polyhedral-constrained nonlinear systems and compare their performances with the proximal Gauss-Newton method in [26]. In the box-constrained case, we also compare performances of the new methods with the inexact Gauss–Newton trust-region method in [25].

The organization of the paper is as follows. In Section 2, we list some notations and basic results used in our presentation. A concept of approximate projection and some of its properties are discussed in Subsection 2.1. Section 3 describes the Gauss-Newton method with approximate projections (GNM-AP) and presents its main local convergence theorem, whose proof is postponed to Subsection 3.1. Also in Section 3, two applications of the main theorem are presented. In Section 4, we propose and analyze a global version of the GNM-AP. Finally, Section 5 presents some numerical experiments of the proposed schemes.

2 Notation and preliminary results

This section presents some definitions, notation and basic results used in this paper.

The open ball in $\mathbb{X}$ with center a and radius r is denoted by $B(a,r)$. For simplicity, given $x \in \mathbb{X}$, we use the short notation

$$\sigma(x) := \|x - x_*\|.$$

Denote $D^+ f(0)$ as the left-hand derivative of a convex function $f : [0, \mathbb{R}) \to \mathbb{R}$. We use $\mathcal{L}(\mathbb{X},\mathbb{Y})$ to denote the space of bounded linear operators from $\mathbb{X}$ to $\mathbb{Y}$ and $\mathrm{I}_{\mathbb{X}}$ corresponds to the identity operator on $\mathbb{X}$. If $A \in \mathcal{L}(\mathbb{X},\mathbb{Y})$, then $Ker(A)$ and $im(A)$ are the kernel and image of A, respectively, and A^* its adjoint operator. Let $H : \mathbb{X} \to \mathbb{X}$ be a continuously, positive definite and self-adjoint, bounded from below and, therefore, invertible operator. Then, we have a new scalar product on $\mathbb{X}$ by setting $\langle x,z\rangle_H = \langle x,Hz\rangle$. Hence, the corresponding induced norm $\|.\|_H$ is equivalent to the given norm on $\mathbb{X}$, since the following inequalities hold

$$\frac{1}{\|H^{-1}\|}\|x\|^2 \le \|x\|_H^2 \le \|H\|\|x\|^2. \tag{4}$$

Let $A \in \mathcal{L}(\mathbb{X},\mathbb{Y})$ with a closed image. The Moore-Penrose inverse of A is the linear operator $A^\dagger \in \mathcal{L}(\mathbb{Y},\mathbb{X})$ which satisfies:

$$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^* = AA^\dagger, \quad (A^\dagger A)^* = A^\dagger A.$$

From definition of the Moore-Penrose inverse, it is easy to see that

$$A^\dagger A = \mathrm{I}_{\mathbb{X}} - \Pi_{Ker(A)}, \qquad AA^\dagger = \Pi_{im(A)}, \tag{5}$$

where Π_E denotes the projection of $\mathbb{X}$ onto subspace E.

If A is injective or A^*A is invertible in $\mathcal{L}(\mathbb{X},\mathbb{X})$, then

$$A^\dagger = (A^*A)^{-1}A^*, \qquad A^\dagger A = I_{\mathbb{X}}, \qquad \|A^\dagger\|^2 = \|(A^*A)^{-1}\|. \tag{6}$$

We end this section by recalling that a point $x_* \in C$ is a stationary point of (1) iff

$$\langle F'(x_*)^* F(x_*), x - x_* \rangle \geq 0, \quad \forall\, x \in C. \tag{7}$$

2.1 Approximate projections

In this section, we introduce a concept of approximate projection and establish some useful properties, which will be fundamental in the course of this work. It is worth pointing out that for some sets, computing the exact projection onto them can be extremely difficult.

Definition 1. *Let $H : \mathbb{X} \to \mathbb{X}$ be a self-adjoint and positive definite operator. For given $x \in \mathbb{X}$ and $\varepsilon \geq 0$, we say that $\tilde{P}_C^H(x)$ is an ε–approximate projection of x onto C iff*

$$\tilde{P}_C^H(x) \in C \qquad \text{and} \qquad \langle x - \tilde{P}_C^H(x), y - \tilde{P}_C^H(x) \rangle_H \leq \varepsilon, \quad \forall\, y \in C. \tag{8}$$

If necessary (depending on definitions of C and H), an iterative method can be used in order to obtain an approximate projection in the sense of Definition 1; for example, when C is bounded, one can use the conditional gradient method [8, 12]. Indeed, given $z_0 \in C$, the conditional gradient method, applied to $\min_{y\in C} \|y-x\|_H^2/2$, generates a sequence $\{z_j\} \subset C$, where $z_{j+1} = z_j + \alpha_j(\bar{z}_j - z_j)$, with $\alpha_j \in (0,1)$, and $\bar{z}_j$ is the solution of subproblem

$$\begin{aligned} \min \quad & \langle z_j - x, z - z_j \rangle_H \\ \text{s.t.} \quad & z \in C. \end{aligned} \tag{9}$$

If this procedure is stopped when

$$\langle z_j - x, \bar{z}_j - z_j \rangle_H \geq -\varepsilon,$$

then condition (8) holds with $\tilde{P}_C^H(x) = z_j$.

Note that, if $\tilde{P}_C^H(x)$ is a *zero*–approximate projection of y onto C, then (8) implies that it is the unique exact solution for the generalized projection problem

$$\min_{y\in C} \frac{1}{2}\|y-x\|_H^2.$$

We will denote this unique exact projection by $P_C^H(x)$. It is easy to prove that the operator $P_C^H(\cdot)$ is nonexpansive in the norm $\|\cdot\|_H$, i.e.

$$\|P_C^H(x) - P_C^H(\hat{x})\|_H \leq \|x - \hat{x}\|_H, \quad x, \hat{x} \in \mathbb{X}. \tag{10}$$

Moreover, for every $x \in \mathbb{X}$ and $\varepsilon \geq 0$, the following relationship between P_C^H and $\tilde{P}_C^H$ holds

$$\|\tilde{P}_C^H(x) - P_C^H(x)\|_H \leq \sqrt{\varepsilon}. \tag{11}$$

Indeed, since $P_C^H(x) \in C$ and $\tilde{P}_C^H(x) \in C$, it follows from Definition (1) that

$$\langle \tilde{P}_C^H(x) - x, \tilde{P}_C^H(x) - P_C^H(x) \rangle_H \leq \varepsilon, \qquad \langle x - P_C^H(x), \tilde{P}_C^H(x) - P_C^H(x) \rangle_H \leq 0$$

The statement follows now by adding the last two inequalities. In Figure (1), an admissible approximation of $P_C^H(x)$ with $H \equiv I_{\mathbb{X}}$ is depicted.

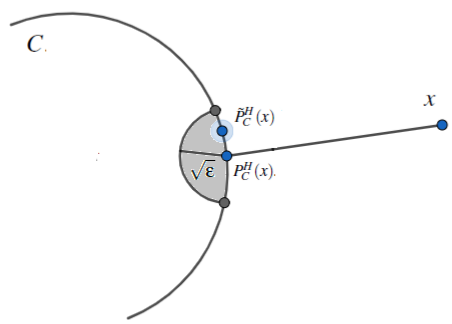

Figure 1: ε–approximate projection

Note also that $P_C^{H_*}(x_* - F'(x_*)^\dagger F(x_*)) = x_*$ iff x_* is a stationary point of (1) (i.e. $\langle F'(x_*)^* F(x_*), x - x_* \rangle \geq 0$, for all $x \in C$). In order to investigate the convergence of the method to be proposed here, we need to establish some relationships between exact and inexact projections when H varies.

Lemma 1. *Let H and H_* be the two continuous, self-adjoint and positive definite operators on $\mathbb{X}$. Then,*

$$\|P_C^H(x) - P_C^{H_*}(x)\|_H \leq \|H^{-1}\|^{1/2} \|(H - H_*)(P_C^{H_*}(x) - x)\|, \quad \forall x \in \mathbb{X}.$$

Proof. Denote $z = P_C^H(x)$ and $\hat{z} = P_C^{H_*}(x)$. Hence, it follows from Definition (1) that

$$\langle H(z - x), \hat{z} - z \rangle \geq 0, \qquad \langle H_*(\hat{z} - x), z - \hat{z} \rangle \geq 0. \tag{12}$$

Combining the last two inequalities, we obtain

$$\langle H(z - \hat{z}), z - \hat{z} \rangle \leq \langle (H_* - H)(\hat{z} - x), z - \hat{z} \rangle,$$

which, combined with Cauchy-Schwarz inequality, yields

$$\|z - \hat{z}\|_H^2 \leq \|(H - H_*)(\hat{z} - x)\| \|z - \hat{z}\| \leq \|H^{-1}\|^{1/2} \|(H - H_*)(\hat{z} - x)\| \|z - \hat{z}\|_H.$$

Therefore, the desired inequality now follows from last one. □

Lemma 2. *Let H and H_* be the two continuous, self-adjoint and positive definite operators on $\mathbb{X}$. Then, for every $x, \hat{x} \in \mathbb{X}$ and $\varepsilon \geq 0$, we have*

$$\|\tilde{P}_C^H(x) - P_C^{H_*}(\hat{x})\|_H \leq \|x-\hat{x}\|_H + \|H^{-1}\|^{1/2}\|(H-H_*)(P_C^{H_*}(\hat{x})-\hat{x})\| + \sqrt{\varepsilon}.$$

Proof. By Lemma (1), we obtain

$$\begin{aligned}\|\tilde{P}_C^H(x) - P_C^{H_*}(\hat{x})\|_H &\leq \|\tilde{P}_C^H(x) - P_C^H(x)\|_H + \|P_C^H(x) - P_C^H(\hat{x})\|_H + \|P_C^H(\hat{x}) - P_C^{H_*}(\hat{x})\|_H \\ &\leq \|\tilde{P}_C^H(x) - P_C^H(x)\|_H + \|P_C^H(x) - P_C^H(\hat{x})\|_H + \|H^{-1}\|^{1/2}\|(H-H_*)(P_C^{H_*}(\hat{x})-\hat{x})\|.\end{aligned}$$

Combining last inequality with (10) and (11), we find

$$\|\tilde{P}_C^H(x) - P_C^{H_*}(\hat{x})\|_H \leq \sqrt{\varepsilon} + \|x-\hat{x}\|_H + \|H^{-1}\|^{1/2}\|(H-H_*)(P_C^{H_*}(\hat{x})-\hat{x})\|,$$

which is equivalent to the desired inequality. □

3 The method and its local convergence

This section describes and investigates a Gauss-Newton method with approximate projections (GNM-AP) for solving (1). Basically, the method consists of computing the Gauss-Newton step and then an approximate projection of it onto the feasible set C. The GNM-AP is formally described as follows.

Gauss-Newton method with approximate projections (GNM-AP)

Step 0 (Initialization). Let $x_0 \in C$, $\{\theta_k\} \subset [0, \infty)$ be given, and set $k = 0$.

Step 1 (Gauss-Newton step). Compute $s_k \in \mathbb{X}$ and $y_k \in \mathbb{X}$ such that

$$F'(x_k)^*F'(x_k)s_k = -F'(x_k)^*F(x_k), \quad y_k = x_k + s_k. \tag{13}$$

Step 2 (Computation of new iterative). Define $H_k = F'(x_k)^*F'(x_k)$. Compute $x_{k+1} \in C$ such that

$$\langle y_k - x_{k+1}, x - x_{k+1}\rangle_{H_k} \leq \varepsilon_k := \theta_k^2\|x_{k+1} - x_k\|_{H_k}^2, \quad \forall\, x \in C, \tag{14}$$

i.e., x_{k+1} is an ε_k–approximate projection of y_k onto C.

Step 3 (Termination criterion and update). If $x_{k+1} = x_k$, then **stop**; Otherwise, set $k \leftarrow k+1$ and go to step 1.

end

Remark 1. *(i) Since the Gauss-Newton step y_k may be infeasible for the constraint set C, it is necessary to compute an ε_k–approximate projection of it onto C. As already mentioned, such an approximate projection can be efficiently computed, for example, by the conditional gradient method when C is bounded. Indeed, we can apply the conditional gradient method to $\min_{x\in C}\frac{1}{2}\|x-y_k\|^2_{H_k}$ in order to obtain a point x_{k+1} satisfying* (14)*. We refer the reader to [16, 17] for some implementations of the conditional gradient method for computing approximate projections. (ii) In Step 3, if $x_{k+1}=x_k$, it follows from Step 2 and* (13) *that*

$$0 \geq \langle y_k - x_{k+1}, x - x_{k+1}\rangle_{H_k} = \langle -[F'(x_k)^*F'(x_k)]^{-1}F'(x_k)^*F(x_k), x-x_k\rangle_{H_k} = \langle -F'(x_k)^*F(x_k), x-x_k\rangle,$$

for all $x \in C$, i.e. x_k is a stationary point of (1)*. (iii) The definition of x_{k+1} as an approximate projection of the unconstrained Gauss-Newton step with respect to the norm $\|\cdot\|_{H_k}$ is essential in order to establish the convergence of the method as well as its fast convergence rate. For example, if x_{k+1} in* (14) *is defined as*

$$\langle y_k - x_{k+1}, x - x_{k+1}\rangle \leq \varepsilon_k := \theta_k^2\|x_{k+1}-x_k\|^2, \quad \forall\, x \in C,$$

i.e., x_{k+1} is an approximate solution of $\min_{x\in C}\frac{1}{2}\|x-y_k\|^2$, and H_k is not a multiple of $I_{\mathbb{X}}$, it is not even possible to show that x_k is a stationary point of (1) *when $x_{k+1}=x_k$.*

In order to analyze GNM-AP, we suppose that following assumptions hold:

(A1) The point x_* satisfies the first-order necessary condition for (7), i.e. $\langle F'(x_*)^*F(x_*), x-x_*\rangle \geq 0$, for all $x \in C$, and $F'(x_*)$ is injective;

(A2) The sequence $\{\theta_k\}$ satisfies $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0,1)$.

For simplicity, let us consider the following constants

$$c := \|F(x_*)\|, \qquad \beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \beta\left\|F'(x_*)\right\| \qquad \delta := \sup\{t \in [0,R) : B(x_*,t) \subset \Omega\}. \tag{15}$$

where $R > 0$ is a given scalar.

We first state a local convergence theorem for GNM-AP under a majorant condition. For technical reasons and for the convenience of the reader, the proof of the next theorem will be given in next subsection.

Theorem 3. *Suppose that there exists a continuously differentiable function $f : [0,\, R) \to \mathbb{R}$ such that*

$$\beta\left\|F'(x) - F'(x_* + \tau(x-x_*))\right\| \leq f'(\sigma(x)) - f'(\tau\sigma(x)), \tag{16}$$

where $x \in B(x_*,\delta)$, $\tau \in [0,1]$ and $\sigma(x) = \|x-x_*\|$, and

h1) *$f(0)=0$ and $f'(0)=-1$;*

h2) *f' is convex and strictly increasing;*

h3) $c\beta((1+\sqrt{2})\kappa+1)D^+f'(0)+\kappa\bar{\theta}<1-\bar{\theta}$.

Let be given positive constants $\nu := \sup\{t\in[0,R) : f'(t)<0\}$,

$$\rho := \sup\left\{t\in(0,\nu) : \frac{[f'(t)+1+\kappa]\left[(1-\bar{\theta})tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta[f'(t)+1]}{(1-\bar{\theta})t[f'(t)]^2}<1\right\}, \qquad (17)$$

$$r := \min\{\rho,\delta\}.$$

Then GNM-AP with starting point $x_0\in C\cap B(x_*,r)\backslash\{x_*\}$ is well-defined, the generated $\{x_k\}$ is contained in $B(x_*,r)\cap C$, converges to x_* and satisfies

$$\|x_{k+1}-x_*\|<\|x_k-x_*\| \qquad (18)$$

and

$$\|x_{k+1}-x_*\|\le\frac{[f'(\sigma(x_0))+1+\kappa]\,[\sigma(x_0)f'(\sigma(x_0))-f(\sigma(x_0))]}{(1-\theta_k)[\sigma(x_0)f'(\sigma(x_0))]^2}\|x_k-x_*\|^2$$
$$+\frac{\left[(1+\sqrt{2})c\beta\,[f'(\sigma(x_0))+1]-\theta_k\sigma(x_0)f'(\sigma(x_0))\right][f'(\sigma(x_0))+1+\kappa]+c\beta\,[f'(\sigma(x_0))+1]}{(1-\theta_k)\sigma(x_0)\,[f'(\sigma(x_0))]^2}\|x_k-x_*\|, \qquad (19)$$

for all $k=0,1,\ldots$.

Remark 2. *(i) Since $\|x_k-x_*\|<\sigma(x_0)=\|x_0-x_*\|$ (see* (18)*), it follows from* (19) *and* **(A2)** *that*

$$\|x_{k+1}-x_*\|\le\left[\frac{[f'(\sigma(x_0))+1+\kappa]\left[(1-\bar{\theta})\sigma(x_0)f'(\sigma(x_0))-f(\sigma(x_0))+c\beta(1+\sqrt{2})(f'(\sigma(x_0))+1)\right]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2}\right.$$
$$\left.+\frac{c\beta\,[f'(\sigma(x_0))+1]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2}\right]\|x_k-x_*\|$$

which, combined with (17) *and the fact that $x_0\in C\cap B(x_*,r)\backslash\{x_*\}$, implies that GNM-AP is linearly convergent to x_*. (ii) Note that, if $c=0$ and $\limsup_{k\to+\infty}\theta_k=0$, then* (19) *implies that GNM-AP converges quadratically to x_*. (iii) If scalar $\bar{\theta}$ in* **(A2)** *is equal to zero (in particular, $\theta_k=0$ for all $k\ge 0$), then iterative x_{k+1} in Step 2 of GNM-AP corresponds to the exact projection $P_C^{H_k}(y_k)$. In this case, Theorem 3 is similar to [9, Theorem 7], which is related to the Gauss-Newton method for solving unconstrained nonlinear least squares problems.*

Before specializing Theorem 3 for two important classes of functions, we present an example in which all conditions of Theorem 3 hold. The following result, which gives a simpler condition to check than condition (16) whenever the functions under consideration are twice continuously differentiable, is needed.

Lemma 4. *Let $x_* \in \Omega$ and $R > 0$ be given, and assume that F is twice continuously differentiable on Ω. If there exists a $f : [0,R) \to \mathbb{R}$ twice continuously differentiable and satisfying*

$$\beta \|F''(x)\| \leqslant f''(\|x - x_*\|), \qquad x \in B(x_*, R),$$

then F and f satisfy (16).

Proof. The proof follows the same pattern as outlined in [9, Lemma 22]. □

Example 1. *Consider the constrained nonlinear least squares problem* (1) *with $C = \mathbb{R}^3_+$ and*

$$F(x) = \frac{9}{50}\left(\|x\|^{5/3}x - 64(3,2,\sqrt{3})\right).$$

Note that $x_* = 2(3,2,\sqrt{3})$ is a stationary point of (1) in this case. Let us apply Theorem 3 for this instance. First, from (15)*, we have $c = 0$, $\beta = (25/1152)\sqrt{137}$, $\kappa = (48/25)\beta\sqrt{82}$. Moreover, since the second derivative of F is given by*

$$F''(x)(v,v) = \frac{9}{50}\left[-\frac{5}{9}\|x\|^{-7/3}\langle x,v\rangle^2 x + \frac{5}{3}\|x\|^{-1/3}\|v\|^2 x + \frac{10}{3}\|x\|^{-1/3}\langle x,v\rangle v\right],$$

for every $x, v \in \mathbb{R}^3$ and $x \neq 0$, and $F''(0) = 0$, we obtain

$$\|F''(x)\| \leq \|x\|^{2/3}, \quad x \in \mathbb{R}^n,$$

or, equivalently,

$$\beta\|F''(x)\| \leqslant f''(\|x - x_*\|), \qquad x \in \mathbb{R}^n,$$

where $f : [0,\infty) \to \mathbb{R}$ is given by

$$f(t) = \frac{9}{40}\beta t^{8/3} - t.$$

Hence, it follows from Lemma 4 that F and f satisfy (16)*. In particular, as $f(0) = 0$, $f'(t) = (3\beta/5)t^{5/3} - 1$, $f'(0) = -1$ and $f''(t) = \beta t^{2/3} > 0$, we obtain f satisfies* **h1** *and* **h2***. Therefore, if $\bar{\theta} < [1/(1+\kappa)] \approx 0.2$ (i.e.,* **h3** *holds), it follows from Theorem 3 that GNM-AP with starting point $x_0 \in \mathbb{R}^3_+ \cap B(x_*, r)\backslash\{x_*\}$, where*

$$r = \left[\frac{5(15\kappa + 48 - 24\bar{\theta}(1+\kappa)) - 5\sqrt{(24\bar{\theta}(1+\kappa) - 15\kappa - 48)^2 + 864(\bar{\theta}(1+\kappa) - 1)}}{54\beta}\right]^{3/5}, \qquad (20)$$

is well-defined, the generated $\{x_k\}$ is contained in $B(x_*, r) \cap \mathbb{R}^3_+$, converges to x_* and satisfies

$$\|x_{k+1} - x_*\| \leq \frac{5}{8(1-\bar{\theta})}\left[\frac{9\beta^2\sigma(x_0)^{7/3} + 15\beta\kappa\sigma(x_0)^{2/3}}{9\beta^2\sigma(x_0)^{10/3} + 30\beta\sigma(x_0)^{5/3} + 5}\right]\|x_k - x_*\|^2, \qquad k = 0, 1, \ldots.$$

For example, if $\bar{\theta} = 0.1$*, then the radius of convergence* r *in* (20) *is approximately equal to* 1.

We next specialize Theorem 3 for two important classes of functions. In the first one, F' satisfies a Lipschitz-like condition [14, 15, 19] and, in the second one, F is an analytic function satisfying a Smale condition [27, 28].

Corollary 5. *Suppose that there exists a* $L > 0$ *such that*

$$\lambda = \frac{[(1+\sqrt{2})\kappa + 1]\, c\,\beta L + \kappa\bar{\theta}}{(1-\bar{\theta})} < 1, \qquad \beta \left\| F'(x) - F'(x_* + \tau(x - x_*)) \right\| \leq L(1-\tau)\sigma(x), \tag{21}$$

where $x \in B(x_*, \delta)$*,* $\tau \in [0,1]$ *and* $\sigma(x) = \|x - x_*\|$*. Let be given the positive constant*

$$r := \min\left\{ \frac{4 + \kappa - 2\bar{\theta}(1+\kappa) + 2c(1+\sqrt{2})\beta L - \sqrt{\left[4 + \kappa - 2\bar{\theta}(1+\kappa) + 2c(1+\sqrt{2})\beta L\right]^2 - 8(1-\lambda)(1-\bar{\theta})}}{2L}, \delta \right\}.$$

Then GNM-AP with starting point $x_0 \in C \cap B(x_*, r) \backslash \{x_*\}$ *is well-defined, the generated* $\{x_k\}$ *is contained in* $B(x_*, r) \cap C$*, converges to* x_* *and satisfies*

$$\|x_{k+1} - x_*\| < \|x_k - x_*\| \tag{22}$$

and

$$\|x_{k+1} - x_*\| \leq \frac{\kappa L + L^2\sigma(x_0)}{2(1-\theta_k)[1 - L\sigma(x_0)]^2}\|x_k - x_*\|^2 + \frac{[(1+\sqrt{2})\kappa + 1]c\beta L + c(1+\sqrt{2})\beta L^2\sigma(x_0)}{(1-\theta_k)[1 - L\sigma(x_0)]^2}\|x_k - x_*\| + \frac{\theta_k(L\sigma(x_0) + k)}{(1-\theta_k)[1 - L\sigma(x_0)]}\|x_k - x_*\|, \quad \forall k = 0, 1, \ldots,$$

Proof. It is immediate to prove that F, x_* and $f : [0,\delta) \to \mathbb{R}$ defined by $f(t) = Lt^2/2 - t$, satisfy inequality (16), conditions **h1** and **h2**. Since $[(1+\sqrt{2})\kappa + 1]\, c\,\beta L + \kappa\bar{\theta} < 1 - \bar{\theta}$, the condition **h3** also holds. In this case, it is easy to see that the constants ν and ρ as defined in Theorem 3, satisfy

$$0 < \rho = \frac{\mu - \sqrt{\mu^2 - 8(1-\bar{\theta})(1-\lambda)}}{2L} \leq \nu = 1/L,$$

where $\mu := 4 + \kappa - 2\bar{\theta}(1+\kappa) + 2c(1+\sqrt{2})\beta L$. As a consequence, $0 < r = \min\{\delta, \rho\}$. Therefore, as F, r, f and x_* satisfy all of the hypotheses of Theorem 3, taking $x_0 \in C \cap B(x_*, r) \backslash \{x_*\}$ the statements of the theorem follow from Theorem 3. □

We next specialize Theorem (3) for the class of analytic functions satisfying a Smale condition.

Corollary 6. *Suppose that*

$$\gamma := \sup_{n>1} \beta \left\| \frac{F^{(n)}(x_*)}{n!} \right\|^{1/(n-1)} < +\infty \quad \textit{and} \quad 2\gamma c\beta((1+\sqrt{2})\kappa+1)+\kappa\bar{\theta} < 1-\bar{\theta}.$$

Let constants $a = \gamma c\beta$, $b = (1+\sqrt{2})\gamma c\beta$,

$$\bar{\rho} := \inf\left\{ s \in (\sqrt{2}/2, 1) : p(s) := \zeta s^4 + \eta s^3 + \iota s^2 + (b-1)s + b < 0 \right\} \tag{23}$$

where $\zeta := -4+(\kappa+1)2\bar{\theta}$, $\eta := 1-\kappa+a+b(\kappa-1)$, *and* $\iota := 3+\kappa-(\kappa+1)\bar{\theta}+a+b(\kappa-1)$, *and*

$$r := \min\left\{(1-\bar{\rho})/\gamma, \delta\right\}.$$

Then GNM-AP with starting point $x_0 \in C \cap B(x_*, r) \backslash \{x_*\}$ *is well-defined, the generated* $\{x_k\}$ *is contained in* $B(x_*, r) \cap C$*, converges to* x_* *and satisfies*

$$\|x_{k+1} - x_*\| < \|x_k - x_*\| \tag{24}$$

and

$$\begin{aligned} \|x_{k+1}-x_*\| \leq {} & \frac{\gamma\left[1+(\kappa-1)(1-\gamma\sigma(x_0))^2\right]}{(1-\theta_k)\left[1-2(1-\gamma\sigma(x_0))^2\right]^2}\|x_k-x_*\|^2 + \frac{c\beta\left[1-(1-\gamma\sigma(x_0))^2\right](1-\gamma\sigma(x_0))^2}{(1-\theta_k)\sigma(x_0)\left[1-2(1-\gamma\sigma(x_0))^2\right]^2}\|x_k-x_*\| \\ & + \frac{\left[(1+\sqrt{2})c\beta(1-(1-\gamma\sigma(x_0))^2)-\theta_k\sigma(x_0)(1-2(1-\gamma\sigma(x_0))^2)\right]\left[1+(\kappa-1)(1-\gamma\sigma(x_0))^2\right]}{(1-\theta_k)\sigma(x_0)\left[1-2(1-\gamma\sigma(x_0))^2\right]^2}\|x_k-x_*\|, \end{aligned} \tag{25}$$

for all $k = 0, 1, \ldots$.

Proof. Consider the real function $f : [0, 1/\gamma) \to \mathbb{R}$ defined by

$$f(t) = \frac{t}{1-\gamma t} - 2t.$$

It is straightforward to show that f is analytic and that

$$f(0) = 0, \quad f'(t) = 1/(1-\gamma t)^2 - 2, \quad f'(0) = -1, \quad f''(t) = (2\gamma)/(1-\gamma t)^3, \quad f^n(0) = n!\gamma^{n-1},$$

for $n \geq 2$. It follows from last equalities that f satisfies **h1** and **h2**. Since $2\gamma c\beta((1+\sqrt{2})\kappa+1)+\kappa\bar{\theta} < 1-\bar{\theta}$, condition **h3** also holds. Now, note that

$$\beta\|F''(x)\| \leqslant f''(\|x-x_*\|),$$

for all $x \in B(x_*, 1/\gamma) \cap \Omega$, the proof of this fact follows the same pattern as outlined in [9, Lemma 21]. As

$f''(t) = (2\gamma)/(1-\gamma t)^3$, we conclude, from Lemma 4, that F and f satisfy (16) with $R = 1/\gamma$. In this case,

$$\nu = (2-\sqrt{2})/2\gamma < 1/\gamma.$$

Now, we will obtain the constant ρ as defined in Theorem 3. For simplicity, consider the following change of variable $s = 1-\gamma t$, which implies that $t = (1-s)/\gamma$. Moreover, if t satisfies $0 < t < \nu = (2-\sqrt{2})2\gamma$, then $\sqrt{2}/2 < s < 1$. Hence, to determine the constant ρ as defined in Theorem 3 is equivalent to determine the constant s such that

$$\bar{\rho} := \inf\left\{ s \in (\sqrt{2}/2, 1) : p(s) := \zeta s^4 + \eta s^3 + \iota s^2 + (b-1)s + b < 0 \right\},$$

where $a = \gamma c\beta$, $b = (1+\sqrt{2})\gamma c\beta$, $\zeta := -4 + (\kappa+1)2\bar{\theta}$, $\eta := 1-\kappa+a+b(\kappa-1)$, and $\iota := 3+\kappa-(\kappa+1)\bar{\theta}+a+b(\kappa-1)$. Thus, taking into account the change of variable, we have $\rho = (1-\bar{\rho})/\gamma$ and

$$r = \min\left\{(1-\bar{\rho})/\gamma,\, \delta\right\}.$$

Therefore, as F, r, f and x_* satisfy all hypothesis of Theorem 3, taking $x_0 \in C \cap B(x_*, r) \backslash \{x_*\}$, the statements of the theorem follow from Theorem 3. $\square$

We end this section by presenting a numerical example, adapted from Dedieu and Shub [5], in which all conditions of Corollary 5 hold.

Example 2. *Let $F : \mathbb{R} \to \mathbb{R}^2$ such that $F(x) = (x, x^2 + a)^*$, where $a \in \mathbb{R}$ is given, and consider*

$$\min_{x\in[-2,2]} \|F(x)\|^2 = x^4 + (2a+1)x^2 + a^2. \tag{26}$$

Note that $x_* = 0$ is a stationary point of (26). Let us apply Corollary 5 for this instance. First, from (15), we have $c = |a|$, $\beta = 1$, $\kappa = 1$. Moreover, since $\beta\|F'(x) - F'(\tau x)\| = (1-\tau)2|x|$, for all $x \in \mathbb{R}$ and $\tau \in [0,1]$, we obtain the Lipschitz-Like constant L is 2. Therefore, if $2[(2+\sqrt{2})|a| + \bar{\theta}] < 1$ (i.e., the first inequality in (21) holds), it follows from Corollary 5 that GNM-AP with starting point $x_0 \in [-2,2] \cap B(x_*, r)\backslash\{x_*\}$, where

$$r = \frac{5-4\bar{\theta}+4|a|(1+\sqrt{2}) - \sqrt{\left[5-4\bar{\theta}+4|a|(1+\sqrt{2})\right]^2 - 8(1-2\bar{\theta}-2(2+\sqrt{2})|a|)}}{4}, \tag{27}$$

is well-defined, the generated $\{x_k\}$ is contained in $B(x_*, r) \cap [-2,2]$, converges to x_* and satisfies

$$\|x_{k+1} - x_*\| \le \frac{1+2\sigma(x_0)}{(1-\bar{\theta})[1-2\sigma(x_0)]^2}\|x_k - x_*\|^2 + \frac{2|a|(2+\sqrt{2}+2(1+\sqrt{2})\sigma(x_0))}{(1-\bar{\theta})[1-2\sigma(x_0)]^2}\|x_k - x_*\|$$
$$+ \frac{\bar{\theta}(1+2\sigma(x_0))}{(1-\bar{\theta})[1-2\sigma(x_0)]}\|x_k - x_*\|, \quad \forall k = 0, 1, \ldots,$$

For example, if $a = 0$ and $\bar\theta = 0.1$, then the radius of convergence r in (27) is approximately equal to 0.2.

3.1 Proof of Theorems 3

Our goal in this subsection is to prove Theorem 3. To this end, we first present some auxiliary results, which establish positiveness of the constants δ, ν and ρ, as well as some useful relationships between the majorant function and the nonlinear function F.

First of all, note that constant δ in (15) is positive, because Ω is an open set and $x_* \in \Omega$.

Proposition 7. *The constant ν as in Theorem (3) is positive and $f'(t) < 0$ for all $t \in (0,\nu)$. Furthermore, the following functions defined on the interval $(0,\nu)$*

$$t \mapsto -\frac{1}{f'(t)}, \qquad t \mapsto -\frac{[f'(t)+1+\kappa]}{f'(t)}, \qquad t \mapsto \frac{[tf'(t)-f(t)]}{t^2}, \qquad t \mapsto \frac{f'(t)+1}{t}, \tag{28}$$

are positive and increasing.

Proof. First, as f' is continuous in $(0,R)$ and $f'(0) = -1$, there exists $\varepsilon > 0$ such that $f'(t) < 0$ for all $t \in (0,\varepsilon)$. Hence, $\nu > 0$. Moreover, using **h2** and the definition of ν, it follows that $f'(t) < 0$ for all $t \in (0,\nu)$. Note now that the first two functions in (28) are positive and increasing due to the facts that $-1 < f'(t) < 0$, for all $t \in [0,\nu)$, and f' is strictly increasing. Finally, for the proofs that the last two functions in (28) are positive and increasing, see items **ii** and **iii** of [9, Proposition 10]. □

We next prove, in particular, that constant ρ in (17) is positive.

Proposition 8. *The constant ρ is positive and there holds*

$$\frac{[f'(t)+1+\kappa]\left[(1-\bar\theta)tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2} < 1, \quad t \in (0,\rho). \tag{29}$$

Proof. Using **h1** and some algebraic manipulation, we obtain

$$\frac{tf'(t)-f(t)}{t} = \left[f'(t) - \frac{f(t)-f(0)}{t-0}\right], \qquad \frac{f'(t)+1}{t} = \frac{f'(t)-f'(0)}{t-0},$$

which, combined with the fact that $f'(0) = -1$, yields

$$\lim_{t\to 0}[tf'(t)-f(t)]/t = 0, \qquad \lim_{t\to 0}[f'(t)+1]/t = D^+f'(0), \tag{30}$$

where the existence of the right derivative $D^+f'(0)$ is guaranteed due to the fact that f' is convex. Note now that equation (29) is equivalent to

$$\frac{[f'(t)+1+\kappa]\,[tf'(t)-f(t)]}{(1-\bar\theta)[tf'(t)]^2}t - \frac{\bar\theta\,[f'(t)+1+\kappa]}{(1-\bar\theta)f'(t)} + \frac{c\beta(1+\sqrt{2})\,[f'(t)+1+\kappa]\,(f'(t)+1)}{(1-\bar\theta)t[f'(t)]^2} + \frac{c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2}. \tag{31}$$

Hence, using $f'(0)=-1$, it follows from (31) and (30) that

$$\lim_{t\to 0}\left[\frac{[f'(t)+1+\kappa]\left[(1-\bar\theta)tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2}\right]$$
$$=\frac{\kappa\bar\theta+c\beta(1+\sqrt{2})\kappa D^+f'(0)+c\beta D^+f'(0)}{(1-\bar\theta)}=\frac{c\beta\left[(1+\sqrt{2})\kappa+1\right]D^+f'(0)+\kappa\bar\theta}{(1-\bar\theta)}.$$

Therefore, since **(h3)** implies that $[c\beta[(1+\sqrt{2})\kappa+1]D^+f'(0)+\kappa\bar\theta]/[(1-\bar\theta)]<1$, we conclude that there exists an $\varepsilon>0$ such that

$$\frac{[f'(t)+1+\kappa]\left[(1-\bar\theta)tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2}<1,\qquad t\in(0,\varepsilon),$$

So, $\varepsilon\le\rho$, which proves the first statement.

Again, since (29) is equivalent to (31), the proof of the last part of proposition trivially follows from definition of ρ and last part of Proposition (7). □

The next two lemmas present some useful relationships between operator F and majorant function f.

Lemma 9. *Let $x\in\Omega$. If $\sigma(x)<\min\{\nu,\delta\}$, then following statements hold:*

- **i)** $\beta\|F(x_*)-[F(x)+F'(x)(x_*-x)]\|\le f(0)-[f(\sigma(x))+f'(\sigma(x))(0-\sigma(x))]:=e_f(\sigma(x),0)$;
- **ii)** *the linear operator $H(x)=F'(x)^*F'(x)$ is invertible and*

$$\|F'(x)^\dagger\|\le\frac{-\beta}{f'(\sigma(x))},\qquad \|F'(x)^\dagger-F'(x_*)^\dagger\|<\frac{-\sqrt{2}\beta[f'(\sigma(x))+1]}{f'(\sigma(x))}.$$

In particular, $H(x)=F'(x)^*F'(x)$ is invertible in $B(x_*,r)$.

Proof. The proof follows the pattern of the proofs of Lemmas 13 and 14 in [9] (see also Lemma 7 in [15]). □

Lemma 10. *Let $x\in\Omega$. If $\sigma(x)<\min\{\nu,\delta\}$, then the following inequalities hold:*

- **i)** $\|H(x)\|^{1/2}\le[f'(\sigma(x))+1+\kappa]/\beta$;
- **ii)** $\|H(x)^{-1}\|^{1/2}\le-\beta/[f'(\sigma(x))]$;
- **iii)** $\beta\|(H(x)-H(x_*))F'(x_*)^\dagger\|\le(f'(\sigma(x))+2+\kappa)(f'(\sigma(x))+1)$,

*where $H(x)$ is defined as in Lemma 9(**ii**).*

Proof. (i) Using inequality in (16) and definition of κ in (15), we have

$$\beta\|F'(x)\| \le \beta\|F'(x)-F'(x_*)\| + \beta\|F'(x_*)\| \le f'(\sigma(x)) + 1 + \kappa. \tag{32}$$

As $\|H(x)\|^{1/2} = \|F'(x)^*F'(x)\|^{1/2} = \|F'(x)\|$, the desired inequality follows.

(ii) To show item **ii**, use the definition of H, the last inequality in (6) and Lemma 9(ii).

(iii) Note that the definition of $H(x)$, some algebraic manipulations and (5) gives

$$\begin{aligned}\beta\|(H(x)-H(x_*))F'(x_*)^\dagger\| &= \beta\|F'(x)^*(F'(x)-F'(x_*))F'(x_*)^\dagger + (F'(x)-F'(x_*))^*\Pi_{im(F'(x_*))}\| \\ &\le (\|F'(x)\|\|F'(x_*)^\dagger\|+1)\beta\|F'(x)-F'(x_*)\|,\end{aligned}$$

which, combined with definition of β in (15) and inequalities in (16) and (32), yields the desired inequality. □

Lemma 9 implies that H is invertible for any $x \in B(x_*, r)$ and hence $F'(x)^\dagger$ and $\tilde{P}_C^H(x - F'(x)^\dagger F(x))$ are well-defined in this region. Therefore, since the starting point $x_0 \in C \cap B(x_*, r)$, we have x_1 is well-defined, but we do not show that $x_1 \in C \cap B(x_*, r)$ and, therefore, if the next iteration x_2 will be well-defined. In the next lemma, we ensure that sequence $\{\|x_k - x_*\|\}$ is strictly decreasing and, hence, that $\{x_k\}$ is well-defined and contained in $C \cap B(x_*, r)$.

Lemma 11. *Let $x_k \in C \cap B(x_*, r)$. Then, for every $k \ge 0$,*

$$\begin{aligned}\|x_{k+1}-x_*\| \le{}& \frac{[f'(\sigma(x_k))+1+\kappa]\,[\sigma(x_k)f'(\sigma(x_k)) - f(\sigma(x_k))]}{(1-\theta_k)\,[\sigma(x_k)f'(\sigma(x_k))]^2}\|x_k-x_*\|^2 \\ &+ \frac{[f'(\sigma(x_k))+1+\kappa]\left[(1+\sqrt{2})c\beta\,[f'(\sigma(x_k))+1] - \theta_k\sigma(x_k)f'(\sigma(x_k))\right] + c\beta\,[f'(\sigma(x_k))+1]}{(1-\theta_k)\sigma(x_k)\,[f'(\sigma(x_k))]^2}\|x_k-x_*\|. \end{aligned} \tag{33}$$

As a consequence,

$$\|x_{k+1}-x_*\| < \|x_k - x_*\|. \tag{34}$$

Proof. Since x_* is a stationary point of (1) (see **(A1)**), we trivially have $P_C^{H_*}(x_* - F'(x_*)^\dagger F(x_*)) = x_*$. Hence, it follows from Lemma 2 with $x = x_k - F'(x_k)^\dagger F(x_k)$, $\hat{x} = x_* - F'(x_*)^\dagger F(x_*)$ and $\varepsilon = \theta_k^2\|x_k - x_{k+1}\|_{H_k}^2$ that

$$\begin{aligned}\|\tilde{P}_C^{H_k}(x_k - F'(x_k)^\dagger F(x_k)) - x_*\|_{H_k} \le{}& \|H_k^{-1}\|^{1/2}\|(H_*-H_k)(F'(x_*)^\dagger F(x_*))\| \\ &+ \|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\|_{H_k} + \theta_k\|x_k - x_{k+1}\|_{H_k}.\end{aligned}$$

For simplicity, the notation defines the following terms:

$$A(x_k, x_*) = \|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\|_{H_k} \tag{35}$$

and

$$B(x_k, x_*) = \|H_k^{-1}\|^{1/2}\|(H - H_*)F'(x_*)^\dagger\|\|F(x_*)\|. \tag{36}$$

So, from the three latter inequalities, we obtain

$$\|x_{k+1}-x_*\|_{H_k} \leq A(x_k,x^*)+B(x_k,x^*)+\theta_k\|x_k-x_{k+1}\|_{H_k}.$$

Hence, since $\|x_k-x_{k+1}\|_{H_k} \leq \|x_{k+1}-x_*\|_{H_k}+\|H_k\|^{1/2}\|x_k-x_*\|$, we obtain

$$(1-\theta_k)\|x_{k+1}-x_*\|_{H_k} \leq A(x_k,x^*)+B(x_k,x^*)+\theta_k\|H_k\|^{1/2}\|x_k-x_*\|.$$

Since $\theta_k<1$, for all $k\geq 0$, (see **(A2)**), the last inequality and (4) imply that

$$\|x_{k+1}-x_*\| \leq \frac{\|H_k^{-1}\|^{1/2}}{(1-\theta_k)}A(x_k,x^*)+\frac{\|H_k^{-1}\|^{1/2}}{(1-\theta_k)}B(x_k,x^*)+\frac{\theta_k\left[\|H_k^{-1}\|\|H_k\|\right]^{1/2}}{(1-\theta_k)}\|x_k-x_*\|. \tag{37}$$

Now, we will obtain upper bounds of $A(x_k,x^*)$ and $B(x_k,x^*)$. First, some algebraic manipulations and the second equality in (6) yield

$$\begin{aligned}\|x_k-F'(x_k)^{\dagger}F(x_k)-x_*&+F'(x_*)^{\dagger}F(x_*)\|\\ &=\|F'(x_k)^{\dagger}[F'(x_k)(x_k-x_*)-F(x_k)+F(x_*)]+(F'(x_*)^{\dagger}-F'(x_k)^{\dagger})F(x_*)\|\\ &\leq\|F'(x_k)^{\dagger}\|\|F(x_*)-\left[F(x_k)+F'(x_k)(x_*-x_k)\right]\|+\|F'(x_*)^{\dagger}-F'(x_k)^{\dagger}\|\|F(x_*)\|.\end{aligned}$$

Combining last inequality, Lemma 9 and definition of c in (15), we have

$$\|x_k-F'(x_k)^{\dagger}F(x_k)-x_*+F'(x_*)^{\dagger}F(x_*)\| = \frac{e_f(\sigma(x_k),0)}{-f'(\sigma(x_k))}+\frac{\sqrt{2}c\beta[f'(\sigma(x_k))+1]}{-f'(\sigma(x_k))},$$

which, combined with (35), the fact that $\|\cdot\|_{H_k}\leq\|H_k\|^{1/2}\|\cdot\|$ and Lemma 10(i), yields

$$\begin{aligned}A(x_k,x^*) &\leq \|H_k\|^{1/2}\|x_k-F'(x_k)^{\dagger}F(x_k)-x_*+F'(x_*)^{\dagger}F(x_*)\|\\ &\leq \frac{(f'(\sigma(x_k))+1+\kappa)}{-\beta f'(\sigma(x_k))}\Big(e_f(\sigma(x_k),0)+\sqrt{2}c\beta[f'(\sigma(x_k))+1]\Big).\end{aligned} \tag{38}$$

On the other hand, from definition in (36) and Lemma 10(ii)–(iii), we have

$$B(x_k,x_*) \leq \frac{c}{-f'(\sigma(x_k))}(f'(\sigma(x_k))+2+\kappa)(f'(\sigma(x_k))+1). \tag{39}$$

Hence, using (37)–(39) and Lemma (10)(i)–(ii), we obtain

$$\|x_{k+1}-x_*\| \le \frac{[f'(\sigma(x_k))+1+\kappa]\, e_f(\sigma(x_k),0)+(1+\sqrt{2})c\beta\,[f'(\sigma(x_k))+1]^2}{(1-\theta_k)\,[f'(\sigma(x_k))]^2} + \frac{c\beta\left[(1+\sqrt{2})\kappa+1\right][f'(\sigma(x_k))+1]}{(1-\theta_k)\,[f'(\sigma(x_k))]^2} - \frac{\theta_k(f'(\sigma(x_k))+1+\kappa)}{(1-\theta_k)f'(\sigma(x_k))}\sigma(x_k),$$

which, combined with definition of $e_f(\sigma(x_k),0)$ in Lemmas 9(i) and **h1**, proves (33).

Now, using $\theta_k < \bar{\theta}$, for all $k \ge 0$, (see **(A2)**), we obtain that the right-hand side of (33) is equivalent to

$$\left[\frac{[f'(\sigma(x_k))+1+\kappa]\left[(1-\bar{\theta})\sigma(x_k)f'(\sigma(x_k))-f(\sigma(x_k))+c\beta(1+\sqrt{2})(f'(\sigma(x_k))+1)\right]+c\beta\,[f'(\sigma(x_k))+1]}{(1-\bar{\theta})\sigma(x_k)\,[f'(\sigma(x_k))]^2}\right]\sigma(x_k).$$

Therefore, as $x_k \in C\cap B(x_*,r)/\{x_*\}$, it follows from Proposition 8 with $t=\sigma(x_k)$ that the quantity in the bracket above is less than one and hence (34) follows. □

Proof of Theorem (3)**:** Since $x_0 \in C\cap B(x_*,r)/\{x_*\}$, combining Lemma (9)(ii), inequality (34) and an induction argument, we have that (18) holds and $\{x_k\}$ is well-defined and remains in $C\cap B(x_*,r)$. Our goal is now to show that $\{x_k\}$ converges to x_*. Using the second part of Lemma (11), we find

$$\sigma(x_k)=\|x_k-x_*\|<\|x_0-x_*\|=\sigma(x_0), \qquad k=1,2\ldots. \tag{40}$$

Hence, by combining (33) with last part of Proposition (7), we obtain

$$\|x_{k+1}-x_*\| \le \frac{[f'(\sigma(x_0))+1+\kappa]\,[\sigma(x_0)f'(\sigma(x_0))-f(\sigma(x_0))]}{(1-\theta_k)[\sigma(x_0)f'(\sigma(x_0))]^2}\|x_k-x_*\|^2 - \frac{\theta_k(f'(\sigma(x_0))+1+\kappa)}{(1-\theta_k)f'(\sigma(x_0))}\|x_k-x_*\| + \frac{(1+\sqrt{2})c\beta\,[f'(\sigma(x_0))+1+\kappa]\,[f'(\sigma(x_0))+1]+c\beta\,[f'(\sigma(x_0))+1]}{(1-\theta_k)\sigma(x_0)\,[f'(\sigma(x_0))]^2}\|x_k-x_*\|, \quad k=0,1,\ldots,$$

which is equivalent to (19). Combining last inequality with (40) and **(A2)**, we obtain

$$\|x_{k+1}-x_*\| \le \left[\frac{[f'(\sigma(x_0))+1+\kappa]\left[(1-\bar{\theta})\sigma(x_0)f'(\sigma(x_0))-f(\sigma(x_0))+c\beta(1+\sqrt{2})(f'(\sigma(x_0))+1)\right]+c\beta\,[f'(\sigma(x_0))+1]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2}\right]\|x_k-x_*\|$$

for all $k=0,1,\ldots$. Hence, applying Proposition (8) with $t=\sigma(x_0)$, we conclude that $\{\|x_k-x_*\|\}$ converges to zero. So, $\{x_k\}$ converges to x_*. □

4 Globalized method

We now present a globalized version of GNM-AP. The globalization strategy used here is based on the non-monotone line search in [18]. Since the Gauss-Newton step can not be defined in some regions, our global method uses, in these cases, the projected gradient step. The method is formally described as follows.

Global GNM-AP (G-GNM-AP)

Step 0 (Initialization). Let $x_0 \in C$, $\tau \in (0,1)$, an integer $M \geq 1$ and $\{\theta_k\} \subset [0,\infty)$ be given, and set $k = 0$.

Step 1 (Gauss-Newton or projected gradient step). If $F'(x_k)^* F'(x_k)$ is non-singular, then $H_k = F'(x_k)^* F'(x_k)$. Otherwise, $H_k = I_{\mathbb{X}}$. Compute $y_k \in \mathbb{X}$ such that

$$y_k = x_k - H_k^{-1} F'(x_k)^* F(x_k).$$

Step 2 (Computation of the approximate projection). Compute $\tilde{P}_C^{H_k}(y_k) \in C$ such that

$$\langle y_k - \tilde{P}_C^{H_k}(y_k), x - \tilde{P}_C^{H_k}(y_k) \rangle_{H_k} \leq \varepsilon_k := \theta_k^2 \|\tilde{P}_C^{H_k}(y_k) - x_k\|_{H_k}^2, \quad \forall\, x \in C, \tag{41}$$

i.e., $\tilde{P}_C^{H_k}(y_k)$ is an ε_k–approximate projection of y_k onto C.

Step 3 (Backtracking). Define $d_k = \tilde{P}_C^{H_k}(y_k) - x_k$ and $G_{max} = \max\{G(x_{k-j});\, 0 \leq j \leq \min\{k, M-1\}\}$. Set $\alpha \leftarrow 1$.

Step 3.1 Set $x_+ = x_k + \alpha d_k$.

Step 3.2 If

$$G(x_+) \leq G_{max} + \tau\alpha \langle F'(x_k)^* F(x_k), d_k \rangle, \tag{42}$$

then $\alpha_k = \alpha$, $x_{k+1} = x_+$, and go to step 4. Otherwise, set $\alpha \leftarrow \alpha/2$ and go to step 3.2.

Step 4 (Termination criterion and update). If $x_{k+1} = x_k$, then **stop**; otherwise, set $k \leftarrow k+1$ and go to step 1.

end

The following theorem, which is an extension to the constrained case of [18, Theorem 1], summarizes the convergence properties of G-GNM-AP method.

Theorem 12. *Assume that there exist constants $c, d \in \mathbb{R}$ such that, for every $k \geq 0$,*

$$0 < c \leq \lambda_{min}(H_k) \leq \lambda_{max}(H_k) \leq d, \tag{43}$$

where $\lambda_{min}(H_k)$ and $\lambda_{max}(H_k)$ are, respectively, the smallest and largest eigenvalues of matrix H_k. Furthermore, assume that level set $\Omega_0 := \{x \in C : G(x) \leq G(x_0)\}$ is bounded and sequence $\{\theta_k\}$ satisfies $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0,1)$. Then, either G-GNM-AP stops at some stationary point x_k or every limit point of the generated sequence is stationary.

Proof. By definitions of d_k and y_k, and the inequality in (41), we have

$$\langle -d_k - H_k^{-1} F'(x_k)^* F(x_k), x - x_k + d_k \rangle_{H_k} \leq \theta_k^2 \|d_k\|_{H_k}^2, \quad \forall k \geq 0. \tag{44}$$

If G-GNM-AP stops, then $x_{k+1} = x_k$, which in turn implies that $d_k = 0$. Hence, it follows from (44) that

$$\langle -H_k^{-1} F'(x_k)^* F(x_k), x - x_k \rangle_{H_k} \leq 0, \quad \forall x \in C,$$

or, equivalently,

$$\langle F'(x_k)^* F(x_k), x - x_k \rangle \geq 0, \quad \forall x \in C,$$

i.e., x_k is a stationary point of (1).

Our goal is now to show that every limit point of $\{x_k\}$ is a stationary point of (1). We first prove that d_k is a descent direction for the objective function $G(x) = \|F(x)\|^2/2$ at x_k. From (43), we have H_k is positive definite and

$$\|H_k\| \leq d \quad \text{and} \quad \|H_k^{-1}\| \leq \frac{1}{c}, \quad \forall k \geq 0. \tag{45}$$

It follows from (44) with $x = x_k$ that

$$\langle -d_k - H_k^{-1} F'(x_k)^* F(x_k), -d_k \rangle_{H_k} \leq \theta_k^2 \|d_k\|_{H_k}^2,$$

which, combined with (45) and the fact that $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, yields

$$\langle F'(x_k)^* F(x_k), d_k \rangle \leq -(1 - \bar{\theta}^2) d \|d_k\|^2. \tag{46}$$

Thus, if $x_{k+1} \neq x_k$ (i.e., $\|d_k\| \neq 0$), we obtain, from last inequality and the fact that $\bar{\theta} < 1$, that d_k is a descent direction for G at x_k. In particular, we can conclude that the backtracking process given in Step 3 is well-defined, and, as a consequence, G-GNM-AP is also well-defined.

Let $l(k)$ be an integer such that $k - \min\{k, M-1\} \leq l(k) \leq k$ and

$$G(x_{l(k)}) = \max_{0 \leq j \leq \min\{k, M-1\}} G(x_{k-j}).$$

Using the first part of the proof of the theorem in [18], it can be shown that $\{G(x_{l(k)})\}$ is monotonically nonincreasing, and from the boundness of Ω_0 we have that $\{G(x_{l(k)})\}$ admits a limit for $k \to \infty$. From (42), it follows that, for $k > M-1$,

$$G(x_{l(k)}) \leq G(x_{l(l(k)-1)}) + \tau \alpha_{(l(k)-1)} \langle F'(x_{(l(k)-1)})^* F(x_{(l(k)-1)}), d_{(l(k)-1)} \rangle. \tag{47}$$

Now, since $\alpha_k > 0$ and $\langle F'(x_k)^*F(x_k), d_k\rangle < 0$, by taking limits in (47), it follows that

$$\lim_{k\to\infty} \alpha_{(l(k)-1)}\langle F'(x_{(l(k)-1)})^*F(x_{(l(k)-1)}), d_{(l(k)-1)}\rangle = 0.$$

Moreover, from (46), we conclude that

$$\lim_{k\to\infty} \alpha_{(l(k)-1)}\|d_{(l(k)-1)}\| = 0,$$

and following the idea in the proof of [18, Theorem 1], we can prove that

$$\lim_{k\to\infty} \alpha_k\|d_k\| = 0. \tag{48}$$

Let $x^* \in C$ a limit point of $\{x_k\}$. Relabel $\{x_k\}$ a subsequence converging to x_*. From (48), there exists a subsequence of indices $K \subset \mathbb{N}$ such that: (i) $\lim_{k\in K}\|d_k\| = 0$ or (ii) $\lim_{k\in K}\alpha_k = 0$.
Case (i): By (45) we can extract a subsequence of indices such $K_1 \subset K$ such that

$$\lim_{k\in K_1} H_k = H_*$$

and H_* also satisfies (43). Hence, we obtain, by continuity and definition of d_k, that

$$\|\tilde{P}_C^{H_*}(x_* - H_*^{-1}F'(x_*)^*F(x_*)) - x_*\| = 0$$

or, equivalently,

$$\tilde{P}_C^{H_*}(x_* - H_*^{-1}F'(x_*)^*F(x_*)) = x_*,$$

which, from Definition 1), implies that x_* is a stationary point of (1).
Case (ii): Let α_k chosen as in step 3.2 such as $\alpha_k = \bar{\alpha}_k/2$, where $\bar{\alpha}_k$ was the last step that failed in (42), i.e.,

$$G(x_k + \bar{\alpha}_k d_k) > \max_{0\le j\le \min\{k,M-1\}} G(x_{k-j}) + \tau\bar{\alpha}_k\langle F'(x_k)^*F(x_k), d_k\rangle \ge G(x_k) + \tau\bar{\alpha}_k\langle F'(x_k)^*F(x_k), d_k\rangle. \tag{49}$$

Now define $s_k = \bar{\alpha}_k d_k$. By the mean value theorem, there exists $\mu_k \in [0,1]$ such that the relation in (49) can be written as

$$\langle F'(x_k + \mu_k s_k)^*F(x_k + \mu_k s_k), s_k\rangle = G(x_k + s_k) - G(x_k) > \tau\langle F'(x_k)^*F(x_k), s_k\rangle. \tag{50}$$

On the other hand, by the fact that $P_C^{H_k}(x_k) = x_k$ and Lemma (2) with $\varepsilon = \theta_k^2\|d_k\|_{H_k}^2$, we have

$$\|d_k\|_{H_k} = \|\tilde{P}_C^{H_k}(y_k) - x_k\|_{H_k} = \|\tilde{P}_C^{H_k}(y_k) - P_C^{H_k}(x_k)\|_{H_k} \le \|y_k - x_k\|_{H_k} + \theta_k\|d_k\|_{H_k},$$

which, combined with the fact that $\theta_k \le \bar{\theta} < 1$ for all $k \ge 0$, (4), (45) and some algebraic manipulations, yields

$$\|d_k\| \leq \sqrt{c}\|d_k\|_{H_k} \leq \frac{\sqrt{c}}{1-\theta_k}\|H_k^{-1}F'(x_k)^*F(x_k)\|_{H_k} \leq \frac{\sqrt{d}}{(1-\bar{\theta})\sqrt{c}}\|F'(x_k)^*F(x_k)\|. \tag{51}$$

Now as $\{x_k\}$ is bounded and G is continuously differentiable, we have $\{F'(x_k)^*F(x_k)\}$ is bounded. Therefore, $\{d_k\}$ is bounded. Hence, as $s_k = 2\alpha_k d_k$ and $\lim_{k\in K}\alpha_k = 0$, we obtain that s_k goes to zero as $k \in K$ goes to infinity. So, from (50), we have

$$\langle F'(x_k+\mu_k s_k)^*F(x_k+\mu_k s_k), \frac{s_k}{\|s_k\|}\rangle > \tau\langle F'(x_k)^*F(x_k), \frac{s_k}{\|s_k\|}\rangle. \tag{52}$$

By taking limit in the last inequality as $k \in K_2$ goes to infinity, where $K_2 \subset K$ is such that $lim_{k\in K_2}\{s_k/\|s_k\|\}$ converges to s, we obtain $(1-\tau)\langle F'(x_*)^*F(x_*), s\rangle \geq 0$. Since $(1-\tau) > 0$, we have

$$\langle F'(x_*)^*F(x_*), s\rangle \geq 0. \tag{53}$$

Now, as d_k is a descent direction for $G(x)$ at x_k and $s_k = \bar{\alpha}_k d_k$, we find

$$\langle F'(x_k)^*F(x_k), \frac{s_k}{\|s_k\|}\rangle < 0.$$

Hence, $\langle F'(x_*)^*F(x_*), s\rangle \leq 0$, which, combined with (53), implies that $\langle F'(x_*)^*F(x_*), s\rangle = 0$. Using (46) and definitions of s_k, we have

$$\langle F'(x_k)^*F(x_k), \frac{s_k}{\|s_k\|}\rangle \leq -(1-\bar{\theta}^2)d\|d_k\|.$$

By (45) we can extract a subsequence of indices such $K_3 \subset K_2$ such that $\lim_{k\in K_3} H_k = H_*$ and H_* satisfying (43). Therefore, by definition of d_k and taking limit in the last inequality as $k \in K_3$ goes to infinity, we have

$$0 = \langle F'(x_*)^*F(x_*), s\rangle \leq -(1-\bar{\theta}^2)d\|\tilde{P}_C^{H_*}(x_* - H_*^{-1}F'(x_*)^*F(x_*)) - x_*\|,$$

which, combined with the fact $\bar{\theta} < 1$, yields $x_* = \tilde{P}_C^{H_*}(x_* - H_*^{-1}F'(x_*)^*F(x_*))$. Therefore, from Definition 1), we conclude that x_* is a stationary point of (1). □

5 Numerical experiments

This section summarizes the results of the numerical experiments we carried out in order to verify the effectiveness of GNM-AP and G-GNM-AP methods. The algorithms were tested on some box- and polyhedral-constrained nonlinear least squares problems. We took $\theta_k = 1/3$, for every k, in both algorithms. Moreover, the ε_k–approximate projection of point y_k onto C was computed by the conditional gradient method, which stopped when either the stopping criterion given in step 2 was satisfied or a maximum of 300 iterations was performed. In order to avoid an excessive number of inner iterations, input ε_k was replaced by $\max\{\theta_k^2\|x_{k+1}-x_k\|_{H_k}^2, 10^{-2}\}$. Linear optimization subproblems in the conditional gradient method (see (9)) were solved via the MATLAB command *linprog*. Other initialization parameters of G-GNM-AP method were set $\tau = 10^{-4}$ and $M = 10$. Nonmonotone parameter $M = 10$ was the best from $\{1,5,10,15,20,25\}$ for an initial small number of problems. For a comparison purpose, we also run the proximal Gauss-Newton (Prox-GN) method in [26], which, applied to (1), corresponds to our GNM-AP method with exact projections (i.e., $\theta_k = 0$ for every k). In the latter method, exact projections were computed by the MATLAB command *quadprog*. In the box-constrained case, we also compare the performance of G-GNM-AP method with the inexact Gauss–Newton trust-region algorithm (ITREBO) of [25]. ITREBO is an algorithm designed for solving nonlinear least-squares problems with simple bounds. For GNM-AP, G-GNM-AP and Prox-GN methods, we used the same termination condition $\|x_{k+1}-x_k\|_{H_k} < 10^{-4}$, whereas in ITREBO we used $\|P_C(x_k - \nabla f(x_k)) - x_k\| < 10^{-4}$. For all algorithms, a failure was declared if the number of iterations was greater than 300 or no progress was detected. The computational results were obtained using MATLAB R2016a on a 2.4GHz Intel(R) i5 with 8GB of RAM and Windows 10 ultimate system.

5.1 Box-constrained nonlinear least squares problems

In this section, our aim is to illustrate the behavior of the algorithms to solve 23 problems of the form (1) with $C = \{x \in \mathbb{R}^n;\ c \le x \le d\}$, where $c, d \in R^n$; see Table (5.1). The first four problems were taken from [26]. The others are originally unconstrained problems for which box constrains were added.

We firstly chose 10 initial points of the form $x_0(\gamma) = c + (\gamma/11)(d-c)$ for $\gamma = 1,2,\ldots,10$. We report in Figure 2 the numerical results of GNM-AP, G-GNM-AP and Prox-GN methods for solving the 23 problems using performance profiles [7]. We adopted the CPU time as performance measurement. It is worth pointing out that the efficiency is related to the percentage of problems for which the method was the fastest, whereas robustness is related to the percentage of problems for which the method found a solution. In the performance profile, efficiency and robustness can be accessed on the left and right extremes of the graphic, respectively.

From Figure 2, we see that GNM-AP method was more robust and efficient in terms of saving time than Prox-GN method. This fact illustrates the advantages of allowing inexactness in the calculation of projections. On the other hand, we also see, as expected, that G-GNM-AP method was more robust than the local methods. Its robustness rate was approximately 95%, whereas for GNM-AP (resp. Prox-GN) the robustness rate was approximately 85% (resp. 71%).

Since our schemes and ITREBO use different stopping criteria, in order to provide a fair comparison, we report in Table (5.1) the performance of G-GNM-AP and ITREBO with three initial point of the form $x_0(\gamma) = c + 0.25\gamma(d-c)$, where $\gamma > 0$, for solving the 23 box-constrained nonlinear least squares problems aforementioned. As can be seen, G-GNM-AP and ITREBO successfully ended 60 and 51 times, respectively, on a total of 69 runs. Moreover, G-GNM-AP (resp. ITREBO) was faster in 31 (resp. 14) cases. Therefore, we can say that our global scheme outperformed ITREBO for the instances considered.

Table 1: Test problems

Problem	Function($F(x)$) and source	n	m	Box
Pb 1	Rosenbrock [26, Problem 1]	2	2	As [26]
Pb 2	Osborne1 [26, Problem 3]	5	33	As [26]
Pb 3	Osborne2 [26, Problem 4]	11	65	As [26]
Pb 4	Twoeq6 [26, Problem 5]	2	2	As [26]
Pb 5	Freudenstein [23, Problem 2]	2	2	$[1,5]$
Pb 6	Powell badly scaled [23, Problem 3]	2	2	$[0,9.106]$
Pb 7	Brown badly scaled [23, Problem 4]	2	3	$[0,10^6]$
Pb 8	Beale [23, Problem 5]	2	3	$[0,3]$
Pb 9	Jennrich and Sampson [23, Problem 6]	2	10	$[-2,1]$
Pb 10	Bard [23, Problem 8]	3	15	$[-10,1]$
Pb 11	Gaussian [23, Problem 9]	3	15	$[-1,1.02]$
Pb 12	Box three-dimensional [23, Problem 12]	3	100	$[0,10]$
Pb 13	Powell singular [23, Problem 13]	4	4	$[-3,3]$
Pb 14	Biggs EXP6 [23, Problem 18]	6	10	$[-1,10]$
Pb 15	Penalty I [23, Problem 23]	4	5	$[-10,1]$
Pb 16	Penalty I [23, Problem 23]	10	11	$[-10,1]$
Pb 17	Variably dimensioned [23, Problem 25]	100	102	$[-1,2]$
Pb 18	Variably dimensioned [23, Problem 25]	450	452	$[-1,2]$
Pb 19	Trigonometric [23, Problem 26]	6	6	$[-2,3]$
Pb 20	Broyden tridiagonal [23, Problem 30]	10	10	$[-2,2]$
Pb 21	Broyden tridiagonal [23, Problem 30]	1000	1000	$[-2,2]$
Pb 22	Example 6.1.10 [11, Chap. 6]	1	2	$[-10,20]$
Pb 23	Example 10.2.4 [6, Chap. 10]	1	3	$[-2,1]$

5.2 Polyhedral-constrained nonlinear least squares problems

In this section, we are interested in solving 23 test problems of the form (1) with $C = \{x \in \mathbb{R}^n;\ c \leq x \leq d,\ Ax \leq b\}$, where $c,d \in R^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$. Our test problems are the box-constrained nonlinear least squares problems of Subsection 5.1, for which randomly generated constraints $Ax \leq b$ were added. In this application, we considered 5 different initial points belonging to the feasible set C.

As in Subsection 5.1, we reported in Figure 3 numerical comparisons of the obtained results using performance profiles. Illustrating again the advantages of allowing inexactness in the calculation of projections, we observe, from Figure 3, that GNM-AP method was more robust and efficient in terms of saving time than Prox-GN method. Moreover, G-GNM-AP was more robust than GNM-AP, which, on the other hand, was more robust than Prox-GN method.

Finally, we conclude that the proposed schemes seems to be promising tools for solving box- and polyhedral-constrained nonlinear least squares problems.

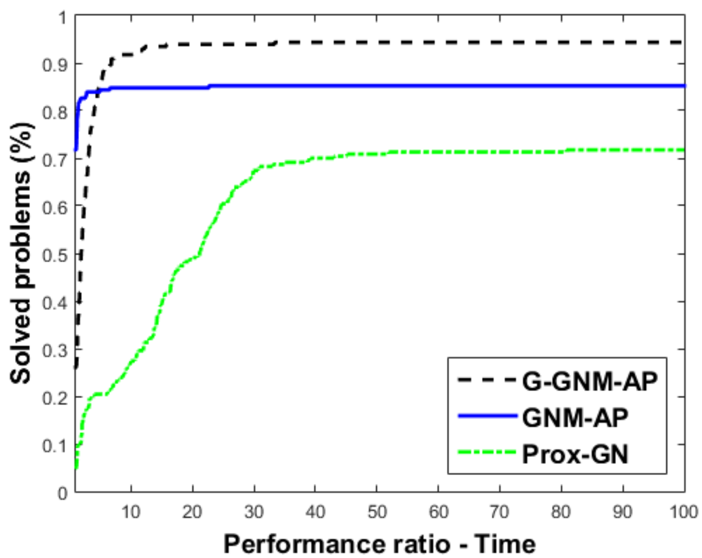

Figure 2: Performance of G-GNM-AP, GNM-AP and Prox-GN methods

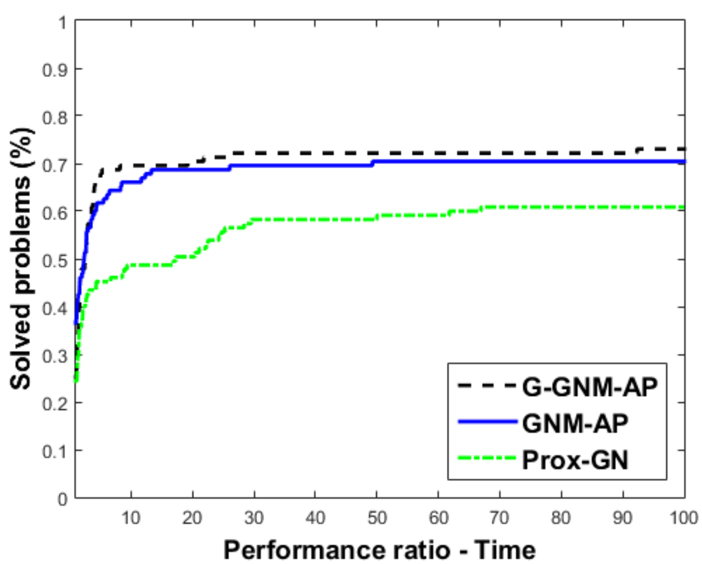

Figure 3: Performance of G-GNM-AP, GNM-AP and Prox-GN methods

Table 2: Performance of G-GNM-AP and ITREBO

		G-GNM-AP	ITREBO			G-GNM-AP	ITREBO
Pb	γ	it/time/Fnorm	it/time/Fnorm	Pb	γ	it/time/Fnorm	it/time/Fnorm
Pb 1	1	273/5.7e+0/1.5e-1	*	Pb 13	1	11/1.1e-2/2.7e-5	8/1.5e-2/3.8e-4
	2	6/4.1e-3/1.3e-1	*		2.5	10/7.9e-3/3.5e-5	7/1.1e-2/6.0e-4
	3	5/3.9e-3/1.3e-1	*		3	11/1.2e-2/4.6e-5	8/1.4e-2/3.8e-4
Pb 2	1	12/9.1e-2/9.0e-2	*	Pb 14	1	186/7.8e-1/4.4e-1	7/1.5e-2/5.4e-1
	2	13/1.0e-1/8.8e-2	*		2	195/6.4e-1/4.4e-1	*
	3	12/9.6e-2/9.0e-2	*		3	31/1.0e-1/4.2e-1	*
Pb 3	1	7/2.9e-2/6.8e-1	*	Pb 15	1	9/6.8e-3/7.9e-3	9/1.2e-2/7.9e-3
	2	8/3.2e-2/6.8e-1	*		2	8/3.6e-3/7.9e-3	8/9.7e-3/7.9e-3
	3	11/4.6e-2/6.8e-1	*		3	7/3.2e-3/7.9e-3	6/8.6e-3/7.9e-3
Pb 4	1	11/6.0e-3/7.1e-5	*	Pb 16	1	9/6.1e-3/1.1e-2	11/1.4e-2/1.1e-2
	2	12/6.5e-3/7.1e-5	*		2	9/5.2e-3/1.1e-2	9/1.3e-2/1.1e-2
	3	16/1.1e-2/1.0e-5	*		3	7/4.4e-3/1.1e-2	7/7.2e-3/1.1e-2
Pb 5	1	6/3.5e-3/3.5e-10	7/9.0e-3/1.2e-7	Pb 17	1	17/4.0e-2/9.1e-6	18/4.7e-2/4.8e-9
	2	6/2.8e-3/2.6e-10	5/6.8e-3/5.3e-8		2	16/3.5e-2/7.7e-8	16/4.2e-2/8.6e-12
	3	2/1.9e-3/0.0e+0	3/5.0e-3/1.8e-7		3	15/3.3e-2/7.7e-8	14/3.4e-2/4.6e-7
Pb 6	1	11/8.6e-3/9.8e-1	11/1.2e-2/9.8e-1	Pb 18	1	30/5.7e-1/9.9e-6	23/3.8e-1/6.1e-7
	2	12/8.4e-3/9.8e-1	14/1.2e-2/9.8e-1		2	63/1.2e+0/9.5e-5	21/3.6e-1/8.7e-10
	3	12/8.9e-3/9.8e-1	15/1.2e-2/9.8e-1		3	17/3.5e-1/9.9e-6	19/3.1e-1/4.9e-8
Pb 7	1	18/3.3e-2/0.0e+0	36/2.8e-2/0.0e+0	Pb 19	1	7/3.7e-3/5.3e-8	6/7.3e-3/2.2e-7
	2	19/3.2e-2/0.0e+0	35/2.9e-2/2.2e+0		2	*	14/1.6e-2/1.6e-2
	3	20/3.7e-2/0.0e+0	37/2.8e-2/1.6e+0		3	*	17/1.6e-2/1.6e-2
Pb 8	1	5/3.6e-3/6.0e-5	7/8.7e-3/2.4e-7	Pb 20	1	4/3.6e-3/9.1e-5	4/6.8e-3/3.4e-8
	2	6/3.5e-3/6.0e-5	9/1.0e-2/7.2e-7		2	5/2.4e-2/4.5e-5	7/1.5e-2/1.3e-7
	3	11/7.3e-3/6.4e-5	10/1.0e-2/7.8e-8		3	*	26/2.8e-2/1.1e+0
Pb 9	1	*	10/1.4e-2/1.1e+1	Pb 21	1	5/4.9e-1/1.0e-9	6/4.1e-1/6.9e-6
	2	35/5.1e-1/1.1e+1	7/1.4e-2/1.1e+1		2	*	189/1.7e+1/9.6e+0
	3	*	5/1.3e-2/1.1e+1		3	139/1.4e+1/1.2e+0	16/1.2e+0/1.0e+0
Pb 10	1	*	*	Pb 22	1	6/2.3e-2/1.4e+0	6/5.8e-2/1.4e+0
	2	*	*		2	7/2.4e-3/1.4e+0	7/9.4e-3/1.4e+0
	3	*	*		3	8/4.8e-3/1.4e+0	9/1.5e-2/1.4e+0
Pb 11	1	9/2.1e-2/1.0e-1	51/6.7e-2/1.0e-1	Pb 23	1	7/8.6e-3/8.7e-6	7/1.0e-2/7.6e-8
	2	7/6.7e-3/1.0e-1	5/1.0e-2/1.0e-1		2	6/4.7e-3/1.9e-7	5/1.3e-2/7.6e-8
	3	4/4.1e-3/1.0e-1	3/8.4e-3/1.0e-1		3	5/2.9e-3/1.9e-7	5/6.9e-3/7.6e-8
Pb 12	1	2/1.2e-2/1.7e-15	*				
	2.5	4/3.1e-2/6.0e-6	4/3.9e-2/1.3e-4				
	3	8/3.9e-2/4.2e-7	5/3.4e-2/5.8e-12				

6 Final remark

In this paper, we proposed Gauss-Newton methods with inexact projections for solving constrained nonlinear least squares problems. For the local method, we were able to show, under a majorant condition, that the generated sequence converges locally linearly. In zero-residual problems, quadratic convergence rate can be achieved with a stronger condition on the inexactness of the projections. As special cases of the majorant condition, convergence results for the method with F' satisfying a Lipschitz-like condition and F being an analytic function satisfying a Smale condition were also discussed. For the global method, under suitable conditions, global convergence of the algorithm to a stationary point of the problem was established. The numerical experiments showed that the new algorithms work quite well and compare favorably with the proximal Gauss-Newton method in [26] and the inexact Gauss–Newton trust-region method in [25].

References

[1] N. M. Alsaifi and P. Englezos. Prediction of multiphase equilibrium using the pc-saft equation of state and simultaneous testing of phase stability. *Fluid Phase Equilibria*, 302(1):169–178, 2011.

[2] R. Andreani, A. Friedlander, M. P. Mello, and S. A. Santos. Box-constrained minimization reformulations of complementarity problems in second-order cones. *J. Global Optim.*, 40(4):505–527, 2008.

[3] I. K. Argyros and A. A. Magreñán. Local convergence analysis of proximal Gauss-Newton method for penalized nonlinear least squares problems. *Appl. Math. Comput.*, 241:401–408, 2014.

[4] L. Bencini, R. Fantacci, and L. Maccari. Analytical model for performance analysis of ieee 802.11 dcf mechanism in multi-radio wireless networks. In *2010 IEEE International Conference on Communications*, pages 1–5. IEEE, 2010.

[5] J.-P. Dedieu and M. Shub. Newton's method for overdetermined systems of equations. *Math. Comput.*, 69(231):1099–1115, 2000.

[6] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1996.

[7] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. program.*, 91(2):201–213, 2002.

[8] J. C. Dunn. Convergence rates for conditional gradient sequences generated by implicit step length rules. *SIAM J. Control Optim.*, 18(5):473–487, 1980.

[9] O. P. Ferreira, M. L. N. Gonçalves, and P. R. Oliveira. Local convergence analysis of the Gauss-Newton method under a majorant condition. *J. Complexity*, 27(1):111–125, 2011.

[10] O. P. Ferreira, M. L. N. Gonçalves, and P. R. Oliveira. Convergence of the Gauss-Newton method for convex composite optimization under a majorant condition. *SIAM J. Optim.*, 23(3):1757–1783, 2013.

[11] R. Fletcher. *Practical methods of optimization*. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, second edition, 1987.

[12] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Res. Logist. Quart.*, 3(1-2):95–110, 1956.

[13] H. Fu, H. Liu, B. Han, Y. Yang, and Y. Hu. A proximal iteratively regularized Gauss-Newton method for nonlinear inverse problems. *J. Inverse Ill-Posed Probl.*, 25(3):341–356, 2017.

[14] M. L. N. Gonçalves. Local convergence of the Gauss-Newton method for injective-overdetermined systems of equations under a majorant condition. *Comput. Math. Appl.*, 66(4):490–499, 2013.

[15] M. L. N. Gonçalves. Inexact Gauss-Newton like methods for injective-overdetermined systems of equations under a majorant condition. *Numer. Algorithms*, 72(2):377–392, 2016.

[16] M. L. N. Gonçalves and J. G. Melo. A Newton conditional gradient method for constrained nonlinear systems. *J. Comput. Appl. Math.*, 311:473–483, 2017.

[17] M. L. N. Gonçalves and F. R. Oliveira. An inexact Newton-like conditional gradient method for constrained nonlinear systems. *Appl. Numer. Math.*, 132:22–34, 2018.

[18] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, 23(4):707–716, 1986.

[19] Z. Huang. The convergence ball of Newton's method and the uniqueness ball of equations under Hölder-type continuous derivatives. *Comput. Math. Appl.*, 47(2):247–251, 2004.

[20] C. Li, N. Hu, and J. Wang. Convergence behavior of Gauss-Newton's method and extensions of the Smale point estimate theory. *J. Complexity*, 26(3):268–295, 2010.

[21] Z. F. Li, M. R. Osborne, and T. Prvan. Adaptive algorithm for constrained least-squares problems. *J. Optim. Theory Appl.*, 114(2):423–441, 2002.

[22] N. Mahdavi-Amiri and R. H. Bartels. Constrained nonlinear least squares: An exact penalty approach with projected structured quasi-Newton updates. *ACM Trans. Math. Software*, 15(3):220–242, 1989.

[23] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. Testing unconstrained optimization software. *ACM Trans. Math. Software*, 7(1):17–41, 1981.

[24] P. M. Pardalos and M. G. C. Resende, editors. *Handbook of applied optimization*. Oxford University Press, Oxford, 2002.

[25] M. Porcelli. On the convergence of an inexact Gauss-Newton trust-region method for nonlinear least-squares problems with simple bounds. *Optim. Lett.*, 7(3):447–465, 2013.

[26] S. Salzo and S. Villa. Convergence analysis of a proximal Gauss-Newton method. *Comput. Optim. Appl.*, 53(2):557–589, 2012.

[27] W. Shen and C. Li. Smale's α-theory for inexact Newton methods under the γ-condition. *J. Math. Anal. Appl.*, 369(1):29–42, 2010.

[28] S. Smale. Newton's method estimates from data at one point. In *The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985)*, pages 185–196. Springer, New York, 1986.