LOCAL CONVERGENCE OF THE GAUSS–NEWTON METHODS FOR CONSTRAINED NONLINEAR EQUATIONS*

Alexey F. Izmailov[1] and Mikhail V. Solodov[2]

January 11, 2026

ABSTRACT

We derive, from unified principles, local convergence and rate-of-convergence results for the classical Gauss-Newton method in a variety of settings. These include overdetermined and underdetermined systems of equations, constrained and unconstrained, possibly with inexact solution of subproblems, as well as the projected variant in the constrained case. Moreover, by a counter-example we show that contrary to some results claimed in the literature, the projected Gauss-Newton method in general does not converge superlinearly under any reasonable assumptions. We then establish its linear rate of convergence.

Key words: (constrained) nonlinear equation; Gauss-Newton method; semistability; hemistability; nonisolated solutions; superlinear convergence; linear convergence.
AMS subject classifications. 65J15, 49M15.

---

* Research of the second author was supported in part by CNPq Grant 306775/2023-9 and by FAPERJ Grant E-26/200.347/2023.

[1] VMK Faculty, OR Department, Lomonosov Moscow State University (MSU), Uchebniy Korpus 2, Leninskiye Gory, 119991 Moscow, Russia.
Email: izmaf@cs.msu.ru

[2] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: `solodov@impa.br`

1 Introduction

We consider the constrained equation

$$\Phi(u) = 0, \quad u \in P, \tag{1.1}$$

where $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ is a given smooth mapping, and $P \subset \mathbb{R}^p$ is a given nonempty closed convex set. Constrained equations often arise as reformulations of systems containing complementarity conditions; see, e.g., [18]. Some examples are the first-order optimality systems for optimization and variational problems, including mathematical programs with complementarity constraints [13], and generalized Nash equilibrium problems [12]. Some other applications are discussed in [20, 22].

Without imposing any relations between the number of variables and equations (p and q) and any assumptions on the structure of the set P, a natural Newton-type method for solving (1.1) is the *constrained Gauss–Newton method*. For a current iterate $u \in P$, the next iterate is $u+v$, where v is some minimizer of the (squared) residual of the linearized equation from (1.1) over the set $P - u$, i.e., v is a solution of the optimization problem

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 \quad \text{subject to} \quad u + v \in P. \tag{1.2}$$

Due to the Frank–Wolfe Theorem [16] (alternative proofs of which can be found in [7, 11] and in [23, Theorem 2.1]), the subproblem (1.2) with a nonnegative (thus bounded below) quadratic objective function always has a solution when P is polyhedral, but it need not be unique. The solution is necessarily unique when

$$\ker \Phi'(u) = \{0\}, \tag{1.3}$$

as in this case the matrix $(\Phi'(u))^\top \Phi'(u)$ of the quadratic objective function in (1.2) is positive definite, and hence, this function is strongly convex. Moreover, in this case the solution exists even if P is not polyhedral. In the unconstrained case, i.e., when $P = \mathbb{R}^p$, this solution is explicitly available from setting the gradient of the objective function in (1.2) equal to zero and resolving the corresponding linear equation. This yields the classical Gauss–Newton method for overdetermined systems of equations [17, 21, 25], most commonly introduced in the context of solving nonlinear least-squares problems (i.e., minimizing the squared residual $\|\Phi(\cdot)\|^2$ over $\mathbb{R}^p$).

However, (1.3) may not hold along all the iterations. In fact, it never holds in the case of underdetermined systems of equations, i.e., when $p > q$. Then there is no reason to expect that the subproblem (1.2) is uniquely solvable, and for any convergence rates analysis one has to complement the iteration by some rule for choosing a specific solution of the subproblem (1.2), or demonstrate that any solution of (1.2) is acceptable. A natural option is to consider the minimal-norm solution of (1.2). In the unconstrained case, under reasonable assumptions, the minimal-norm solution of (1.2) can be found explicitly, as will be recalled in Section 3 below. This yields the classical Gauss–Newton method for undetermined systems of equations [21].

In the constrained case when P is polyhedral, finding the minimal-norm solution of (1.2) amounts to solving two quadratic programming (QP) problems. The first computes *some* solution of (1.2), which is a QP in this case. The second QP computes the minimal-norm point of the solution set of (1.2), which is a polyhedral set and can be explicitly characterized as explained in [24], once some solution of (1.2) is known.

However, the approach just described requires solving two QPs per iteration, where the matrix of the objective function of the first QP is positive semidefinite but not positive definite, while the second QP involves $p+1$ extra equality constraints in addition to those defining P. When computing the metric projection $\pi_P(\cdot)$ onto P is cheap (e.g., when P is a box), it makes sense to consider instead the *projected Gauss–Newton method* [1, 22]: for $u \in P$, it defines the next iterate as $\pi_P(u+v(u))$, where $v(u)$ is the minimal-norm solution of the unconstrained counterpart of (1.2), with P replaced by $\mathbb{R}^p$.

The main purpose of this paper is to derive local convergence properties of the Gauss–Newton method and its variants in a variety of settings (unconstrained, constrained, overdertermined, underdetermined, inexact) from a unified perspective (in particular, from the abstract framework developed in [15]). Apart from giving a general view and insights, this also leads to some new results for the constrained case; specifically, for the constrained Gauss-Newton method in the case when the solution in question is isolated, and for the projected Gauss-Newton method in the cases of square or underdetermined systems.

The rest of the paper is organized as follows. In Section 2 we provide the auxiliary tools needed for our analysis. One is the general local convergence framework from [15]. Section 3 deals with local superlinear convergence of the Gauss–Newton method for the unconstrained equations, mainly addressing the case of square and underdetermined systems. The case of possibly overdetermined or undetermined systems is treated in Section 4 for the more general constrained setting, and for the constrained Gauss–Newton method. In particular, apart from results concerned with the variant of the method employing the minimal-norm solutions, we provide assumptions ensuring that *any* solution of subproblem (1.2) yields superlinear local convergence. Finally, in Section 5, by a counter-example we demonstrate that under reasonable assumptions, one cannot expect superlinear convergence of the projected Gauss–Newton method. This is in contrast to some previous results that do claim superlinear convergence. Once this issue is settled, we establish local R-linear convergence for the square or underdetermined constrained systems of equations.

Our notation is quite standard, but some comments are in order, to avoid any misunderstandings. For a convex P, by $\operatorname{ri} P$ we denote its relative interior, by $T_P(u)$ the tangent cone to P at $u \in P$, i.e., the closure of the cone $\{v \in \mathbb{R}^p \mid \exists\, t > 0 : \ u + tv \in P\}$ of feasible directions to P at $\bar{u}$, and by $N_P(u)$ the normal cone to P at $u \in \mathbb{R}^p$, i.e., $N_P(u) = \{v \in \mathbb{R}^p \mid \langle v, \widetilde{u} - u\rangle \leq 0 \ \forall\, \widetilde{u} \in P\}$ if $u \in P$, and $N_P(u) = \emptyset$ otherwise, where $\langle u, v\rangle$ is the Euclidean inner product of $u, v \in \mathbb{R}^p$. Let $\|\cdot\|$ stand for the Euclidean norm throughout. For P closed and convex, $\pi_P(u)$ is the unique Euclidean projection of u onto P. For a set $U \subset \mathbb{R}^p$ and a point $u \in \mathbb{R}^p$, let $\operatorname{dist}(u, U) = \inf_{v\in U} \|v - u\|$ stand for the distance from u to U, and let $B(u, \delta) = \{v \in \mathbb{R}^p \mid \|v - u\| \leq \delta\}$ be the (closed) ball of radius $\delta \geq 0$ centered at u. For a sequence $\{u^k\} \subset \mathbb{R}^p$ convergent to some $\bar{u} \in \mathbb{R}^p$, we say that the rate of convergence is R-linear if there exist $c > 0$ and $\rho \in (0, 1)$ such that $\|u^k - \bar{u}\| \leq c\rho^k$ for all k. Furthermore, we say that convergence is of Q-order $\theta > 1$ if there exists $c > 0$ such that $\|u^{k+1} - \bar{u}\| \leq c\|u^k - \bar{u}\|^\theta$ for all k large enough. Such rate of convergence is superlinear, and it is at least quadratic if $\theta \geq 2$.

2 Preliminaries

Our analysis will make use of the abstract local convergence framework, recently proposed in [15]. Consider a scalar constrained equation

$$\varphi(u) = 0, \quad u \in P, \tag{2.1}$$

with $\varphi : \mathbb{R}^p \to \mathbb{R}_+$ and $P \subset \mathbb{R}^p$ for now assumed just closed. Let U stand for the solution set of (2.1).

Observe that the constrained equation (1.1) can be equivalently stated in the form (2.1) by taking, say,

$$\varphi(u) = \|\Phi(u)\|. \tag{2.2}$$

We consider an abstract iterative process updating the current iterate $u \in P$ to $\Psi(u)$, where $\Psi : P \to P$ is a given mapping. The following is a simplified version of [15, Theorem 2.1].

Theorem 2.1 *Let $\varphi : \mathbb{R}^p \to \mathbb{R}_+$ be a continuous function, $P \subset \mathbb{R}^p$ be a closed set, $\bar{u} \in U$. Assume that*

$$\varphi(u) = O(\mathrm{dist}(u, U)) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{2.3}$$

Moreover, let $\Psi : P \to P$ be a mapping such that

$$\Psi(u) - u = O(\varphi(u)) \quad \text{as } u \in P \text{ tends to } \bar{u}, \tag{2.4}$$

and, with some $\theta > 1$,

$$\varphi(\Psi(u)) = O((\varphi(u))^\theta) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{2.5}$$

Then, for every $\delta > 0$ small enough, and every $u^0 \in P$ close enough to $\bar{u}$, the sequence $\{u^k\}$ defined by $u^{k+1} = \Psi(u^k)$ for all k is contained in $B(\bar{u}, \delta)$ and converges to some $u^* \in U$, with the rate of convergence being superlinear with the Q-order θ.

From now on, let U stand for the solution set of (1.1). Then the assumption (2.3) evidently holds for φ defined in (2.2) provided Φ is Lipschitz-continuous on P near $\bar{u}$, and this will be the case under the assumptions of all the results that follow.

Observe further that any $u \in U$ is a (global) solution of the optimization problem

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u)\|^2 \quad \text{subject to} \quad u \in P,$$

and if Φ is differentiable at u, the objective function of this problem is also differentiable at u, with the gradient being $(\Phi'(u))^\top \Phi(u)$. Therefore, any such u must satisfy the first-order necessary optimality condition given by the variational inequality (VI)

$$u \in P, \quad \langle (\Phi'(u))^\top \Phi(u),\, \widetilde{u} - u\rangle \geq 0 \quad \forall\, \widetilde{u} \in P,$$

which in the sequel we shall use in its equivalent form of

$$(\Phi'(u))^\top \Phi(u) + N_P(u) \ni 0. \tag{2.6}$$

The analysis in Section 4 will rely upon two key assumptions. The first is semistability of $\bar{u}$ as a solution of the VI (2.6), as defined in [5] (see also [17, Definition 1.29]). The second assumption is an adaptation of hemistability, also defined in [5] (see also [17, Definition 3.1]). The relations of those assumptions to some others, including less abstract ones, and those typically used in the context of unconstrained Gauss–Newton methods, will be explored in due course.

Definition 2.1 A solution $\bar{u}$ of the VI (2.6) is *semistable* if for any $\omega \in \mathbb{R}^p$, and for any solution u of the perturbed VI

$$(\Phi'(u))^\top \Phi(u) + N_P(u) \ni \omega, \tag{2.7}$$

with this solution u close enough to $\bar{u}$, it holds that

$$u - \bar{u} = O(\|\omega\|) \quad \text{as } \omega \to 0. \tag{2.8}$$

Theorem 2.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a convex set. Assume that Φ is differentiable near a solution $\bar{u}$ of* (1.1)*, and its derivative is continuous at $\bar{u}$ with respect to P (that is, $\Phi'(u) \to \Phi'(\bar{u})$ as $u \in P$ tends to $\bar{u}$).*

Then the following three properties are equivalent:

(a) The point $\bar{u}$ is semistable as a solution of the VI (2.6).

(b) The constrained error bound holds:

$$u - \bar{u} = O(\|\Phi(u)\|) \quad \textit{as } u \in P \textit{ tends to } \bar{u}. \tag{2.9}$$

(c) It holds that

$$\ker \Phi'(\bar{u}) \cap T_P(\bar{u}) = \{0\}. \tag{2.10}$$

Proof. We first show that (a) implies (b). To that end, let $\bar{u}$ be semistable. Take any $u \in P$ and set $\omega = (\Phi'(u))^\top \Phi(u)$. Then u is a solution of (2.7), and under the stated smoothness assumptions it holds that $\omega \to 0$ as $u \to \bar{u}$. Therefore, by (2.8),

$$u - \bar{u} = O(\|\omega\|) = O(\|(\Phi'(u))^\top \Phi(u)\|) = O(\|\Phi(u)\|) \quad \text{as } u \in P \text{ tends to } \bar{u},$$

i.e., (2.9) holds.

We proceed with showing that (b) implies (c). Let (2.9) be satisfied, and suppose that (2.10) is violated, that is, there exists $v \in \ker \Phi'(\bar{u}) \cap T_P(\bar{u})$, $\|v\| = 1$. As $v \in T_P(\bar{u})$, there exists a sequence $\{u^k\} \subset P \setminus \{\bar{u}\}$ convergent to $\bar{u}$ such that the sequence $\{v^k\}$ with $v^k = (u^k - \bar{u})/\|u^k - \bar{u}\|$ converges to v. Therefore, as $\Phi'(\bar{u})v^k \to \Phi'(\bar{u})v = 0$ as $k \to \infty$, we obtain that

$$\begin{aligned} \Phi(u^k) &= \Phi(\bar{u}) + \Phi'(\bar{u})(u^k - \bar{u}) + o(\|u^k - \bar{u}\|) \\ &= \|u^k - \bar{u}\| \Phi'(\bar{u})v^k + o(\|u^k - \bar{u}\|) \\ &= o(\|u^k - \bar{u}\|) \quad \text{as } k \to \infty. \end{aligned}$$

Hence, by (2.9),

$$u^k - \bar{u} = o(\|u^k - \bar{u}\|) \quad \text{as } k \to \infty,$$

which is a contradiction.

Finally, we show that (c) implies (a). Let (2.10) be satisfied, and suppose $\bar{u}$ is not semistable, that is, there exist sequences $\{u^k\} \subset P \setminus \{\bar{u}\}$ converging to $\bar{u}$, and $\{\omega^k\} \subset \mathbb{R}^p$, such that for every k, the point u^k is a solution of (2.7) with $\omega = \omega^k$, and

$$\omega^k = o(\|u^k - \bar{u}\|) \quad \text{as } k \to \infty.$$

Then, substituting $\Phi(u^k)$ in (2.7) again by its expansion $\Phi'(\bar{u})(u^k - \bar{u}) + o(\|u^k - \bar{u}\|)$, we obtain that

$$(\Phi'(u^k))^\top \Phi'(\bar{u})(u^k - \bar{u}) + N_P(u^k) \ni o(\|u^k - \bar{u}\|) \quad \text{as } k \to \infty.$$

Dividing both sides of this inclusion by $\|u^k - \bar{u}\|$, defining $v^k = (u^k - \bar{u})/\|u^k - \bar{u}\|$, and recalling that $N_P(u^k)$ is a cone (and hence, dividing by a positive number does not change it), we obtain that for each k, there exists $\eta^k \in N_P(u^k)$ such that

$$(\Phi'(u^k))^\top \Phi'(\bar{u})v^k + \eta^k \to 0 \quad \text{as } k \to \infty.$$

Multiplying both sides by v^k, we then obtain that

$$\langle \Phi'(\bar{u})v^k, \Phi'(u^k)v^k \rangle + \frac{\langle \eta^k, u^k - \bar{u} \rangle}{\|u^k - \bar{u}\|} \to 0 \quad \text{as } k \to \infty.$$

Since $\bar{u} \in P$ and $\eta^k \in N_P(u^k)$, the second term in the left-hand side of the relation above is nonnegative. Assuming without loss of generality that $\{v^k\}$ converges to some $v \in \mathbb{R}^p$, $\|v\| = 1$, and passing onto the limit as $k \to \infty$, we then obtain that $\|\Phi'(\bar{u})v\|^2 = 0$, i.e., $v \in \ker \Phi'(\bar{u})$. At the same time, the definition of v^k implies that $v \in T_P(\bar{u})$. Since $v \neq 0$, this yields a contradiction with (2.10). ∎

We mention that assuming twice differentiability of Φ at $\bar{u}$, the fact that (a) implies (c), as well as the converse implication under the additional assumption that P is a polyhedral set, can also be obtained from [5, Theorem 3.1, Remark 3.2].

Since semistability implies the constrained error bound (2.9), it follows, in particular, that a semistable solution is an isolated point in U. We note also that semistability does not assert the existence of solutions of the perturbed VIs (2.7), or solvability of the iteration subproblems (1.2). Existence of solutions of (1.2) is part of hemistability, defined next.

Definition 2.2 Assuming that Φ is differentiable near a solution $\bar{u}$ of (1.1), we say that $\bar{u}$ is *hemistable* if for any $u \in P$ close enough to $\bar{u}$, the subproblem (1.2) has a solution $v(u)$ such that $v(u) \to 0$ as $u \to \bar{u}$.

The first-order necessary and sufficient optimality condition for the convex problem (1.2) has the form

$$(\Phi'(u))^\top (\Phi(u) + \Phi'(u)v) + N_P(u + v) \ni 0, \tag{2.11}$$

so one can think of $v(u)$ as a solution of (2.11). Note that this is not exactly hemistability for the VI (2.11) as defined in [5], as the latter would require involving second derivatives of Φ, the existence of which we never assume in this paper. Rather, the definition is adapted to the structure of the Gauss–Newton iteration subproblem (1.2), or more precisely, specifically to that of the VI (2.11).

3 The unconstrained case

Within this short section $P = \mathbb{R}^p$. The case of square or overdetermined systems (when $p \leq q$) is the most classical. It is considered, e.g., in [8, Corollary 10.2.2], [25, Exercise 10.2.10], [21, Theorem 2.4.1] and references therein. The standard regularity assumption for convergence is

$$\ker \Phi'(\bar{u}) = \{0\}. \tag{3.1}$$

We shall cover these results in Section 4, in the more general constrained case.

The rest of this section is concerned with square or underdetermined systems, i.e., those with $p \geq q$. The solutions of (1.2) with $P = \mathbb{R}^p$ are characterized by the gradient of the objective function in (1.2) being equal to zero, which yields the linear equation $(\Phi'(u))^\top(\Phi(u) + \Phi'(u)v) = 0$. Then computing the minimal-norm solution of (1.2) is the following QP:

$$\text{minimize} \quad \frac{1}{2}\|v\|^2 \quad \text{subject to} \quad (\Phi'(u))^\top(\Phi(u) + \Phi'(u)v) = 0. \tag{3.2}$$

If

$$\operatorname{rank} \Phi'(u) = q \tag{3.3}$$

(and hence $\ker(\Phi'(u))^\top = \{0\}$, so that the matrix $\Phi'(u)(\Phi'(u))^\top$ is nonsingular), the constraint in (3.2) reduces to $\Phi(u) + \Phi'(u)v = 0$. Then the subproblem (3.2) further simplifies to

$$\text{minimize} \quad \frac{1}{2}\|v\|^2 \quad \text{subject to} \quad \Phi(u) + \Phi'(u)v = 0. \tag{3.4}$$

Moreover, the unique solution of (3.4) can be explicitly obtained from its Lagrange optimality conditions, resulting in

$$v(u) = -(\Phi'(u))^\top(\Phi'(u)(\Phi'(u))^\top)^{-1}\Phi(u). \tag{3.5}$$

Assume that $\bar{u} \in U$ satisfies the regularity condition

$$\operatorname{rank} \Phi'(\bar{u}) = q. \tag{3.6}$$

Assume further that Φ is differentiable near $\bar{u}$, and with some $\tau > 0$, it holds that

$$\Phi'(u^1) - \Phi'(u^2) = O(\|u^1 - u^2\|^\tau) \quad \text{as } u^1,\, u^2 \to \bar{u}, \tag{3.7}$$

i.e., Φ' is Hölder-continuous near $\bar{u}$ with the exponent τ. Then (3.3) holds for all $u \in \mathbb{R}^p$ close enough to $\bar{u}$, and since the inverse matrices to those close enough to a nonsingular square matrix are uniformly bounded, we conclude that $v(u)$ in (3.5) is well-defined for all $u \in \mathbb{R}^p$ close enough to $\bar{u}$, and moreover, for φ defined in (2.2), we have that

$$v(u) = O(\|\Phi(u)\|) \quad \text{as } u \to \bar{u}. \tag{3.8}$$

This implies that (2.4) is satisfied as well, if we set $\Psi(u) = u + v(u)$.

Furthermore, by the constraint in (3.4) and the Mean-Value Theorem (e.g., [17, Theorem A.10]), employing (3.7) we derive that

$$\begin{aligned} \|\Phi(u + v(u))\| &= \|\Phi(u + v(u)) - \Phi(u) - \Phi'(u)v(u)\| \\ &\leq \sup_{t \in [0,\, 1]} \|\Phi'(u + tv(u)) - \Phi'(u)\| \|v(u)\| \\ &= O(\|v(u)\|^{1+\tau}) \\ &= O(\|\Phi(u)\|^{1+\tau}) \quad \text{as } u \to \bar{u}, \end{aligned} \tag{3.9}$$

where the last estimate is by (3.8). According to (2.2), this yields (2.5) with $\theta = 1 + \tau$, and hence, Theorem 2.1 is applicable. This proves the following results, covering [21, Theorem 2.4.2].

Proposition 3.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be differentiable near a solution $\bar{u}$ of the equation*

$$\Phi(u) = 0, \tag{3.10}$$

with the derivative of Φ satisfying (3.7) *with some $\tau > 0$. Let the regularity condition* (3.6) *be satisfied.*

Then, for every $\delta > 0$ small enough, and every $u^0 \in \mathbb{R}^p$ close enough to $\bar{u}$, there exists the unique sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ equals $v(u^k)$ defined by (3.5) *with $u = u^k$, which is the minimal-norm solution of the problem*

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2, \quad u \in \mathbb{R}^p, \tag{3.11}$$

(that is, of (1.2) *with $P = \mathbb{R}^p$), this sequence is contained in $B(\bar{u}, \delta)$ and converges to some solution u^* of* (3.10)*, with the rate of convergence being superlinear with the Q-order $1+\tau$. In particular, if the derivative of Φ is Lipschitz-continuous near $\bar{u}$, then the rate of convergence is quadratic.*

The last claim follows by observing that under the Lipschitz-continuity of Φ' near $\bar{u}$, (3.7) automatically holds with $\tau = 1$.

4 Constrained Gauss–Newton method

The analysis in this section starts with a local convergence result under semistability and hemistability of the solution in question.

Theorem 4.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a closed convex set. Assume that Φ is differentiable near a solution $\bar{u}$ of* (1.1)*, and*

$$\Phi'(u^1) - \Phi'(u^2) = O(\|u^1 - u^2\|^\tau) \quad \text{as } u^1,\, u^2 \in P \text{ tend to } \bar{u} \tag{4.1}$$

with some $\tau > 0$ (i.e., Φ' is Hölder-continuous on P near $\bar{u}$ with the exponent τ). Let $\bar{u}$ be semistable and hemistable.

Then, for every $u^0 \in P$ close enough to $\bar{u}$, there exists the unique sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is the minimal-norm solution of the problem (1.2) *with $u = u^k$, this sequence converges to $\bar{u}$, and the rate of convergence is superlinear with the Q-order $1 + \tau$. In particular, if the derivative of Φ is Lipschitz-continuous on P near $\bar{u}$, then the rate of convergence is quadratic.*

Proof. By the hemistability of $\bar{u}$, for any $u \in P$ close enough to $\bar{u}$, the subproblem (1.2) has a solution tending to 0 as $u \to \bar{u}$. Hence, the solution set of this convex problem is nonempty, closed and convex. Therefore, it has the unique minimal-norm solution $v(u)$, and $v(u) \to 0$ as $u \to \bar{u}$. As in Section 3, we aim to apply Theorem 2.1 with φ defined in (2.2), and with $\Psi(u) = u + v(u)$. To begin with, from the differentiability of Φ at $\bar{u}$, it evidently follows that the assumption (2.3) in Theorem 2.1 is satisfied in the form

$$\varphi(u) = O(\|u - \bar{u}\|) \quad \text{as } u \to \bar{u}.$$

Furthermore, $u + v(u)$ is a solution of the VI (2.7) with

$$\begin{aligned} \omega &= (\Phi'(u + v(u)))^\top \Phi(u + v(u)) - (\Phi'(u))^\top (\Phi(u) + \Phi'(u)v(u)) \\ &= \left((\Phi'(u + v(u)))^\top - (\Phi'(u))^\top\right) \Phi(u) \\ &\quad + \left((\Phi'(u + v(u)))^\top - (\Phi'(u))^\top\right) (\Phi(u + v(u)) - \Phi(u)) \\ &\quad + (\Phi'(u))^\top (\Phi(u + v(u)) - \Phi(u) - \Phi'(u)v(u)). \end{aligned} \tag{4.2}$$

Taking into account that $u \in P$ and $u + v(u) \in P$, and employing again the Mean-Value Theorem and (4.1), we then obtain the estimate

$$\omega = O(\|v(u)\|^\tau \|\Phi(u)\|) + O(\|v(u)\|^{1+\tau}) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{4.3}$$

In particular, $\omega \to 0$ as $u \to \bar{u}$, and by the semistability of $\bar{u}$ it holds that

$$u + v(u) - \bar{u} = O(\|v(u)\|^\tau \|\Phi(u)\|) + O(\|v(u)\|^{1+\tau}) \quad \text{as } u \in P \text{ tends to } \bar{u}.$$

Furthermore, since $v(u) \to 0$ as $u \to \bar{u}$, it holds that

$$u + v(u) - \bar{u} = o(\|\Phi(u) - \Phi(\bar{u})\|) + o(\|v(u)\|) = o(\|u - \bar{u}\|) + o(\|v(u)\|) \quad \text{as } u \in P \text{ tends to } \bar{u}.$$

Employing now the error bound (2.9) from Theorem 2.2, this yields the estimate

$$v(u) = O(\|u - \bar{u}\|) = O(\|\Phi(u)\|) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{4.4}$$

This justifies (2.4).

Finally, by the differentiability of Φ at $\bar{u}$, and by (4.3) and (4.4),

$$\Phi(u+v(u)) = \Phi(u+v(u)) - \Phi(\bar{u}) = O(\|u+v(u)-\bar{u}\|) = O(\|\Phi(u)\|^{1+\tau}) \quad \text{as } u \in P \text{ tends to } \bar{u},$$

justifying (2.5) with $\theta = 1 + \tau$.

The needed conclusions now follow by applying Theorem 2.1. ∎

We next discuss some verifiable sufficient conditions ensuring semistability and hemistability of $\bar{u}$. Assume first that $\bar{u} \in U$ satisfies the regularity condition (3.1) (or equivalently, $\operatorname{rank} \Phi'(\bar{u}) = p$), which is of course only possible when $p \leq q$. What follows extends the classical results on local convergence of the Gauss–Newton method for square and overdetermined systems of equations (see [17, 21, 25] and references therein), from the unconstrained to the constrained case. In this case, there is no need to refer to the minimal-norm solution of (1.2), as the solution of this problem is unique, at least for $u \in P$ near $\bar{u}$. Furthermore, since (3.1) evidently implies (2.10), from Theorem 2.2 we have that semistability of $\bar{u}$ holds under (3.1). Hemistability is addressed in the next lemma.

Lemma 4.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a closed convex set. Assume that Φ is differentiable near a solution $\bar{u}$ of* (1.1)*, and its derivative is continuous at $\bar{u}$ with respect to P. Let the regularity condition* (3.1) *be satisfied.*

Then $\bar{u}$ is hemistable. More precisely, for every $u \in P$ close enough to $\bar{u}$, the problem (1.2) *has the unique solution $v(u)$, and $v(u) \to 0$ as $u \to \bar{u}$.*

Proof. As Φ is differentiable on P near $\bar{u}$, and its derivative is continuous at $\bar{u}$ with respect to P, from (3.1) it also follows that (1.3) holds for all $u \in P$ close enough to $\bar{u}$, and hence, for such u, the subproblem (1.2) has the unique solution $v(u)$. It remains to show that $v(u) \to 0$ as $u \to \bar{u}$. But this readily follows by noting that $v = 0$ is feasible in problem (1.2) for any $u \in P$, and hence, $\|\Phi(u) + \Phi'(u)v(u)\| \leq \|\Phi(u)\|$. Since $\Phi(u) \to 0$ as $u \to \bar{u}$, (3.1) implies that this is only possible when $v(u) \to 0$ as $u \to \bar{u}$. ∎

Combining Theorem 4.1 with Theorem 2.2 and Lemma 4.1 yields the following result.

Corollary 4.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a closed convex set. Assume that Φ is differentiable near a solution $\bar{u}$ of* (1.1)*, and* (4.1) *holds with some $\tau > 0$. Let the regularity condition* (3.1) *be satisfied.*

Then, for every $u^0 \in P$ close enough to $\bar{u}$, there exists the unique sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is the solution of (1.2) *with $u = u^k$, this sequence converges to $\bar{u}$, and the rate of convergence is superlinear with the Q-order $1+\tau$. In particular, if the derivative of Φ is Lipschitz-continuous on P near $\bar{u}$, then the rate of convergence is quadratic.*

In the discussion that follows, we need some more (standard) notation. Given a set $S \subset \mathbb{R}^p$, we denote by $\operatorname{span} S$ the linear subspace spanned by S, by $\operatorname{aff} S$ the affine hull of S, and by $\operatorname{int} S$ the interior of S. For an affine S, let $\operatorname{Lin} P$ stand for a linear subspace parallel to S.

One seemingly natural possibility to relax the regularity assumption (3.1) in Corollary 4.1, while preserving the property that the solutions of subproblems are unique, would be to involve strong metric regularity of $\bar{u}$ as a solution of the VI (2.6). According to [9, Section 3.G], this property means that for any $\omega \in \mathbb{R}^p$ close enough to 0, the perturbed VI (2.7) has near $\bar{u}$ the unique solution $u(\omega)$, and $u(\cdot)$ is Lipschitz-continuous. This evidently implies semistability of $\bar{u}$.

Furthermore, considering now $u \in P$ as a parameter, and setting $\widetilde{u} = u + v$, the VI (2.11) with respect to $v = \widetilde{u} - u$ can be treated as a parametric perturbation of the VI (2.6) with u substituted by $\widetilde{u}$. Assuming that the derivative of Φ is Lipschitz-continuous near $\bar{u}$, and employing again the Mean-Value Theorem, one can see by [9, Theorem 3G.4] (which is a variant of Robinson's implicit function theorem [26]) that for $u \in P$ close enough to $\bar{u}$, (2.11) has near 0 the unique solution $v(u)$, and $v(\cdot)$ is Lipschitz-continuous. In particular, this gives the hemistability of $\bar{u}$, with subproblems uniquely solvable for $u \in P$ near $\bar{u}$.

Observe also that under the stated smoothness assumptions, from [9, Theorem 3G.3] it follows that strong metric regularity of $\bar{u}$ as a solution of (2.6) is equivalent to its strong metric regularity as a solution of the VI

$$(\Phi'(\bar{u}))^\top \Phi'(\bar{u})(u - \bar{u}) + N_P(u) \ni 0 \tag{4.5}$$

with an affine base mapping, while the latter is the same as strong regularity as defined in [26].

However, it turns out that if $\operatorname{int} P \neq \emptyset$, condition (3.1) is necessary for strong regularity of a solution $\bar{u}$ of (4.5). In order to show this, consider the general VI with an affine base mapping

$$a + Au + N_P(u) \ni 0, \tag{4.6}$$

and assume that $a + A\bar{u} = 0$ for a given $\bar{u} \in P$, implying, in particular, that $\bar{u}$ is a solution of (4.6). Observe that this setting covers (4.5) with $a = -(\Phi'(\bar{u}))^\top \Phi'(\bar{u})\bar{u}$, $A = (\Phi'(\bar{u}))^\top \Phi'(\bar{u})$. Then the condition

$$\ker A \cap \operatorname{Lin} \operatorname{aff} P = \{0\} \tag{4.7}$$

is necessary for strong regularity of the solution $\bar{u}$ of (4.6).

Indeed, suppose that there exists $v \in (\ker A \cap \operatorname{Lin} \operatorname{aff} P) \setminus \{0\}$. By the line segment principle [27, Theorem 6.1], one can take $\widetilde{u} \in \operatorname{ri} P$ arbitrarily close to $\bar{u}$, and then $\omega = a + A\widetilde{u}$ can be made arbitrarily close to 0. For a fixed such $\widetilde{u}$, and for any real t close enough to zero, it holds that $\widetilde{u} + tv \in P$, and hence, $N_P(\widetilde{u} + tv)$ is nonempty (contains 0 at least). Therefore,

$$a + A(\widetilde{u} + tv) + N_P(\widetilde{u} + tv) \ni a + A\widetilde{u} = \omega,$$

that is, $\widetilde{u} + tv$ is a solution of the perturbed VI

$$a + Au + N_P(u) \ni \omega.$$

Hence, for ω arbitrarily close to 0, this VI has more than one solution arbitrarily close to $\bar{u}$. Thus, strong regularity of $\bar{u}$ as a solution of (4.6) does not hold.

If $\operatorname{int} P \neq \emptyset$, it holds that $\operatorname{aff} P = \mathbb{R}^p$, and (4.7) takes the form $\ker A = \{0\}$. In our case of interest when $A = (\Phi'(\bar{u}))^\top \Phi'(\bar{u})$, this is equivalent to (3.1).

We complete this discussion by mentioning that by the critical (super)face criterion in [9, Theorem 4H.9] (coming from [10, Theorem 2]), for a polyhedral P, making use of the equality

$$\operatorname{Lin} \operatorname{aff} P = \operatorname{span} T_P(\bar{u}) \quad \forall\, \bar{u} \in P,$$

it can be derived that (4.7) is not only necessary but also sufficient for strong regularity of the solution $\bar{u}$ of (4.6), satisfying $a + A\bar{u} = 0$.

We next demonstrate that in the case of P polyhedral, the weaker than (3.1) condition (2.10) still implies hemistability (as well as semistability, according to Theorem 2.2). We emphasize that (2.10) may hold for any p and q. In this case, solutions of (1.2) need not be unique, even for $u \in P$ near $\bar{u}$. But as will be demonstrated below, there is still no need to refer to the minimal-norm solutions: taking any solution results in local superlinear convergence.

Lemma 4.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a polyhedral set. Assume that Φ is differentiable near a solution $\bar{u}$ of* (1.1)*, and its derivative is continuous on P near $\bar{u}$. Let condition* (2.10) *be satisfied.*

Then $\bar{u}$ is hemistable. More precisely, the solution set $S(u)$ of the problem (1.2) *is nonempty for all $u \in P$, and it holds that*

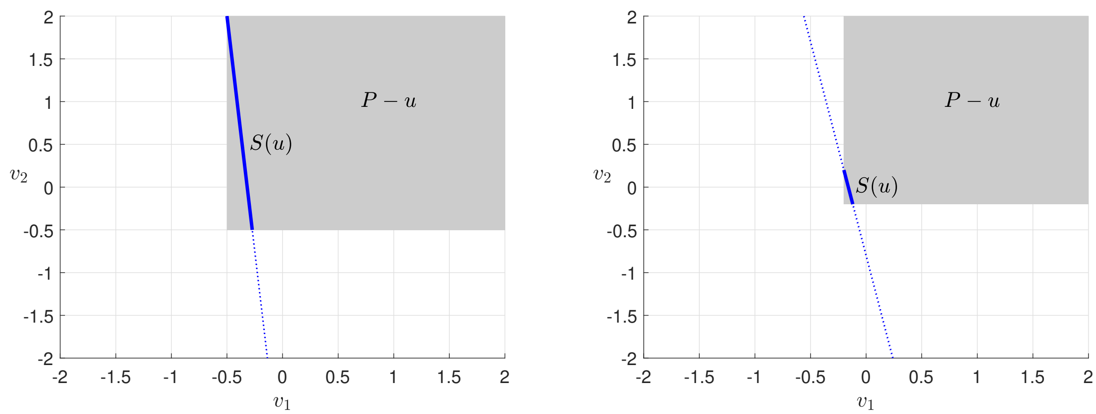
$$\sup_{v \in S(u)} \|v\| \to 0 \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{4.8}$$

Proof. In the case of polyhedral P, the set $S(u)$ is nonempty for all $u \in P$ by the Frank–Wolfe Theorem. Moreover, the condition (2.10) evidently implies that $S(\bar{u}) = \{0\}$, and (4.8) will follow, e.g., by applying [6, Proposition 4.4] to the problem (1.2) with $u \in P$ regarded as a parameter, once the assumptions therein are verified. To that end, define the function $f : P \times \mathbb{R}^p \to \mathbb{R}_+$, $f(u, v) = \|\Phi(u) + \Phi'(u)v\|$, and the multifunction $F : P \to 2^{\mathbb{R}^p}$, $F(u) = P - u$. Under the stated smoothness assumptions, f is continuous on $(P \cap O) \times \mathbb{R}^p$, where O is some neighborhood of $\bar{u}$, yielding assumption (i) in [6, Proposition 4.4]. Further, from closedness of P it follows that F is closed (i.e., its graph is closed), yielding assumption (ii) in [6, Proposition 4.4]. Furthermore, $F(u) \ni 0$ for all $u \in P$, which verifies assumption (iv). The remaining assumption (iii) in [6, Proposition 4.4] is the inf-compactness condition that in the current setting reduces to saying that there exists $\alpha > 0$ such that the level sets

$$L_\alpha(u) = \{v \in P - u \mid \|\Phi(u) + \Phi'(u)v\| \leq \alpha\}$$

are nonempty and uniformly bounded for all $u \in P$ near $\bar{u}$. Fix any $\alpha > 0$. Then, by the continuity of Φ at $\bar{u}$, for any $u \in P$ close enough to $\bar{u}$, the set $L_\alpha(u)$ contains 0, and in particular, it is nonempty. Suppose there exist sequences $\{u^k\} \subset P$ and $\{v^k\} \subset \mathbb{R}^p \setminus \{0\}$ such that $u^k \in L_\alpha(u^k)$ (i.e., $u^k + v^k \in P$ and $\|\Phi(u^k) + \Phi'(u^k)v^k\| \leq \alpha$) for all k, and $\|v^k\| \to \infty$ as $k \to \infty$. Without loss of generality, we may assume that $\{v^k/\|v^k\|\}$ converges to some $v \in \mathbb{R}^p \setminus \{0\}$. For all k, by the convexity of P, it holds that

$$u^k + t\frac{v^k}{\|v^k\|} \in P \quad \forall\, t \in [0,\, \|v^k\|],$$



(a) $u = (0.5,\, 0.5)$ (b) $u = (0.2,\, 0.2)$

Figure 1: Example 4.1.

and

$$\left\| \frac{\Phi(u^k)}{\|v^k\|} + \Phi'(u^k)\frac{v^k}{\|v^k\|} \right\| \le \frac{\alpha}{\|v^k\|}.$$

Passing in the last two relations onto the limit as $k \to \infty$, since P is closed, we obtain that

$$\bar{u} + tv \in P \quad \forall\, t \ge 0, \quad \Phi'(\bar{u})v = 0.$$

In particular, $0 \ne v \in \ker \Phi'(\bar{u}) \cap T_P(\bar{u})$, in violation of (2.10). ∎

From Theorem 2.2 and Lemma 4.2, by evident modifications of the proof of Theorem 4.1 consisting of making use of an arbitrary solution of the problem (1.2) instead of the minimal-norm solution, we obtain the following result.

Corollary 4.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a polyhedral set. Assume that Φ is differentiable near a solution $\bar{u}$ of* (1.1)*, and* (4.1) *holds with some $\tau > 0$. Let the condition* (2.10) *be satisfied.*

Then, for every $u^0 \in P$ close enough to $\bar{u}$, there exists a sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is a solution of (1.2) *with $u = u^k$, and any such sequence converges to $\bar{u}$, and the rate of convergence is superlinear with the Q-order $1+\tau$. In particular, if the derivative of Φ is Lipschitz-continuous on P near $\bar{u}$, then the rate of convergence is quadratic.*

The next example demonstrates that under (2.10), the solution set $S(u)$ of the subproblem (1.2) may be not a singleton for $u \in P$ arbitrarily close to $\bar{u}$, but (4.8) holds, and any choice of $v \in S(u)$ results in superlinear convergence.

Example 4.1 Let $p = 2$, $q = 1$, $P = \mathbb{R}^2_+$, $\Phi(u) = u_1 + u_2 + \chi u_1^2$, where $\chi \ge 0$ is a scalar parameter. Then $U = \{\bar{u}\}$, with $\bar{u} = 0$, and $\ker \Phi'(\bar{u}) = \{v \in \mathbb{R}^2 \mid v_1 + v_2 = 0\}$, while $T_P(\bar{u}) = P = \mathbb{R}^2$, and condition (2.10) is satisfied.

One can directly verify that for every $u \in P$, the set of unconstrained minimizers of the objective function of the problem (1.2) is the straight line spanned by the line segment connecting the points $(-u_1, -u_2+\chi u_1^2)$ and $(-(u_1+\chi u_1^2)/(1+2\chi u_1), -u_2)$ on the boundary of the feasible set $P - u = \{v \in \mathbb{R}^2 \mid v_1 \geq -u_1, v_2 \geq -u_2\}$ of that problem. In particular, the solution set $S(u)$ of (1.2) is precisely this line segment, and it shrinks to 0 as $u \to \bar{u}$, but for $u \neq 0$, it is not a singleton. It can be easily seen that taking any $v \in S(u)$ yields the next iterate satisfying $u + v = O(\|u - \bar{u}\|^2)$ as $u \in P$ tends to $\bar{u}$.

Figure 1 shows the feasible set of the problem (1.2), the straight line corresponding to the set of unconstrained minimizers of the objective function of that problem, and the set $S(u)$, for $\chi = 10$, and for two points $u \in P$: for $u = (0.5, 0.5)$ in Figure 1a, and for $u = (0.2, 0.2)$ in Figure 1b.

We next characterize the level of inexactness that can be allowed when solving the subproblems (1.2), for the local convergence and rate of convergence properties of the Gauss–Newton method established in Theorem 4.1 to be preserved. Let the process of solving the subproblem (1.2) be terminated once (2.11) is satisfied approximately, in the following sense:

$$(\Phi'(u))^\top(\Phi(u) + \Phi'(u)v) + N_P(u+v) \ni r, \tag{4.9}$$

where $r \in \mathbb{R}^p$ is smaller by norm than some given tolerance. As is natural in inexact Newton-type methods, inexactness must be related to the residual of the equation in (1.1); in our case, as follows:

$$r = O(\|\Phi(u)\|^{1+\tau}) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{4.10}$$

Observe that (4.9) is a necessary and sufficient optimality condition for the following (tilt) perturbation of the subproblem (1.2):

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 - \langle r, v\rangle \quad \text{subject to} \quad u + v \in P. \tag{4.11}$$

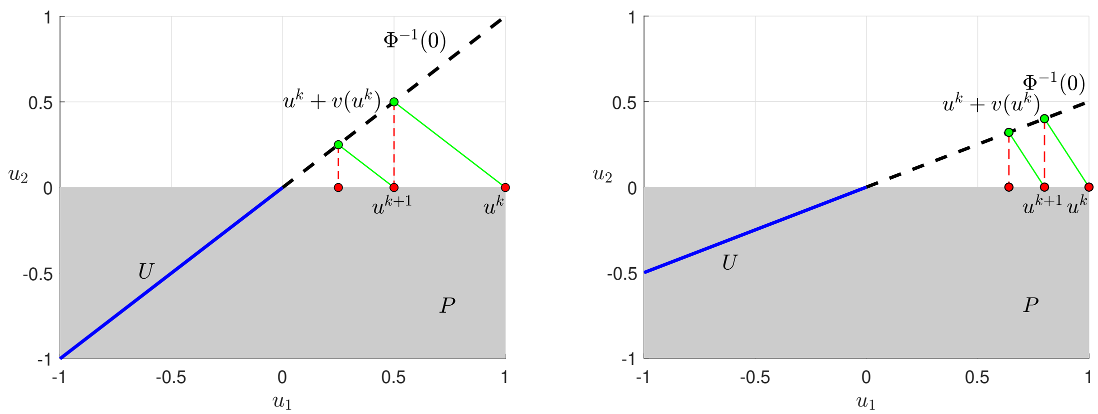
One can now repeat almost literally the proof of Theorem 4.1, with the following modification: ω defined in (4.2) should be substituted by $\omega + r$. Then, employing (4.10), for this new ω we have the estimate

$$\omega = O(\|v(u)\|^\tau\|\Phi(u)\|) + O(\|v(u)\|^{1+\tau}) + O(\|r\|) \quad \text{as } u \in P \text{ tends to } \bar{u}.$$

This yields the following result.

Theorem 4.2 *Under the assumptions of Theorem 4.1, let the function $\psi : P \to \mathbb{R}_+$ satisfy $\psi(u) = O(\|\Phi(u)\|^{1+\tau})$ as $u \in P$ tends to $\bar{u}$.*

Then, for every $u^0 \in P$ close enough to $\bar{u}$, there exists a sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is the minimal-norm solution of the problem (4.11) *with $u = u^k$, with some $r \in \mathbb{R}^p$ satisfying $\|r\| \leq \psi(u^k)$, and any such sequence converges to $\bar{u}$, and the rate of convergence is superlinear with the Q-order $1 + \tau$. In particular, if the derivative of Φ is Lipschitz-continuous on P near $\bar{u}$, then the rate of convergence is quadratic.*



(a) $\chi = 1$ (b) $\chi = 0.5$

Figure 2: The iterates of the projected Gauss–Newton method in Example 5.1.

The corresponding extensions of Corollaries 4.1 and 4.2 to the case of inexact solutions of subproblems are straightforward.

To end this section, we mention that under appropriate assumptions, it must be possible to extend its material to the piecewise smooth case, along the lines of the analysis in [19] for the Levenberg–Marquardt method.

5 Projected Gauss–Newton method

Local convergence analysis of the constrained Gauss–Newton method in Section 4 concerns the case when the solution $\bar{u}$ in question is isolated. In this section, we deal with local convergence of the projected Gauss–Newton method, and our assumptions allow for nonisolated solutions. Specifically, we shall make use of the regularity condition (3.6) and the constrained error bound

$$\operatorname{dist}(u, U) = O(\|\Phi(u)\|) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{5.1}$$

For the current iterate $u \in P$, let now $v(u)$ be the solution of (3.2) (if it exists, i.e., if this subproblem is feasible), and let the next iterate be defined as $\pi_P(u + v(u))$.

The example below demonstrates that under the assumptions (3.6) and (5.1), and for very simple constraints (P in the example is a half-space), the projected Gauss–Newton method may not converge superlinearly (but only linearly). This contradicts the claims of superlinear convergence in [22] and in [1, Theorem 4.1].

Example 5.1 Let $p = 2$, $q = 1$, $P = \{u \in \mathbb{R}^2 \mid u_2 \leq 0\}$, $\Phi(u) = \chi u_1 - u_2$, where χ is a scalar parameter. Then $U = \{(t, \chi t) \mid \chi t \leq 0\}$, and the regularity condition (3.6) holds at every $\bar{u} \in U$. Moreover, the constrained error bound (5.1) holds as well.

For every $u \in \mathbb{R}^2$, the unconstrained Gauss–Newton displacement is correctly defined by

(3.5):

$$v(u) = \frac{\chi u_1 - u_2}{1+\chi^2}(-\chi,\, 1),$$

and hence,

$$u + v(u) = \frac{1}{1+\chi^2}(u_1 + \chi u_2)(1,\, \chi).$$

If $\chi(u_1 + \chi u_2) \geq 0$, then

$$\pi_P(u + v(u)) = \frac{1}{1+\chi^2}(u_1 + \chi u_2,\, 0).$$

Therefore, for the subsequent iterates of the projected Gauss–Newton method, it holds that $u_2 = 0$, and

$$\pi_P(u + v(u)) = \frac{1}{1+\chi^2}u.$$

If $\chi \neq 0$, this yields linear convergence to $\bar{u} = 0$, with the common ratio $\rho = 1/(1+\chi^2)$. This behavior is illustrated in Figure 2 for two values of χ.

Observe that $\rho \to 1$ as $\chi \to 0$, thus making the linear convergence arbitrarily slow. At the same time, $\rho \to 0$ as $\chi \to \infty$, thus making the linear convergence arbitrarily fast. This dependence of ρ on χ agrees with the theory developed below, because the constant in $O(\|\Phi(u)\|)$ in the right-hand side of the constrained error bound (5.1), employed in this analysis, is no less than $1/|\chi|$.

We mention, in the passing, that the same example also demonstrates linear convergence of the projected Levenberg–Marquardt method [14], in which the subproblem (3.2) is replaced by its regularized version

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{2}\sigma(u)\|v\|^2, \quad u \in \mathbb{R}^p,$$

where $\sigma(u) \geq 0$ is a regularization parameter. Indeed, at $u \in \mathbb{R}^2$ such that $\sigma(u) > 0$, the displacement defined by this subproblem is

$$v(u) = \frac{\chi u_1 - u_2}{1+\chi^2+\sigma(u)}(-\chi,\, 1),$$

and hence,

$$u + v(u) = \frac{1}{1+\chi^2+\sigma(u)}((1+\sigma(u))u_1 + \chi u_2,\, \chi u_1 + (\chi^2+\sigma(u))u_2).$$

If $\chi u_1 + (\chi^2 + \sigma(u))u_2 \geq 0$, then

$$\pi_P(u + v(u)) = \frac{1}{1+\chi^2+\sigma(u)}((1+\sigma(u))u_1 + \chi u_2,\, 0).$$

Therefore, for the subsequent iterates of the projected Levenberg–Marquardt method, it holds that $u_2 = 0$, and then

$$\pi_P(u + v(u)) = \frac{1+\sigma(u)}{1+\chi^2+\sigma(u)}u,$$

yielding linear convergence to $\bar{u} = 0$ if $\chi \neq 0$, with the asymptotic common ratio $\rho = 1/(1+\chi^2)$, assuming only that $\sigma(u) > 0$ for $u \in P \setminus U$, and $\sigma(u) \to 0$ as $u \to \bar{u}$. This agrees with the local convergence results in [3, 4].

Having settled the issue of the absence of superlinear convergence, we proceed with establishing the linear rate.

Lemma 5.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a closed convex set. Assume that Φ is differentiable near $\bar{u} \in U$, and its derivative is continuous at $\bar{u}$ with respect to P. Let the regularity condition* (3.6) *and the constrained error bound* (5.1) *be satisfied.*

Then there exist $\delta > 0$ and $\rho \in (0, 1)$ such that

$$\operatorname{dist}(\pi_P(u + v(u), U) \leq \rho \operatorname{dist}(u, U) \quad \forall\, u \in P \cap B(\bar{u}, \delta), \tag{5.2}$$

where $v(u)$ is well-defined by (3.5).

Proof. Under the stated smoothness assumptions, from (3.6) it follows that if $\delta > 0$ is taken small enough, then $v(u)$ is well-defined by (3.5) for all $u \in P \cap B(\bar{u}, \delta)$, and there exist $\gamma > 0$ and $\Gamma > 0$ such that

$$\gamma\|\Phi(u)\| \leq \|v(u)\| \leq \Gamma\|\Phi(u)\| \quad \forall\, u \in P \cap B(\bar{u}, \delta).$$

Employing now (5.1), reducing $\delta > 0$ and $\gamma > 0$ if necessary, and enlarging $\Gamma > 0$ if necessary, we obtain that

$$\gamma \operatorname{dist}(u, U) \leq \|v(u)\| \leq \Gamma \operatorname{dist}(u, U) \quad \forall\, u \in P \cap B(\bar{u}, \delta). \tag{5.3}$$

Let $\widehat{u} \in U$ be some projection of $u \in \mathbb{R}^p$ on U, that is,

$$\|u - \widehat{u}\| = \operatorname{dist}(u, U). \tag{5.4}$$

Since $\pi_P(\cdot)$ is nonexpansive, and since $\widehat{u} \in P$, for $u \in P \cap B(\bar{u}, \delta)$ we have that

$$\|\pi_P(u + v(u)) - \widehat{u}\| = \|\pi_P(u + v(u)) - \pi_P(\widehat{u})\| \leq \|u + v(u) - \widehat{u}\|.$$

To establish the needed property (5.2), it is sufficient to show that there exist $\rho \in (0, 1)$ such that

$$\|u + v(u) - \widehat{u}\| \leq \rho\|u - \widehat{u}\| \quad \forall\, u \in P \cap B(\bar{u}, \delta), \tag{5.5}$$

provided $\delta > 0$ is taken small enough. To do this, some extra preparations are needed.

Set

$$\bar{v}(u) = -(\Phi'(\bar{u}))^\top(\Phi'(u)(\Phi'(u))^\top)^{-1}\Phi(u) \in \operatorname{im}(\Phi'(\bar{u}))^\top = (\ker \Phi'(u))^\perp.$$

According to (3.5), it holds that

$$v(u) = \bar{v}(u) + O(\|\Phi'(u) - \Phi'(\bar{u})\|\|\Phi(u)\|) = \bar{v}(u) + o(\|u - \widehat{u}\|) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{5.6}$$

On the other hand, employing again the Mean-Value Theorem, we obtain that

$$\begin{aligned}\|\Phi(u+v(u))-\Phi'(\bar{u})(u+v(u)-\widehat{u})\| &= \|\Phi(u+v(u))-\Phi(\widehat{u})-\Phi'(\bar{u})(u+v(u)-\widehat{u})\| \\ &= o(\|u+v(u)-\widehat{u}\|) \\ &= o(\|u-\widehat{u}\|) \quad \text{as } u\in P \text{ tends to } \bar{u}, \qquad (5.7)\end{aligned}$$

where the last estimate is by the second inequality in (5.3) and by (5.4). Moreover, as in (3.9), under the current smoothness assumptions we obtain that

$$\Phi(u+v(u)) = o(\|v(u)\|) = o(\|u-\widehat{u}\|) \quad \text{as } u\in P \text{ tends to } \bar{u},$$

where the last estimate is again by the second inequality in (5.3) and by (5.4). Combining this with (5.7), we conclude that

$$\Phi'(\bar{u})(u+v(u)-\widehat{u}) = o(\|u-\widehat{u}\|) \quad \text{as } u\in P \text{ tends to } \bar{u}.$$

Therefore, since there evidently exists $c>0$ such that

$$\operatorname{dist}(v,\, \ker\Phi'(\bar{u})) \le c\|\Phi'(\bar{u})v\| \quad \forall\, v\in\mathbb{R}^p,$$

there exists $\widetilde{v}(u)\in\ker\Phi'(\bar{u})$ such that

$$u+v(u)-\widehat{u} = \widetilde{v}(u)+o(\|u-\widehat{u}\|) \quad \text{as } u\in P \text{ tends to } \bar{u}. \qquad (5.8)$$

Observe that from (5.6) and (5.8), and again from the second inequality in (5.3), and from (5.4), it follows that

$$\bar{v}(u) = O(\|u-\widehat{u}\|), \quad \widetilde{v}(u) = O(\|u-\widehat{u}\|) \quad \text{as } u\in P \text{ tends to } \bar{u}. \qquad (5.9)$$

Since thus defined $\bar{v}(u)$ and $\widetilde{v}(u)$ are orthogonal, by (5.6), (5.8), (5.9), and by (5.4), we obtain that

$$\begin{aligned}\langle v(u),\, u+v(u)-\widehat{u}\rangle &= \langle \bar{v}(u)+o(\|u-\widehat{u}\|),\, \widetilde{v}(u)+o(\|u-\widehat{u}\|)\rangle \\ &= o(\|u-\widehat{u}\|^2) \quad \text{as } u\in P \text{ tends to } \bar{u}.\end{aligned}$$

Therefore,

$$\begin{aligned}\|u-\widehat{u}\|^2 &= \|v(u)-(u+v(u)-\widehat{u})\|^2 \\ &= \|v(u)\|^2+\|u+v(u)-\widehat{u}\|^2-2\langle v(u),\, u+v(u)-\widehat{u}\rangle \\ &= \|v(u)\|^2+\|u+v(u)-\widehat{u}\|^2+o(\|u-\widehat{u}\|^2) \quad \text{as } u\in P \text{ tends to } \bar{u}. \quad (5.10)\end{aligned}$$

We now get back to establishing (5.5). If $u\in U$, then $v(u)=0$ and $\widehat{u}=u$, and (5.5) evidently holds with any ρ. Therefore, it remains to establish (5.5) for $u\not\in U$. We argue by contradiction: suppose there exists a sequence $\{u^k\}\subset P\setminus U$ such that it converges to $\bar{u}$, and

$$\lim_{k\to\infty}\frac{\|u^k+v(u^k)-\widehat{u}^k\|}{\|u^k-\widehat{u}^k\|} \ge 1. \qquad (5.11)$$

Substituting u in (5.10) by u^k, and dividing both sides by $\|u^k - \widehat{u}^k\|^2$, we obtain that

$$\frac{\|v(u^k)\|^2}{\|u^k - \widehat{u}^k\|^2} + \frac{\|u^k + v(u^k) - \widehat{u}^k\|^2}{\|u^k - \widehat{u}^k\|^2} - 1 \to 0 \quad \text{as } k \to \infty.$$

By the first inequality in (5.3) and by (5.4), the first term in the left-hand side of the last relation above is staying no smaller than $\gamma > 0$, while according to (5.11), the limit of the second term as $k \to \infty$ is greater or equal to 1. Therefore, the limit of the left-hand side above is greater or equal to γ, yielding a contradiction. ■

The proof of the next theorem is along the lines of that in [3, Theorem 1], which deals with the projected Levenberg–Marquardt method. Nevertheless, we need to give the proof in full, to work out all the details specifically for the algorithm considered here, i.e., the projected Gauss–Newton method. In particular, our proof employs Lemma 5.1 instead of [3, Lemma 4].

Theorem 5.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a closed convex set. Assume that Φ is differentiable near $\bar{u} \in U$, and its derivative is continuous at $\bar{u}$ with respect to P. Let the regularity condition* (3.6) *and the constrained error bound* (5.1) *be satisfied.*

Then, for any $\delta > 0$, and every $u^0 \in P$ close enough to $\bar{u}$, there exists the unique sequence $\{u^k\}$ such that for every k, it holds that $u^{k+1} = \pi_P(u^k + v(u^k))$, where $v(u^k)$ is the solution of (3.2) *with $u = u^k$, this sequence is contained $B(\bar{u}, \delta)$ and converges to some $u^* \in U$, and the rate of convergence is R-linear.*

Proof. Let $\delta > 0$ and $\rho \in (0, 1)$ be chosen according to Lemma 5.1. Observe that this allows to take $\delta > 0$ arbitrarily small, and if the assertion of the theorem is true for some $\delta > 0$, it is evidently true for any larger δ. Moreover, if $\delta > 0$ is taken small enough, under the stated smoothness assumption, from (3.5) and (3.6) it follows that there exists $C > 0$ such that

$$\|v(u)\| = O(\|\Phi(u)\|) \leq C \operatorname{dist}(u, \Phi^{-1}(0)) \leq C \operatorname{dist}(u, U) \quad \forall\, u \in P \cap B(\bar{u}, \delta). \tag{5.12}$$

Fix any $\varepsilon > 0$ satisfying

$$\varepsilon \leq \left(1 + \frac{C}{1 - \rho}\right)^{-1} \delta. \tag{5.13}$$

We first prove by induction that if $u^0 \in B(\bar{u}, \varepsilon)$, then the algorithm specified in the statement of the theorem generates a well-defined unique sequence $\{u^k\} \subset B(\bar{u}, \delta)$.

Suppose that the iterates $u^1, u^2, \ldots, u^k \in B(\bar{u}, \delta)$ are already generated. Then, by the choice of δ, we have that $u^{k+1} = \pi_P(u^k + v(u^k))$ is also well-defined, and

$$\|u^{k+1} - \bar{u}\| \leq \|u^0 - \bar{u}\| + \sum_{i=0}^{k} \|u^{i+1} - u^i\|. \tag{5.14}$$

For every $i \in \{0, 1, \ldots, k\}$, since $u^i \in P$, we have that

$$\|u^{i+1} - u^i\| = \|\pi_P(u^i + v(u^i)) - u^i\| = \|\pi_P(u^i + v(u^i)) - \pi_P(u^i)\| \leq \|v(u^i)\| \leq C \operatorname{dist}(u^i, U), \tag{5.15}$$

where the last estimate is by (5.12). Combined with (5.14), and employing (5.2), this yields

$$\begin{aligned} \|u^{k+1} - \bar{u}\| &\leq \|u^0 - \bar{u}\| + C \sum_{i=0}^{k} \operatorname{dist}(u^i, U) \\ &= \|u^0 - \bar{u}\| + C \operatorname{dist}(u^0, U) + C \sum_{i=1}^{k} \operatorname{dist}(\pi_P(u^{i-1} + v(u^{i-1})), U) \\ &\leq \|u^0 - \bar{u}\| + C \operatorname{dist}(u^0, U) \sum_{i=0}^{k} \rho^i \\ &\leq \left(1 + \frac{C}{1-\rho}\right) \|u^0 - \bar{u}\|. \end{aligned} \tag{5.16}$$

Since $u^0 \in B(\bar{u}, \varepsilon)$, employing (5.13) we now obtain that $u^{k+1} \in B(\bar{u}, \delta)$, thus proving the claim that the unique sequence $\{u^k\} \subset B(\bar{u}, \delta)$ is well-defined.

In particular, (5.15) holds for all i, and similarly to (5.16) we derive that for any k and l

$$\begin{aligned} \|u^{k+l} - u^k\| &\leq \sum_{i=k}^{k+l-1} \|u^{i+1} - u^i\| \\ &\leq C \sum_{i=k}^{k+l-1} \operatorname{dist}(u^i, U) \\ &\leq C\rho^k \operatorname{dist}(u^0, U) \sum_{i=0}^{l-1} \rho^i \\ &\leq \frac{C}{1-\rho} \|u^0 - \bar{u}\| \rho^k \\ &\leq \frac{C}{1-\rho} \varepsilon \rho^k. \end{aligned} \tag{5.17}$$

As the right-hand side in (5.17) tends to 0 as $k \to \infty$, this implies that $\{u^k\}$ is a Cauchy sequence, and hence, it converges to some $u^* \in B(\bar{u}, \delta)$.

Moreover, the choice of δ ensuring (5.2) in Lemma 5.1 implies the estimate $\operatorname{dist}(u^{k+1}, U) \leq \rho \operatorname{dist}(u^k, U)$ for all k, yielding $\operatorname{dist}(u^k, U) \to 0$ as $k \to \infty$, and hence, $u^* \in U$.

Finally, passing onto the limit in (5.17) as $l \to \infty$ yields the estimate

$$\|u^k - u^*\| \leq \frac{C}{1-\rho} \varepsilon \rho^k$$

for all k. This completes the proof of the R-linear convergence rate. ∎

As demonstrated by [3, Examples 1, 2], assumptions (3.6) and (5.1) are independent, i.e., one does not imply the other. In [4], it was also established that for the projected Levenberg–Marquardt method (with exact projections!), assumption (3.6) can be avoided, i.e., only the constrained error bound (5.1) is needed, apparently unlike the case of inexact projections as in [3], where the unconstrained error bound

$$\operatorname{dist}(u,\, \Phi^{-1}(0)) = O(\|\Phi(u)\|) \quad \text{as } u \to \bar{u},$$

implied by (3.6), is also needed.

Remark 5.1 Local superlinear convergence of the projected Gauss–Newton method can be established under the regularity condition (3.6) complemented by the very restrictive (for the constrained case) error bound of the form

$$\operatorname{dist}(u,\, U) = O(\|\Phi(u)\|) \quad \text{as } u \to \bar{u}, \tag{5.18}$$

which is much stronger than (5.1). In particular, (5.18) implies that $\Phi^{-1}(0) \subset U$, the property that does not hold in Example 5.1.

Indeed, set $\Psi(u) = \pi_P(u + v(u))$, where $v(u)$ is the solution of (3.2) (given by (3.5) if $\operatorname{rank} \Phi'(u) = q$). Under the smoothness assumptions as in Theorem 4.1,

$$\|\Psi(u) - u\| = \|\pi_P(u + v(u)) - \pi_P(u)\| \leq \|v(u)\| = O(\varphi(u)) \quad \text{as } u \in P \text{ tends to } \bar{u}, \tag{5.19}$$

where the last estimate is by (3.8). Hence, (2.4) is satisfied. Furthermore, let $\widehat{u}$ now stand for any projection of $u + v(u)$ onto U. Then

$$\begin{aligned} \|\pi_P(u + v(u)) - \widehat{u}\| &= \|\pi_P(u + v(u)) - \pi_P(\widehat{u})\| \\ &\leq \|u + v(u) - \widehat{u}\| \\ &= \operatorname{dist}(u + v(u),\, U) \\ &= O(\|\Phi(u + v(u))\|) \\ &= O((\varphi(u))^{1+\tau}) \quad \text{as } u \in P \text{ tends to } \bar{u}, \end{aligned}$$

where the next-to-last estimate is by (5.18), while the last one is by (3.9). Therefore,

$$\begin{aligned} \varphi(\Psi(u)) &= \|\Phi(\pi_P(u + v(u))) - \Phi(\widehat{u})\| \\ &= O(\|\pi_P(u + v(u)) - \widehat{u}\|) \\ &= O((\varphi(u))^{1+\tau}) \quad \text{as } u \in P \text{ tends to } \bar{u}, \end{aligned}$$

yielding (2.5) with $\theta = 1 + \tau$. Local superlinear convergence with Q-order $1 + \tau$ now follows from Theorem 2.1.

Remark 5.2 Local superlinear convergence of the projected Gauss–Newton method is also guaranteed when $p \leq q$ and (3.1) holds. In this case, assuming that the derivative of Φ is continuous at $\bar{u}$ with respect to P, for $u \in P$ close enough to $\bar{u}$, it holds that $v(u)$ is the unique solution of the problem (3.11). Setting $\Psi(u) = \pi_P(u + v(u))$, the needed result can be again derived from Theorem 2.1. However, it also readily follows from the estimate

$$\|\pi_P(u + v(u)) - \bar{u}\| = \|\pi_P(u + v(u)) - \pi_P(\bar{u})\| \leq \|u + v(u) - \bar{u}\|$$

and from the above-mentioned classical results for square or overdetermined unconstrained systems, contained in Corollary 4.1 applied with $P = \mathbb{R}^p$. Under the smoothness assumptions as in Theorem 4.1, this yields local superlinear convergence with Q-order $1 + \tau$.

Declarations

Conflicts of Interest. The authors declare that they have no conflict of interest of any kind related to the manuscript.

Data Availability Statement. Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

[1] R. Behling. The Method and the Trajectory of Levenberg–Marquardt. PhD thesis. IMPA - Instituto Nacional de Matemática Pura e Aplicada. Rio de Janeiro, 2011.

[2] R. Behling and A. Fischer. A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim. Lett. 6 (2012), 927–940.

[3] R. Behling, A. Fischer, M. Herrich, A. Iusem, and Y. Ye. A Levenberg-Marquardt method with approximate projections. Comput. Optim. Appl. 59 (2014), 2–26.

[4] R. Behling, A. Fischer, G. Haeser, A. Ramos, and K. Schonefeld. On the constrained error bound condition and the projected Levenberg-Marquardt method. Optimization 66 (2017), 1397–1411.

[5] J.F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. 29 (1994), 161–186.

[6] J.F. Bonnans and A. Shapiro. Perturbation Analysis of Optimization Problems. Springer–Verlag, New York, 2000.

[7] E. Blum and W. Oettli. Direct proof of the existence theorem for quadratic programming. Oper. Res. 20 (1972), 165–167.

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

[9] A.L. Dontchev and R.T. Rockafellar. Implicit Functions and Solution Mappings. Second edition. Springer, New York, 2014.

[10] A.L. Dontchev and R.T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. 6 (1996), 1087–1105.

[11] B.C. Eaves. On quadratic programming. Management Sci. 11 (1971), 689–711.

[12] A. Fischer, M. Herrich, A.F. Izmailov, and M.V. Solodov. Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. Comput. Optim. Appl. 63 (2016), 425–459.

[13] A. Fischer, A.F. Izmailov, M. Punke, and Z. Yan. Error bounds and Newton-type methods for reformulations of Karush-Kuhn-Tucker systems of mathematical programs with complementarity constraints. Comput. Optim. Appl. 2025. https://doi.org/10.1007/s10589-025-00729-1.

[14] A. Fischer, A.F. Izmailov, and M.V. Solodov. The Levenberg–Marquardt method: An overview of modern convergence theories and more. Computational Optimization and Applications 89 (2024), 33–67.

[15] A. Fischer and N. Strasdat. An extended convergence framework applied to complementarity systems with degenerate and nonisolated solutions. Pure Appl. Func. Analys. 8 (2023), 1039–1054.

[16] M. Frank and P. Wolfe. An algorithm for quadratic programming. Naval Research Logistics Quarterly 3 (1956), 95–110.

[17] A.F. Izmailov and M.V. Solodov. Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering, Cham, 2014.

[18] A.F. Izmailov and M.V. Solodov. A general perturbed Newtonian framework and critical solutions of nonlinear equations. Set-Valued Var. Anal. 33 (2025), 3.

[19] A.F. Izmailov, E.I. Uskov, and Z. Yan. Newton method vs. semismooth Newton method for singular solutions of nonlinear complementarity problems. Adv. Syst. Sci. Appl. 23 (2023), 16–26.

[20] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J. Comput. Appl. Math. 172 (2004), 375–397.

[21] C.T. Kelley. Iterative Methods for Optimization. SIAM, Philadelphia, 1999.

[22] M. Macconi, B. Morini, and M. Porcelli. A Gauss-Newton method for solving bound-constrained underdetermined nonlinear systems. Optim. Meth. Software 24 (2009), 219–235.

[23] G.M. Lee, N.N. Tam, and N.D. Yen. Quadratic Programming and Affine Variational Inequalities. Springer, New York, 2005.

[24] O.L. Mangasarian. A simple characterization of solution sets of convex programs. Oper. Res. Lett. 7 (1988), 21–26.

[25] J.M. Ortega and W.C. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables. Academiv Press, New York, London, 1970.

[26] S.M. Robinson. Strongly regular generalized equations. Math. Oper. Res. 5 (1980), 43–62.

[27] R.T. Rockafellar. Convex Analysis. Princeton University Press, New Jersey, 1970.