PH.D. DISSERTATION

INSTITUTE OF APPLIED MATHEMATICS
CHINESE ACADEMY OF SCIENCES

Algorithms and Convergence Analysis for Nonsmooth Optimization and Nonsmooth Equations

Defeng Sun

Institute of Applied Mathematics
Academia Sinica, P.O. Box 2734
Beijing 100080, P.R. China

ADVISOR: Prof. Jiye Han

December 1994

Contents

Acknowledgements

I would like to express my sincere appreciation to Professor Jiye Han, my advisor, for his earnest guidance and encouragement. In fact, the completion of this dissertation is due to his constant help and support.

I am also indebted to Professor Jong-Shi Pang and Professor Liqun Qi for their many helpful suggestions on this subject. I have learned a lot from them during they visited China.

I would like to thank Professor Bingsheng He and Dr. Xiaojun Chen for providing me their latest papers.

Finally, I thank my wife Bairong Zhou for her long-term understanding and support.

摘要

本论文共分6章，主要讨论作者攻读学位期间在非光滑优化和非光滑方程组方面所取得的成果。为了本论文的完整性和协调性，作者在共轭梯度方面和拟牛顿方面所取得成果不再出现在本论文中；这两方面的内容可见附录中作者的相应文章。下面简单介绍一下各章的内容

Ch. 1

引言

本章主要介绍目前国际上关于非光滑优化（包括LC^1优化问题）及非光滑方程组(包括非线性互补问题，变分不等式问题和非线性规划的Karush-Kuhn-Tucker方程）中的一些新兴课题的由来，发展及最新成果。同时对作者在这一领域所获得的成果也作了简单的介绍。

Ch. 2

一类求解非线性投影方程组的迭代方法

本章给出了一类求解投影方程组

$$x - \Pi_X[x - F(x)] = 0 \tag{1}$$

的迭代算法。在映射F伪单调及连续性条件下证明了算法的全局收敛性。而国外的算法，如Korpelevich [5]，Fukushima [1] 及Harker和Pang [2]，是在假设F（强）单调及Lipschitz连续的条件下给出的，并且算法本身强烈依赖并不易获得的Lipschitz常数。另外我们得到了一个关于解集非空的充要条件。此充要条件可用来解决He及Stoer [3] 提出的一个公开问题。本章给出的数值例子充分显示了我们的算法的有效性。

Ch. 3

一类非光滑方程组及其相关问题的牛顿法和拟牛顿法

本章通过引进一种广义 Jacobian 的概念及一种新的逼近思想给出了求解一类非光滑方程组如 (1) 中的 X 取框形约束及其相关问题如由广义互补问题得到的映射

$$H(x) := \min(F(x), G(x)) = 0 \tag{2}$$

的牛顿方法及拟牛顿方法。在广义 Jacobian 非奇异的条件下我们证明了算法的局部超线性收敛性。新的拟牛顿算法每步仅需解一个线性方程组，而以前的文献 [4，9] 中的拟牛顿法每步需解一线性互补问题（这是一非线性问题）。其次我们以非线性互补问题为例，对新的拟牛顿方法讨论了如何从当前迭代矩阵的 QR 分解生成下一步迭代矩阵的 QR 分解。从而对 Harker 和 Pang [2] 提出的一个公开问题给出了一个确定的回答。与本章内容有关的国外的最新报告可见 [8]，那里处理的是非线性规划的 Karush-Kuhn-Tucker 方程。这是本章考虑的投影方程组的一种特殊情况。

Ch. 4

LC^1 优化问题的逼近牛顿法在无严格互补条件下的超线性收敛性

LC^1 优化问题的 SQP 方法和逼近牛顿方法的超线性收敛性在严格互补性条件下最近由 Qi [7] 得到。本章通过提供一种新的处理途径在不假设严格互补性条件下证明了上述方法的超线性收敛性并且指出该途径还可用来证明其它算法的收敛性。

Ch. 5

线性约束下正则映射的牛顿法和拟牛顿法

本章给出了求解正则映射

$$F(\Pi_K(x)) + x - \Pi_K(x) = 0 \tag{3}$$

的牛顿和拟牛顿法，其中 K 是一凸多面体。本章与第三章的出发点一致：每步只需求解一线性方程组并且对拟牛顿法通过校正迭代矩阵

的 QR 分解获得线性方程组的解，但两章内容互不包含，各有特色。而国外文献中的同类方法[4, 9]每步需解一线性变分不等式问题(此为一非线性非凸问题)。我们的超线性(二次)收敛性的证明是在目前已知文献中最弱的条件下建立起来的。在 K 为一般凸多面体的情况下，未见国外有任何类似的结果。

Ch. 6

一类非线性投影方程组的自修正牛顿法

本章首先把第5章的牛顿方法推广到 F 为不连续可微但半光滑的情形，然后结合第2章的一种全局收敛性方法在 F 伪单调的条件下给出了一种全局收敛、局部超线性收敛的自修正牛顿法。当 F 是某一凸函数 f 的梯度即 $F = \nabla f$ 时，Pang 和 Qi [6] 得到了一种全局收敛，局部超线性收敛的牛顿方法。当 F 不为某一凸函数 f 的梯度时，未见类似报道。如果 $F \in C^1$ 及强单调时，Taji, Fukushima 和 Ibaraki [12] 建立了一种全局收敛，局部二次收敛的牛顿方法。他们的方法每步需解一线性变分不等式问题(非线性非凸)，而我们的方法每步仅需解一线性方程组。另外，[12] 中的二次收敛性是建立在较强的严格互补条件下的并且[12] 中的结果也不能推广到 F 不可微的情形。

References

[1] M. Fukushima, "Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems", *Mathematical Programming* 53 (1992) 99-110.

[2] P.T. Harker and J.-S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications", *Mathematical Programming* 48 (1990) 339-357.

[3] B.S. He and J. Stoer, "Solutions of projection problems over polytopes", *Numerische Mathematik* 61 (1992) 73-90.

[4] N.H. Josephy, "Newton's method for generalized equations", Technical Summary Report No. 1965, Mathematical Research Center, University of Wisconsin (Madison, WI, 1979).

[5] G.M. Korpelevich, "Ekstragradientnyi method dlia otyskamia sedlovykh tchek i drugikh zadach", *Ekonomica I Mathematicheski Metody* 12 (1976) 947-956.

[6] J.-S. Pang and L. Qi, "A globally convergent Newton method for convex SC^1 minimization problems", to appear in *Journal of Optimization Theory and Applications.*

[7] L. Qi, "Superlinearly convergent approximate Newton methods for LC^1 optimization problems", *Mathematical Programming* 64 (1994) 277-294.

[8] L. Qi and H. Jiang, "Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations", Applied Mathematics Report 94/5, School of Mathematics, The University of New South Wales, Sydney, Australia (Revised in November 1994).

[9] S.M. Robinson, "Generalized equations", in A. Bachem, M. Gröschel and B. Korte, eds., *Mathematical Programming: The State of the Art* (Springer-Verlag, Berlin, 1983) 346-367.

[10] D. Sun, "A projection and contraction method for the nonlinear complementarity problem and its extensions", *Mathematica Numerica Sinica* 16 (1994) 183-194.

[11] D. Sun, "A new step-size skill for solving a class of nonlinear projection equations", *Journal of Computational Mathematics* 13 (1995) forthcoming.

[12] K. Taji, M. Fukushima, and T. Ibaraki, "A globally convergent Newton method for solving strongly monotone variational inequalities", *Mathematical Programming* 58 (1993) 369-383.

Chapter 1

Introduction

The nonlinear complementarity problem (NCP) is formally defined as follows. Given a mapping $F : D \supseteq \Re^n_+ \to \Re^n$, this problem, denoted $NCP(F)$, is to find a vector $x \in \Re^n_+$ such that

$$x \geq 0, \ F(x) \geq 0, \ x^T F(x) = 0. \tag{1}$$

There are many generalizations of NCP. For example, the general nonlinear complementarity, denoted $NCP(F,G)$, is to find $x \in \Re^n$ such that

$$F(x) \geq 0, \ G(x) \geq 0, \ F(x)^T G(x) = 0, \tag{2}$$

where $F, G : \Re^n \to \Re^n$. For a comprehensive review of complementarity problems, see [12].

The complementarity problem $NCP(F)$ is a special case of variational inequality problem which is defined as follows. Let K be a given subset of $\Re^n$ and $F : D \supseteq K \to \Re^n$. This problem, denoted $VI(K,F)$, is to find a vector $x \in K$ such that

$$(y - x)^T F(x) \geq 0 \quad \forall \ y \in K. \tag{3}$$

Complementarity problems and variational inequality problems arise from a diversity of sources and disciplines, such as mathematical programs, economic equilibrium problems, and engineering applications. For example, consider the standard nonlinear program (NLP):

$$\begin{aligned} &\min \theta(x) \\ &\text{s.t. } g_i(x) \leq 0, \ i = 1, ..., p \\ &\qquad h_j(x) = 0, \ j = 1, ..., q \end{aligned} \tag{4}$$

where the given functions θ, g_i, $h_j : \Re^n \to \Re$ are all continuously differentiable. The Lagrangian function for (4) can be defined as

$$L(x, \lambda, \mu) = \theta(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \mu_j h_j(x). \tag{5}$$

The well-known Karush-Kukn-Tucker (K-K-T) optimality conditions for the above problem is [9]:

$$\begin{cases} \nabla_x L(x, \lambda, \mu) = 0, \\ \lambda \geq 0,\ g(x) \leq 0,\ \lambda^T g(x) = 0, \\ h(x) = 0. \end{cases} \tag{6}$$

We shall call (x, λ, μ) a K-K-T triple of the NLP (4) if it satisfies the above K-K-T conditions; in this case, the corresponding vector x called a K-K-T point. Let $z = (x, \lambda, \mu) \in \Re^n \times \Re^p \times \Re^q$ and define

$$F(z) = \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ -g(x) \\ -h(x) \end{pmatrix} \tag{7}$$

and

$$K = \{z = (x, \lambda, \mu) |\ \lambda_i \geq 0,\ i = 1, ..., p\}.$$

Then $z = (x, \lambda, \mu)$ is a solution of system (6) if and only if z is a solution of $VI(K, F)$. So the K-K-T optimality conditions is a special case of variational inequality problems.

The case of a linearly constrained NLP is particularly of interest. This special case may be expressed as

$$\begin{array}{l} \min\ \theta(x) \\ \text{s.t. } x \in K, \end{array} \tag{8}$$

where K is a polyhedral set in $\Re^n$. It is well known that a vector x is a K-K-T point if and only if it is a stationary point, i.e., if and only if $x \in K$ and satisfies the so-called variational principles:

$$(y - x)^T \nabla\theta(x) \geq 0 \quad \forall\ y \in K.$$

This latter problem is precisely the $VI(K, \nabla\theta)$. For the relations of variational inequality problems with equilibrium problems and engineering applications, see [12]

A general approach for solving the variational inequality $VI(X, F)$ consists of creating a sequence $x^k \subset X$ such that each x^{k+1} solves the problem $VI(K, F^k)$:

$$F^k(x^{k+1})^T(y - x^{k+1}) \geq 0 \quad \forall\ y \in K, \tag{9}$$

where $F^k(x)$ is some approximation to $F(x)$. The two basic choices for this approximation are that F^k is either a linear or nonlinear function. For the linear approximations:

$$F^k(x) = F(x^k) + A(x^k)(x - x^k), \tag{10}$$

where $A(x^k)$ is an $n \times n$ matrix, several methods exist which differ in the choice of $A(x^k)$:

$$\begin{aligned} A(x^k) &= F'(x^k) \text{ (Newton's method)} \\ &\approx F'(x^k) \text{ (Quasi} - \text{Newton methods)} \\ &= E \text{ (Projection method)}, \end{aligned}$$

where E is a fixed, symmetric, positive definite matrix.

Before giving the convergence of Newton method and quasi-Newton methods, we must discuss the notion of a *regular solution* which was introduced by Robinson under the context of generalized equations.

Definition 1 [20]. Let x^* be a solution of the problem $VI(K,F)$. Then x^* is called regular if there exist a neighborhood N of x^* and a scalar $\delta > 0$ such that for every vector y with $||y|| < \delta$, there is a unique vector $x(y) \in N$ that solves the perturbed linearized variational inequality problem $VI(X, F^y)$, where $F^y : \Re^n \to \Re^n$ is defined by

$$F^y(x) = F(x^*) + y + F'(x^*)(y - x^*);$$

moreover, as a function of the perturbed vector y, the solution $x(y)$ is Lipschitz continuous; i.e., there exists a constant $L > 0$ such that whenever $||y|| < \delta$ and $||z|| < \delta$, one has

$$||x(y) - x(z)|| \leq L||y - z||.$$

It is easy to see that when $K = R^n$, the regularity of a solution x^* of the $VI(K,F)$ is equivalent to the nonsingularity of the Jacobian matrix $F'(x^*)$. For the details of regular solution, see Robinson [20].

Theorem 1 [6]. Let K be a nonempty, closed and convex subset of $\Re^n$, $F : \Re^n \to \Re^n$ be once continuously differentiable, and x^* be a regular solution of $VI(K,F)$. Then there exists a neighborhood N of x^* such that whenever the initial vector x^0 is chosen in N, the entire sequence $\{x^k\}$ generated by Newton's method is well-defined and converges to x^*. Furthermore if $F'(x)$ is Lipschitz continuous around x^*, then the convergence is quadratic; i.e., there exists a constant $c > 0$ such that for all k sufficiently large,

$$||x^{k+1} - x^*|| \leq c||x^k - x^*||^2.$$

In [7], Josephy considered such quasi-Newton methods that in the linear approximation scheme the matrix $A(x^k)$ is updated from one iteration to the next by a simple small-rank matrix. These quasi-Newton methods reduce the work to evaluate $F'(x^k)$, but do not ease the computational effort involved in solving the resulting subproblems, which are nonlinear and nonconvex problems in general.

Definition 2 [10]. The mapping $F : \Re^n \to \Re^n$ is said to be

(i) monotone over a set K if

$$[F(x) - F(y)]^T(x - y) \geq 0 \quad \forall\, x, y \in K; \tag{11}$$

(ii) pseudomonotone over K if

$$F(y)^T(x - y) \geq 0 \text{ implies } F(x)^T(x - y) \geq 0 \quad \forall\, x, y \in K; \tag{12}$$

(iii) strongly monotone over K if there exists an $\alpha > 0$ such that

$$[F(x) - F(y)]^T(x - y) \geq \alpha||x - y||^2 \quad \forall\, x, y \in K.$$

In Pang and Chan [13], the convergence of the Projection method is presented.

Theorem 2 [13]. Let K be a nonempty, closed and convex subset of $\Re^n$ and let $F : \Re^n \to \Re^n$ be given. Suppose that F is Lipschitz continuous and strongly monotone with positive constants β and γ respectively; i.e., for all vectors $x, y \in K$,

$$||F(x) - F(y)|| \leq \beta||x - y||,$$

$$[F(x) - F(y)]^T(x - y) \geq \gamma||x - y||^2.$$

Let E be a symmetric positive definite matrix with smallest and largest eigenvalues given by κ^{-1}, η respectively. If $\kappa^2\beta^2 < 2\gamma/\eta$, then for any initial vector x^0 the sequence $\{x^k\}$ generated by the Projection algorithm with the matrix E will converge to the unique solution of the $VI(K, F)$.

Consider the $VI(K, F)$ with a closed convex set K and a continuous mapping F. Denote $\Pi_K(z)$ be the projection of a vector $z \in \Re^n$ onto the set K under the Euclidean norm, then we can easily show that a vector $x \in \Re^n$ solves the $VI(K, F)$ if and only if x is a zero of the following projection equations

$$H(z) := z - \Pi_K[z - F(z)] = 0. \tag{13}$$

With a change of variable, we can show that if x solves $VI(K, F)$, then $y := x - F(x)$ is a zero of the following equations

$$\widetilde{H}(y) := F(\Pi_K(y)) + y - \Pi_K(y) = 0; \tag{14}$$

conversely, if y is a zero of $\widetilde{H}$, then $x := \Pi_K(y)$ is a solution of $VI(K, F)$. Letting F and G be continuously mappings, then we can show that x is solution of $NCP(F, G)$ if and only if x is a zero of the following mapping

$$\overline{H}(x) := \min(F(x), G(x)) = 0, \tag{15}$$

where "min" denotes the componentwise minimum operator of two vectors in $\Re^n$.

In general, the mappings H, $\widetilde{H}$, and $\overline{H}$ are not Frechét differentiable even if F and G are continuously differentiable. In a recent paper [22], Robinson cointed the term "normal maps" for $\widetilde{H}$. Since the advent of the path-breaking work of Pang [11], there have appeared a large number of literatures on solving nonsmooth equations and related problems, such as LC^1 optimization problem. An LC^1 optimization problem is such optimization problem that the objective function and constrained functions are not C^2 functions but LC^1 functions, i.e., they are once continuously differentiable and their derivatives are locally Lipschitzian but not necessarily F-differentiable. For example, the extended linear-quadratic program, which arise from stochastic programming and optimal control [23], is such a problem in the fully quadratic case. The augmented Lagrangian of a C^2 nonlinear program is also a LC^1 function [19].

Definition 3 [21]. A function $H : \Re^n \to \Re^n$ is said to be B-differentiable at a point z if there exists a function $BH(z) : \Re^n \to \Re^n$, called the B-derivative of H at z, which is positively homogeneous of degree 1 (i.e., $BH(z)(tv) = tBH(z)v$ for all $v \in \Re^n$ and all $t \geq 0$), such that

$$\lim_{v \to 0} \frac{H(z+v) - H(z) - BH(z)v}{||v||} = 0. \tag{16}$$

If H is B-differentiable at all points in a set S, then H is said to be B-differentiable in S. It was proved by Shapiro [24] that if $H : \Re^n \to \Re^n$ is locally Lipschitzian at a vector z, then H is B-differentiable at z if and only if H is directionally differentiable at z; i.e., for any $h \in \Re^n$

$$H'(x; h) = \lim_{t \downarrow 0} \frac{H(x+th) - H(x)}{t}. \tag{17}$$

Basing on the B-derivative, Pang [11] gave the following modified Newton method for solving

$$H(x) = 0. \tag{18}$$

Newton's method with line search. Let z^0 be an arbitrary vector. Let s, β and σ be given scalars with $s > 0$, $\beta \in (0,1)$ and $\sigma \in (0, 1/2)$. In general, given z^k with $H(z^k) \neq 0$, solve the generalized Newton equations

$$H(z^k) + BH(z^k)d = 0 \tag{19}$$

for a direction d^k. Let $\alpha_k = \beta^{m_k} s$ where m_k is the first nonnegative integer m for which

$$g(z^k) - g(z^k + \beta^m s d^k) \geq -\sigma \beta^m s g'(z^k; d^k), \tag{20}$$

where

$$g(z) = \frac{1}{2} H(z)^T H(z) \tag{21}$$

and

$$g'(z; d) = H(z)^T BH(z)d.$$

Set $z^{k+1} = z^k + \alpha_k d^k$.

Some limited global convergence for Newton's method with line search is obtained in Pang [11]. Since (19) is a nonlinear problem, it is difficult to solve. In order to ease the difficulty of computing (19), Pang and Gabriel [14] proposed an NE/SQP method for solving nonlinear complementarity problem. NE/SQP stands for Nonsmooth Equation Sequential Quadratic Programming. Pang and Gabriel's method needs to solve a convex quadratic problem to get the direction d^k. Global convergence is discussed and locally quadratic convergence is obtained in [14]. See Pang and Qi [15] for such method's extensions. For variational inequality problem $VI(K, F)$, when $F \in C^1$, Fukushima [3] gave a differentiable merit function

$$\gamma(x) = \frac{1}{2} F(x)^T F(x) - \frac{1}{2} ||x - F(x) - \Pi_K[x - F(x)]||^2. \tag{22}$$

Basing on this differentiable merit function, when F is strongly monotone , Taji, Fukushima, and Ibaraki [26] gave a globally convergent Newton method, which in the k-th step needs to solve a linear variational inequality problem $VI(F^k(x), K)$, where

$$F^k(x) = F(x^k) + F'(x^k)(x - x^k).$$

The quadratic convergence is established under the generalized strict complementarity condition, which is somewhat restrictive.

Suppose that $H : \Re^n \to \Re^n$ is locally Lipschitzian. H is said to be semismooth at $x \in \Re^n$ if the following limit exists for any $h \in \Re^n$

$$\lim_{\substack{V \in \partial H(x+th') \\ h' \to h, t \downarrow 0}} \{Vh'\}. \tag{23}$$

If H is semismooth at x, then H is directionally differentiable at x and $H'(x;h)$ is equal to the limit in (23). For the semismoothness, see [19]. Basing on the concept of the semismoothness, Qi and Sun [19] gave the following generalized Newton method

$$x^{k+1} = x^k - V_k^{-1} H(x^k), \tag{24}$$

where $V_k \in \partial H(x^k)$.

Suppose that x^* is a zero of (18), then under the conditions of semismoothness of H at x^* and the nonsingularity assumption of $V \in \partial H(x^*)$, Qi and Sun [19] established the superlinear convergence of the iterative form (24). In order to reduce the nonsingularity assumption of $\partial H(x)$, $\partial_B H(x)$ was introduced in [17, 15].

$$\partial_B H(x) = \{ \lim_{\substack{x^k \in D_H \\ x^k \to x}} F'(x^k)\}, \tag{25}$$

where D_H is the set where H is differentiable. Then

$$\partial H(x) = \text{co } \partial_B H(x), \tag{26}$$

where co S is the convex hull of a set S. So in the generalized Newton method we can restrict $V_k \in \partial_B H(x^k)$ [17, 15]. When H is of the special form (15), Qi [17] gave a method how to choose an element of $\partial_B H(x)$. But for (13) and (14), there exist no results on how to compute $\partial_B H(x)$ even if K is just a polyhedral set. In a certain sense, various generalized Newton methods for solving nonsmooth equations are satisfactory. But for quasi-Newton methods, there exist few satisfactory results. Ip and Kyparisis [5] considered quasi-Newton methods directly applied to nonsmooth equations. The superlinear convergence is established on the assumption that H is strongly differentiable [10] at the solution point. This is restrictive. Chen and Qi's results [1] for quasi-Newton methods are not too far away from this. Kojima and Shindo [8] considered quasi-Newton methods for piecewise smooth functions. When the iterative sequence moves to a new C^1 piece, a new starting approximation matrix is needed. Thus a potentially large number of starting matrices need to be computed and stored. Qi and Jiang [18] considered quasi-Newton methods for solving various K-K-T systems of NLP. This is a special case of (13) or (15).

The rest of this dissertation is organized as follows. In Chapter 2, we give a globally convergent iterative method for solving (13) when F is pseudomonotone and continuous. In Chapter 3, we give a Newton method and a quasi-Newton method for solving (13) with K being a box constraint set, and (15). The superlinear convergence property is established under very mild conditions. In particular, our methods need to solve a linear equations in each step. Moreover, for quasi-Newton method we discuss how to update the QR factorization of the present iterative matrix to the QR factorization of the next. In Chapter 4, we prove the superlinear convergence of the approximate Newton methods for solving LC^1 optimization problem without assuming the strict complementarity. In Chapter 5, we give a Newton method and a quasi-Newton method for solving (14) with K being a polyhedral set. The new resulting methods in each step need to solve a linear equations whereas the corresponding algorithms in the literatures need to solve a variational inequality problem defined on K. Also the computational cost is discussed for quasi-Newton method. In Chapter 6, by combining the result of Chapter 2 and the extensions of the results of Chapter 5, we give a globally and superlinearly convergent safeguarded Newton method for solving (13) when K is a polyhedral set and F is locally Lipschitzian, semismooth over $\Re^n$ and pseudomonotone over K.

References

[1] X. Chen and L. Qi, "A parametrized Newton method and a Broyden-like method for solving nonsmooth equations", *Computational Optimization and Applications* 3 (1994) 157-179.

[2] F.H. Clarke, *Optimization and Nonsmooth Analysis* (John Wiley and Sons, New York, 1983).

[3] M. Fukushima,"Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems", *Mathematical Programming* 53 (1992) 99-110.

[4] P.T. Harker and J.-S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications", *Mathematical Programming* 48 (1990) 339-357.

[5] C.-M. Ip and T. Kyparisis, "Local convergence of quasi-Newton methods for B-differentiable equations", *Mathematical Programming* 56 (1992) 71-89.

[6] N.H. Josephy, "Newton's method for generalized equations", Technical Summary Report No. 1965, Mathematical Research Center, University of Wisconsin (Madison, WI, 1979).

[7] N.H. Josephy, "Quasi-Newton methods for generalized equations", Technical Summary Report No. 1966, Mathematical Research Center, University of Wisconsin (Madison, WI, 1979).

[8] M. Kojima and S. Shindo, "Extensions of Newton and quasi-Newton methods to systems of PC^1 equations", *Journal of the Operations Research Society of Japan* 29 (1986) 352-374.

[9] O.L. Mangasarian, *Nonlinear Programming* (McGraw-Hill, New York, 1969).

[10] J.M. Ortega and W.C. Rheinboldt, *Iterative solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).

[11] J.-S. Pang, "Newton's method for B-differentiable equations", *Mathematics of Operations Research* 15 (1990) 311-341.

[12] J.-S. Pang, "Complementarity problems", in R. Horst and P. Pardolas, eds., *Handbook on Global Optimization* (Klumer Academic Publishers, B.V., Boston, 1994).

[13] J.-S. Pang and D. Chan, "Iterative methods for variational and complementarity problems", *Mathematical Programming* 24 (1984) 284-313.

[14] J.-S. Pang and S.A. Gabriel, "NE/SQP: a robust algorithm for the nonlinear complementarity problem", Mathematical Programming 60 (1993) 295-337.

[15] J.-S. Pang and L. Qi, "Nonsmooth equations: motivation and algorithms", *SIAM Journal on Optimization* 3 (1993) 443-465.

[16] J.-S. Pang and L. Qi, "A globally convergent Newton method for convex SC^1 minimization problems", to appear in *Journal of Optimization Theory and Applications.*

[17] L. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations", *Mathematics of Operations Research* 18 (1993) 227-244.

[18] L. Qi and H. Jiang, "Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations", Applied Mathematics Report 94/5, School of Mathematics, The University of New South Wales, Sydney, Australia (Revised in November 1994).

[19] L. Qi and J. Sun, "A nonsmooth version of Newton's method", *Mathematical Programming* 58 (1993) 353-368.

[20] S.M. Robinson, "Strongly regular generalized equations", *Mathematics of Operations Research* 5 (1980) 43-62.

[21] S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity", *Mathematical Programming Study* 30 (1987) 45-66.

[22] S.M. Robinson, "Normal maps induced by linear transformation", *Mathematics of Operations Research* 17 (1992) 691-714.

[23] R.T. Rockafellar, "Computational schemes for solving large-scale problems in extended linear-quadratic programming", *Mathematical Programming* 48 (1990) 447-474.

[24] A. Shapiro, "On concepts of directional differentiability", *Journal of Optimization Theory and Applications* 66 (1990) 477-487.

[25] D. Sun,"A projection and contraction method for the nonlinear complementarity problem and its extensions", *Mathematica Numerica Sinica* 16 (1994) 183-194.

[26] K. Taji, M. Fukushima and T. Ibaraki, "A globally convergent Newton method for solving strongly monotone variational inequalities", *Mathematical Programming* 58 (1993) 369-383.

Chapter 2

A Class of Iterative Methods for Solving Nonlinear Projection Equations

Abstract

A class of globally convergent iterative methods for solving nonlinear projection equations are provided under the continuity condition of the mapping F. When F is pseudomonotone, a necessary and sufficient condition on the nonemptyness of the solution set is obtained.

Chapter 2
A Class of Iterative Methods for Solving Nonlinear
Projection Equations

1. Introduction

Assume that the mapping $F: X \subset R^n \to R^n$ is continuous and X is a closed convex subset of R^n, we will consider the solution of the following projection equations:

$$x - \Pi_X[x - F(x)] = 0, \tag{1}$$

where for any $y \in R^n$,

$$\Pi_X(y) = \text{argmin}\{x \in X | \; ||x - y||\}. \tag{2}$$

Here $||\cdot||$ denotes the l_2-norm of R^n or its induced matrix norm. The complementarity problem, variational inequality problem, and the Karush-Kuhn-Tucker systems of the nonlinear programming problems can all be cast as a special case of (1); see Eaves (Ref. 3) for a proof. For any $\beta > 0$, define

$$E_X(x, \beta) = x - \Pi_X[x - \beta F(x)]. \tag{3}$$

Without causing any confusion, we will use $E(x, \beta)$ to represent $E_X(x, \beta)$. It is easy to see that x is a solution of (1) if and only if $E(x, \beta) = 0$ for some or any $\beta > 0$. Denote

$$X^* = \{x \in X | \; x \text{ is a solution of (1)}\}. \tag{4}$$

Definition 1.1. The mapping $F: R^n \to R^n$ is said to
(i) be monotone over a set D if

$$[F(x) - F(y)]^T(x - y) \geq 0, \quad \text{for all } x, \; y \in D; \tag{5}$$

(ii) be pseudomonotone over a set D relative to a set $Y (\subset D)$ if

$$F(y)^T(x - y) \geq 0 \text{ implies } F(x)^T(x - y) \geq 0, \text{ for all } x \in D, \; y \in Y. \tag{6}$$

Remark 1.1. When $Y = D$, the pseudomonotonicity of F over a set D relative to Y is the usual pseudomonotonicity, and in this case we will say directly that F is pseudomonotone over D.

For solving projection equations (1) and related problems, there is a long history in mathematical programming field; see the comprehensive articles by Pang and Chan (Ref. 24), Harker and Pang (Ref. 7), and Pang and Qi (Ref. 26) for a detail. Among the algorithms on solving (1), Newton's method is the basic method when the derivative of F exists and is easy to implement. In this chapter, we will investigate a globally convergent method for solving (1) only with assuming the continuity of the mapping F.

When F is monotone and Lipschitzian continuous over X (i.e., there exists a positive number L such that $\|F(x) - F(y)\| \leq L\|x - y\|$, for all $x,\ y \in X$), Korpelevich (Ref. 19) proposed the following extragradient (EG) method:

$$\begin{cases} \bar{x}^k = \Pi_X[x^k - \beta F(x^k)], \\ x^{k+1} = \Pi_X[x^k - \beta F(\bar{x}^k)], \end{cases}$$

where $\beta \in (0,\ 1/L)$ is a constant. By introducing an inexact line search, Sun (Ref. 28) proposed the following improved extragradient (IEG) method:

Given constants $\eta \in (0,\ 1)$, $\alpha \in (0,\ 1)$, and $s \in (0,\ +\infty)$. The iterative form is as follows

$$\begin{cases} \bar{x}^k = \Pi_X[x^k - \beta_k F(x^k)], \\ x^{k+1} = \Pi_X[x^k - \beta_k F(\bar{x}^k)], \end{cases}$$

where $\beta_k = s\alpha^{m_k}$ and m_k is the smallest nonnegative integer m such that

$$\|F(\Pi_X[x^k - s\alpha^m F(x^k)]) - F(x^k)\| \leq \eta \frac{\|\Pi_X[x^k - s\alpha^m F(x^k)] - x^k\|}{s\alpha^m}$$

holds. The improved algorithm needs not the Lipschitzian constant. For algorithms with strong monotonicity and Lipschitzian continuity assumptions, see Fukushima (Ref. 4) and Pang and Chan (Ref. 24).

When F is an affine map, i.e., $F(x) = Mx + c$, where $M \in R^{n\times n}$ and $c \in R^n$, He (Refs. 9, 11-12) and He and Stoer (Ref. 10) proposed a projection and contraction (PC) method for solving (1). The numerical results show that PC method behaves much better than EG method or IEG method in linear cases (i.e., $F(x) = Mx + c$). This stimulates us to investigate such algorithms that not only can compete with the PC method in the linear cases but also behave much better than EG method or IEG method in the nonlinear cases. By introducing some parameters, Sun (Ref. 29) made a first step towards this. In this chapter, we will propose a class of iterative methods for solving (1) without choosing these parameters. When $F(x) = Mx + c$ and M is a skew-symmetric matrix (i.e., $M^T = -M$), our algorithms are also discussed by He (Refs. 12-13).

In section 2, we will give some preliminaries. In section 3, we give a class of abstract search directions and the corresponding algorithms. In section 4, we discuss two forms of search directions which satisfy the requirements. In section 5, we establish a necessary and sufficient condition on the nonemptyness of the solution set when F is pseudomonotone. Numerical results are presented in section 6. In section 7, we give some discussions.

2. Basic Preliminaries

Throughout this chapter, we will assume that X is a nonempty convex subset of R^n and F is continuous over X.

Lemma 2.1 [Moré (Ref. 22)]. If F is continuous over a nonempty compact convex set Y, then there exists $y^* \in Y$ such that

$$F(y^*)^T(y - y^*) \geq 0, \quad \text{for all } y \in Y.$$

Lemma 2.2 [Zarantonello (Ref. 32)]. For the projection operator $\Pi_X(\cdot)$, we have
(i) when $y \in X$, $[z - \Pi_X(z)]^T[y - \Pi_X(z)] \leq 0$, for all $z \in R^n$;
(ii) $\|\Pi_X(z) - \Pi_X(y)\| \leq \|z - y\|$, for all $y,\ z \in R^n$.

Lemma 2.3 [Gafni and Bertsekas (Ref. 5) and Calamai and Moré (Ref. 2)]. Given $x \in R^n$ and $d \in R^n$, then the function θ defined by

$$\theta(\beta) = \|\Pi_X(x + \beta d) - x\|/\beta, \quad \beta > 0$$

is antitone (nonincreasing).

Choose an arbitrary constant $\eta \in (0,\ 1)$ (e.g., $\eta = 1/2$). When $x \in X$, define

$$\eta(x) = \begin{cases} \max\{\eta, 1 - \dfrac{t(x)}{\|E(x,1)\|^2}\}, & \text{if } t(x) > 0 \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

and

$$s(x) = \begin{cases} [1 - \eta(x)]\dfrac{\|E(x,1)\|^2}{t(x)}, & \text{if } t(x) > 0 \\ 1, & \text{otherwise} \end{cases}, \tag{8}$$

where $t(x) = \{F(x) - F(\Pi_X[x - F(x)])\}^T E(x,1)$. It is easy to see that $0 < s(x) \leq 1$.

Theorem 2.1. Suppose that F is continuous over X and $\eta \in (0,\ 1)$ is a constant. If $S \subset X\backslash X^*$ is a compact set, then there exists a positive constant δ (≤ 1) such that for all $x \in S$ with $s(x) < 1$ and $\beta \in (0,\ \delta]$, we have

$$\{F(x) - F(\Pi_X[x - \beta F(x)])\}^T E(x,\beta) \leq [1 - \eta(x)]\|E(x,\beta)\|^2/\beta. \tag{9}$$

Proof. Note that for any $x \in X\backslash X^*$ with $s(x) < 1$, we have

$$t(x) > 0 \text{ and } \eta(x) > 1 - \frac{t(x)}{\|E(x,1)\|^2},$$

which, and the definition of $\eta(x)$, means that $\eta(x) = \eta$.

Since $S \subset X\backslash X^*$ is a compact set and F is continuous over X, there exists a constant $\delta_0 > 0$ such that for all $x \in S$, we have

$$\|\Pi_X[x - F(x)] - x\| \geq \delta_0 > 0. \tag{10}$$

From Lemma 2.3 and (10), for all $\beta \in (0,\ 1]$ and $x \in S$ we have

$$\|x - \Pi_X[x - \beta F(x)]\|/\beta \geq \|x - \Pi_X[x - F(x)]\| \geq \delta_0. \tag{11}$$

From the continuity of F we know that F is uniformly continuous over compact sets. So from (ii) of Lemma 2.2 we know that there exists a positive constant δ (≤ 1) such for all $x \in S$ with $s(x) < 1$ and $\beta \in (0, \delta]$ that

$$\|F(\Pi_X[x - \beta F(x)]) - F(x)\| \leq (1 - \eta)\delta_0. \tag{12}$$

Combining (11) and (12), for all $x \in S$ and $\beta \in (0, \delta]$ we have

$$\begin{aligned}
&\{F(x) - F(\Pi_X[x - \beta F(x)])\}^T E(x, \beta) \\
&\quad \le \|F(x) - F(\Pi_X[x - \beta F(x)])\| \|E(x, \beta)\| \\
&\quad \le (1 - \eta)\|E(x, \beta)\|^2/\beta \\
&\quad = [1 - \eta(x)]\|E(x, \beta)\|^2/\beta,
\end{aligned}$$

which completes the proof of (9). □

3. Algorithms and Convergence

Suppose that $g : R^n \times R^1_{++} \to R^n$ is a continuous mapping. We will use $g(x, \beta)$ as a search direction in this section. The various forms of $g(x, \beta)$ will be given in section 4. First we will describe our algorithms (in the abstract form of $g(x, \beta)$).

Projection and Contraction (PC) Methods

Given $x^0 \in X$, positive constants η, $\alpha \in (0,\ 1)$, and $0 < \Delta_1 \le \Delta_2 < 2$.

For $k = 0, 1, ...,$ if $x^k \notin X^*$, then do

1. Calculate $\eta(x^k)$ and $s(x^k)$. If $s(x^k) = 1$, let $\beta_k = 1$; otherwise determine $\beta_k = s(x^k)\alpha^{m_k}$, where m_k is the smallest nonnegative integer m such that

$$\begin{aligned}
&\{F(x^k) - F(\Pi_X[x^k - s(x^k)\alpha^m F(x^k)])\}^T E(x^k, s(x^k)\alpha^m) \\
&\qquad \le [1 - \eta(x^k)]\|E(x^k, s(x^k)\alpha^m)\|^2/(s(x^k)\alpha^m)
\end{aligned} \tag{13}$$

holds.

2. Calculate $g(x^k, \beta_k)$.
3. Calculate

$$\rho_k = E(x^k, \beta_k)^T g(x^k, \beta_k)/\|g(x^k, \beta_k)\|^2. \tag{14}$$

4. Take $\gamma_k \in [\Delta_1,\ \Delta_2]$ and set

$$\bar{x}^k = x^k - \gamma_k \rho_k g(x^k, \beta_k), \tag{15}$$

$$x^{k+1} = \Pi_X(\bar{x}^k). \tag{16}$$

Remark 3.1. Theorem 2.1 ensures that β_k can be obtained in finite number of trials if $s(x^k) < 1$. When $s(x^k) = 1$, (13) holds for $m = 0$.

For $\beta > 0$, define

$$\psi(x, \beta) = \eta(x)\|E(x, \beta)\|^2/\beta. \tag{17}$$

Theorem 3.1. Suppose that F, g are continuous over X, $X \times R^1_{++}$, respectively. If $X^* \neq \emptyset$ and there exists $x^* \in X^*$ such that the infinite sequence $\{x^k\}$ generated by PC methods satisfies

$$(x^k - x^*)^T g(x^k, \beta_k) \ge E(x^k, \beta_k)^T g(x^k, \beta_k) \ge \psi(x^k, \beta_k), \tag{18}$$

then

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \gamma_k(2 - \gamma_k)\psi^2(x^k, \beta_k)/\|g(x^k, \beta_k)\|^2. \tag{19}$$

Proof. From (ii) of Lemma 2.2 and (18), we have

$$
\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|\Pi_X[x^k - \gamma_k \rho_k g(x^k, \beta_k)] - x^*\|^2 \\
&\le \|x^k - \gamma_k \rho_k g(x^k, \beta_k) - x^*\|^2 \\
&= \|x^k - x^*\|^2 - 2\gamma_k \rho_k (x^k - x^*)^T g(x^k, \beta_k) + \gamma_k^2 \rho_k^2 \|g(x^k, \beta_k)\|^2 \\
&\le \|x^k - x^*\|^2 - 2\gamma_k \rho_k E(x^k, \beta_k)^T g(x^k, \beta_k) + \gamma_k^2 \rho_k^2 \|g(x^k, \beta_k)\|^2 \\
&= \|x^k - x^*\|^2 - \gamma_k (2 - \gamma_k)[E(x^k, \beta_k)^T g(x^k, \beta_k)]^2 / \|g(x^k, \beta_k)\|^2 \\
&\le \|x^k - x^*\|^2 - \gamma_k (2 - \gamma_k) \psi^2(x^k, \beta_k) / \|g(x^k, \beta_k)\|^2,
\end{aligned}
$$

which verifies (19).

Define

$$
\operatorname{dist}(x, X^*) = \inf\{\|x - x^*\| \mid x^* \in X^*\}. \tag{20}
$$

Theorem 3.2. Suppose that the conditions of Theorem 3.1 hold. Then the infinite sequence $\{x^k\}$ generated by PC methods is bounded and $\liminf\limits_{k \to \infty} \operatorname{dist}(x^k, X^*) = 0$. Furthermore, if (18) holds for any $x^* \in X^*$, then there exists $\bar{x} \in X^*$ such that $x^k \to \bar{x}$ as $k \to \infty$.

Proof. For the sake of simplicity, we take $\gamma_k = 1$.

From (19) we know that $\{\|x^k - x^*\|\}$ is a decreasing sequence. So the sequence $\{x^k\}$ generated by PC methods is bounded and the sequence $\{\operatorname{dist}(x^k, X^*)\}$ is also bounded. Suppose that there exists a positive constant ε such that

$$
\operatorname{dist}(x^k, X^*) \ge \varepsilon > 0, \quad \text{for all } k.
$$

Define

$$
S = \{x \in X \mid \operatorname{dist}(x, X^*) \ge \varepsilon,\ \|x - x^*\| \le \|x^0 - x^*\|\}.
$$

Then $S \subset X \backslash X^*$ is a compact set and $\{x^k\} \subset S$. From Theorem 2.1 we know that there exists a positive constant δ (≤ 1) such for all $x \in S$ with $s(x) < 1$ and $\beta \in (0,\ \delta]$ that [illegible] holds. Hence for each k with $s(x^k) < 1$, we have

$$
\beta_k \ge \min\{\alpha\delta, s(x^k)\}. \tag{21}
$$

From the definition of $s(x^k)$, we know that if $s(x^k) < 1$, then

$$
\{F(x^k) - F(\Pi_X[x^k - F(x^k)])\}^T E(x^k, 1) > 0, \quad \eta(x^k) = \eta,
$$

and

$$
\begin{aligned}
s(x^k) &= (1 - \eta) \frac{\|E(x^k, 1)\|^2}{\{F(x^k) - F(\Pi_X[x^k - F(x^k)])\}^T E(x^k, 1)} \\
&\ge (1 - \eta) \frac{\|E(x^k, 1)\|}{\|F(x^k)\| + \|F(\Pi_X[x^k - F(x^k)])\|}.
\end{aligned} \tag{22}
$$

From the continuity of F and $\{x^k\} \subset S \subset X \backslash X^*$, we know that

$$\inf_k ||E(x^k, 1)|| > 0. \tag{23}$$

From (21-23), there exists a positive constant $\bar{\delta}$ (≤ 1) such that

$$\beta_k \geq \bar{\delta} > 0.$$

If $s(x^k) = 1$, then $\beta_k = 1$. Hence

$$1 \geq \beta_k \geq \bar{\delta} > 0, \quad \text{for all } k. \tag{24}$$

Therefore,

$$\inf_k \psi(x^k, \beta_k)/||g(x^k, \beta_k)|| = \varepsilon_0 > 0,$$

which, and (19) (note that we just take $\gamma_k = 1$), means that

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 - {\varepsilon_0}^2.$$

Taking limits in both sides of the above inequality, we can derive a contradiction since $\{||x^k - x^*||\}$ is a convergent sequence. So we have

$$\liminf_{k \to \infty} \text{dist}(x^k, X^*) = 0. \tag{25}$$

Furthermore, if (18) holds for all $x^* \in X^*$. We can conclude that there exists $\bar{x} \in X^*$ such that $x^k \to \bar{x}$ as $k \to \infty$. In fact, since X^* is closed, (25) and the boundedness of $\{x^k\}$ mean that there exist $\bar{x} \in X^*$ and a subsequence $\{x^{k_j}\}$ such that $x^{k_j} \to \bar{x}$ as $j \to \infty$. Since $\{||x^k - \bar{x}||\}$ is a decreasing sequence and $x^{k_j} \to \bar{x}$ as $j \to \infty$, the whole sequence $\{x^k\}$ also converges to $\bar{x}$. □

When X is of the following form

$$X = \{x \in R^n |\ l \leq x \leq u\}, \tag{26}$$

where l and u are two vectors of $\{R \cup \{\infty\}\}^n$, we can give an improved form of the PC methods. For any $x \in X$ and $\beta > 0$, denote

$$N(x, \beta) = \{i |\ (x_i = l_i \text{ and } (g(x, \beta))_i \geq 0) \text{ or } (x_i = u_i \text{ and } (g(x, \beta))_i \leq 0)\},$$

$$B(x, \beta) = \{1, ..., n\} \backslash N(x, \beta). \tag{27}$$

Denote $g_N(x, \beta)$ and $g_B(x, \beta)$ as follows

$$(g_N(x, \beta))_i = \begin{cases} 0, & \text{if } i \in B(x, \beta) \\ (g(x, \beta))_i, & \text{otherwise} \end{cases},$$

$$(g_B(x, \beta))_i = (g(x, \beta))_i - (g_N(x, \beta))_i, \quad i = 1, ..., n. \tag{28}$$

Then for any $x^* \in X^*$ and $x \in X$,

$$(x - x^*)^T g_B(x, \beta) \geq (x - x^*)^T g(x, \beta). \tag{29}$$

So if in the PC methods we set

$$x^{k+1} = \Pi_X[x^k - \gamma_k \rho_k g_B(x^k, \beta_k)], \tag{30}$$

where

$$\rho_k = E(x^k, \beta_k)^T g(x^k, \beta_k) / \|g_B(x^k, \beta_k)\|^2,$$

then the convergence Theorems 3.1-3.2 hold for the improved PC methods. In practice, we will use the iterative form (30) whenever X is of the form (26).

4. The Search Directions

In this section, under some conditions, we will give two forms of search directions which satisfy the assumptions of Theorems 3.1-3.2.

For any $\beta > 0$, define

$$g(x, \beta) = F(\Pi_X[x - \beta F(x)]) \tag{31}$$

or

$$g(x, \beta) = F(\Pi_X[x - \beta F(x)]) - F(x) + E(x, \beta)/\beta. \tag{32}$$

Theorem 4.1. Suppose that F is continuous over X, X^* is nonempty, and $g(x, \beta)$ is of the form (31) or (32). If F is pseudomonotone over X relative to $x^* \in X^*$ and there exists $\beta > 0$ such that (9) holds for some $x \in X \backslash X^*$, then

$$(x - x^*)^T g(x, \beta) \geq E(x, \beta)^T g(x, \beta) \geq \psi(x, \beta). \tag{33}$$

Furthermore, if F is pseudomonontone over X relative to X^*, then (33) holds for all $x^* \in X^*$.

Proof. Since F is pseudomonotone over X relative to $x^* \in X^*$, for all $z \in X$ we have

$$(z - x^*)^T F(z) \geq 0.$$

In particular, we have

$$\{\Pi_X[x - \beta F(x)] - x^*\}^T F(\Pi_X[x - \beta F(x)]) \geq 0. \tag{34}$$

First we consider the case that $g(x, \beta)$ takes the form (31). Considering (34), we have

$$\begin{aligned}
(x - x^*)^T g(x, \beta) &= (x - x^*)^T F(\Pi_X[x - \beta F(x)]) \\
&= E(x, \beta)^T g(x, \beta) + \{\Pi_X[x - \beta F(x)] - x^*\}^T F(\Pi_X[x - \beta F(x)]) \\
&\geq E(x, \beta)^T g(x, \beta) \\
&= E(x, \beta)^T \{F(\Pi_X[x - \beta F(x)]) - F(x)\} + E(x, \beta)^T F(x) \\
&\geq -[1 - \eta(x)]\|E(x, \beta)\|^2/\beta + E(x, \beta)^T F(x),
\end{aligned}$$

where the last inequality follows from (9). By taking $z = x - \beta F(x)$ and $y = x$ in (i) of Lemma 2.2, we have

$$\beta E(x,\beta)^T F(x) \geq \|E(x,\beta)\|^2,$$

which, and the above formulas, means that

$$\begin{aligned}(x - x^*)^T g(x,\beta) &\geq E(x,\beta)^T g(x,\beta) \\ &\geq -[1-\eta(x)]\|E(x,\beta)\|^2/\beta + \|E(x,\beta)\|^2/\beta \\ &= \eta(x)\|E(x,\beta)\|^2/\beta \\ &= \psi(x,\beta).\end{aligned}$$

Next we will consider the case that $g(x,\beta)$ takes the form (32). By taking $z = x - \beta F(x)$ and $y = x^*$ in (i) of Lemma 2.2, we have

$$\{x^* - \Pi_X[x - \beta F(x)]\}^T\{x - \beta F(x) - \Pi_X[x - \beta F(x)]\} \leq 0.$$

By rearrangement,

$$(x - x^*)^T E(x,\beta) \geq \beta\{\Pi_X[x - \beta F(x)] - x^*\}^T F(x) + \|E(x,\beta)\|^2.$$

Therefore,

$$\begin{aligned}&(x - x^*)^T g(x,\beta) \\ &= (x - x^*)^T F(\Pi_X[x - \beta F(x)]) - (x - x^*)^T F(x) + (x - x^*)^T E(x,\beta)/\beta \\ &\geq (x - x^*)^T F(\Pi_X[x - \beta F(x)]) - (x - x^*)^T F(x) \\ &\quad + \{\Pi_X[x - \beta F(x)] - x^*\}^T F(x) + \|E(x,\beta)\|^2/\beta \\ &\geq E(x,\beta)^T F(\Pi_X[x - \beta F(x)]) - E(x,\beta)^T F(x) + \|E(x,\beta)\|^2/\beta \quad \text{(using (34))} \\ &= E(x,\beta)^T g(x,\beta).\end{aligned}$$

Substituting (9) into the above formulas, we have

$$\begin{aligned}(x - x^*)^T g(x,\beta) &\geq E(x,\beta)^T g(x,\beta) \\ &= E(x,\beta)^T\{F(\Pi_X[x - \beta F(x)]) - F(x)\} + \|E(x,\beta)\|^2/\beta \\ &\geq -[1-\eta(x)]\|E(x,\beta)\|^2/\beta + \|E(x,\beta)\|^2/\beta \\ &= \eta(x)\|E(x,\beta)\|^2/\beta \\ &= \psi(x,\beta).\end{aligned}$$

□

Remark 4.1. Assume that $F(x) = Mx + c$ and M is skew-symmetric (i.e., $M^T = -M$). If $g(x, \beta)$ takes the form (31), then

$$\beta_k = 1 \ \text{ and } \ g(x_k, \beta_k) = M^T E(x^k, 1) + (Mx^k + c),$$

which means that for linear programming (translated into an equivalent linear complementarity problem), our method reduced to the same discussed by He (Ref. 11). If $g(x, \beta)$ takes the form (32), then

$$\beta_k = 1 \ \text{ and } \ g(x^k, \beta_k) = M^T E(x^k, 1) + E(x^k, 1),$$

which was also appeared in He (Ref. 13).

5. A Theorem on the Existence of the Solution(s)

When F is continuous and pseudomonotone over X, there exist some results on the existence of the solution(s) of equations (1); see Harker and Pang (Ref. 7). Here we will give a necessary and sufficient condition on the existence of the solution(s).

Theorem 5.1. Suppose that $g(x, \beta)$ takes form (31) or (32). If F is continuous and pseudomonotone over X, then $X^* \neq \emptyset$ if and only if some or any sequence $\{x^k\}$ generated by PC methods is bounded.

Proof. We just discuss the case that $g(x, \beta)$ takes the form (31). The proof on taking the form (32) is similar.

When $X^* \neq \emptyset$, then from Theorems 3.2 and 4.1, any sequence $\{x^k\}$ generated by PC methods is bounded.

For the converse part of the theorem, we suppose that there exists a bounded sequence $\{x^k\}$ generated by the PC methods. From the boundedness of $\{x^k\}$ and the continuity of F, there exists a positive constant r such that

$$\|x^k\| \leq r, \quad \|F(x^k)\| \leq r, \quad \text{for all } k.$$

From (ii) of Lemma 2.2, for all k and $\beta \in [0, \ 1]$ we have

$$\|\Pi_X[x^k - \beta F(x^k)]\| \leq 2r.$$

Choosing an arbitrary fixed vector $v \in X$, define

$$Y = \{x \in R^n | \ \|x\| \leq 2r + \|v\|\} \cap X,$$

then Y is a nonempty compact convex set, and for all k and $\beta \in [0, \ 1]$ we have

$$x^k, \ \ \Pi_X[x^k - \beta F(x^k)] \in Y.$$

Hence from the definition of Y and the properties of the projection operators $\Pi_X(\cdot)$ and $\Pi_Y(\cdot)$, for all k we have

$$\Pi_Y[x^k - \beta F(x^k)] = \Pi_X[x^k - \beta F(x^k)], \quad \text{for all } \beta \in [0, \ 1] \tag{35}$$

and

$$x^{k+1} = \Pi_X[x^k - \gamma_k \rho_k g(x^k, \beta_k)] = \Pi_Y[x^k - \gamma_k \rho_k g(x^k, \beta_k)]. \tag{36}$$

For any $x \in Y$ and $\beta > 0$, define

$$\bar{\eta}(x) = \begin{cases} \max\{\eta, 1 - \dfrac{\bar{t}(x)}{\|E_Y(x,1)\|^2}\}, & \text{if } \bar{t}(x) > 0 \\ 1, & \text{otherwise} \end{cases},$$

$$\bar{s}(x) = \begin{cases} [1 - \bar{\eta}(x)]\dfrac{\|E_Y(x,1)\|^2}{\bar{t}(x)}, & \text{if } \bar{t}(x) > 0 \\ 1, & \text{otherwise} \end{cases},$$

and

$$\bar{\psi}(x,\beta) = \bar{\eta}(x)\|E_Y(x,\beta)\|^2/\beta,$$

where $\bar{t}(x) = \{F(x) - F(\Pi_Y[x - F(x)])\}^T E_Y(x,1)$. For each k, if $\bar{s}(x^k) = 1$, let $\bar{\beta}_k = 1$; otherwise determine $\bar{\beta}_k = \bar{s}(x^k)\alpha^{m_k}$, where m_k is the smallest nonnegative integer m such that

$$\{F(x^k) - F(\Pi_Y[x^k - \bar{s}(x^k)\alpha^m F(x^k)])\}^T E_Y(x^k, \bar{s}(x^k)\alpha^m)$$
$$\le [1 - \bar{\eta}(x^k)]\|E_Y(x^k, \bar{s}(x^k)\alpha^m)\|^2/(\bar{s}(x^k)\alpha^m).$$

From (35) we know that

$$\bar{\eta}(x^k) = \eta(x^k), \quad \bar{s}(x^k) = s(x^k), \tag{37}$$

and for all $\beta \in [0,\ 1]$,

$$E_Y(x^k,\beta) = E_X(x^k,\beta). \tag{38}$$

Therefore, for all k we have

$$\bar{\beta}_k = \beta_k. \tag{39}$$

Define

$$\bar{g}(x,\beta) = F(\Pi_Y[x - \beta F(x)])$$

and

$$\bar{\rho}_k = E_Y(x^k, \beta_k)^T \bar{g}(x^k, \bar{\beta}_k)/\|\bar{g}(x^k, \bar{\beta}_k)\|^2.$$

Then from (35) and (37)-(39), we have

$$\bar{g}(x^k, \bar{\beta}_k) = g(x^k, \beta_k) \text{ and } \bar{\rho}_k = \rho_k. \tag{40}$$

Hence from (36) and (40), we have

$$\begin{aligned} x^{k+1} &= \Pi_X[x^k - \gamma_k \rho_k g(x^k, \beta_k)] \\ &= \Pi_Y[x^k - \gamma_k \rho_k g(x^k, \beta_k)] \\ &= \Pi_Y[x^k - \gamma_k \bar{\rho}_k \bar{g}(x^k, \bar{\beta}_k)], \end{aligned}$$

which means that $\{x^k\}$ can be regarded as such a sequence that generated by applying the PC methods to solve

$$E_Y(x,1)=0. \tag{41}$$

Since Y is a nonempty compact convex subset of R^n, from Lemma 2.1 and Eaves (Ref. 3) we know that the solution set

$$Y^* = \{y \in Y|\ y \text{ is a solution of (41)}\}$$

is nonempty. According to Theorems 3.2 and 4.1, there exists $x^* \in Y^*$ such that

$$x^k \to x^* \quad \text{as } k \to \infty.$$

Since $x^* \in Y^*$ and $v \in Y$, from Eaves (Ref. 3) we know that

$$F(x^*)^T(v-x^*) \geq 0.$$

Since v is an arbitrary fixed point of X and x^* is the limit point of $\{x^k\}$, we have

$$F(x^*)^T(x-x^*) \geq 0, \quad \text{for all } x \in X,$$

which, again from Eaves (Ref. 3), means that $E_X(x^*,1)=0$, i.e., X^* is nonempty and $x^* \in X^*$.

□

Remark 5.1. When X is of the form (26), Theorem 5.1 also holds for the improved PC methods. The proof is similar and the detail is omitted.

Remark 5.2. The skill introduced here can be used to give a positive answer to an open problem proposed by He and Stoer (Ref. 10); also see Sun (Ref. 30) for a proof on this open problem.

6. Numerical Experiments

In the following examples, we will take $\eta = \alpha = 0.5$, and $\Delta_1 = \Delta_2 = 1.95$ (the algorithms behave better when γ_k approaches 2.0) and use $\varphi(x,1) = F(x)^T E(x,1) \leq \varepsilon^2$ (note that $\varphi(x,1) \geq \|E(x,1)\|^2$ for all $x \in X$) as a stop criteria, where ε is a small nonnegative number. The projection and contraction method for solving nonlinear projection equations with taking forms (31) and (32) will be abbreviated as "NPC1" and "NPC2", respectively. The projection and contraction method for solving linear projection equations by He (Ref. 11) will be abbreviated as "LPC". In the above algorithms, we will use the improved search direction $g_B(x,\beta)$ instead of $g(x,\beta)$.

Example 1. This example is a 4-dimensional nonlinear complementarity problem, discussed by Kojima and Shindo (Ref. 18), where $X = R^n_+$ and

$$F(x) = \begin{pmatrix} 3x_1^2+2x_1x_2+2x_2^2+x_3+3x_4-6 \\ 2x_1^2+x_1+x_2^2+10x_3+2x_4-2 \\ 3x_1^2+x_1x_2+2x_2^2+2x_3+9x_4-9 \\ x_1^2+3x_2^2+2x_3+3x_4-3 \end{pmatrix}.$$

We take $\varepsilon^2 = 10^{-16}$.

Tabular 1
Results for example 1 with starting point (0,0,0,0)

Algorithms	Number of iterations	Number of inner iterations
NPC1	54	2
NPC2	58	2

Example 2. This example, discussed by Ahn (Ref. 1), is of the form $F(x) = Dx + c$, where c is an n-vector and D is an $n \times n$ nonsymmetric matrix

$$D = \begin{pmatrix} 4 & -2 & & & & \\ 1 & 4 & -2 & & & \\ & 1 & 4 & -2 & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & -2 \\ & & & & & 1 & 4 \end{pmatrix}.$$

$X = [l,\ u]$, where $l = (0,0,...,0)^T$ and $u = (1,1,...,1)^T$. We take $\varepsilon^2 = n10^{-14}$, where n is the dimension of the problem.

Tabular 2
Results for example 2 with starting point (0,0,...,0).

Algorithms	Number of iterations (left) and number of inner iterations (right)									
	n=10		n=50		n=100		n=200		n=500	
LPC	39		39		39		39		39	
NPC1	19	13	16	6	15	5	17	9	16	11
NPC2	16	8	17	11	14	4	14	4	13	4

Example 3. This example is a linear complementarity problem for which Lemke's algorithm is known to run in exponential time (see Murty (Ref. 23, chapter 6)). This problem is also discussed by Harker and Pang (Ref. 6) and Harker and Xiao (Ref. 8). The mapping $F(x) = Dx + c$, where

$$D = \begin{pmatrix} 1 & 2 & 2 & \cdot & \cdot & \cdot & 2 \\ 0 & 1 & 2 & \cdot & \cdot & \cdot & 2 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & 2 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

and $c = (-1, -1, ..., -1)^T$. We take $\varepsilon^2 = n10^{-14}$, where n is the dimension of the problem.

Tabular 3

Results for example 3 with starting point (0,0,...,0)

Algorithms	Number of iterations (left) and number of inner iterations (right)									
	n=10		n=50		n=100		n=200		n=500	
LPC	10		5		11		7		12	
NPC1	11	3	17	3	19	8	24	8	34	10
NPC2	11	3	17	4	19	5	25	8	31	5

Example 4. This problem is discussed by Sun (Ref. 29). Consider $F(x) = F_1(x) + F_2(x)$, $x = (x_1, ..., x_n)^T$, $x_0 = x_{n+1} = 0$, $F_1(x) = (f_1(x), ..., f_n(x))^T$ and $F_2(x) = Dx + c$, where $f_i(x) = x_{i-1}^2 + x_i^2 + x_{i-1}x_i + x_i x_{i+1}$, $i = 1, ..., n$, and D and c are the same to those of example 2. We take $X = [l,\ u]$ and $\varepsilon^2 = n10^{-14}$, where $l = (0, 0, ..., 0)^T$, $u = (1, 1, ..., 1)^T$, and n is the dimension of the problem.

Tabular 4

Results for example 4 with starting point (0,0,...,0)

Algorithms	Number of iterations (left) and number of inner iterations (right)									
	n=10		n=50		n=100		n=200		n=500	
NPC1	9	0	9	0	9	0	9	0	10	2
NPC2	9	0	9	0	9	0	10	0	10	0

7. Some Discussions

In this chapter, a class of globally convergent algorithms for solving nonlinear projection equations (1) are provided. Here the convergence rate of the given methods is not discussed, since we think that the best convergence rate is Q-linear. The basic reason for this is that the derivative of F is not assumed. However, the methods given here can converge to the neighborhood of the solution set very fast. In practice, a suitable choice is that when the iterative point is far away from the solution set, the PC methods can be used to make the iterative sequence to reach the neighborhood of the solution set; and when the iterative sequence approaches the solution set close enough, more rapid locally convergent methods, such as Newton and quasi-Newton methods, can be used. For Newton and quasi-Newton methods for solving equations (1), See Josephy (Refs. 14-15), Pang (Ref. 25), Qi and Sun (Ref. 27), Pang and Qi (Ref. 26), and Sun and Han (Ref. 31).

In section 4, two forms of search directions are given to satisfy the requirements. In fact, more search directions can be given. For example, the convex combination of the forms (31) and (32) is also a suitable choice. For various forms of search directions for solving linear projection equations, see He (Refs. 9, 11-13) and He and Stoer (Ref. 10).

When F is Lipschitzian continuous over X, we can prove that the steplength is bounded away from zero if $g(x, \beta)$ takes the form (32). This result doesn't hold for the form (31). But from the computational experience, there is no difference between choosing (31) and (32).

References

1 Ahn, Byong-hun, *Iterative Methods for Linear Complementarity Problem with Upperbounds and Lowerbounds*, Mathematical Programming, Vol. 26, pp. 295-315, 1983.

2 Calamai, P. H., and Moré, J. J., *Projected Gradient Method for Linearly Constrained Problems*, Mathematical Programming, Vol. 39, pp. 93-116, 1987.

3 Eaves, B. C., *On the Basic Theorem of Complementarity*, Mathematical Programming, Vol. 1, pp. 68-75, 1971.

4 Fukushima, M., *Equivalent Differentiable Optimization Problems and Descent Methods for Asymmetric Variational Inequality Problems*, Mathematical Programming, Vol. 53, pp. 99-110, 1992.

5 Gafni, E. H., and Bertsekas, D. P., *Two-Metric Projection Methods for Constrained Optimization*, SIAM Journal on Control and Optimization, Vol. 22, pp. 936-964, 1984.

6 Harker, P. T., and Pang, J. -S., *A Damped-Newton Method for Linear Complementarity Problem*, in G. Allgower and K. Georg, eds., Computational Solutions of Nonlinear Systems of Equations, Lectures in Applied Mathematics, Vol. 26 (American Mathematical Society, Province, RI, 1990), pp. 265-284.

7 Harker, P. T., and Pang, J. -S., *Finite-Dimensional Variational Inequality and Nonlinear Complementarity Problems: A Survey of Theory, Algorithms and Applications*, Mathematical Programming, Vol. 48, pp. 161-220, 1990.

8 Harker, P. T., and Xiao, B., *Newton's Method for the Nonlinear Complementarity Problem: A B-Differentiable Equation Approach*, Mathematical Programming, Vol. 48, pp. 339-357, 1990.

9 He, B., *A Saddle Point Algorithm for Linear Programming*, Shu Xue Banian Kan, Vol. 6, pp. 42-48, 1989.

10 He, B., and Stoer, J., *Solutions of Projection Problems over Polytopes*, Numerische Mathematik, Vol. 61, pp. 73-90, 1992.

11 He, B., *A Projection and Contraction Method for a Class of Linear Complementarity Problems and Its Application in Convex Quadratic Programming*, Applied Mathematics and Optimization, Vol. 25, pp. 247-262, 1992.

12 He, B., *On a Class of Iterative Projection and Contraction Methods for Linear Programming*, Journal of Optimization Theory and Applications, Vol. 78, pp. 247-266, 1993.

13 He, B., *Solving a Class of Linear Projection Equations*, to Appear in Mathematical Programming.

14 Josephy, N. H., *Newton's Method for Generalized Equations*, Technical Summary Report No. 1965, Mathematical Research Center, Univerty of Wisconsin-Madison, 1979.

15 Josephy, N. H., *Quasi-Newton Methods for Generalized Equations*, Technical Summary Report No. 1966, Mathematical Research Center, University of Wisconsin-Madison, 1979.

16 Karamardian, S., *Generalized Complementarity Problems*, Journal of Optimization Theory and Applications, Vol. 8, pp. 747-756, 1971.

17 Karamardian, S., and Schaible, S., *Seven Kinds of Monotone Maps*, Journal of Optimization Theory and Applications, Vol. 66, pp. 37-46, 1990.

18 Kojima, M., and Shindo, S., *Extensions of Newton and Quasi-Newton Methods to Systems of PC^1 Equations*, Journal of Operations Research Society of Japan, Vol. 29, pp. 352-374, 1986.

19 Korpelevich, G. M., *Ekstragradientnyi Method Dlia Otyskaniia Sedlovykh Tchek I Drugikh Zadach*, Ekonomica I Matematicheski Metody, Vol. 12, pp. 947-956, 1976.

20 Lemke, C. E., *On Complementarity Pivot Theory*, in Mathematics of the Decision Sciences, G. B. Dantzig and A. F. Veinott, eds., 1968.

21 Mathiesen, L., *An Algorithm Based on a Sequence of Linear Complementarity Problems Applied to a Walrasian Equilibrium Model: An Example*, Mathematical Programming, Vol. 37, pp. 1-18, 1987.

22 Moré, J. J., *Coercivity Conditions in Nonlinear Complementarity Problems*, SIAM Review, Vol. 16, pp. 1-16, 1974.

23 Murty, K. G., *Linear Complementarity, Linear and Nonlinear Programming*, Helderman, Berlin, 1988.

24 Pang, J. -S., and Chan, D., *Iterative Methods for Variational and Complementarity Problems*, Mathematical Programming, Vol. 24, pp. 284-313, 1982.

25 Pang, J. -S., *Newton's Method for B-Differentiable Equations*, Mathematics of Operations Research, Vol. 15, pp. 311-341, 1990.

26 Pang, J. -S., and Qi., L., *Nonsmooth Equations: Motivation and Algorithms*, SIAM Journal on Optimization, Vol. 3, pp. 443-465, 1993.

27 Qi, L., and Sun, J., *A Nonsmooth Version of Newton's Method*, Mathematical Programming, Vol. 58, pp. 353-368, 1993.

28 Sun, D., *An Iterative Method for Solving Variational Inequality Problems and Complementarity Problems*, Numerical Mathematics, A Journal of Chinese Universities, Vol. 16, pp. 145-153, 1994.

29 Sun, D., *A Projection and Contraction Method for the Nonlinear Complementarity Problem and Its Extensions*, Mathematica Numerica Sinica, Vol. 16, pp. 183-194, 1994.

30 Sun, D., *On the Convergence Properties of a Projection and Contraction Method*, to Appear in Numerical Mathematics, A Journal of Chinese Universities (English Series).

31 Sun, D., and Han, J., *Newton and Quasi-Newton Methods for a Class of Nonsmooth Equations and Related Problems*, Technical Report No. 026, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, China, 1994

32 Zarantonello, E. H., *Projections on Convex Sets in Hilbert Space and Spectral Theory*, in E. H. Zarantonello, ed., Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971.

Chapter 3

Newton and Quasi-Newton Methods for a Class of Nonsmooth Equations and Related Problems

Abstract

This chapter presents a Newton method and a quasi-Newton method for solving some nonsmooth equations (NE). In order to construct a convergent and practical quasi-Newton method for solving a class of nonsmooth equations, which arises from complementarity problem, variational inequality problem, the Karush-Kuhn-Tucker (KKT) system of nonlinear programming, and related problems, a concept $\partial_b F(x)$ and an approximation idea are introduced in this chapter. The Q-superlinear convergence of the Newton method and the quasi-Newton method is established under suitable assumptions, in which the existence of $F'(x^*)$ is not assumed. The new algorithms only need to solve a linear equations in each step. For complementarity problem, the QR factorization on the quasi-Newton method is discussed.

Chapter 3

Newton and Quasi-Newton Methods for a Class of Nonsmooth Equations and Related Problems

1. Introduction

In the recent years, many authors have considered various forms of Newton method for solving nonsmooth equations (NE) (see, e.g., [16, 17, 18, 19, 11, 12, 20, 21, 22, 25]). Some authors have also considered the application of quasi-Newton methods to nonsmooth equations. In Kojima and Shindo [11], quasi-Newton method was applied to piecewise smooth equations. When the iteration sequence moves to a new C^1-piece, a new approximate starting matrix is needed. Ip and Kyparisis [9] considered the local convergence of quasi-Newton methods directly applied to B-differentiable equations (in the sense of Robinson [24]). The superlinearly convergent theorems are established under the assumption that F is strongly F- differentiable [14] at the solution.

The main object of this chapter is to construct a practical quasi-Newton method for nonsmooth equations, especially for the nonsmooth equations, which is of concrete background. In order to complete this, we first give a slight modification of the generalized Newton method [20, 21]. Basing on the modified generalized Newton method, we give a quasi-Newton method for solving a class of nonsmooth equations, which arises from complementarity problem, variational inequality problem, the Karush-Kuhn-Tucker (KKT) system of nonlinear programming, and related problems. In each step, we only need to solve a linear equations. The Q-superlinear convergence is established under mild conditions. Although we don't know how to give a convergent quasi-Newton method for general nonsmooth equations, the general convergent theorems in abstract forms are established. These theorems will be helpful in constructing new methods for solving nonsmooth equations, which is of some special form.

The characteristics of the quasi-Newton method for solving (4.1) established in §4 include: (i) without assuming the existence of $F'(x^*)$, we prove the Q-superlinearly convergent property, (ii) only one approximate starting matrix is needed, and (iii) from the QR factorization of the kth iterate matrix we need at most $O((I(k)+1)n^2)$ arithmetic operations to get the QR factorization of the $(k+1)$th iterate matrix (for the definition of $I(k)$ see (5.8)).

The remainder of this chapter is organized as follows. In §2, we give some preliminaries on nonsmooth functions. In §3, we propose a modified generalized Newton method. In §4, we give a quasi-Newton method for solving a class of nonsmooth equations. In §5, we discuss the implementation of the quasi-Newton method for the nonlinear complementarity problem. The KKT system of variational inequality problem with upper and lower bounds are discussed in §6. The computational results are given in §7.

2. Preliminaries

In general, assume that $F : R^n \rightarrow R^m$ is locally Lipschitzian. In order to reduce the nonsingularity assumption of the generalized Newton method [21], the concept $\partial_B F(x)$ was introduced by Qi [20]

$$\partial_B F(x) = \Big\{ \lim_{\substack{x^k \rightarrow x \\ x^k \in D_F}} F'(x^k) \Big\}, \tag{2.1}$$

where D_F is the set where F is differentiable. Let ∂F be the generalized Jacobian of F in the sense of Clarke [4]. Then $\partial F(x)$ is the convex hull of $\partial_B F(x)$,

$$\partial F(x) = \text{co } \partial_B F(x). \tag{2.2}$$

For $m = 1$, $\partial_B F(x)$ was introduced by Shor [26]. Here, we denote

$$\partial_b F(x) = \partial_B F_1(x) \times \partial_B F_2(x) \times \cdots \times \partial_B F_m(x). \tag{2.3}$$

When $m = 1$, $\partial_b F(x) = \partial_B F(x)$.

Suppose that $f,\ g : R^n \rightarrow R^1$ are continuously differentiable functions. Let $h(x) = \min(f(x), g(x))$, then

$$\partial_b h(x) = \begin{cases} \{\nabla f(x)^T\} & \text{if } f(x) < g(x), \\ \{\nabla f(x)^T, \nabla g(x)^T\} & \text{if } f(x) = g(x), \\ \{\nabla g(x)^T\} & \text{if } f(x) > g(x). \end{cases}$$

In fact, when $f(x) = g(x)$ but $\nabla f(x) \neq \nabla g(x)$, we can prove that in an arbitrary neighborhood $N(x)$ of x there exist $y,\ z \in N(x)$ such that $f(y) < g(y)$ and $f(z) > g(z)$.

We say that F is *semismooth* at x, if

$$\lim_{\substack{V \in \partial F(x+th') \\ h' \rightarrow h,\ t \downarrow 0}} \{Vh'\} \tag{2.4}$$

exists for any $h \in R^n$. If F is semismooth at x, then F is directionally differentiable at x and $F'(x;h)$ is equal to the limit in (2.4) (see [21]). Semismoothness was originally introduced by Mifflin [13] for functionals. Convex functions, smooth functions, and piecewise linear functions are examples of semismooth functions. Scalar productions and sums of semismooth functions are still semismooth functions (see [13]). In [22], Qi and Sun extended the definition of semismooth functions to $F : R^n \rightarrow R^m$. It was proved in [22] that F is semismooth at x if and only if all its component functions are so.

LEMMA **2.1** **[21]**. *Suppose that $F : R^n \rightarrow R^m$ is a locally Lipschitzian function and semismooth at x. Then*

(1) *for any $V \in \partial F(x+h)$, $h \rightarrow 0$,*

$$Vh - F'(x;h) = o(\|h\|); \tag{2.5}$$

(2) *for any* $h \to 0$,

$$F(x+h) - F(x) - F'(x;h) = o(\|h\|). \tag{2.6}$$

LEMMA 2.2. *Suppose that* $F : R^n \to R^n$ *is a locally Lipschitzian function. If all* $V \in \partial_b F(x)$ *are nonsingular. Then there exists a positive constant* C *such that*

$$\|V^{-1}\| \leq C$$

for any $V \in \partial_b F(x)$*. Furthermore, there exists a neighborhood* $N(x)$ *of* x *such that for any* $y \in N(x)$*, all* $W \in \partial_b F(y)$ *are nonsingular and satisfy*

$$\|W^{-1}\| \leq \frac{10C}{9}. \tag{2.7}$$

Proof. From the definition of $\partial_b F(x)$ we have

$$\partial_b F(x) \subset \partial F_1(x) \times \partial F_2(x) \times \cdots \times \partial F_n(x).$$

Since F_i is locally Lipschitzian, ∂F_i is bounded in a neighborhood of x. Therefore, $\partial_b F$ is also bounded in a neighborhood of x. The closeness of $\partial_b F(x)$ can be easily derived from the definition of $\partial_b F$. Since all $V \in \partial_b F(x)$ are nonsingular, and $\partial_b F(x)$ is bounded and closed, there is a positive number C such that

$$\|V^{-1}\| \leq C$$

for any $V \in \partial_b F(x)$.

In order to complete the second part of the Lemma, for given $\varepsilon = \dfrac{1}{20nC}$, we claim that for each $i \in \{1, 2, \ldots, n\}$ there exists a neighborhood $N_i(x)$ of x such that for any $y \in D_{F_i} \cap N_i(x)$,

$$F_i'(y) \subset \partial_b F_i(x) + \varepsilon B, \tag{2.8}$$

where B is the unit ball of R^n. If this claim is not true, then there exists some $i \in \{1, 2, \ldots, n\}$ and a sequence $\{y^k\} \to x$, $y^k \in D_{F_i}$ such that

$$\|F_i'(y^k) - V^i\| > \varepsilon \tag{2.9}$$

for any $V^i \in \partial_b F_i(x)$. Since $\partial_b F_i$ is locally bounded and $F_i'(y^k) \in \partial_b F_i(y^k)$, by passing to a subsequence if necessary, we may assume that $F_i'(y^k) \to W^i$. Then from the definition of $\partial_b F_i(x)$ we have $W^i \in \partial_b F_i(x)$, which contradicts (2.9). Hence, (2.8) holds. From the definition of $\partial_b F_i$ and (2.8), we can prove by contradiction that there exists a neighborhood $N(x)$ of x such that

$$\partial_b F_i(y) \subset \partial_b F_i(x) + 2\varepsilon B \tag{2.10}$$

for any $y \in N(x)$ and $i \in \{1, 2, \ldots, n\}$. Therefore, for any $W \in \partial_b F(y)$, $y \in N(x)$, there exists $V \in \partial_b F(x)$ such that

$$\|W - V\| \leq 2n\varepsilon = \frac{1}{10C}.$$

Then from Theorem 2.3.2 of [14] we know that W is nonsingular and

$$\|W^{-1}\| \leq \frac{\|V^{-1}\|}{1 - \|V^{-1}(W - V)\|} \leq \frac{C}{1 - \frac{C}{10C}} = \frac{10}{9}C. \qquad \square$$

3. Newton Method for Nonsmooth Equations

Suppose that $F : R^n \to R^n$ is locally Lipschitzian. We are interested in finding a solution of the equations

$$F(x) = 0. \tag{3.1}$$

Qi and Sun [21] and Qi [20] considered various forms of Newton method for solving (3.1) when F is not F-differentiable. Here we will consider the following slightly modified Newton method

$$x^{k+1} = x^k - V_k^{-1}F(x^k), \quad k = 0, 1, ..., \tag{3.2}$$

where $V_k \in \partial_b F(x^k)$. This method is useful to establish the superlinear convergence of quasi-Newton methods given in §4. Similar to that of [20, 21], we can give the following convergent theorem.

THEOREM **3.1.** *Suppose that x^* is a solution of* (3.1)*, F is locally Lipschitzian and semismooth at x^*, and all $V_* \in \partial_b F(x^*)$ are nonsingular. Then the iteration method* (3.2) *is well defined and converges to x^* Q-superlinearly in a neighborhood of x^*.*

Proof. By Lemma 2.2, (3.2) is well defined in a neighborhood of x^* for the first step $k = 0$. Since $V_k \in \partial_b F(x^k)$, the ith row V_k^i of V_k satisfies

$$V_k^i \in \partial_b F_i(x^k).$$

From the semismoothness of F we know that F_i is semismooth at x^*. By Lemma 2.1,

$$V_k^i(x^k - x^*) - F_i'(x^*; x^k - x^*) = o(||x^k - x^*||), \ i = 1, \ldots, n.$$

Therefore,

$$V_k(x^k - x^*) - F'(x^*; x^k - x^*) = o(||x^k - x^*||). \tag{3.3}$$

From Lemma 2.1 and (3.3) we have

$$\begin{aligned} ||x^{k+1} - x^*|| &= ||x^k - x^* - V_k^{-1}F(x^k)|| \\ &\le ||V_k^{-1}[F(x^k) - F(x^*) - F'(x^*; x^k - x^*)]|| \\ &\quad + ||V_k^{-1}[V_k(x^k - x^*) - F'(x^*; x^k - x^*)]|| \\ &= o(||x^k - x^*||). \end{aligned}$$

□

4. Quasi-Newton Method for Some Nonsmooth Equations

In this section, we will first consider the following nonsmooth equations, which arises from complementarity problem, variational inequality problem, and the KKT system of nonlinear programming:

$$F(x) = x - P_X[x - f(x)] = 0, \tag{4.1}$$

where $f : R^n \to R^n$ is a continuously differentiable function, $P_Y(\cdot)$ is the orthogonal projection operator onto a nonempty closed convex set Y, and $X = \{x \in R^n | \ l \le x \le u\}$, where $l, u \in \{R \cup \{\infty\}\}^n$. To solve equations (4.1) is the original motivation in investigating nonsmooth equations. When $f \in C^1$, F is a semismooth function. The results of Newton method for solving (4.1) are fruitful, but not for the quasi-Newton method. In this section, we will give a new quasi-Newton method for solving equations (4.1), and generalize the convergent theory to general nonsmooth equations. We don't know how to construct a superlinearly convergent quasi-Newton method for general nonsmooth equations under mild conditions, but the skill introduced here will be helpful in devising quasi-Newton methods for some other special nonsmooth equations. We also give examples to demonstrate this.

We will give a quasi-Newton method for solving equations (4.1).

Quasi-Newton Method (Broyden's Case)

Given $f : R^n \to R^n$, $x^0 \in R^n$, $A_0 \in R^{n\times n}$

Do for $\quad k = 0, 1, \dots :$

Define

$$\begin{aligned} f^k(x) &= f(x^k) + A_k(x - x^k) \\ F^k(x) &= x - P_X[x - f^k(x)] \end{aligned} \tag{4.2}$$

Choose $\quad V_k \in \partial_b F^k(x^k)$

Solve $\quad V_k s^k + F(x^k) = 0$ for s^k

$$x^{k+1} = x^k + s^k$$

$$y^k = f(x^{k+1}) - f(x^k)$$

$$A_{k+1} = A_k + \frac{(y^k - A_k s^k){s^k}^T}{{s^k}^T s^k}. \tag{4.3}$$

For any matrix $B \in R^{n\times n}$, let B^i be the ith row of B. For an arbitrary function $f \in C^1$, if $V \in \partial_b F(x)$, then V satisfies

$$V^i = \begin{cases} I^i & \text{if } x_i - f_i(x) < l_i \text{ (or} > u_i), \\ \lambda_i I^i + (1 - \lambda_i) f_i'(x) & \text{if } x_i - f_i(x) = l_i \text{ (or} = u_i), \\ f_i'(x) & \text{if } l_i < x_i - f_i(x) < u_i, \end{cases} \tag{4.4}$$

where $\lambda_i \in \{0, 1\}$, I is the unit matrix of $R^{n\times n}$. On the other hand, any V of the above form is an element of $\partial_b F(x)$.

THEOREM **4.1.** *Suppose that $f : R^n \to R^n$ is continuously differentiable, x^* is a solution of (4.1), $f'(x)$ is Lipschitz continuous in a neighborhood of x^* and the Lipschitz constant is γ. Suppose that all $W_* \in \partial_b F(x^*)$ are nonsingular. There exist positive constants ε, δ such that if $\|x^0 - x^*\| \le \varepsilon$ and $\|A_0 - f'(x^*)\| \le \delta$, then the sequence $\{x^k\}$ generated by the Quasi-Newton Method (Broyden's Case) is well defined and converges Q-superlinearly to x^*.*

Proof. From Lemma 2.2, there exists a positive constant β such that $\|W_*^{-1}\| \le \beta$ for all $W_* \in \partial_b F(x^*)$ and there exists a neighborhood $N_0(x^*)$ of x^* such that

$$\|W^{-1}\| \le \frac{10}{9}\beta$$

for any $y \in N_0(x^*)$, $W \in \partial_b F(y)$. Choose ε_1 and δ such that

$$\|f'(y) - f'(x^*)\| \le \gamma\|y - x^*\|, \tag{4.5}$$

$$12\beta\delta \le 1, \tag{4.6}$$

$$3\gamma\varepsilon_1 \le 2\delta, \tag{4.7}$$

$$\|W^{-1}\| \le \frac{10\beta}{9} \tag{4.8}$$

for any $y \in N_1(x^*) = \{x \mid \|x - x^*\| \le \varepsilon_1\}$, $W \in \partial_b F(y)$. From (1) and (2) of Lemma 2.1, if F_i is semismooth at x^*, then for any $W^i \in \partial_b F_i(x)$, $x \to x^*$

$$\|F_i(x) - F_i(x^*) - W^i(x - x^*)\| = o(\|x - x^*\|).$$

The semismoothness of F_i is obvious. Therefore, for any $W \in \partial_b F(x)$, $x \to x^*$, we have

$$\|F(x) - F(x^*) - W(x - x^*)\| = o(\|x - x^*\|).$$

Then we can choose positive constant ε_2 such that for any $y \in N_2(x^*) = \{x \mid \|x - x^*\| \le \varepsilon_2\}$, $W \in \partial_b F(y)$, we have

$$\|F(y) - F(x^*) - W(y - x^*)\| \le 2\delta\|y - x^*\|. \tag{4.9}$$

Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ and $N(x^*) = N_1(x^*) \cap N_2(x^*)$. Then (4.5), (4.8) and (4.9) hold for any $y \in N(x^*)$, $W \in \partial_b F(y)$. Denote $e^k = x^k - x^*$.

The local Q-linear convergence proof consists of showing by induction that

$$\|A_k - f'(x^*)\| \le (2 - 2^{-k})\delta, \tag{4.10}$$

$$\|V_k^{-1}\| \le \frac{3}{2}\beta, \tag{4.11}$$

$$\|e^{k+1}\| \le \frac{1}{2}\|e^k\|, \tag{4.12}$$

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

For $k = 0$, (4.10) is trivially true. The proof of (4.11) and (4.12) is identical to the proof at the induction step, so we omit it here.

Now assume that (4.10), (4.11) and (4.12) hold for $k = 0, \ldots, i-1$. For $k = i$, we have from Lemma 8.2.1 of [6] (also see [5]), and the induction hypothesis that

$$\begin{aligned}\|A_i - f'(x^*)\| &\le \|A_{i-1} - f'(x^*)\| + \frac{\gamma}{2}(\|e^i\| + \|e^{i-1}\|)\\ &\le (2 - 2^{-(i-1)})\delta + \frac{3\gamma}{4}\|e^{i-1}\|.\end{aligned} \tag{4.13}$$

From (4.12) and $\|e^0\| \le \varepsilon$ we get

$$\|e^{i-1}\| \le 2^{-(i-1)}\|e^0\| \le 2^{-(i-1)}\varepsilon.$$

Substituting this into (4.13) and using (4.7) gives

$$\begin{aligned}\|A_i - f'(x^*)\| &\le (2 - 2^{-(i-1)})\delta + \frac{3\gamma}{4}\varepsilon \cdot 2^{-(i-1)} \\ &\le (2 - 2^{-(i-1)})\delta + \frac{3\gamma}{4}\varepsilon_1 \cdot 2^{-(i-1)} \\ &\le (2 - 2^{-(i-1)} + 2^{-i})\delta = (2 - 2^{-i})\delta,\end{aligned}$$

which verifies (4.10).

To verify (4.11), we must first show that V_i is invertible. From the definition of $F^i(x)$ and (4.4) the jth row V_i^j of V_i satisfies

$$V_i^j = \begin{cases} I^j & \text{if } x_j^i - f_j^i(x^i) < l_j \text{ (or } > u_j), \\ \lambda_j^i I^j + (1-\lambda_j^i)A_i^j & \text{if } x_j^i - f_j^i(x^i) = l_j \text{ (or } = u_j), \\ A_i^j & \text{if } l_j < x_j^i - f_j^i(x^i) < u_j, \end{cases} \tag{4.14}$$

where $\lambda_j^i \in \{0, 1\}$. For such constants λ_j^i we define a companion matrix W_i such that the jth row W_i^j of W_i satisfies

$$W_i^j = \begin{cases} I^j & \text{if } x_j^i - f_j^i(x^i) < l_j \text{ (or } > u_j), \\ \lambda_j^i I^j + (1-\lambda_j^i)f_j'(x^i) & \text{if } x_j^i - f_j^i(x^i) = l_j \text{ (or } = u_j), \\ f_j'(x^i) & \text{if } l_j < x_j^i - f_j^i(x^i) < u_j. \end{cases} \tag{4.15}$$

From $f(x^i) = f^i(x^i)$ and (4.15) we get

$$W_i \in \partial_b F(x^i). \tag{4.16}$$

From (4.8) we get

$$\|W_i^{-1}\| \le \frac{10\beta}{9}. \tag{4.17}$$

From (4.14) and (4.15) for any $x \in R^n$ we get

$$|(W_i^j - V_i^j)x| \le |(A_i^j - f_j'(x^i))x|.$$

Therefore,

$$\begin{aligned}\|W_i - V_i\| &\le \|A_i - f'(x^i)\| \\ &\le \|A_i - f'(x^*)\| + \|f'(x^i) - f'(x^*)\|.\end{aligned} \tag{4.18}$$

Using (4.10) for $k = i$ and the Lipschitz condition (4.5) gives

$$\|W_i - V_i\| \leq (2 - 2^{-i})\delta + \gamma\|x^i - x^*\|$$

$$= (2 - 2^{-i})\delta + \gamma\|e^i\|. \tag{4.19}$$

From (4.12), $\|e^0\| \leq \varepsilon$, and (4.7)

$$\gamma\|e^i\| \leq 2^{-i}\varepsilon\gamma \leq 2^{-i}\varepsilon_1\gamma \leq \frac{2}{3} \cdot 2^{-i}\delta,$$

which, substituted into (4.19), gives

$$\|W_i - V_i\| \leq (2 - 2^{-i})\delta + \frac{2}{3} \cdot 2^{-i}\delta < 2\delta. \tag{4.20}$$

From (4.17), (4.20) and (4.6) we get

$$\|W_i^{-1}(W_i - V_i)\| \leq \frac{10\beta}{9} \cdot 2\delta \leq \frac{20}{9} \cdot \frac{1}{12} = \frac{5}{27} < 1, \tag{4.21}$$

so we have from Theorem 2.3.2 of [14] that V_i is invertible and

$$\|V_i^{-1}\| \leq \frac{\|W_i^{-1}\|}{1 - \|W_i^{-1}(W_i - V_i)\|} \leq \frac{\frac{10}{9}\beta}{1 - \frac{5}{27}} < \frac{3}{2}\beta,$$

which verifies (4.11).

To complete the induction, we verify (4.12). From $F(x^i) + V_i(x^{i+1} - x^i) = 0$ we have

$$F(x^i) + V_i(x^{i+1} - x^* + x^* - x^i) = 0,$$

$$V_i e^{i+1} = -F(x^i) + V_i e^i$$

$$= F(x^*) - F(x^i) + V_i e^i.$$

Therefore,

$$\|e^{i+1}\| \leq \|V_i^{-1}\|\|F(x^i) - F(x^*) - V_i e^i\|$$

$$\leq \|V_i^{-1}\|[\|F(x^i) - F(x^*) - W_i e^i\| + \|W_i - V_i\|\|e^i\|]. \tag{4.22}$$

From (4.9), (4.11), (4.6), (4.20) and (4.22) we get

$$\|e^{i+1}\| \leq \frac{3\beta}{2}[2\delta\|e^i\| + 2\delta\|e^i\|] = 6\beta\delta\|e^i\| \leq \frac{\|e^i\|}{2}.$$

This proves (4.12) and completes the proof of Q-linear convergence.

Next, we will prove the Q-superlinear convergence of $\{x^k\}$ under the assumptions. Let $E_k = A_k - f'(x^*)$. From the last part of the proof of Theorem 8.2.2 of [6] (also see we get

$$\lim_{k\to\infty} \frac{\|E_k s^k\|}{\|s^k\|} = 0. \tag{4.23}$$

From $F(x^k) + V_k(x^{k+1} - x^k) = 0$ we have

$$F(x^k) + W_k(x^{k+1} - x^k) + (V_k - W_k)(x^{k+1} - x^k) = 0,$$

$$W_k e^{k+1} = (W_k - V_k)(x^{k+1} - x^k) - [F(x^k) - F(x^*) - W_k e^k].$$

Therefore,

$$\|e^{k+1}\| \le \|W_k^{-1}\| \{\|F(x^k) - F(x^*) - W_k e^k\| + \|(V_k - W_k)s^k\|\}. \tag{4.24}$$

From (4.14), (4.15) and (4.5) we get

$$\begin{aligned}\|(V_k - W_k)s^k\| &\le \|(A_k - f'(x^k))s^k\| \\ &\le \|(A_k - f'(x^*))s^k\| + \|(f'(x^k) - f'(x^*))s^k\| \\ &\le \|E_k s^k\| + \gamma\|e^k\|\|s^k\|.\end{aligned}$$

Substituting this, and (4.17) into (4.24) gives

$$\|e^{k+1}\| \le \frac{10\beta}{9}\{\|F(x^k) - F(x^*) - W_k e^k\| + \|E_k s^k\| + \gamma\|e^k\|\|s^k\|\}. \tag{4.25}$$

From (4.12) and (4.25) we get

$$\begin{aligned}\frac{\|e^{k+1}\|}{\|e^k\|} &\le \frac{10}{9}\beta\Big\{\frac{\|F(x^k) - F(x^*) - W_k e^k\|}{\|e^k\|} + \frac{\|E_k s^k\|}{\|s^k\|}\frac{\|s^k\|}{\|e^k\|} + \gamma\|s^k\|\Big\} \\ &\le \frac{10}{9}\beta\Big\{\frac{\|F(x^k) - F(x^*) - W_k e^k\|}{\|e^k\|} + \frac{3}{2}\frac{\|E_k s^k\|}{\|s^k\|} + \frac{3}{2}\gamma\|e^k\|\Big\}.\end{aligned} \tag{4.26}$$

From Lemma 2.1 and the Q-linear convergence of $\{x^k\}$ we have

$$\lim_{k\to\infty} \|e^k\| = 0, \tag{4.27}$$

$$\lim_{k\to\infty} \frac{\|F(x^k) - F(x^*) - W_k e^k\|}{\|e^k\|} = 0. \tag{4.28}$$

Substituting (4.23), (4.27) and (4.28) into (4.26) gives

$$\lim_{k\to\infty} \frac{\|e^{k+1}\|}{\|e^k\|} = 0, \tag{4.29}$$

which completes the proof of Q-superlinear convergence. □

Remark. For nonlinear complementarity problem, the nonsingularity assumption of $\partial F(x^*)$ is equivalent to the b-regularity assumption in [18].

For general nonsmooth equations, we will consider the following method's convergence

$$x^{k+1} = x^k - A_k^{-1}F(x^k),\ A_k \in R^{n\times n},\ k = 0, 1, \dots. \tag{4.30}$$

THEOREM **4.2.** *Suppose that $F : R^n \to R^n$ is a locally Lipschitzian function in the open convex set $D \subset R^n$ and $x^* \in D$ is a solution of $F(x) = 0$. Suppose that F is semismooth at x^* and all $W_* \in \partial_b F(x^*)$ are nonsingular. There exists positive constants ε, δ such that if $x^0 \in D$, $\|x^0 - x^*\| \le \varepsilon$ and there exists $W_k \in \partial_b F(x^k)$ such that*

$$\|A_k - W_k\| \le \delta, \tag{4.31}$$

then the sequence of points generated by (4.30) is well defined and converges to x^* Q-linearly in a neighborhood of x^*.

Proof. From the proof of Lemma 2.2, Theorems 3.1 and 4.1 we can obtain the result of this theorem without difficulty. The detail is omitted here. □

In [19], Pang and Qi extended Theorem 2.2 in Dennis and Moré [5] to nonsmooth equations. Here, we can also do a similar extension and point out that some algorithms can be cast in our frame form.

THEOREM **4.3.** *Suppose that $F : R^n \to R^n$ is a locally Lipschitzian function in the open convex set $D \subset R^n$. Assume that F is semismooth at some $x^* \in D$ and all $W_* \in \partial_b F(x^*)$ are nonsingular. Let $\{A_k\}$ be a sequence of nonsingular matrices in $R^{n \times n}$, and suppose for some x^0 in D that the sequence of points generated by (4.30) remains in D, and satisfies $x^k \ne x^*$ for all k, and $\lim_{k\to\infty} x^k = x^*$. Then $\{x^k\}$ converges Q-superlinearly to x^*, and $F(x^*) = 0$ if and only if there exists $W_k \in \partial_b F(x^k)$ such that*

$$\lim_{k\to\infty} \frac{\|(A_k - W_k)s^k\|}{\|s^k\|} = 0, \tag{4.32}$$

where $s^k = x^{k+1} - x^k$.

Proof. The proof of the theorem is similar to that of Theorem 2 in Pang and Qi [19]. If we also notice of Lemma 2.2, there is no difficulty. So we omit the detail here. □

The Q-superlinear convergence of our algorithm discussed in this section is an application of Theorem 4.3, but not a special case discussed in Pang and Qi [19]. Besides the algorithms discussed in this chapter, we will also give two examples to demonstrate the applications of Theorems 4.2 and 4.3. One example is dicussed by Ip and Kyparisis [9], the other is a new algorithm.

Example 1. In [9], Ip and Kyparisis discussed the local convergence of the following quasi-Newton method (Broyden's method [1])

$$x^{k+1} = x^k + s^k, \quad s^k = -A_k^{-1} F(x^k), \tag{4.33}$$

$$A_{k+1} = A_k + \frac{(t^k - A_k s^k){s^k}^T}{{s^k}^T s^k}, \quad t^k = F(x^{k+1}) - F(x^k)$$

for solving nonsmooth equations. The Q-superlinear convergence is established under the strong condition that F is strongly F-differentiable at the solution point x^*. Under their conditions, we can easily verify that (4.32) is satisfied (actually, in this case $\partial_b F(x^*) = \partial_B F(x^*) = \{F'(x^*)\}$). So Theorem 4.3 (in this case also Theorem 2 in Pang and Qi [19]) generalizes the result obtained by Ip and Kyparisis [9].

Example 2. Consider the following nonsmooth equations

$$F(x) = \min(f(x), g(x)) = 0, \tag{4.34}$$

where $f,\ g : R^n \to R^n$ are continuously differentiable and the "min" operator denotes the componentwise minimum of two vectors. Such a system arises from nonsmooth partial differentiable equations [3, 2, 14] and implicit complementarity problem (see, e.g., [15]).

Consider the following quasi-Newton method (Broyden's Case)

Given $x^0 \in R^n,\ A_0,\ B_0 \in R^{n\times n}$

Do for $k = 0, 1, \dots :$

Define

$$f^k(x) = f(x^k) + A_k(x - x^k)$$

$$g^k(x) = g(x^k) + B_k(x - x^k)$$

$$F^k(x) = \min(f^k(x), g^k(x))$$

Choose $V_k \in \partial_b F^k(x^k)$

Solve $V_k s^k + F(x^k) = 0$ for s^k

$$x^{k+1} = x^k + s^k$$

$$y^k = f(x^{k+1}) - f(x^k)$$

$$z^k = g(x^{k+1}) - g(x^k)$$

$$A_{k+1} = A_k + \frac{(y^k - A_k s^k){s^k}^T}{{s^k}^T s^k}$$

$$B_{k+1} = B_k + \frac{(z^k - B_k s^k){s^k}^T}{{s^k}^T s^k}.$$

The Q-superlinear convergence of the sequence of points generated by this algorithm can be obtained from Theorem 4.3 under the stated assumptions.

5. Implementation of the Quasi-Newton Method

The implementation of the quasi-Newton method discussed in §4 for solving equations (4.1) has no difference to the smooth case except for the implementation of the QR factorization of the iterate matrix V_k. The entire QR factorization of V_k costs $O(n^3)$ arithmetic operations. If we do this in every step, then the advantage of quasi-Newton method loses a lot. In this section, we will show how to update the QR factorization of V_k into the QR factorization of V_{k+1} at most in $O((I(k)+1)n^2)$ operations (see (5.8) for the definition of $I(k)$). For the simplicity, we will assume that $X = R^n_+$.

For a given vector $x \in R^n$, denote the index sets

$$\alpha(x) = \{i : x_i > f_i(x)\},$$

$$\beta(x) = \{i : x_i = f_i(x)\},$$

$$\gamma(x) = \{i : x_i < f_i(x)\}.$$

Suppose for each k that we choose $V_k \in \partial_b F^k(x^k)$ such that the ith row V_k^i of V_k satisfies

$$V_k^i = \begin{cases} A_k^i & \text{if } \ i \in \alpha(x^k), \\ I^i & \text{if } \ i \in \beta(x^k) \cup \gamma(x^k). \end{cases} \tag{5.1}$$

Denote a matrix $\overline{V}_k$ such that its ith row $\overline{V}_k^i$ satisfies

$$\overline{V}_k^i = \begin{cases} A_{k+1}^i & \text{if } \ i \in \alpha(x^k), \\ I^i & \text{if } \ i \in \beta(x^k) \cup \gamma(x^k). \end{cases} \tag{5.2}$$

From (5.1), (5.2) and (4.3) we get

$$\overline{V}_k = V_k + \frac{(\overline{y}^k - V_k s^k) s^{kT}}{s^{kT} s^k}, \tag{5.3}$$

where $\overline{y}^k$ satisfies

$$\overline{y}_i^k = \begin{cases} y_i^k & \text{if } \ i \in \alpha(x^k), \\ s_i^k & \text{if } \ i \in \beta(x^k) \cup \gamma(x^k). \end{cases} \tag{5.4}$$

It is well known that we can update the QR factorization of V_k into the QR factorization of $\overline{V}_k$ in $O(n^2)$ operations (see, e.g., [7, 8]).

The ith row V_{k+1}^i of V_{k+1} satisfies

$$V_{k+1}^i = \begin{cases} A_{k+1}^i & \text{if } \ i \in \alpha(x^{k+1}), \\ I^i & \text{if } \ i \in \beta(x^{k+1}) \cup \gamma(x^{k+1}). \end{cases} \tag{5.5}$$

Therefore,

$$V_{k+1} = \overline{V}_k + \Delta \overline{V}_k, \tag{5.6}$$

where $\Delta \overline{V}_k$ satisfies

$$\Delta \overline{V}_k^i = \begin{cases} 0 & \text{if } \ i \in \alpha(x^k) \cap \alpha(x^{k+1}), \\ 0 & \text{if } \ i \in \{\beta(x^k) \cup \gamma(x^k)\} \cap \{\beta(x^{k+1}) \cup \gamma(x^{k+1})\}, \\ V_{k+1}^i - \overline{V}_k^i & \text{otherwise.} \end{cases} \tag{5.7}$$

Denote

$$I(k) = n - (|\alpha(x^k) \cap \alpha(x^{k+1})| + |\{\beta(x^k) \cup \gamma(x^k)\} \cap \{\beta(x^{k+1}) \cup \gamma(x^{k+1})\}|). \qquad (5.8)$$

Since the number of the nonzero rows of $\Delta \overline{V}_k$ is at most $I(k)$, we can update the QR factorization of $\overline{V}_k$ into the QR factorization of V_{k+1} at most in $O(I(k)n^2)$ operations (see, e.g., [7, 8]).

Therefore, we get

THEOREM **5.1.** *The cost of updating the QR factorization of V_k into the QR factorization of V_{k+1} is at most $O((I(k)+1)n^2)$ arithmetic operations.*

Josephy [10] considered the quasi-Newton method for solving generalized equations (see Robinson [23]). For nonlinear complementarity problem, in every step his method needs to solve a linear complementarity problem, which requires more cost than solving a linear equations. Kojima and Shindo [11] extended the quasi-Newton method to piecewise smooth equations. They applied the classical Broyden's method as the points x^k stayed within a given C^1-piece. When the points x^k arrived a new piece, a new starting matrix was used and it was needed to perform the entire QR factorization (or other factorizations) in $O(n^3)$ operations in general. Thus a potentially large number of matrices need to be stored and to be performed entire QR factorization (or other factorizations). Here, our method needs only one approximate matrix, and except for the first step we only need less effort to solve a linear equations, which may be solved in much less than $O(n^3)$ operations. The smaller the measure of $I(k)$ is, the less computing effort is needed at $(k-1)$th step (note that $I(k)$ is related to the nonsmoothness of F). Ip and Kyparisis [?] discussed the local convergence of Broyden's method (4.33) for solving nonsmooth equations. Although the form of (4.33) is very simple, the convergence remains open without assuming the existence of $F'(x^*)$.

6. The KKT System of Variational Inequality Problem

For a given closed set $X \subseteq R^n$ and a mapping $f : X \to R^n$, the variational inequality problem which denoted by $\mathrm{VI}(X, f)$ is to find a vector $x^* \in X$ such that

$$(x - x^*)^T f(x^*) \geq 0, \quad \text{for all } x \in X.$$

If $X = R^n_+$, then $\mathrm{VI}(X, f)$ is equivalent to the complementarity problem which is to find $x^* \in R^n_+$ such that

$$f(x^*) \in R^n_+ \text{ and } {x^*}^T f(x^*) = 0.$$

When f is a gradient mapping, say $f(x) = \nabla\theta(x)$ for some real-valued function θ, $\mathrm{VI}(X, f)$ is equivalent to the problem of finding a stationary point for the minimization problem:

$$\text{minimize } \theta(x)$$

$$\text{subject to } x \in X.$$

Here we shall assume that X has the form

$$X = \{x \in R^n |\ g(x) \leq 0,\ h(x) = 0,\ l \leq x \leq u\}, \tag{6.1}$$

where $g : R^n \to R^m$ and $h : R^n \to R^p$ are assumed to be twice continuously differentiable, and $l, u \in \{R \cup \{\infty\}\}^n$. By introducing multipliers $(\lambda, \mu, v, w) \in R^{m+p+2n}$ corresponding to the constraints in X, the (VI) Lagrangian (vector-valued) function (see, e.g., Tobin [27]) can be defined by

$$L(x, \lambda, \mu, v, w) = f(x) + \sum_{i=1}^{m} \nabla g_i(x)\lambda_i + \sum_{j=1}^{p} \nabla h_j(x)\mu_j - v + w.$$

If $l_i = -\infty$ (or $u_i = +\infty$) for some i, the corresponding v_i (w_i respectively) is absent in the above formular. Then the KKT system of VI(X, f) can be written as

$$\begin{cases} L(x, \lambda, \mu, v, w) = 0, \\ \lambda \geq 0,\ -g(x) \geq 0,\ \text{and } \lambda^T g(x) = 0, \\ -h(x) = 0, \\ v \geq 0,\ x - l \geq 0,\ \text{and } v^T(x - l) = 0, \\ w \geq 0,\ u - x \geq 0,\ \text{and } w^T(x - u) = 0. \end{cases} \tag{6.2}$$

Define

$$\tilde{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \nabla g_i(x)\lambda_i + \sum_{j=1}^{p} \nabla h_j(x)\mu_j$$

and

$$H(x, \lambda, \mu) = \begin{pmatrix} x - P_{[l,u]}[x - \tilde{L}(x, \lambda, \mu)] \\ \lambda - P_{R^n_+}[\lambda - (-g(x))] \\ -h(x) \end{pmatrix}. \tag{6.3}$$

Suppose that $(x^*, \lambda^*, \mu^*, v^*, w^*) \in R^{n+m+p+2n}$ is a solution of the KKT system (6.2), then (x^*, λ^*, μ^*) satisfies $H(x^*, \lambda^*, \mu^*) = 0$; conversely if $(x^*, \lambda^*, \mu^*) \in R^{n+m+p}$ is a solution of $H(x, \lambda, \mu) = 0$, then $(x^*, \lambda^*, \mu^*, v^*, w^*)$ is a solution of the KKT system (6.2), where v^*, w^* are defined as

$$v^* = P_{R^n_+}[\tilde{L}(x^*, \lambda^*, \mu^*)] \text{ and } w^* = P_{R^n_+}[-\tilde{L}(x^*, \lambda^*, \mu^*)]. \tag{6.4}$$

So find a solution of the KKT system of VI is equivalent to solve $H(x, \lambda, \mu) = 0$. Let $z = (x, \lambda, \mu)$, $K = [l, u] \times R^n_+ \times R^p$, and

$$\tilde{f}(z) = \begin{pmatrix} \tilde{L}(z) \\ -g(x) \\ -h(x) \end{pmatrix}.$$

Then $H(x, \lambda, \mu) = 0$ can be written as

$$H(z) = z - P_K[z - \tilde{f}(z)] = 0, \tag{6.5}$$

which is a special form of (4.1).

Now suppose that z^* is a solution of $H(z) = 0$, and f is continuously differentiable at x^*, we will discuss a sufficient condition on the nonsingularity assumption of $\partial_b H(z^*)$. Let

$$I(z^*) = \{i|\ 1 \le i \le m,\ g_i(x^*) = 0\},$$
$$I^+(z^*) = \{i \in I(z^*)|\ \lambda_i^* > 0\},$$
$$G^+(z^*) = \{d \in R^n| \quad \nabla g_i(x^*)^T d = 0 \text{ for } i \in I^+(z^*)$$
$$\text{and } \nabla h_i(x^*)^T d = 0 \text{ for } i = 1, ..., p\},$$

and

$$R(z^*) = \{d \in R^n|\ d_i = 0 \text{ if } x_i^* = l_i \text{ (or } u_i) \text{ and } (\tilde{L}(z^*))_i \ne 0 \text{ for } i = 1, ..., n\}.$$

THEOREM **6.1.** *Suppose that z^* is a solution of $H(z) = 0$, and satisfies $d^T \nabla_{xx}^2 \tilde{L}(z^*) d > 0$ for all $d \in G^+(z^*) \cap R(z^*) \backslash \{0\}$. If $\{\nabla g_i(x^*),\ i \in I(z^*)\}$ and $\{\nabla h_i(x^*),\ i = 1, ..., p\}$ are linearly independent, then all $V \in \partial_b H(z^*)$ are nonsingular.*

Proof. Combining (4.4) and the proof of Theorem 4.1 in Robinson [23], we can get the result. □

7. Numerical Examples

In this section, we report computational results obtained for two small nonlinear complementarity problems using the above Newton method and quasi-Newton method. For quasi-Newton method, the initial matrices are generated by the difference approximation method. In Table 1, "N" and "QN" represent Newton method and quasi-Newton method, respectively; and "P 1" and "P 2" represent Problem 1 and Problem 2, respectively.

Problem 1 (A Nondegenerate Nonlinear Complementarity Problem, [10, 9]). Consider the following problem: find $x \in R^4$ such that $x \ge 0,\ f(x) \ge 0$, and $x^T f(x) = 0$, where $f : R^4 \to R^4$ is given by

$$f_1(x) = 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6,$$

$$f_2(x) = 2x_1^2 + x_1 + x_2^2 + 3x_3 + 2x_4 - 2,$$

$$f_3(x) = 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1,$$

$$f_4(x) = x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3.$$

This problem has a solution

$$x^* = (\frac{1}{2}\sqrt{6} \approx 1.2247, 0, 0, 0.5), \quad f(x^*) = (0, 2 + \frac{1}{2}\sqrt{6} \approx 3.2247, 5, 0).$$

Since $\beta(x^*) = \emptyset$, x^* is nondegenerate (see [9]) and it is easy to check that $F'(x^*)$ (here $\partial_b F'(x^*) = \{F'(x^*)\}$) is nonsingular.

Problem 2 (A Degenerate Nonlinear Complementarity Problem, [11, 9]). Consider the following problem: find $x \in R^4$ such that $x \geq 0$, $f(x) \geq 0$, and $x^T f(x) = 0$, where $f : R^4 \to R^4$ is given by

$$f_1(x) = 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6,$$

$$f_2(x) = 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2,$$

$$f_3(x) = 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9,$$

$$f_4(x) = x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3.$$

This problem has two solutions

$$x_D^* = (\frac{1}{2}\sqrt{6} \approx 1.2247, 0, 0, 0.5), \quad f(x_D^*) = (0, 2 + \frac{1}{2}\sqrt{6} \approx 3.2247, 0, 0),$$

and

$$x_{ND}^* = (1, 0, 3, 0), \quad f(x_{ND}^*) = (0, 31, 0, 4).$$

Since $\beta(x_{ND}^*) = \emptyset$ for the solution x_{ND}^*, it is a nondegerate solution (see [9]). On the other hand, $\beta(x_D^*) = \{3\}$ for the solution x_D^*, so it is a degenerate solution (see [9]). It is easy to check that $\partial_b F(x_{ND}^*)$ and $\partial_b F(x_D^*)$ are nonsingular.

TABLE 1
Results for problems 1 and 2

Algorithm	Starting point	Number of Iterations		sum of $I(k)$	
		P 1	P 2	P 1	P 2
N	(1,0,0,0)	3	3(D)		
QN	(1,0,0,0)	4	4(D)	0	2
N	(1,0,1,0)	4	1(ND)		
QN	(1,0,1,0)	5	1(ND)	1	0
N	(1,0,0,1)	4	4(D)		
QN	(1,0,0,1)	5	5(D)	1	2
N	(1,0.2,0.5,1)	4	4(D)		
QN	(1,0.2,0.5,1)	6	6(D)	0	2
N	(1,0,1,-1)	3	3(D)		
QN	(1,0,1,-1)	5	5(D)	1	2
N	(1.5,-0.5,4.5,-1.0)	4	4(D)		
QN	(1.5,-0.5,4.5,-1.0)	6	6(D)	1	0
N	(1.1,-0.1,3.1,-0.1)	4	3(ND)		
QN	(1.1,-0.1,3.1,-0.1)	5	4(ND)	1	0
N	(0.85,0.2,0.5,1)	4	5(D)		
QN	(0.85,0.2,0.5,1)	7	7(D)	1	2

D=degenerate solution, ND=nondegenerate solution.

From Table 1 we see that even for problem 2 when the starting point is close to a solution, the sequence will converge to the corresponding solution no matter wether it is degenerate or not.

In this chapter two small examples are used to show the effectiveness of the Newton method and the quasi-Newton method for solving some nonsmooth equations. More examples are needed to show the efficiency of the above algorithms. For problem (4.1) with a general convex set X, especially when X is a polyhedral set, how to construct appropriate Newton methods and quasi-Newton methods is our further research topic.

REFERENCES

[1] C.G. BROYDEN, *A class of methods for solving nonlinear simultaneous equations*, Math. Comp., 19 (1965), pp. 577-593.

[2] X. CHEN, AND T. YAMAMOTO, *On the convergence of some quasi-Newton methods for solving nonlinear equations with nondifferentiable operators*, Computing, 49 (1992), pp. 87-94.

[3] X. CHEN, AND L. QI, *A parameterized Newton method and a quasi-Newton method for solving nonsmooth equations*, Comp. Opti. Appl., 3 (1994), pp. 157-179.

[4] F.H. CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

[5] J.E. DENNIS, AND J.J. MORÉ, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp., 28 (1974), pp. 549-560.

[6] J.E. DENNIS, AND R.B. SCHNABLE, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-hall, Englewood Cliffs, N.J., 1983.

[7] P.E. GILL, AND and M. MURRAY, *Quasi-Newton methods for unconstrained optimization*, J. Inst. Math. Appl., 9 (1972), pp. 91-108.

[8] G.H. GOLUB, AND C. VAN LOAN, *Matrix Computations*, the Johns Hopkins University Press, 1983.

[9] C.-M. IP, AND T. KYPARISIS, *Local convergence of quasi-Newton methods for B-differentiable equations*, Math. Programming, 56 (1992), pp. 71-89.

[10] N.H. JOSEPHY, *Quasi-Newton methods for generalized equations*, Technical Summary Report No. 1966, Mathematical Research Center, University of Wisconsin, Madison, WI, 1979.

[11] M. KOJIMA, AND S. SHINDO, *Extensions of Newton and quasi-Newton methods to systems of PC^1 equations*, J. Oper. Res. Soc. Japan, 29 (1986), pp. 352-374.

[12] B. KUMMER, *Newton's method for non-differentiable functions*, in J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach and K. Zimmerman, eds., *Advances in Mathematical Programming*, Academi Verlag, Berlin, 1988, pp. 114-125.

[13] M. MIFFLIN, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 957-972.

[14] J.M. ORTEGA, AND W.C. RHEINBOLDT, *Iterative solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[15] J.-S. PANG, *The implicit complementarity problem*, in O.L. Mangasarian, S.M. Robinson, and P.R. Meyer, eds., *Nonlinear Programming* 4, Academic Press, New York,

1981, pp. 487-518.

[16] ———, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311-341.

[17] ———, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Programming, 51 (1991), pp. 101-131.

[18] J.-S. PANG, AND S.A. GABRIEL, *NE/SQP: A robust algorithm for the nonlinear complementarity problem*, Math. Programming, 60 (1993), pp. 295-337.

[19] J.-S. PANG, AND L. QI, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optimization, 3 (1993), pp. 443-465.

[20] L. QI, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227-244.

[21] L. QI, AND J. SUN, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353-368.

[22] ———, *A nonsmooth version of Newton's method and an interior point algorithm for convex programming*, Applied Mathematics Preprint 89/33, School of Mathematics, The University of New South Wales, Sydney, Australia, Revised in 1991.

[23] S.M. ROBINSON, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43-62.

[24] ———, *Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity*, Math. Programming Study, 30 (1987), pp. 45-66.

[25] ———, *Newton's methods for a class of nonsmooth functions*, Industrial Engineering Working Paper, University of Wisconsin, Madison, USA, 1988.

[26] N.Z. SHOR, *A class of almost-differentiable functions and a minimization method for functions of this class*, Kibernetic, 4 (1972), pp. 65-70.

[27] R.L.TOBIN, *Sensitivity analysis for variational inequalities*, J. Optim. Theory Appl., 48 (1986), pp.191

Chapter 4

Superlinear Convergence of Approximate Newton Methods for LC^1 Optimization Problems without Strict Complementarity

Abstract

In this chapter, the Q-superlinear convergence property of the approximate Newton or SQP methods for solving LC^1 optimization problems is established under the assumptions that the derivatives of the objective and constraint functions are semismooth, the strong second-order sufficiency condition is satisfied and the gradients to the active constraints are linearly independent. The strong second-order sufficiency condition is weaker than the second-order sufficiency condition and the strict complementarity condition.

Chapter 4
Superlinear Convergence of Approximate Newton Methods for LC^1 Optimization Problems without Strict Complementarity

1. Introduction

Consider the standard nonlinear programming

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \\ & h(x) = 0, \end{array} \tag{1.1}$$

where f, g and h are differentiable functions from R^n into R, R^p and R^q respectively. One method for solving (1.1) is to solve the following linearly constrained quadratic program Q_k

$$\begin{array}{ll} \text{minimize} & \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T G_k (x - x^k) \\ \text{subject to} & g(x^k) + \nabla g(x^k)^T(x - x^k) \leq 0, \\ & h(x^k) + \nabla h(x^k)^T(x - x^k) = 0 \end{array} \tag{1.2}$$

successively. Here G_k is an $n \times n$ matrix. This method is called an approximate Newton method or a SQP (sequential quadratic programming) method. If G_k is exactly the second-order derivative of the Lagrangian at x^k, this is Wilson's method. See Garcia Palomares and Mangasarian (Ref. 4) and Robinson (Refs. 21-22).

Before the advent of the very recent chapter by Qi (Ref. 19), the proof of the superlinear convergence of such approximate Newton or SQP methods for solving nonlinear programming problems requires twice smoothness of the objective and constrained functions. Sometimes, the second-order derivatives of those functions are required to be Lipschitzian, for example, see Garcia Palomares and Mangasarian (Ref. 4), Han (Ref. 5), McCormick (Ref. 9) and Robinson (Refs. 21-22). However, the second-order differentiability may not hold for some problems. For example, the extended linear-quadratic programming problem, recently emerged in stochastic programming and optimal control, even in the fully quadratic case, does not possess twice differentiable objective functions. However, their objective functions are differentiable and their derivatives are Lipschitzian in that case. See Rockafellar (Ref. 24) or Rockafellar and Wets (Ref. 25) for a detail. We call a function $F : R^n \to R^m$ a LC^1 function, if it is differentiable and its derivative function is locally Lipschitzian. We call a nonlinear programming problem a LC^1 optimization problem if its objective and constrained functions are LC^1 functions. For the detail of LC^1 functions and LC^1 optimization problems, see Qi (Ref. 17). In Qi (Ref. 19), the Q-superlinear convergence of the approximate Newton or SQP methods for solving LC^1 optimization problems was established under the assumption that the derivatives of the objective and constrained functions are semismooth and the three key assumptions that the second-order sufficiency condition, the strict complementarity slackness and linear independence of the gradients to the active constraints are satisfied under the context of LC^1 optimization problems. Basing on generalized equations' theory established by Robinson (Ref. 23), Josephy (Refs. 7-8) provided a proof to the local superlinear (quadratic) convergence of quasi-Newton (Newton) methods without assuming the strict complementarity slackness condition when the second-order differentiability is available. Also basing on Robinson's generalized equations' theory (Ref. 23), without assuming the strict complementarity condition Lescrenier (Ref. 29) provided a proof to the convergence of a class of trust region methods proposed by Conn, Gould, and Toint (Ref. 30) for optimization problem with simple bounds constraints when the objective function is twice continuously differentiable. In this chapter, we will discuss the superlinear convergence of approximate Newton or SQP methods for solving LC^1 optimization problems without assuming the existence of the second-order differentiability and the strict complementarity slackness condition.

In a certain sense, our results in this chapter are the LC^1 version of the results in Josephy (Refs. 7-8) or a generalization of the results in Qi (Ref. 19) without the strict complementarity slackness. To achieve this, our technique is different from that of Josephy (Refs. 7-8) or Qi (Ref. 19). First we consider the superlinear convergence of a generalized approximate Newton type method for solving nonsmooth equations, recently developed in Pang (Ref. 14) and Qi (Refs. 16-17). Then, we prove that the approximate Newton or SQP methods are special cases of such generalized approximate Newton method.

In section 2, we discuss the strong second-order sufficiency condition and linear independence under the context of LC^1 optimization. The Q-superlinear convergence of approximate Newton or SQP methods for LC^1 optimization is established in section 3. In section 4, we give some discussions.

2. The Strong Second-Order Sufficiency Condition

Throughout this chapter, we assume that f, g and h in (1.1) are LC^1 functions.

The Lagrangian of (1.1) is $L(x,u,v) = f(x) + u^T g(x) + v^T h(x)$. Denote the gradient of L with respect to x by $F_{u,v}$. Then

$$F_{u,v}(x) = \nabla f(x) + \nabla g(x)u + \nabla h(x)v$$

is a locally Lipschitzian function.

In Josephy (Refs. 7-8) or Robinson (Ref. 23), the two key assumptions other than second-order differentiability are the strong second-order sufficiency condition and linear independence of the gradients to the active constraints. We still need these two assumptions. However the strong second-order sufficiency condition needs to be modified because we will not assume the second-order differentiability of f, g and h.

In general, assume that $F: R^n \to R^m$ is locally Lipschitzian. By Rademacher's Theorem, F is differentiable almost everywhere. Let D_F be the set where F is differentiable. Let ∂F be the generalized Jacobian of F in the sense of Clarke (Ref. 2). Then

$$\partial F(x) = \mathrm{co}\{ \lim_{\substack{x^k \in D_F \\ x^k \to x}} F'(x^k)\}, \tag{2.1}$$

where co$\{A\}$ is a convex hull of a set A.

In Qi (Ref. 16) and Pang and Qi (Ref. 15), the concept $\partial_B F(x)$ was introduced

$$\partial_B F(x) = \{ \lim_{\substack{x^k \in D_F \\ x^k \to x}} F'(x^k) \}.$$

Then

$$\partial F(x) = \mathrm{co}\,\partial_B F(x).$$

For $m = 1$, $\partial_B F(x)$ was introduced by Shor (Ref. 26). Let F_i denote the ith component of F. Sun and Han (Ref. 27) introduced

$$\partial_b F(x) = \partial_B F_1(x) \times \partial_B F_2(x) \times \cdots \times \partial_B F_m(x).$$

Then $\partial_B F(x) \subseteq \partial_b F(x)$ and the converse relation does not hold in general. For example if $F : R^1 \to R^2$ has the form

$$F(x) = \begin{pmatrix} \min(x, x^2) \\ \min(-x, -x^2) \end{pmatrix},$$

then

$$\partial_B F(0) = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}, \ \partial_b F(0) = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\},$$

and $\partial_B F(0) \subset \partial_b F(0)$. But when $m = 1$, $\partial_b F(x) = \partial_B F(x)$.

From the results of Clarke (Ref. 2), Qi (Ref. 16), and Sun and Han (Ref. 27) we know that $\partial F(x)$, $\partial_B F(x)$ and $\partial_b F(x)$ are nonempty compact subsets of $R^{m\times n}$, and the maps ∂_F, $\partial_B F$ and $\partial_b F$ are upper semi-continuous (Ref. 1). In fact if we note that $\partial F(x)$ and $\partial_i F(x)$ are compact subsets, and that the maps ∂F and $\partial_i F$ are upper semi-continuous (Ref. 2), we can draw the same conclusions for the maps $\partial_B F$ and $\partial_b F$ through the standard analysis. In this chapter we use $\mathcal{M}(x, F)$ to represent one of $\partial F(x)$, $\partial_B F(x)$ and $\partial_b F(x)$ and use the multifunction $\mathcal{M}(\cdot, F)$ to represent one of ∂F, $\partial_B F$ and $\partial_b F$. Therefore, $\mathcal{M}(x, F)$ is a nonempty compact subset of $R^{m\times n}$, and the map $\mathcal{M}(\cdot, F)$ is upper semi-continuous.

Suppose that $f_1, f_2 : R^n \to R^1$ are continuously differentiable functions. Let $f_0(x) = \min(f_1(x), f_2(x))$, then

$$\partial_b f_0(x) = \begin{cases} \{\nabla f_1(x)^T\} & \text{if } f_1(x) < f_2(x), \\ \{\nabla f_1(x)^T, \nabla f_2(x)^T\} & \text{if } f_1(x) = f_2(x), \\ \{\nabla f_2(x)^T\} & \text{if } f_1(x) > f_2(x). \end{cases}$$

This formulae will be used later in this chapter.

The first-order Kuhn-Tucker conditions for (1.1) are

$$\begin{aligned} F_{u,v}(x) = \nabla f(x) + \nabla g(x) u + \nabla h(x) v = 0, \\ u \geq 0, \quad g(x) \leq 0, \\ u_i g_i(x) = 0, \quad \text{for } i = 1, \ldots, p, \\ h(x) = 0. \end{aligned} \tag{2.2}$$

Let

$$H(z) = \begin{pmatrix} \nabla f(x) + \nabla g(x)u + \nabla h(x)v \\ \min(u, -g(x)) \\ -h(x) \end{pmatrix}, \tag{2.3}$$

where the 'min' operator denotes the componentwise minimum. Then the first-order Kuhn-Tucker conditions are equivalent to $H(z) = 0$. Denote $H_1(z) = \nabla f(x) + \nabla g(x)u + \nabla h(x)v$, $H_2(z) = \min(u, -g(x))$ and $H_3(z) = -h(x)$. Then

$$H(z) = \begin{pmatrix} H_1(z) \\ H_2(z) \\ H_3(z) \end{pmatrix}.$$

For every $z = (x, u, v) \in R^n \times R^p \times R^q$, denote

$$\partial_Q H(z) = \mathcal{M}(z, H_1) \times \partial_b H_2(z) \times \{\nabla H_3(z)^T\}.$$

It is easy to see that $\partial_Q H(z)$ is a nonempty compact subset of $R^{m\times m}$, and the map $\partial_Q H$ is upper semi-continuous, where $m = n + p + q$. For any $A \in \mathcal{M}(z, H_1)$, there exists $V \in R^{n\times n}$ such that $A = (V \;\nabla g(x)\; \nabla h(x))$. Denote

$$\mathcal{V}_x(z) = \{V \in R^{n\times n} |\; (V \;\nabla g(x)\; \nabla h(x)) \in \mathcal{M}(z, H_1)\}.$$

From the definition of the map $\mathcal{M}(\cdot,\cdot)$, it is easy to see that for any $z = (x, u, v) \in R^n \times R^p \times R^q$, we have

$$\mathcal{M}(x, F_{u,v}) \subseteq \mathcal{V}_x(z).$$

Suppose that $z = (x, u, v) \in R^n \times R^p \times R^q$ is a Kuhn-Tucker point of (1.1). Let

$$I(z) = \{i |\; 1 \le i \le p,\; g_i(x) = 0\},$$

$$I^+(z) = \{i \in I(z) |\; u_i > 0\},$$

$$I^0(z) = \{i \in I(z) |\; u_i = 0\},$$

$$G(z) = \{d \in R^n |\; f'(x; d) = 0,\; g_i'(x; d) = 0 \text{ for } i \in I^+(z),\; g_i'(x; d) \le 0 \text{ for } i \in I^0(z) \text{ and } h_i'(x; d) = 0 \text{ for } i = 1, ..., q\}$$

and

$$G^+(z) = \{d \in R^n |\; f'(x; d) = 0,\; g_i'(x; d) = 0 \text{ for } i \in I^+(z) \text{ and } h_i'(x; d) = 0 \text{ for } i = 1, ..., q\}.$$

A point $z = (x, u, v) \in R^n \times R^p \times R^q$ is said to satisfy the second-order sufficiency conditions (strong second-order sufficiency conditions) for (1.1) if it satisfies the first-order Kuhn-Tucker conditions and if $d^T V d > 0$ for all $d \in G(z)\backslash 0$ ($d \in G^+(z)\backslash 0$), $V \in \mathcal{V}_x(z)$.

Suppose that $z = (x, u, v) \in R^n \times R^p \times R^q$ is a Kuhn-Tucker point of (1.1). We say that z satisfies the linear independence condition if $\{\nabla g_i(x),\; i \in I(z)\}$ and $\{\nabla h_i(x),\; i = 1, ..., q\}$ are linearly independent. We say that z satisfies the strict complementarity slackness condition if $I^0(z) = \emptyset$. When the strict complementarity condition is satisfied (i.e., $I^0(z) = \emptyset$), then $G(z) = G^+(z)$. Therefore, second-order sufficiency conditions and the strict complementarity slackness condition mean strong second-order sufficiency conditions. In general, strong second-order sufficiency conditions mean the second-order sufficiency conditions, but don't mean the strict complementarity slackness condition. The strict complementarity slackness condition may not hold in nonlinear optimization problems. Therefore, we will consider the superlinear convergence properties of approximate Newton or SQP methods for LC^1 optimization problems without assuming the strict complementarity condition.

First, we shall consider the nonsingularity of matrices $W \in \partial_Q H(z)$ at a solution of $H(z) = 0$. If the components of such a solution are denoted by x_0, u_0, v_0, we can partition the vector $g(x_0)$ into smaller vectors $g^+(x_0)$, $g^0(x_0)$ and $g^-(x_0)$, of dimensions r, s and t, respectively, and partition u_0 conformably into u_0^+, u_0^0 and u_0^- so that

$$\begin{array}{ll} g^+(x_0) = 0, & u_0^+ > 0, \\ g^0(x_0) = 0, & u_0^0 = 0, \\ g^-(x_0) < 0, & u_0^- = 0, \end{array} \tag{2.4}$$

where the ordering is componentwise. After suitable arrangement, (2.3) can be written as

$$H\begin{pmatrix} x \\ u^+ \\ u^0 \\ u^- \\ v \end{pmatrix} = \begin{pmatrix} \nabla f(x) + \nabla g(x)u + \nabla h(x)v \\ \min(u^+, -g^+(x)) \\ \min(u^0, -g^0(x)) \\ \min(u^-, -g^-(x)) \\ -h(x) \end{pmatrix}. \tag{2.5}$$

Theorem 2.1. Suppose that $z_0 = (x_0, u_0, v_0) \in R^n \times R^p \times R^q$ satisfies the strong second-order sufficiency conditions and the linear independence condition of (1.1). Then all $W \in \partial_Q H(z_0)$ are nonsingular.

Proof. According to the definition of $\partial_Q H(z_0)$, we only need to prove for $i = 0, 1, ..., s$, the nonsingularity of the following matrices

$$W_{(i)} = \begin{pmatrix} V & G_0^{+T} & G_0^{0IT} & G_0^{0JT} & G_0^{-T} & H_0^T \\ -G_0^+ & 0 & 0 & 0 & 0 & 0 \\ -G_0^{0I} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{j\times j} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{t\times t} & 0 \\ -H_0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $V \in \mathcal{V}_{x_0}(z_0)$, H_0 denotes $\nabla h(x_0)^T$, G_0^+ denotes $\nabla g^+(x_0)^T$, etc, $I = \{1, ..., i\}$ (when $i = 0$, $I = \emptyset$), $J = \{1, ..., s\} \backslash I$, $j = |J|$, G_0^{0I} is a matrix of the I rows of G_0^0, G_0^{0J} is a matrix of the J rows of G_0^0, and $I_{j\times j}$ and $I_{t\times t}$ are the unit matrices of $R^{j\times j}$ and $R^{t\times t}$

respectively. Suppose that $a,\ b,\ c,\ d,\ e$ and l are such that

$$\begin{array}{llll} Va & + {G_0^+}^T b + {G_0^{0I}}^T c + {G_0^{0J}}^T d + {G_0^-}^T e + {H_0}^T l & = 0, & \\ -G_0^+ a & & = 0, & \\ -G_0^{0I} a & & = 0, & (2.6) \\ & I_{j\times j} d & = 0, & \\ & \qquad I_{t\times t} e & = 0, & \\ -H_0 a & & = 0. & \end{array}$$

Therefore, we get

$$\begin{array}{lll} Va & + {G_0^+}^T b + {G_0^{0I}}^T c + {H_0}^T l & = 0, \\ -G_0^+ a & & = 0, \\ -G_0^{0I} a & & = 0, \\ -H_0 a & & = 0. \end{array} \tag{2.7}$$

Premultiplying the equations in (2.7) by $a^T,\ b^T,\ c^T$ and l^T, respectively, and adding the result we find that $a^T V a = 0$. This, together with the second and fourth equations of (2.7) and the strong second-order sufficiency conditions, implies that $a = 0$; the first equation of (2.7) and the linear independence assumption now imply that $b,\ c$ and l are also zero. The fourth and fifth equations of (2.6) means that d and e are zero. Thus the matrix $W_{(i)}$ is nonsingular. This completes the proof. □

Corollary 2.1. Under the conditions of Theorem 2.1, there exist $\delta > 0$ and $C > 0$ such that for any $\hat{z} = (\hat{x}, \hat{u}, \hat{v}) \in R^n \times R^p \times R^q$, satisfying $\|\hat{z} - z_0\| \leq \delta$, and any $W \in \partial_Q H(\hat{z})$, W is invertible and $\|W^{-1}\| \leq C$.

Proof. Applying Theorem 2.1 of this chapter, and that $\partial_Q H(\hat{z})$ is a nonempty compact subset and the map $\partial_Q H$ is upper semi-continuous, we can easily obtain the conclusion. □

We say that a locally Lipschitzian function $F:\ R^n \to R^m$ is semismooth at x if

$$\lim_{\substack{V \in \partial F(x+th') \\ h' \to h,\ t \downarrow 0}} \{Vh'\} \tag{2.8}$$

exists for any $h \in R^n$. If F is semismooth at x, then F is directionally differentiable at x and $F'(x; h)$ is equal to the limit in (2.8). Semismoothness was first introduced by Mifflin (Ref. 10) for functional. Convex functions, continuously piecewise linear functions, smooth functions and subsmooth functions are examples of semismooth functions. Scalar products and sums of semismooth functions are also semismooth functions. In Qi (Ref. 16) and Qi and Sun (Ref. 18), the definition of semismoothness was extended to F : $R^n \to R^m$. It was proved in Qi (Ref. 17) that F is semismooth at x if and only if each of its components is semismooth at x.

3. Superlinear Convergence Property

To establish the superlinear convergence of approximate Newton or SQP methods, we need the following two properties of semismoothness:

Suppose that $F: R^n \to R^m$ is locally Lipschitzian and semismooth at x. Then (1) F is B-differentiable at x, i.e., $F'(x;h)$ exists for all $h \in R^n$, and

$$F(x+h) = F(x) + F'(x;h) + o(\|h\|), \tag{3.1}$$

(2) For any $V \in \partial F(x+h)$, $h \to 0$

$$Vh - F'(x;h) = o(\|h\|). \tag{3.2}$$

See Theorem 2.3 of Qi and Sun (Ref. 18).

The approximate Newton method (ANM) for solving (1.1) is as follows:

Start at a point $z^0 = (x^0, u^0, v^0) \in R^n \times R^p \times R^q$. Having $z^k = (x^k, u^k, v^k)$, find a Kuhn-Tucker point $z^{k+1} = (x^{k+1}, u^{k+1}, v^{k+1})$ of the quadratic subproblem Q_k described by (1.2). If z^{k+1} is not unique, choose any Kuhn-Tucker point z^{k+1} which is closest to z^k in terms of distance $\|z^{k+1} - z^k\|$.

Suppose that $z^* = (x^*, u^*, v^*) \in R^n \times R^p \times R^q$ is a solution of $H(z) = 0$ (i.e., z^* is a Kuhn-Tucker point of (1.1)). For every $z = (x, u, v) \in R^n \times R^p \times R^q$, denote

$$\alpha(z) = \{i \mid u_i > -g_i(x)\},\ \beta(z) = \{i \mid u_i = -g_i(x)\} \text{ and } \gamma(z) = \{i \mid u_i < -g_i(x)\}.$$

For $i \in I^\beta \equiv \{1, ..., 2^{|\beta(z^*)|}\}$, define

$$H^{(i)}(z) = \begin{pmatrix} \nabla f(x) + \nabla g(x)u + \nabla h(x)v \\ p^{(i)}(z) \\ -h(x) \end{pmatrix}, \tag{3.3}$$

where $p^{(i)}(z) \in P(z)$ and $P(z)$ consists of all the following functions $p(z)$,

$$p_j(z) = \begin{cases} -g_j(x) & \text{if } j \in \alpha(z^*), \\ u_j \text{ or } -g_j(x) & \text{if } j \in \beta(z^*), \\ u_j & \text{if } j \in \gamma(z^*), \end{cases}$$

$j = 1, ..., p$ and define

$$\partial_Q H^{(i)}(z) = \mathcal{M}(z, H_1) \times \{\nabla p^{(i)}(z)^T\} \times \{\nabla H_3(z)^T\}.$$

Lemma 3.1. Suppose that $z^* = (x^*, u^*, v^*) \in R^n \times R^p \times R^q$ is a Kuhn-Tucker point of (1.1) and satisfies the conditions of Theorem 2.1. Then there exist positive constants δ and C such that for any $\hat{z} = (\hat{x}, \hat{u}, \hat{v}) \in R^n \times R^p \times R^q$ with $\hat{z} \in \{z \mid \|z - z^*\| \leq \delta\}$, and any $i \in I^\beta$, all $W_{(i)} \in \partial_Q H^{(i)}(\hat{z})$ are invertible and $\|W_{(i)}^{-1}\| \leq C$.

Proof. From the definition of $H^{(i)}(z)$ and $\partial_Q H^{(i)}(z)$ we know that

$$H^{(i)}(z^*) = 0 \quad \forall\, i \in I^\beta$$

and

$$\partial_Q H^{(i)}(z^*) \subseteq \partial_Q H(z^*) \quad \forall\, i \in I^\beta.$$

From Theorem 2.1 we know that all matrices $W \in \partial_Q H(z^*)$ are nonsingular. This means that all matrices $W_{(i)} \in \partial_Q H^{(i)}(z^*)$, $i \in I^\beta$ are nonsingular. It is easy to see that all $\partial_Q H^{(i)}(z)$, $i \in I^\beta$ are nonempty compact subsets and all the maps $\partial_Q H^{(i)}$, $i \in I^\beta$ are upper semi-continuous. Therefore for each $i \in I^\beta$ there exist a neighborhood $N^{(i)}(z^*)$ of z^* and a positive number C_i such that for any $\hat{z} \in N^{(i)}(z^*)$, all $W_{(i)} \in \partial_Q H^{(i)}(\hat{z})$ are nonsingular and satisfy $\|W_{(i)}^{-1}\| \le C_i$. Since I^β is of finite elements, the conclusion of this lemma holds. □

In order to establish the superlinear convergence of approximate Newton method, we first consider the following generalized approximate Newton method (GANM) for solving $H(z) = 0$:

Given $z^0 = (x^0, u^0, v^0) \in R^n \times R^p \times R^q$.

For $k = 0, 1, ...,$ choose $i \in I^\beta$ and let

$$z^{k+1} = z^k - B_{(i)k}^{-1} H^{(i)}(z^k), \tag{3.4}$$

where $B_{(i)k} = \nabla H^{(i)k}(z^k)^T$ and $H^{(i)k}$ is defined as

$$H^{(i)k}(z) = \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)u + \nabla h(x^k)v + G_k(x - x^k) \\ q^{(i)k}(z) \\ -h(x^k) - \nabla h(x^k)^T(x - x^k) \end{pmatrix}, \tag{3.5}$$

$i \in I^\beta$, where $q^{(i)k}(z)$ is defined as

$$q_j^{(i)k}(z) = \begin{cases} -g_j(x^k) - \nabla g_j(x^k)^T(x - x^k) & \text{if } j \in \alpha(z^*), \\ p_j^{(i)}(z^k) + \nabla p_j^{(i)}(z^k)^T(z - z^k) & \text{if } j \in \beta(z^*), \\ u_j & \text{if } j \in \gamma(z^*), \end{cases} \tag{3.6}$$

$j = 1, ..., p$, and $G_k \in R^{n \times n}$.

Remark 3.1. In practice, we can't use the above method since we don't know z^*. However, the above method provides an approach to prove the Q-superlinear convergence of the approximate Newton method.

Theorem 3.1. Suppose that $z^* = (x^*, u^*, v^*) \in R^n \times R^p \times R^q$ is a Kuhn-Tucker point of (1.1) and satisfies the conditions of Theorem 2.1. Suppose that ∇f, ∇g and ∇h are semismooth at x^*. Let C and δ be the positive constants in Lemma 3.1. If there exists $V_k \in \mathcal{V}_{x^k}(z^k)$ such that

$$\|G_k - V_k\| \le \frac{1}{4C} \quad \forall\ k, \tag{3.7}$$

then the above method GANM is well defined and Q-linearly converges to z^* in a neighborhood of z^*. If furthermore,

$$\lim_{k \to \infty} \frac{\|(G_k - V_k)(x^{k+1} - x^k)\|}{\|z^{k+1} - z^k\|} = 0, \tag{3.8}$$

then the convergence is Q-superlinear. If in the later case $H(z^k) \neq 0$, we have

$$\lim_{k\to\infty} \frac{||H(z^{k+1})||}{||H(z^k)||} = 0. \tag{3.9}$$

Proof. Since ∇f, ∇g and ∇h are semismooth at x^*, H and $H^{(i)}$, $i \in I^\beta$ are semismooth at z^*.

From the definitions of $\mathcal{V}_{x^k}(z^k)$ and $\partial_Q H^{(i)}(z^k)$, $i \in I^\beta$, for each $B_{(i)k}$, $i \in I^\beta$ there exists $W_{(i)k} \in \partial_Q H^{(i)}(z^k)$ such that for any $z = (x, u, v) \in R^n \times R^p \times R^q$

$$||(B_{(i)k} - W_{(i)k})z|| = ||(V_k - G_k)x||. \tag{3.10}$$

In particular, we have

$$||B_{(i)k} - W_{(i)k}|| \leq ||V_k - G_k|| \leq \frac{1}{4C}. \tag{3.11}$$

If $||z^k - z^*|| \leq \delta$, then by Lemma 3.1, $W_{(i)k}^{-1}$ exists and $||W_{(i)k}^{-1}|| \leq C$. By the Perturbation Lemma of Ortega and Rheinboldt (Ref. 12, p. 45), $B_{(i)k}$ is invertible and

$$||B_{(i)k}^{-1}|| \leq \frac{4}{3}C. \tag{3.12}$$

Recall that a map is semismmooth at z^* if and only if each of its components is semismooth at z^* and there are finite elements in the set I^β, so by (3.1) and (3.2), for every $\varepsilon > 0$ there exists a neighborhood $N(z^*)$ of z^* such that when $z \in N(z^*)$ and $W_{(i)} \in \partial_Q H^{(i)}(z)$ (note $W_{(i)}^j \in \partial H_j^{(i)}(z)$) we have

$$\begin{aligned} ||H^{(i)}(z) - H^{(i)}(z^*) - W_{(i)}(z - z^*)|| &\leq \sum_{j=1}^{n+p+q} |H_j^{(i)}(z) - H_j^{(i)}(z^*) - W_{(i)}^j(z - z^*)| \\ &\leq \varepsilon ||z - z^*|| \quad \forall i \in I^\beta. \end{aligned} \tag{3.13}$$

So we may choose $\delta_1 > 0$ sufficiently small such that when $||z^k - z^*|| \leq \delta_1$, for any $i \in I^\beta$ we have

$$||H^{(i)}(z^k) - H^{(i)}(z^*) - W_{(i)k}(z^k - z^*)|| \leq \frac{1}{8C}||z^k - z^*||. \tag{3.14}$$

Let $\bar{\delta} = \min(\delta_1, \delta)$. Then when $||z^k - z^*|| \leq \bar{\delta}$, we have

$$\begin{aligned} ||z^{k+1} - z^*|| &= ||z^k - B_{(i)k}^{-1} H^{(i)}(z^k) - z^*|| \\ &\leq ||B_{(i)k}^{-1}|| ||H^{(i)}(z^k) - H^{(i)}(z^*) - B_{(i)k}(z^k - z^*)|| \\ &\leq ||B_{(i)k}^{-1}|| [||H^{(i)}(z^k) - H^{(i)}(z^*) - W_{(i)k}(z^k - z^*)|| \\ &\quad + ||(B_{(i)k} - W_{(i)k})(z^k - z^*)||]. \end{aligned} \tag{3.15}$$

Substituting (3.11)-(3.12) and (3.14) into (3.15) gives

$$\begin{aligned}||z^{k+1} - z^*|| &\leq \frac{4}{3}C(\frac{1}{4C} + \frac{1}{8C})||z^k - z^*|| \\ &= \frac{1}{2}||z^k - z^*||. \end{aligned} \tag{3.16}$$

This proves that GANM is well defined and Q-linearly converges to z^* in a neighborhood of z^*.

Furthermore if (3.8) holds, by (3.10)-(3.11), (3.13) and (3.15), we have

$$\begin{aligned}||z^{k+1} - z^*|| &\leq \frac{4}{3}C[||H^{(i)}(z^k) - H^{(i)}(z^*) - W_{(i)k}(z^k - z^*)|| \\ &\quad +||(B_{(i)k} - W_{(i)k})(z^{k+1} - z^k)|| + ||(B_{(i)k} - W_{(i)k})(z^{k+1} - z^*)||] \\ &\leq \frac{4}{3}C[o(||z^k - z^*||) + ||(V_k - G_k)(x^{k+1} - x^k)|| + \frac{1}{4C}||z^{k+1} - z^*||] \\ &\leq o(||z^k - z^*||) + o(||z^{k+1} - z^k||) + \frac{1}{3}||z^{k+1} - z^*||. \end{aligned} \tag{3.17}$$

This, and the Q-linear convergence of $\{z^k\}$, turns out to be

$$||z^{k+1} - z^*|| = o(||z^k - z^*||), \tag{3.18}$$

i.e., the convergence of GANM is Q-superlinear.

The proof of (3.9) is similar to the proof of Theorem 3.1 of Qi (Ref. 16). □

Remark 3.2. For unconstrained optimization problem ($f \in C^2$), condition (3.8) is known as the Dennis-Moré type condition (see, e.g., Dennis and Schnabel (Ref. 3)) and that for nonlinear programming (C^2 optimization problem) with equality constraints a generalization of this condition due to Boggs, Tolle, and Wang (Ref. 31) is widely used.

Corollary 3.1. Assume that the conditions of Theorem 3.1 hold. Then there exists a positive number $\varepsilon > 0$ such that when there exists $V_k \in \mathcal{V}_{x^k}(z^k)$ such that

$$||V_k - G_k|| \leq \min(\varepsilon, \frac{1}{4C}) \quad \forall\ k, \tag{3.19}$$

the approximate Newton method described above is well defined and Q-linearly converges to z^* in a neighborhood of z^*. If furthermore (3.8) holds, then the convergence is Q-superlinear. If in the later case, $H(z^k) \neq 0$, then (3.9) holds.

Proof. To complete the proof, we prove that the approximate Newton method is a special case of GANM in a neighborhood of z^*.

Choose a positive number $\delta_2 > 0$ ($\delta_2 \leq \bar{\delta}/3$, $\bar{\delta}$ is defined in the proof of Theorem 3.1) such that when

$$z,\ z^k \in B(z^*; 3\delta_2) \equiv \{z|\ ||z - z^*|| \leq 3\delta_2\},$$

we have

$$\begin{cases} -g_i(x^k) - \nabla g_i(x^k)^T (x - x^k) < u_i^k & \text{if } i \in \alpha(z^*), \\ -g_i(x^k) - \nabla g_i(x^k)^T (x - x^k) > u_i^k & \text{if } i \in \gamma(z^*). \end{cases} \tag{3.20}$$

So when $z^k \in B(z^*; 3\delta_2)$ we have

$$\alpha(z^*) \subseteq \alpha(z^k),\ \gamma(z^*) \subseteq \gamma(z^k) \text{ and } \beta(z^k) \subseteq \beta(z^*). \tag{3.21}$$

The first-order Kuhn-Tucker conditions of the quadratic subproblem Q_k can be written as

$$H^k(z) = 0, \tag{3.22}$$

where $H^k(z)$ is defined as

$$H^k(z) = \begin{pmatrix} \nabla f(x^k) + \nabla g(x^k)u + \nabla h(x^k)v + G_k(x - x^k) \\ \min(u, -g(x^k) - \nabla g(x^k)^T (x - x^k)) \\ -h(x^k) - \nabla h(x^k)^T (x - x^k) \end{pmatrix}. \tag{3.23}$$

We now show that (3.22) has a solution if δ_2 sufficiently small. Similarly to the proof of Theorem 4.1 of Robinson (Ref. 23), we can easily conclude that the following matrix

$$A_* = \begin{pmatrix} V_* & \nabla g_{\alpha(z^*)}(x^*) & \nabla h(x^*) \\ -\nabla g_{\alpha(z^*)}(x^*)^T & 0 & 0 \\ -\nabla h(x^*)^T & 0 & 0 \end{pmatrix}$$

is nonsingular, and the Schur complement

$$B(z^*) = C(z^*)^T A_*^{-1} C(z^*)$$

is a P-matrix (i.e., a matrix with positive principle minors), where $V_* \in \mathcal{V}_{x^*}(z^*)$ and

$$C(z^*) = \begin{pmatrix} \nabla g_{\beta(z^*)}(x^*) \\ 0 \\ 0 \end{pmatrix}.$$

From the definitions of $\mathcal{M}(z, H_1)$ and $\mathcal{V}_x(z)$, for every $\varepsilon > 0$ we can prove that there exists $\delta_3 > 0$ such that when

$$z^k \in B(z^*; \delta_3) \equiv \{z |\ \|z - z^*\| \le \delta_3\},$$

we have

$$\mathcal{V}_{x^k}(z^k) \subseteq \mathcal{V}_{x^*}(z^*) + \varepsilon B(0; 1), \tag{3.24}$$

where $B(0;1) \equiv \{z \in R^n |\ \|z\| \le 1\}$. So we may restrict δ_2 and ε such that for any $z^k \in B(z^*; \delta_2) \equiv \{z |\ \|z - z^*\| \le \delta_2\}$, the matrix

$$A(z^k) = \begin{pmatrix} G_k & \nabla g_{\alpha(z^*)}(x^k) & \nabla h(x^k) \\ -\nabla g_{\alpha(z^*)}(x^k)^T & 0 & 0 \\ -\nabla h(x^k)^T & 0 & 0 \end{pmatrix}$$

is nonsingular, and the Schur complement

$$B(z^k) = C(z^k)^T A(z^k)^{-1} C(z^k)$$

is a P-matrix, where

$$C(z^k) = \begin{pmatrix} \nabla g_{\beta(z^*)}(x^k) \\ 0 \\ 0 \end{pmatrix}.$$

Note that in the matrix

$$\begin{pmatrix} A(z^k) & C(z^k) \\ -C(z^k)^T & 0 \end{pmatrix}, \tag{3.25}$$

the index sets α and β are defined at z^* but the various gradients are evaluated at z^k.

In order to consider the solvability of the system (3.22), we consider the solvability of the following system

$$\begin{cases} F_{u^k,v^k}(x^k) + G_k d^x + \nabla g(x^k) d^u + \nabla h(x^k) d^v = 0, \\ -g_i(x^k) - \nabla g_i(x^k)^T d^x = 0 \quad \text{for } i \in \alpha(z^*), \\ \min(u_i^k + d^{u_i}, -g_i(x^k) - \nabla g_i(x^k)^T d^x) = 0 \quad \text{for } i \in \beta(z^*), \\ u_i^k + d^{u_i} = 0 \quad \text{for } i \in \gamma(z^*), \\ -h(x^k) - \nabla h(x^k)^T d^x = 0. \end{cases} \tag{3.26}$$

The component d^{u_i} is explicit for $i \in \gamma(z^*)$. Simplifying these equations, we deduce that the remaining components of the vector $d = (d^x, d^u, d^v) \in R^n \times R^p \times R^q$ can be obtained by solving the mixed linear complementarity problem

$$\begin{cases} \bar{q}(z^k) + A(z^k) w + C(z^k) d^{u_\beta} = 0, \\ -g_\beta(x^k) - C(z^k)^T w \geq 0, \\ u_\beta^k + d^{u_\beta} \geq 0, \\ [-g_\beta(x^k) - C(z^k)^T w]^T (u_\beta^k + d^{u_\beta}) = 0, \end{cases} \tag{3.27}$$

where

$$w = (d^x, d^{u_\alpha}, d^v), \; \bar{q}(z^k) = (\bar{q}_\beta(z^k), -g_\alpha(x^k), -h(x^k)), \; \bar{q}_\beta(z^k) = F_{u^k,v^k}(x^k) - \nabla g_\gamma(x^k) u_\gamma^k$$

and α, β and γ denotes respectively the index sets $\alpha(z^*)$, $\beta(z^*)$ and $\gamma(z^*)$. From linear complementarity theory (see, e.g., Murty (Ref. 11)), we know that a sufficient condition for the system (3.27) to have a unique solution is (i) the matrix $A(z^k)$ is nonsingular and (ii) the Schur complement $B(z^k) = C(z^k)^T A(z^k)^{-1} C(z^k)$ is a P-matrix. Since we have proved that these two conditions are satisfied, system (3.27) has a unique solution. Then system (3.26) has a unique solution when $z^k \in B(z^*; \delta_2)$. We denote this solution by

$$d^k = (d^{x^k}, d^{u^k}, d^{v^k}) \in R^n \times R^p \times R^q.$$

It is easy to prove that for each k there exists $i \in I^\beta$ such that

$$H^{(i)}(z^k) - B_{(i)k} d^k = 0. \tag{3.28}$$

From the proof of Theorem 3.1, we know that

$$\|z^k + d^k - z^*\| \leq \frac{1}{2}\|z^k - z^*\|. \tag{3.29}$$

Let $\bar{z}^{k+1} = z^k + d^k$. Then $\bar{z}^{k+1} \in B(z^*;\delta_2)$ if $z^k \in B(z^*;\delta_2)$.

We now prove that $H^k(\bar{z}^{k+1}) = 0$, which means that (3.22) has a solution. when z^k, $\bar{z}^{k+1} \in B(z^*;\delta_2)$, we have

$$\begin{aligned}
&\min(\bar{u}_i^{k+1}, -g_i(x^k) - \nabla g_i(x^k)^T(\bar{x}^{k+1} - x^k)) \\
&\qquad = \begin{cases} -g_i(x^k) - \nabla g_i(x^k)^T(\bar{x}^{k+1} - x^k) & \text{if } i \in \alpha(z^*), \\ \bar{u}_i^{k+1} & \text{if } i \in \gamma(z^*). \end{cases}
\end{aligned}$$

Thus if $z^k \in B(z^*;\delta_2)$, then

$$\begin{aligned}
H^k(\bar{z}^{k+1}) &= \begin{pmatrix} F_{u^k,v^k}(x^k) + \nabla g(x^k)d^{u^k} + \nabla h(x^k)d^{v^k} + G_k d^{x^k} \\ \min(u^k + d^{u^k}, -g(x^k) - \nabla g(x^k)^T d^{x^k}) \\ -h(x^k) - \nabla h(x^k)^T d^{x^k} \end{pmatrix} \\
&= \begin{pmatrix} F_{u^k,v^k}(x^k) + \nabla g(x^k)d^{u^k} + \nabla h(x^k)d^{v^k} + G_k d^{x^k} \\ -g_\alpha(x^k) - \nabla g_\alpha(x^k)^T d^{x^k} \\ \min(u_\beta^k + d^{u_\beta^k}, -g_\beta(x^k) - \nabla g_\beta(x^k)^T d^{x^k}) \\ u_\gamma^k + d^{u_\gamma^k} \\ -h(x^k) - \nabla h(x^k)^T d^{x^k} \end{pmatrix} \\
&= 0,
\end{aligned}$$

which means that system $H^k(z) = 0$ has a solution $\bar{z}^{k+1}$ in $B(z^*;\delta_2)$, i.e., $\bar{z}^{k+1}$ is a Kuhn-Tucker point of (1.2). Suppose that $\tilde{z}^{k+1} \in B(z^*;3\delta_2)$ is an arbitrary solution of $H^k(z) = 0$. Since $\tilde{z}^{k+1} \in B(z^*;3\delta_2)$, then

$$\begin{aligned}
&\min(\tilde{u}_i^{k+1}, -g_i(x^k) - \nabla g_i(x^k)^T(\tilde{x}^{k+1} - x^k)) \\
&\qquad = \begin{cases} -g_i(x^k) - \nabla g_i(x^k)^T(\tilde{x}^{k+1} - x^k) & \text{if } i \in \alpha(z^*), \\ \tilde{u}_i^{k+1} & \text{if } i \in \gamma(z^*). \end{cases}
\end{aligned}$$

Therefore $\tilde{d}^k = \tilde{z}^{k+1} - z^k$ is also a solution of system (3.26). From the uniqueness of the solution of system (3.26), we know that $\tilde{z}^{k+1} = \bar{z}^{k+1}$, which shows that $\bar{z}^{k+1}$ is the closest Kuhn-Tucker point to z^k in terms of distance $\|\bar{z}^{k+1} - z^k\|$. So there exists $i \in I^\beta$ such that

$$z^{k+1} = \bar{z}^{k+1} = z^k - B_{(i)k}^{-1} H^{(i)}(z^k),$$

which means that approximate Newton method is a special case of GANM in a neighborhood of z^*. So we complete the proof of Corollary 3.1 by considering Theorem 3.1. □

Remark 3.3. If we choose $G_k \in \mathcal{V}_{x^k}(z^k)$, (3.7) and (3.8) are satisfied.

4. Some Discussions

In this chapter we considered the local convergence of approximate Newton or SQP methods for LC^1 optimization problems without assuming the strict complementarity condition. The global convergent technique used in Qi (Ref. 19) can be applied to this chapter similarly.

GANM is useful in proving the Q-superlinear convergence of approximate Newton or SQP methods, but it can't be used in practice since we don't know $\alpha(z^*)$, $\beta(z^*)$ and $\gamma(z^*)$. The approximate Newton or SQP methods are well used and in each step a quadratic programming is needed to be solved. In the following we give such a method that in each step only a linear equations is needed to be solved.

Given $z^0 = (x^0, u^0, v^0) \in R^n \times R^p \times R^q$.
For $k = 0, 1, ...,$

$$z^{k+1} = z^k - B_k^{-1} H(z^k), \tag{4.1}$$

where $B_k \in \partial_Q H^k(z^k) \equiv \{\nabla L^k(z^k)^T\} \times \partial_b g^k(z^k) \times \{\nabla h^k(z^k)^T\}$, and

$$L^k(z) = \nabla f(x^k) + \nabla g(x^k) u + \nabla h(x^k) v + G_k(x - x^k),$$

$$g^k(z) = \min(u, -g(x^k) - \nabla g(x^k)^T (x - x^k))$$

and

$$h^k(z) = -h(x^k) - \nabla h(x^k)^T (x - x^k).$$

It is easy to see that in a neighborhood of the solution z^* of $H(z) = 0$, the above method is a special case of GANM. So similar convergent properties for (4.1) can be found in Theorem 3.1.

References

1. Aubin, J. D., and Frankowska, F., *Set-Valued Analysis*, Birkhäuser, Boston, 1990.
2. Clarke, F. H., *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
3. Dennis, J. E., and Schnabel, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jercy, 1983.
4. Garcia Palomares, U. C., and Mangasarian, O. L., *Superlinearly Convergent Quasi-Newton Algorithms for Nonlinearly Constrained Optimization Problems*, Mathematical Programming, Vol. 11, pp. 1-13, 1976.
5. Han, S. P., *Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems*, Mathematical Programming, Vol. 11, pp. 263-282, 1976.
6. Hiriart-Urruty, J. B., Strodoit, J. J., and Nguyen, V. H., *Generalized Hessian Matrix and Secord-Order Optimality Conditions for Problems with $C^{1,1}$ Data*, Applied Mathematics and Optimization, Vol. 11, pp. 43-56, 1984.
7. Josephy, N. H., *Newton's Method for Generalized Equations*, Technical Summary Report 1965, Mathematical Research Center, University of Wisconsin-Madison, 1979.

8. Josephy, N. H., *Quasi-Newton Methods for Generalized Equations*, Technical Summary Report 1966, Mathematical Research Center, University of Wisconsin-Madison, 1979.
9. McCormick, G. P., *Penalty Function Versus Non-Penalty Function Methods for Constrained Nonlinear Programming Problems*, Mathematical Programming, Vol. 1, pp. 217-238, 1971.
10. Mifflin, R., *Semismooth and Semiconvex Functions in Constrained Optimization*, SIAM Journal on Control and Optimization, Vol. 15, pp. 957-972, 1972.
11. Murty, K. G., *Linear Complementarity, Linear and Nonlinear Programming*, Helderman-Verlag, Berlin, 1988.
12. Ortega, J. M., and Rheinboldt, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
13. Pang, J. -S., Han, S. P., and Rangaraj, R., *Minimization of Locally Lipschitzian Functions*, SIAM Journal on Optimization, Vol. 1, pp. 57-82, 1991.
14. Pang, J. -S., *A B-Differentiable Equation-Based, Globally and Locally Quadratically Convergent Algorithm for Nonlinear Programs, Complementarity and Variational Inequality Problems*, Mathematical Programming, Vol. 51, pp. 101-131, 1991.
15. Pang, J. -S., and Qi, L., *Nonsmooth Equations: Motivation and Algorithms*, SIAM Journal on Optimization, Vol. 3, pp. 443-465, 1993.
16. Qi, L.,*Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations*, Mathematics of Operations Research, Vol. 18, pp. 227-244, 1993.
17. Qi, L., LC^1 *Functions and* LC^1 *Optimization Problems*, Applied Mathematics Preprint 91/21, School of Mathematics, The University of New South Wales, Sydney, Australia, 1991.
18. Qi, L., and Sun, J., *A Nonsmooth Version of Newton's Method*, Mathematical Programming, Vol. 58, pp. 353-368, 1993.
19. Qi, L., *Superlinearly Convergent Approximate Newton Methods for* LC^1 *Optimization Problems*, Mathematical Programming, Vol. 64, pp. 277-294, 1994.
20. Qi, L., and Womersley, R., *An SQP Algorithm for Solving Extended Linear-Quadratic Problems in Stochastic Programming*, Applied Mathematics Preprint 92/23, School of Mathematics, The University of New South Wales, Sydney, Australia, 1992.
21. Robinson, S. M., *A Quadratically Convergent Algorithm for General Nonlinear Programming Problems*, Mathematical Programming, Vol. 3, pp. 145-156, 1972.
22. Robinson, S. M., *Perturbed Kuhn-Tucker Points and Rates of Convergence for A Class of Nonlinear Programming Algorithms*, Mathematical Programming, Vol. 7, pp. 1-16, 1974.
23. Robinson, S. M., *Strongly Regular Generalized Equations*, Mathematics of Operations Research, Vol. 5, pp. 43-62, 1980.
24. Rockafellar, R. T., *Computational Schemes for Solving Large-Scale Problems in Extended Linear-Quadratic Programming*, Mathematical Programming, Vol. 48, pp. 447-474, 1990.
25. Rockafellar, R. T. , and Wets, R. J. -B., *Generalized Linear-Quadratic Problems of Deterministic and Stochastic Optimal Control in Discrete Time*, SIAM Journal on Control and Optimization, Vol. 28, pp. 810-822, 1990.
26. Shor, N. Z., *A Class of Almost-Differentiable Functions and A Minimization Method*

for Functions of This Class, Kibernetika, Vol. 4, pp. 65-70, 1972.

27. Sun, D., and Han, J., *Newton and Quasi-Newton Methods for A Class of Nonsmooth Equations and Related Problems*, Technical Report No. 026, Institute of Applied Mathematics, Academia Sinica, Beijing, China, 1994.
28. Zhu, C., and Rockafellar, R. T., *Primal-Dual Projected Gradient Algorithm for Extended Linear-Quadratic Programming*, to Appear in SIAM Journal on Optimization.
29. Lescrenier, M., *Convergence of Trust Region Algorithms for Optimization with Bounds when Strict Complementarity Does Not Hold*, SIAM J. Numer. Anal., Vol. 28, pp. 476-495, 1991.
30. Conn, A. R., Gould, N. I. M., and Toint, Ph. L., *Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds*, SIAM J. Numer. Anal., Vol. 25, pp. 433-460, 1988. Erratum in the same Journal, Vol. 26, p.764, 1989.
31. Boggs, P. T., Tolle, J. W., and Wang, P., *On the Local Convergence of Quasi-Newton Methods for Constrained Optimization*, Vol. 20, pp. 161-171, 1982.

Chapter 5

Newton and Quasi-Newton Methods for Normal Maps with Polyhedral Set

Abstract

This chapter presents a Newton method and a quasi-Newton method for solving normal maps $H(x) := F(\Pi_C(x)) + x - \Pi_C(x) = 0$ when C is a polyhedral set. For both Newton and quasi-Newton methods established here the subproblem needed to solve is a linear equations in per iteration. The other characteristics of the quasi-Newton method established in this chapter include: (i) without assuming the existence of $H'(x^*)$, a Q-superlinear convergence theorem is established, (ii) only one initial approximation matrix is needed, (iii) the linear independence condition is not assumed, (iv) the Q-superlinear convergence is established on the original variable x, and (v) from the QR factorization of the k-th iterative matrix we need at most $O((1+2|J_k|+2|L_k|)n^2)$ arithmetic operations to get the QR factorization of the $(k+1)$-th iterative matrix.

Chapter 5
Newton and Quasi-Newton Methods for Normal Maps
with Polyhedral Set

1. Introduction

Let C be nonempty closed convex set in $\Re^n$, and F be the continuous function from $\Re^n$ to itself. A very common problem arising in optimization and equilibrium analysis is that of finding a point x such that x is a solution of the following normal maps [26]

$$H(x) := F(\Pi_C(x)) + x - \Pi_C(x) = 0, \tag{1.1}$$

where Π_C is the Euclidean projector on C. For example, the variational inequality problem defined on C is to find $y \in C$ such that

$$(z - y)^T F(y) \geq 0 \quad \forall z \in C. \tag{1.2}$$

It is easy to verify that if $H(x) = 0$, then the point $y =: \Pi_C(x)$ solves (1.2); conversely if y solves (1.2), then with $x := y - F(y)$ one has $H(x) = 0$. Therefore the equations $H(x) = 0$ is an equivalent way of formulating the variational inequality problem (1.2).

For solving (1.1) or (1.2), the basic methods are Josephy's Newton's method [10] and quasi-Newton methods [11]. In each step, Josephy's methods need to solve a linear variational inequality problem defined on the set C. This is a nonlinear and nonconvex subproblem in general. Kojima and Shindo [12] generalized Newton and quasi-Newton methods to piecewise smooth functions. For quasi-Newton methods, their method needs a new approximate starting matrix when the iteration sequence moves to a new C^1 piece. This may require to store lots of initial matrices. Ip and Kyparisis [9] discussed quasi-Newton methods directly applied to nonsmooth equations. The Q-superlinear convergence of quasi-Newton methods was established by them on the assumption that the mapping is strongly Frechét differentiable [14]. This is too restrictive for (1.1). The results of Chen and Qi [3] are not far from this. Sun and Han [28] considered Newton and quasi-Newton methods for a class of nonsmooth equations and related problems, which include the general nonlinear complementarity problem, the variational inequality problem with simple bound constraints, and the Karush-Kuhn-Tucker (K-K-T) systems of nonlinear programming problem. Sun and Han's methods need one approximate initial matrix and in each step only need to solve a linear equations. Furthermore for quasi-Newton method they discussed how to update the QR factorization of the present iterative matrix to the QR factorization of the next iterative matrix in less than $O(n^3)$ arithmetic operations. However, the skill introduced in [28] can't be used directly to solve (1.1) when C is a general polyhedral set.

In this chapter, we shall assume that C has the form

$$C = \{x | \ Ax \leq a, \ Bx = b\}, \tag{1.3}$$

where $A : \Re^n \to \Re^m$, $B : \Re^n \to \Re^p$, $a \in \Re^m$, and $b \in \Re^p$. Throughout this chapter we will assume that rank (B)=p $(p \leq n)$. In the following we will discuss such kinds of

Newton and quasi-Newton methods that use a linear equations as the subproblem in per iteration.

The main characteristics of the quasi-Newton method established in this chapter include: (i) without assuming the existence of $H'(x^*)$, we establish a Q-superlinear convergence theorem, (ii) only one approximate matrix is needed, (iii) the linear independence condition is not assumed, (iv) the Q-superlinear convergence is established on the original variable x, and (v) from the QR factorization of the k-th iterative matrix we need at most $O((1+2|J_k|+2|L_k|)n^2)$ arithmetic operations to get the QR factorization of the $(k+1)$-th iterative matrix (see (5.6) for the definition of J_k and L_k).

The rest of this chapter is organized as following. In § 2 we discuss some properties of the normal maps (1.1). The Newton and quasi-Newton methods are given in § 3 and § 4, respectively. In § 5 we discuss the implementation aspects of Newton and quasi-Newton methods.

2. Basic Preliminaries

For any $x \in \Re^n$, $\Pi_C(x)$ is the Euclidean projection of x on C and C is of the form (1.3), then there exist multipliers $\lambda \in \Re^m_+$, $\mu \in \Re^p$ such that

$$\begin{cases} \Pi_C(x) - x + A^T\lambda + B^T\mu = 0, \\ \lambda \geq 0,\ a - A\Pi_C(x) \geq 0,\ \lambda^T(a - A\Pi_C(x) = 0, \\ b - B\Pi_C(x) = 0. \end{cases} \tag{2.1}$$

Let $\mathcal{M}(x)$ denote the nonempty set of multipliers $(\lambda,\ \mu) \in \Re^m_+ \times \Re^p$ that satisfy the K-K-T conditions (2.1). For a nonnegative vector $d \in \Re^m$, we shall let $supp(d)$, called the support of d, be the subset of $\{1, ..., m\}$ consisting of the indexes i for $d_i > 0$. Denote

$$I(x) = \{i|\ A_i\Pi_C(x) = a_i,\ i = 1, ..., m\}. \tag{2.2}$$

Define the family $\mathcal{B}(x)$ of indexes of $\{1, ..., m\}$ as follows: $K \in \mathcal{B}(x)$ if and only if $supp(\lambda) \subseteq K \subseteq I(x)$ for some $(\lambda,\ \mu) \in \mathcal{M}(x)$ and the vectors

$$\{A_i^T,\ i \in K\} \cup \{B_j^T,\ j = 1, ..., p\} \tag{2.3}$$

are linearly independent. This family $\mathcal{B}(x)$ is nonempty because $\mathcal{M}(x)$ has an extreme point which easily yields a desired index set K with the stated properties.

Define

$$\mathcal{P}(x) = \{P \in \Re^{n\times n}|\ P = I - (A_K^T\ B^T)\left(\begin{pmatrix} A_K \\ B \end{pmatrix}(A_K^T\ B^T)\right)^{-1}\begin{pmatrix} A_K \\ B \end{pmatrix}, \quad K \in \mathcal{B}(x)\}, \tag{2.4}$$

where I is the unit matrix of $\Re^{n\times n}$ and A_K is the matrix consisting of the K rows of A.

Remark 2.1. The existence of $\left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \ B^T)\right)^{-1}$ comes from the linear independence of the vectors $\{A_i^T,\ i \in K\} \cup \{B_j^T,\ j = 1, ..., p\}$. Note that for all $P \in \mathcal{P}(x)$, we have $P^T = P$, $P^2 = P$, and $\|P\| \le 1$. These simple facts will be used later.

In the following lemma, part (i) is a consequence of Pang and Ralph [18]. For the completeness, we also give the proof.

Lemma 2.1. (i) There exists a neighborhood $N(x)$ of x such that when $y \in N(x)$, we have

$$\mathcal{B}(y) \subseteq \mathcal{B}(x) \text{ and } \mathcal{P}(y) \subseteq \mathcal{P}(x);$$

(ii) when $\mathcal{B}(y) \subseteq \mathcal{B}(x)$, $\Pi_C(y) = \Pi_C(x) + P(y - x) \quad \forall\ P \in \mathcal{P}(y)$.

Proof. According to the definition of $\mathcal{P}(\cdot)$, we only need to prove that there exists a neighborhood $N(x)$ of x such that

$$\mathcal{B}(y) \subseteq \mathcal{B}(x) \quad \forall y \in N(x). \tag{2.5}$$

If not, then there exists a sequence $\{y^k\}$ converging to x such that for all k, there is an index set $K^k \in \mathcal{B}(y^k) \backslash \mathcal{B}(x)$. Since there are only finitely many such index sets, if necessary by taking a subsequence we assume that these index sets K^k are the same for all k. By letting K be the common index set, we have that the vectors

$$\{A_i^T,\ i \in K\} \cup \{B_j,\ j = 1, ..., p\}$$

are linearly independent and there exists $(\lambda^k,\ \mu^k) \in \mathcal{M}(y^k)$ such that $supp(\lambda^k) \subseteq K \subseteq I(y^k)$, but $K \notin \mathcal{B}(x)$. Clearly $K \subseteq I(x)$. The only way for $K \notin \mathcal{B}(x)$ is that there exists no $(\lambda,\ \mu) \in \mathcal{M}(x)$ such that $supp(\lambda) \subseteq K$. But we have

$$\Pi_C(y^k) - y^k + \sum_{i \in k} \lambda_i^k A_i^T + \sum_{j=1}^{p} \mu_j^k B_j^T = 0.$$

Since $y^k \to x$ and $\{A_i^T,\ i \in K\} \cup \{B_j^T,\ j = 1, ..., p\}$ are linearly independent, it follows that $\{\lambda_i^k,\ i \in K\}$ and $\{\mu_j^k,\ j = 1, ..., p\}$ are bounded; thus, the full sequence $\{\lambda^k\} \cup \{\mu^k\}$ must have an accumulation point which must necessary be an element in $\mathcal{M}(x)$ and whose support is a subset of K. This is a contradiction.

(ii) when $\mathcal{B}(y) \subseteq \mathcal{B}(x)$, from the K-K-T conditions (2.1) we know that for any $P \in \mathcal{P}(y)$ there exists $K \in \mathcal{B}(y) \subseteq \mathcal{B}(x)$ such that

$$P = I - (A_K^T \ B^T) \left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \ B^T)\right)^{-1} \begin{pmatrix} A_K \\ B \end{pmatrix}$$

and

$$\Pi_C(y) = Py + c_K,\ \Pi_C(x) = Px + c_K$$

where $c_K = (A_K^T \ B^T) \left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \ B^T)\right)^{-1} \begin{pmatrix} a_K \\ b \end{pmatrix}$. and a_K is the vector consisting of the K components of a.

Thus

$$\Pi_C(y) = \Pi_C(x) + P(y - x) \quad \forall\ P \in \mathcal{P}(y).$$

3. Newton Method

In the following sections suppose that F is continuously differentiable and C is of the form (1.3). Denote

$$\mathcal{W}(x) = \{W \in \Re^{n\times n} | \; W = F'(\Pi_C(x))P + I - P, \; P \in \mathcal{P}(x)\}.$$

The Newton method for solving (1.1) can be described as following:

Given $x^0 \in \Re^n$.
Do for $k = 0, 1, \dots$:
Choose $P_k \in \mathcal{P}(x^k)$ and compute

$$W_k := F'(\Pi_C(x^k))P_k + I - P_k \in \mathcal{W}(x^k)$$

Solve

$$W_k s + H(x^k) = 0 \tag{3.1}$$

for s^k

$$x^{k+1} = x^k + s^k. \tag{3.2}$$

Theorem 3.1. Suppose that $F : \Re^n \to \Re^n$ is continuously differentiable, C is of the form (1.3), and x^* is a solution of (1.1). If all $W_* \in \mathcal{W}(x^*)$ are nonsingular, then there exists a neighborhood N of x^* such that when the initial vector x^0 is chosen in N, the entire sequence $\{x^k\}$ generated by (3.2) is well defined and converges to x^* Q-superlinearly. Furthermore, if $F'(y)$ is Lipschitz continuous around $\Pi_C(x^*)$, then the convergence is quadratic.

Proof. From Lemma 2.1 we know that there exists a neighborhood N of x^* such that $\mathcal{B}(x) \subseteq \mathcal{B}(x^*)$ and $\mathcal{P}(x) \subseteq \mathcal{P}(x^*)$ hold for all $x \in N$. So from the assumption that all $W_* \in \mathcal{W}(x^*)$ are nonsingular and the fact that there are finitely many elements in $\mathcal{P}(x^*)$, we know that there exists a positive number $\beta > 0$ such that

$$\|W^{-1}\| \leq \beta$$

for all $W \in \mathcal{W}(x)$, $x \in N$. So (3.2) is well defined for the first step.

When $x^k \in N$, $\mathcal{B}(x^k) \subseteq \mathcal{B}(x^*)$ holds. So from (ii) of Lemma 2.1 we have

$$\Pi_C(x^k) - \Pi_C(x^*) - P_k(x^k - x^*) = 0.$$

From $W_k s^k + H(x^k) = 0$ we have

$$W_k(x^{k+1} - x^*) + W_k(x^* - x^k) + H(x^k) = 0.$$

Therefore,

$$\begin{aligned}
\|x^{k+1} - x^*\| &\leq \beta \|H(x^k) - H(x^*) - W_k(x^k - x^*)\| \\
&= \beta \|F(\Pi_C(x^k)) + x^k - \Pi_C(x^k) - (F(\Pi_C(x^*)) + x^* - \Pi_C(x^*)) \\
&\quad -(F'(\Pi_C(x^k))P_k + I - P_k)(x^k - x^*)\| \\
&= \beta \|[F(\Pi_C(x^k)) - F(\Pi_C(x^*)) - F'(\Pi_C(x^k))P_k(x^k - x^*)] \\
&\quad +[x^k - x^* - I(x^k - x^*)] - [\Pi_C(x^k) - \Pi_C(x^*) - P_k(x^k - x^*)]\| \\
&= \beta \|F(\Pi_C(x^k)) - F(\Pi_C(x^*)) - F'(\Pi_C(x^k))P_k(x^k - x^*)\| \\
&= \beta \|F(\Pi_C(x^k)) - F(\Pi_C(x^*)) - F'(\Pi_C(x^k))(\Pi_C(x^k) - \Pi_C(x^*))\| \\
&= o(\|\Pi_C(x^k) - \Pi_C(x^*)\|).
\end{aligned}$$

From the property of the projection operator Π_C, we know that

$$\|\Pi_C(x^k) - \Pi_C(x^*)\| \leq \|x^k - x^*\|.$$

So

$$\|x^{k+1} - x^k\| \leq o(\|x^k - x^*\|).$$

If $F'(y)$ is Lipschitz continuous around $\Pi_C(x^*)$, then from the above formulas we can conclude that the convergence is quadratic. $\square$

For the assumption of nonsingularity of $W_* \in \mathcal{W}(x^*)$, we have the following result.

Proposition 3.1. Suppose that $V := F'(\Pi_C(x))$ is strictly copositive on the cone

$$\mathcal{C}(x; C) = \bigcup_K \{v|\ A_K v = 0,\ Bv = 0,\ K \in \mathcal{B}(x)\},$$

i.e.,

$$v^T V v > 0 \quad \forall v \in \mathcal{C}(x; C)\backslash 0, \tag{3.3}$$

then all $W \in \mathcal{W}(x)$ are nonsingular.

Proof. For $W \in \mathcal{W}(x)$, there exists $K \in \mathcal{B}(x)$ such that

$$W = VP + I - P,$$

where $P = I - (A_K^T\ B^T)\left(\begin{pmatrix} A_K \\ B \end{pmatrix}(A_K^T\ B^T)\right)^{-1}\begin{pmatrix} A_K \\ B \end{pmatrix}$ is an element of $\mathcal{P}(x)$.

Assume that v is such that

$$Wv = 0,$$

i.e.,

$$VPv + v - Pv = 0. \tag{3.4}$$

Multiplying $(Pv)^T$ in both sides of (3.4) and noting that $P^T = P$ and $P^2 = P$, we have

$$\begin{aligned} 0 &= (Pv)^T Pv + (Pv)^T v - (Pv)^T Pv \\ &= (Pv)^T V Pv + v^T Pv - v^T P^2 v \\ &= (Pv)^T V Pv + v^T Pv - v^T Pv \\ &= (Pv)^T V Pv. \end{aligned}$$

Therefore,

$$(Pv)^T V Pv = 0. \tag{3.5}$$

But

$$\begin{aligned} \begin{pmatrix} A_K \\ B \end{pmatrix} Pv &= \begin{pmatrix} A_K \\ B \end{pmatrix} v - \begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \; B^T) \left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \; B^T) \right)^{-1} \begin{pmatrix} A_K \\ B \end{pmatrix} v \\ &= \begin{pmatrix} A_K \\ B \end{pmatrix} v - \begin{pmatrix} A_K \\ B \end{pmatrix} v = 0, \end{aligned}$$

which means that

$$Pv \in \mathcal{C}(x; C).$$

From (3.3) and (3.5) we know that

$$Pv = 0.$$

Substituting this into (3.4) gives

$$v = 0,$$

which means that W is nonsingular. $\square$

Remark 3.1. In Proposition 3.1 we needn't the condition of the linear independence of the vectors

$$\{A_i^T,\ i \in I(x)\} \cup \{B_j^T,\ j = 1, ..., p\}.$$

If this linear independence condition is satisfied, then condition (3.3) is equivalent to Robinson's strong sufficiency condition [23], which is implied by the sufficiency condition and the strict complementarity condition (i.e., there exists no $i \in I(x)$ such that $\lambda_i = 0$, where $(\lambda, \mu) \in \mathcal{M}(x)$).

4. Quasi-Newton Method

Basing on the Newton method established in § 3, we can describe the quasi-Newton method for solving (1.1).

Quasi-Newton method (Broyden's case [1])
Given $x^0 \in \Re^n$, $D_0 \in \Re^{n \times n}$ (an approximation of $F'(\Pi_C(x^0))$)
Do for $k = 0, 1, ...$:

Choose $P_k \in \mathcal{P}(x^k)$ and compute

$$V_k := D_k P_k + I - P_k$$

Solve $V_k s + H(x^k) = 0$ for s^k

$$x^{k+1} = x^k + s^k$$

$$\delta^k = \Pi_C(x^{k+1}) - \Pi_C(x^k)$$
$$y^k = F(\Pi_C(x^{k+1})) - F(\Pi_C(x^k))$$
$$D_{k+1} = D_k + \frac{(y^k - D_k\delta^k){\delta^k}^T}{{\delta^k}^T\delta^k}.$$

Theorem 4.1. Suppose that $F : \Re^n \to \Re^n$ is continuously differentiable, x^* is a solution of (1.1), $F'(y)$ is Lipschitz continuous in a neighborhood of $\Pi_C(x^*)$ and the Lipschitz constant is γ. Suppose that all $W_* \in \mathcal{W}(x^*)$ are nonsingular. There exist positive constants ε, δ such that if $\|x^0 - x^*\| \le \varepsilon$ and $\|D_0 - F'(\Pi_C(x^*))\| \le \delta$, then the sequence $\{x^k\}$ generated by the above quasi-Newton method (Broyden's case) is well defined and converges Q-superlinearly to x^*.

Proof. From the proof of Theorem 3.1 we know that there exist a neighborhood $N_0(x^*)$ of x^* and a positive number $\beta > 0$ such that $\mathcal{B}(x) \subseteq \mathcal{B}(x^*)$ and $\|W^{-1}\| \le \beta$ for any $x \in N_0(x^*)$, $W \in \mathcal{W}(x)$.

Choose ε and δ such that

$$\mathcal{B}(x) \subseteq \mathcal{B}(x^*), \tag{4.1}$$

$$\|F'(\Pi_C(x)) - F'(\Pi_C(x^*))\| \le \gamma\|\Pi_C(x) - \Pi_C(x^*)\|, \tag{4.2}$$

$$7\beta\delta \le 1, \tag{4.3}$$

$$3\gamma\varepsilon \le 2\delta, \tag{4.4}$$

$$\|W^{-1}\| \le \beta, \tag{4.5}$$

$$\|F(\Pi_C(x)) - F(\Pi_C(x^*)) - F'(\Pi_C(x))(\Pi_C(x) - \Pi_C(x^*))\|$$

$$\le \frac{\delta}{2}\|\Pi_C(x) - \Pi_C(x^*)\| \tag{4.6}$$

for any $x \in N(x^*) := \{x| \ \|x - x^*\| \le \varepsilon\}$, $W \in \mathcal{W}(x)$. Denote $e^k = x^k - x^*$.

We will first prove that $\{x^k\}$ is locally Q-linearly convergent. The local Q -linear convergence proof consists of showing by induction that

$$\|D_k - F'(\Pi_C(x^*))\| \le (2 - 2^{-k})\delta, \tag{4.7}$$

$$\|V_k^{-1}\| \le \frac{7}{5}\beta, \tag{4.8}$$

$$\|e^{k+1}\| \le \frac{1}{2}\|e^k\|, \tag{4.9}$$

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

For $k = 0$, (4.7) is trivially true. The proof of (4.8) and (4.9) is identical to the proof at the induction step, so we omit it here.

Now assume that (4.7), (4.8), and (4.9) hold for $k = 0, 1, \ldots, i-1$. For $k = i$, we have from Dennis and Moré [5] (also see Lemma 8.2.1 of Dennis and Schnabel [6]), and the induction hypothesis that

$$\begin{aligned}\|D_i - F'(\Pi_C(x^*))\| &\le \|D_{i-1} - F'(\Pi_C(x^*))\| + \frac{\gamma}{2}(\|\Pi_C(x^i) - \Pi_C(x^*)\| \\ &\quad + \|\Pi_C(x^{i-1}) - \Pi_C(x^*)\|) \\ &\le (2 - 2^{-(i-1)})\delta + \frac{\gamma}{2}(\|e^i\| + \|e^{i-1}\|) \\ &\le (2 - 2^{-(i-1)})\delta + \frac{3}{4}\gamma\|e^{-(i-1)}\|. \end{aligned} \tag{4.10}$$

From (4.9) and $\|e^0\| \le \varepsilon$ we get

$$\|e^{-(i-1)}\| \le 2^{-(i-1)}\|e^0\| \le 2^{-(i-1)}\varepsilon.$$

Substituting this into (4.10), and using (4.4), gives

$$\begin{aligned}\|D_i - F'(\Pi_C(x^*))\| &\le (2 - 2^{-(i-1)})\delta + \frac{3}{4}\gamma \cdot 2^{-(i-1)}\varepsilon \\ &\le (2 - 2^{-(i-1)})\delta + 2^{-i}\delta \\ &= (2 - 2^{-i})\delta, \end{aligned}$$

which verifies (4.7).

To verify (4.8), we must first show that V_i is invertible. From the definition of V_i there exists $P_i \in \mathcal{P}(x^i)$ such that

$$V_i = D_i P_i + I - P_i.$$

Denote

$$W_i = F'(\Pi_C(x^i))P_i + I - P_i.$$

Then

$$W_i \in \mathcal{W}(x^i)$$

and

$$\begin{aligned}\|V_i - W_i\| &\le \|D_i - F'(\Pi_C(x^i))\|\|P_i\| \\ &\le \|D_i - F'(\Pi_C(x^i))\| \\ &\le \|D_i - F'(\Pi_C(x^*))\| + \|F'(\Pi_C(x^i)) - F'(\Pi_C(x^*))\|. \end{aligned} \tag{4.11}$$

Using (4.7) for $k = i$ and the Lipschitz condition (4.2) gives

$$\|V_i - W_i\| \le (2 - 2^{-i})\delta + \gamma\|\Pi_C(x^i) - \Pi_C(x^*)\|$$
$$\le (2 - 2^{-i})\delta + \gamma\|e^i\|. \tag{4.12}$$

From (4.9), $\|e^0\| \le \varepsilon$, and (4.4)

$$\gamma\|e^i\| \le 2^{-i}\varepsilon\gamma \le \frac{2}{3} \cdot 2^{-i}\delta,$$

which substituted into (4.12), gives

$$\|V_i - W_i\| \le (2 - 2^{-i})\delta + \frac{2}{3} \cdot 2^{-i}\delta \le 2\delta. \tag{4.13}$$

From (4.5), (4.13), and (4.3) we get

$$\|W_i^{-1}(W_i - V_i)\| \le \beta \cdot 2\delta \le 2/7 < 1.$$

So we have from Theorem 2.3.2 of Ortega and Rheinboldt [14] that V_i is invertible and

$$\|V_i^{-1}\| \le \frac{\|W_i^{-1}\|}{1 - \|W_i^{-1}(W_i - V_i)\|} \le \frac{\beta}{1 - 2/7} = \frac{7}{5}\beta,$$

which verifies (4.8).

To complete the induction, we verify (4.9). From $V_i(x^{i+1} - x^i) + H(x^i) = 0$ we have

$$V_i e^{i+1} = -H(x^i) + V_i e^i$$
$$= -(H(x^i) - H(x^*) - V_i e^i).$$

From Lemma 2.1 and (4.1), we know that

$$\Pi_C(x^i) - \Pi_C(x^*) = P_i(x^i - x^*). \tag{4.14}$$

Therefore,

$$\|e^{i+1}\| \le \|V_i^{-1}\|\|H(x^i) - H(x^*) - V_i e^i\|$$
$$= \|V_i^{-1}\|\|[F(\Pi_C(x^i)) - F(\Pi_C(x^*)) - D_i P_i(x^i - x^*)]$$
$$+[x^i - x^* - I(x^i - x^*)] - [\Pi_C(x^i) - \Pi_C(x^*) - P_i(x^i - x^*)]\|$$
$$= \|V_i^{-1}\|\|F(\Pi_C(x^i)) - F(\Pi_C(x^*)) - D_i(\Pi_C(x^i) - \Pi_C(x^*))\|$$
$$\le \|V_i^{-1}\|[\|F(\Pi_C(x^i)) - F(\Pi_C(x^*)) - F'(\Pi_C(x^i))(\Pi_C(x^i) - \Pi_C(x^*))\|$$
$$+\|(F'(\Pi_C(x^i)) - D_i)P_i(x^i - x^*)\|]$$
$$= \|V_i^{-1}\|[\|F(\Pi_C(x^i)) - F(\Pi_C(x^*)) - F'(\Pi_C(x^i))(\Pi_C(x^i) - \Pi_C(x^*))\|$$
$$+\|(W_i - V_i)(x^i - x^*)\|]. \tag{4.15}$$

From (4.15), (4.8), (4.6), (4.13), and (4.3) we get

$$\begin{aligned}||e^{i+1}|| &\leq \frac{7}{5}\beta[\frac{\delta}{2}||\Pi_C(x^i) - \Pi_C(x^*)|| + 2\delta||e_i||] \\ &\leq \frac{7}{5}\beta[\frac{\delta}{2}||e^i|| + 2\delta||e^i||] \\ &\leq \frac{1}{2}||e^i||.\end{aligned}$$

This proves (4.9) and completes the Q-linear convergence.

Next, we will prove the Q-superlinear convergence of $\{x^k\}$ under the assumptions. Let $E_k = D_k - F'(\Pi_C(x^*))$. From [5] or the last part of the proof of Theorem 8.2.2 of [6], we get

$$\lim_{k\to\infty} \frac{||E_k\delta^k||}{||\delta^k||} = 0. \tag{4.16}$$

So from (4.15), (4.8), (4.14), (4.7), (4.16), and (4.3),

$$\begin{aligned}||e^{k+1}|| \leq\; & \frac{7}{5}\beta[||F(\Pi_C(x^i)) - F(\Pi_C(x^*)) - F'(\Pi_C(x^i))(\Pi_C(x^i) - \Pi_C(x^*))|| \\ & +||(F'(\Pi_C(x^k)) - D_k)(\Pi_C(x^k) - \Pi_C(x^*))||] \\ \leq\; & o(||\Pi_C(x^k) - \Pi_C(x^*)||) + \frac{7}{5}\beta[||(D_k - F'(\Pi_C(x^*)))(\Pi_C(x^k) - \Pi_C(x^*))|| \\ & +||(F'(\Pi_C(x^k)) - F'(\Pi_C(x^*)))(\Pi_C(x^k) - \Pi_C(x^*))||] \\ \leq\; & o(||\Pi_C(x^k) - \Pi_C(x^*)||) + \frac{7}{5}\beta[||(D_k - F'(\Pi_C(x^*)))(\Pi_C(x^{k+1}) - \Pi_C(x^k))|| \\ & +||(D_k - F'(\Pi_C(x^*)))(\Pi_C(x^{k+1}) - \Pi_C(x^*))||] + o(||Pi_C(x^k) - \Pi_C(x^*)||) \\ \leq\; & o(||\Pi_C(x^k) - \Pi_C(x^*)||) + \frac{7}{5}\beta||E_k\delta^k|| + \frac{14}{5}\beta\delta||\Pi_C(x^{k+1}) - \Pi_C(x^*)|| \\ \leq\; & o(||e^k||) + o(||\delta^k||) + \frac{2}{5}||e^{k+1}|| \\ \leq\; & o(||e^k||) + o(||e^k||) + o(||e^{k+1}||) + \frac{2}{5}||e^{k+1}||,\end{aligned}$$

which means that

$$\lim_{k\to\infty} \frac{||e^{k+1}||}{||e^k||} = 0.$$

This completes the Q-superlinear convergence of $\{x^k\}$. □

5. Implementation Aspects

For implementing the Newton method established in this chapter, there is no much difference from the smooth case except for choosing the iterative matrices. For implementing the quasi-Newton method, there exist some differences from the smooth case, especially for the factorization of the iterative matrix V_k. The entire QR factorization of V_k costs $O(n^3)$ arithmetic operations. If we do this in per iteration, then the advantages of quasi-Newton methods lose a lot. In this section, we will discuss how to update the QR factorization of V_k into the QR factorization of V_{k+1} in much less than $O(n^3)$ operations.

Denote

$$\overline{V}_k = D_{k+1}P_k + I - P_k. \tag{5.1}$$

Then

$$\overline{V}_k = V_k + \frac{(y_k - D_k\delta^k){\delta^k}^T P_k}{{\delta^k}^T \delta^k} \tag{5.2}$$

and

$$V_{k+1} = \overline{V}_k + (D_{k+1} - I)(P_{k+1} - P_k). \tag{5.3}$$

It is well known that we can update the QR factorization of V_k into the QR factorization of $\overline{V}_k$ in $O(n^2)$ operations (see, e.g., [7, 8]).

According to the definition of P_k and P_{k+1}, there exist $K \in \mathcal{B}(x^k)$ and $\widetilde{K} \in \mathcal{B}(x^{k+1})$ such that

$$P_k = I - (A_K^T\ B^T)\left(\begin{pmatrix} A_K \\ B \end{pmatrix}(A_K^T\ B^T)\right)^{-1}\begin{pmatrix} A_K \\ B \end{pmatrix} \tag{5.4}$$

and

$$P_{k+1} = I - (A_{\widetilde{K}}^T\ B^T)\left(\begin{pmatrix} A_{\widetilde{K}} \\ B \end{pmatrix}(A_{\widetilde{K}}^T\ B^T)\right)^{-1}\begin{pmatrix} A_{\widetilde{K}} \\ B \end{pmatrix}. \tag{5.5}$$

Denote

$$\overline{K} = K \cap \widetilde{K},\ J_k = K \backslash \overline{K},\ \textbf{and}\ L_k = \widetilde{K} \backslash \overline{K}. \tag{5.6}$$

Define

$$\overline{P}_k = I - (A_{\overline{K}}^T\ B^T)\left(\begin{pmatrix} A_{\overline{K}} \\ B \end{pmatrix}(A_{\overline{K}}^T\ B^T)\right)^{-1}\begin{pmatrix} A_{\overline{K}} \\ B \end{pmatrix} \tag{5.7}$$

and

$$\overline{V}_{k+1} = \overline{V}_k + (D_{k+1} - I)(\overline{P}_k - P_k). \tag{5.8}$$

After simple computations, we can see that $(D_{k+1} - I)(\overline{P}_k - P_k)$ is at most a rank-$2|J_k|$ matrix and $(D_{k+1} - I)(P_{k+1} - \overline{P}_k)$ is at most a rank-$2|L_k|$ matrix. But from (5.3) and (5.8) we know that

$$V_{k+1} = \overline{V}_{k+1} + (D_{k+1} - I)(P_{k+1} - \overline{P}_k).$$

So we can update the QR factorization of $\overline{V}_k$ into the QR factorization of V_{k+1} in $O(2(|J_k| + |L_k|)n^2)$ operations (see, e.g., [7, 8]).

Therefore, we get

Theorem 5.1. The cost of updating the QR factorization of V_k into the QR factorization of V_{k+1} is at most $O((1 + 2|J_k| + 2|L_k|)n^2)$ arithmetic operations.

The numerical results will be reported separately later, since we feel that it is never an easy task to program a good numerical software. As the further research topics, we just mention two points: (i) when C has the general form $C = \{x| \ g_i(x) \le 0, \ i = 1, ..., m, \ h_j(x) = 0, \ j = 1, ..., p\}$, how to give the Newton and quasi-Newton methods accordingly; (ii) how to globalize the Newton and quasi-Newton methods established here.

References

[1] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations", *Mathematics of Computation* 19 (1965) 577-593.

[2] X. Chen and T. Yamamoto, "On the convergence of some quasi-Newton methods for solving nonlinear equations with nondifferentiable operators', *Computing* 49 (1992) 87-94.

[3] X. Chen and L. Qi, "A parameterized Newton method and a Broyden-like method for solving nonsmooth equations", *Computational Optimization and Applications* 3 (1994) 157-179.

[4] F.H. Clarke, *Optimization and Nonsmooth Analysis* (John Wiley and Sons, New York, 1983).

[5] J.E. Dennis and J.J. Moré, "A characterization of superlinear convergence and its application to quasi-Newton methods", *Mathematics of Computation* 28 (1974) 549-560.

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

[7] P.E. Gill and M. Murray, "Quasi-Newton methods for unconstrained optimization", *J. Inst. Maths Applics* 9 (1972) 91-108.

[8] G.H. Golub and C. Van Loan, *Matrix Computations* (the Johns Hopkins University Press, 1983).

[9] C.-M. Ip and T. Kyparisis, "Local convergence of quasi-Newton methods for B-differentiable equations", *Mathematical Programming* 56 (1992) 71-89.

[10] N.H. Josephy, "Newton's method for generalized equations", Technical Summary Report No. 1965, Mathematical Research Center, University of Wisconsin (Madison, WI, 1979).

[11] N.H. Josephy, "Quasi-Newton methods for generalized equations", Technical Summary Report No. 1966, Mathematical Research Center, University of Wisconsin (Madison, WI, 1979).

[12] M. Kojima and S. Shindo, "Extensions of Newton and quasi-Newton methods to systems of PC^1 equations", *Journal of the Operations Research Society of Japan* 29 (1986) 352-374.

[13] B. Kummer, "Newton's method for non-differentiable functions", in J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach and K. Zimmerman, eds., *Advances in Mathematical Programming* (Academi Verlag, Berlin, 1988) 114-125.

[14] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).

[15] J.-S. Pang, "The implicit complementarity problem", in O.L. Mangasarian, S.M. Robinson, and P.R. Meyer, eds., *Nonlinear Programming* 4 (Academic Press, New York, 1981) 487-518.

[16] J.-S. Pang, "Newton's method for B-differentiable equations, *Mathematics of Operations Research* 15 (1990) 311-341.

[17] J.-S. Pang, "A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems", *Mathematical Programming* 51 (1991) 101-131.

[18] J.-S. Pang and Daniel Ralph, "Piecewise smoothness, local invertibility, and parametric analysis of normal maps", Manuscript, Department of Mathematics, The University of Melbourne, Parkville, Victoria, Australia (1993).

[19] J.-S. Pang and L. Qi, "Nonsmooth equations: motivation and algorithms", *SIAM Journal on Optimization* 3 (1993) 443-465.

[20] L. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations", *Mathematics of Operations Research* 18 (1993) 227-244.

[21] L. Qi and J. Sun, "A nonsmooth version of Newton's method", *Mathematical programming* 58 (1993) 353-368.

[22] L. Qi and H. Jiang, "Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations", Applied Mathematics Report 94/5, School of Mathematics, The University of New South Wales, Sydney, Australia (Revised in November 1994).

[23] S.M. Robinson, "Strongly regular generalized equations", *Mathematics of Operations Research* 5 (1980) 43-62.

[24] S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity", *Mathematical Programming Study* 30 (1987) 45-66.

[25] S.M. Robinson, "Newton's methods for a class of nonsmooth functions", Industrial Engineering Working Paper, University of Wisconsin (Madison, USA, 1988).

[26] S.M. Robinson, "Normal maps induced by linear transformation", *Mathematics of Operations Research* 17 (1992) 691-714.

[27] S.M. Robinson, "Nonsingularity and symmetry for linear normal maps", *Mathematical Programming* 62 (1993) 415-425.

[28] D. Sun and J. Han, "Newton and quasi-Newton methods for a class of nonsmooth equations and related problems", Technical Report 94/026, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, P.R. China.

[29] E.H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory", in E.H. Zarantonello, ed., *Contributions to Nonlinear Functional Analysis* (Academic Press, New York, 1971).

Chapter 6

Safeguarded Newton Method for a Class of Nonlinear Projection Equations

Abstract

This chapter presents a globally and superlinearly convergent safeguarded Newton method for solving the projection equations $H(x) := x - \Pi_C[x - F(x)] = 0$, where C is a polyhedral set and F is locally Lipschitzian, semismooth over $\Re^n$, and pseudomonotone over C. In each step, the basic Newton method presented here needs to solve a linear equations, which is easier than to solve a linear complementarity problem and a linear variational inequality problem.

Chapter 6
Safeguarded Newton Method for a Class of Nonlinear
Projection Equations

1. Introduction

We consider the following nonlinear projection equations

$$H(x) := x - \Pi_C[x - F(x)] = 0, \tag{1.1}$$

where C is a closed convex set of $\Re^n$, Π_C is the Euclidean operator on C, and $F : \Re^n \to \Re^n$ is not Fréchet differentiable but locally Lipschitzian and semismooth over $\Re^n$. For the definition of semismoothness, see [22]. Such a class of nonlinear projection equations often arise in optimization and equilibrium analysis. For example, the variational inequality problem defined on C is to find $x \in C$ such that

$$(y - x)^T F(x) \geq 0 \quad \forall y \in C. \tag{1.2}$$

It is easy to see that x is a solution of (1.2) if and only if x is a solution of (1.1). Therefore the equations $H(x) = 0$ is an equivalent way of formulating the variational inequality problem (1.2).

When F is continuously differentiable, there are many kinds of Newton methods for solving (1.1) or (1.2); for examples, see [9, 26, 28, 10, 3, 12, 17, 20, 4, 32]. But when F is just semismooth, there are few results. Pang and Qi [18] considered the following linearly constrained convex minimization problem

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad x \in C, \end{aligned} \tag{1.3}$$

where $f : \Re^n \to \Re$ is a continuously differentiable function and C is a polyhedral set. When ∇f is semismooth over $\Re^n$, a globally and superlinearly convergent Newton method with a line search technique is obtained by Pang and Qi [18]. In [21], Qi and Jiang considered the trust region case, correspondingly. Problem (1.3), which is a special case of (1.2), includes stochastic quadratic programming problems [24] and minimax problems [18, 23]. In [8], Jiang and Qi generalized Josephy's Newton method [9, 26] for solving (1.2) to the case that F is semismooth. They proved the superlinear convergence of their Newton method. But no global convergence result is obtained. This arises a question: can a globally and superlinearly convergent method be obtained for solving (1.2) or (1.1) when F is monotone over $\Re^n$? Recently, Sun [31] obtained a class of globally convergent iterative methods for solving (1.1) when F is pseudomonotone over C. Then a natural way for solving (1.1) is to combine the globally convergent methods of [31] and the superlinearly convergent method of [8]. However, in per iteration Jiang and Qi's method [8] needs to solve a linear variational inequality subproblem defined on C. This subproblem is nonlinear and nonconvex, which may make it difficult to solve. So a more efficient Newton method for solving (1.1) or (1.2) is needed. When C is a polyhedral set, we will first give such a Newton method that in per iteration the subproblem needed to solve is a linear equations, which is relatively easy to solve, and then combine the new resulting Newton method with the global convergent method appeared in [31] to obtain a globally and superlinearly convergent method.

Denote

$$C^* = \{x \in C |\ x \text{ is a solution of } (1.1)\}. \tag{1.4}$$

Definition 1.1. The mapping $F : \Re^n \to \Re^n$ is said to

(i) be monotone over C if

$$[F(x) - F(y)]^T(x - y) \geq 0 \quad \forall\ x,\ y \in C; \tag{1.5}$$

(ii) be strongly monotone over C if there exists a constant $\mu > 0$ such that

$$[F(x) - F(y)]^T(x - y) \geq \mu \|x - y\|^2 \quad \forall\ x,\ y \in C; \tag{1.6}$$

(iii) be pseudomonotone over C if

$$F(x)^T(y - x) \geq 0 \text{ implies } F(y)^T(y - x) \geq 0 \quad \forall\ x,\ y \in C. \tag{1.7}$$

Suppose that $G : \Re^n \to \Re^m$ is locally Lipschitzian. By Rademacher's theorem, G is differentiable almost everywhere. Let D_G be the set where G is differentiable. Let ∂G be the generalized Jacobian of G in the sense of Clarke [2]. Then

$$\partial G(x) = \text{co } \{ \lim_{\substack{x^k \in D_G \\ x^k \to x}} G'(x^k)\}.$$

In order to reduce the nonsingularity assumption of the generalized Jacobian, $\partial_B G(x)$ was introduced in [17, 20]

$$\partial_B G(x) = \{ \lim_{\substack{x^k \in D_G \\ x^k \to x}} G'(x^k)\}. \tag{1.8}$$

Then

$$\partial G(x) = \text{co } \partial_B G(x).$$

In order to construct a class of quasi-Newton methods, $\partial_b G(x)$ was introduced in [32, 4]

$$\partial_b G(x) = \partial_B G_1(x) \times \partial_B G_2(x) \times \cdots \times \partial_B G_m(x). \tag{1.9}$$

Through similar analysis to that of [2], we can also know that $\partial_B G(x)$ and $\partial_b G(x)$ are nonempty compact sets of $\Re^{m\times n}$, and the mappings $\partial_B G(\cdot)$ and $\partial_b G(\cdot)$ are upper semi-continuous [1]. In the next sections, we will use ∂_Q to represent one of ∂, ∂_B, and ∂_b. Then $\partial_Q G(x)$ is a nonempty compact set of $\Re^{m\times n}$ and the mapping $\partial_Q G(\cdot)$ is upper semi-continuous.

In this chapter, unless other specified, we will assume that C has the form

$$C = \{x |\ Ax \leq a,\ Bx = b\}, \tag{1.10}$$

where $A : \Re^n \to \Re^m$, $B : \Re^n \to \Re^p$, $a \in \Re^m$, and $b \in \Re^p$. For the sake of simplicity, we will assume that rank $(B)=p$ $(p \leq n)$.

The rest of this chapter is organized as follows. In § 2, we describe a globally convergent method for solving (1.1) when F is pseudomonotone over C. In § 3, we give such a new Newton method for solving (1.1) that in per iteration only a linear equations is needed to solve. In § 4, by combining the globally convergent method given in § 2 and the superlinearly convergent Newton method given in § 3, we give a globally and superlinearly convergent method, which is called safeguarded Newton method, for solving (1.1) when F is pseudomonotone over C and H is b-regular at a solution point.

2. A Globally Convergent Method

In this section we will describe a globally convergent method recently obtained by Sun [31]. The search direction $g(x,\beta)$ (see (2.5)) used in this chapter is a special case of [31]. Nearly the same time, the search direction $g(x,\beta)$ also appeared in He [7]. In this section, C is not necessarily assumed to be a polyhedral set but a nonempty closed convex set.

Lemma 2.1 [34]. For the projection operator Π_C, we have
(i) when $y \in C$, $[z - \Pi_C(z)]^T[y - \Pi_C(z)] \leq 0 \quad \forall\, z \in \Re^n$;
(ii) $\|\Pi_C(z) - \Pi_C(y)\| \leq \|z - y\| \quad \forall y,\ z \in \Re^n$.

Define

$$E(x,\beta) = x - \Pi_C[x - \beta F(x)]. \tag{2.1}$$

When $\beta = 1$, $E(x,1) = H(x)$.

Choose an arbitrary constant $\eta \in (0,1)$. When $x \notin C^*$, define

$$\eta(x) = \begin{cases} \max\{\eta, 1 - \dfrac{t(x)}{\|E(x,1)\|^2}\}, & \text{if } t(x) > 0 \\ 1, & \text{otherwise} \end{cases} \tag{2.2}$$

and

$$s(x) = \begin{cases} (1-\eta(x))\dfrac{\|E(x,1)\|^2}{t(x)}, & \text{if } t(x) > 0 \\ 1, & \text{otherwise} \end{cases}, \tag{2.3}$$

where $t(x) = \{F(x) - F(\Pi_C[x - F(x)])\}^T E(x,1)$.

Lemma 2.2. Suppose that F is continuous over $\Re^n$ and $\eta \in (0,1)$ is a constant. If $S \subseteq \Re^n \backslash C^*$ is a compact set, then there exists a positive $\delta (\leq 1)$ such that for all $x \in S$ and $\beta \in (0,\delta]$, when $s(x) < 1$, we have

$$\{F(x) - F(\Pi_C[x - \beta F(x)])\}^T E(x,\beta) \leq (1-\eta(x))\|E(x,\beta)\|^2/\beta. \tag{2.4}$$

Proof. Similar to the proof of Theorem 2.1 of [31]. $\square$

Define

$$g(x,\beta) = F(\Pi_C[x - \beta F(x)]) - F(x) + E(x,\beta)/\beta. \tag{2.5}$$

Then we can describe a globally convergent method appeared in [31].

Projection and Contraction (PC) Method

Step 0. Choose an arbitrary vector $x^0 \in \Re^n$ (in [31], x^0 is chosen in C). Choose positive constants η, $\alpha \in (0,\ 1)$, $0 < \Delta_1 \le \Delta_2 < 2$. $k := 0$, go to step 1.

Step 1. Calculate $\eta(x^k)$ and $s(x^k)$. If $s(x^k) = 1$, let $\beta_k = 1$; otherwise determine $\beta_k = s(x^k)\alpha^{m_k}$, where m_k is the smallest nonnegative integer m such that

$$\begin{aligned}\{F(x^k) - F(\Pi_X[x^k - s(x^k)\alpha^m F(x^k)])\}^T E(x^k, s(x^k)\alpha^m)\\ \le (1-\eta(x^k))\|E(x^k, s(x^k)\alpha^m)\|^2/(s(x^k)\alpha^m)\end{aligned} \tag{2.6}$$

holds.

Step 2. Calculate $g(x^k,\beta_k)$.

Step 3. Calculate

$$\rho_k = E(x^k,\beta_k)^T g(x^k,\beta_k)/\|g(x^k,\beta_k)\|^2. \tag{2.7}$$

Step 4. Take $\gamma_k \in [\Delta_1,\ \Delta_2]$ and set

$$x^{k+1} = \Pi_C[x^k - \gamma_k\rho_k g(x^k,\beta_k)]. \tag{2.8}$$

$k := k+1$, go to step 1.

Remark 2.1. If in (2.8) we just take

$$x^{k+1} = x^k - \gamma_k\rho_k g(x^k,\beta_k),$$

then the following Theorem 2.1 also holds.

Suppose that $x^* \in C^*$. By taking $z = x^k - \beta_k F(x^k)$ and $y = x^*$ in (i) of Lemma 2.1, we have

$$\{x^* - \Pi_C[x^k - \beta_k F(x^k)]\}^T\{x^k - \beta_k F(x^k) - \Pi_C[x^k - \beta_k F(x^k)]\} \le 0.$$

Therefore,

$$(x^k - x^*)^T E(x^k,\beta_k) \ge \beta_k\{\Pi_C[x^k - \beta_k F(x^k)] - x^*\}^T F(x^k) + \|E(x^k,\beta_k)\|^2.$$

So if F is pseudomonotone over C, then

$$\{\Pi_C[x^k - \beta_k F(x^k)] - x^*\}^T F(\Pi_C[x^k - \beta_k F(x^k)]) \ge 0$$

and

$$
\begin{aligned}
&(x^k - x^*)^T g(x^k, \beta_k) \\
&\quad = (x^k - x^*)^T F(\Pi_C[x^k - \beta_k F(x^k)]) - (x^k - x^*)^T F(x^k) \\
&\qquad + (x^k - x^*)^T E(x^k, \beta_k)/\beta_k \\
&\quad \geq (x^k - x^*)^T F(\Pi_C[x^k - \beta_k F(x^k)]) - (x^k - x^*)^T F(x^k) \\
&\qquad + \{\Pi_C[x - \beta_k F(x^k)] - x^*\}^T F(x^k) + \|E(x^k, \beta_k)\|^2/\beta_k \\
&\quad \geq E(x^k, \beta_k)^T F(\Pi_C[x^k - \beta_k F(x^k)]) - E(x^k, \beta_k)^T F(x^k) + \|E(x^k, \beta_k)\|^2/\beta_k \\
&\quad = E(x^k, \beta_k)^T g(x^k, \beta_k) \\
&\quad \geq \eta(x^k)\|E(x^k, \beta_k)\|^2/\beta_k.
\end{aligned}
$$

Therefore, we can get

Theorem 2.1 [31]. Suppose that F is continuous over $\Re^n$ and pseudomonotone over C. If $C^* \neq \emptyset$, then the infinite sequence $\{x^k\}$ generated by the above PC method converges to a solution of (1.1).

When C is of the following form

$$C = \{x \in \Re^n |\ l \leq x \leq u\}, \tag{2.9}$$

where l and u are two vectors of $\{R \cup \{\infty\}\}^n$, we can give an improved form of the PC method. For any $x \in \Re^n$ and $\beta > 0$, denote

$$
\begin{aligned}
N(x,\beta) &= \{i |\ (x_i \leq l_i \text{ and } (g(x,\beta))_i \geq 0) \text{ or } (x_i \geq u_i \text{ and } (g(x,\beta))_i \leq 0)\}, \\
B(x,\beta) &= \{1, ..., n\} \backslash N(x,\beta).
\end{aligned} \tag{2.10}
$$

Denote $g_N(x,\beta)$ and $g_B(x,\beta)$ as follows

$$
\begin{aligned}
(g_N(x,\beta))_i &= \begin{cases} 0, & \text{if } i \in B(x,\beta) \\ (g(x,\beta))_i, & \text{otherwise} \end{cases}, \\
(g_B(x,\beta))_i &= (g(x,\beta))_i - (g_N(x,\beta))_i, \quad i = 1, ..., n.
\end{aligned} \tag{2.11}
$$

Then for any $x^* \in C^*$ and $x \in \Re^n$,

$$(x - x^*)^T g_B(x,\beta) \geq (x - x^*)^T g(x,\beta). \tag{2.12}$$

So if in the PC method we set

$$x^{k+1} = \Pi_C[x^k - \gamma_k \bar{\rho}_k g_B(x^k, \beta_k)] \tag{2.13}$$

or

$$x^{k+1} = x^k - \gamma_k \bar{\rho}_k g_B(x^k, \beta_k), \tag{2.14}$$

where

$$\bar{\rho}_k = E(x^k, \beta_k)^T g(x^k, \beta_k) / \|g_B(x^k, \beta_k)\|^2,$$

then the convergence Theorem 2.1 holds for the modified PC method. In practice, we will use the iterative form (2.13) or (2.14) when C is of the form (2.9).

3. Superlinearly Convergent Newton Method

Suppose that $G : \Re^n \to \Re^m$ is locally Lipschitzian. G is said to be semismooth at $x \in \Re^n$ if the following limit exists for any $h \in \Re^n$

$$\lim_{\substack{V \in \partial F(x+th') \\ h' \to h,\ t \downarrow 0}} \{Vh'\}. \tag{3.1}$$

If G is semismooth at x, then G is directionally differentiable at x and $G'(x;h)$ is equal to the limit in (3.1). Semismoothness was first introduced by Mifflin [13] for functionals. In [22, 20], the definition of semismoothness was extended to $G : \Re^n \to \Re^m$. It was proved in [20] that G is semismooth at x if and only if each of its components is semismooth at x.

If $G : \Re^n \to \Re^m$ is semismooth at x, then we have
(i) G is B-differentiable [27] at x, i.e., $G'(x;h)$ exists for all $h \in \Re^n$ and

$$G(x+h) = G(x) + G'(x;h) + o(\|h\|) \tag{3.2}$$

as $h \to 0$. See (2.2) of [20]
(ii) For any $V \in \partial G(x+h)$, $h \to 0$,

$$Vh - G'(x;h) = 0(\|h\|). \tag{3.3}$$

See Theorem 2.3 of [22]

In the following sections we will assume that C has the form (1.10). For any $z \in \Re^n$, $\Pi_C(z)$ is the Euclidean projection of z on C, then there exist multipliers $\lambda \in \Re^m_+$, $\mu \in \Re^p$ such that

$$\begin{cases} \Pi_C(z) - z + A^T\lambda + B^T\mu = 0, \\ \lambda \geq 0,\ a - A\Pi_C(z) \geq 0,\ \lambda^T(a - A\Pi_C(z)) = 0, \\ b - B\Pi_C(z) = 0. \end{cases} \tag{3.4}$$

Let $\mathcal{M}(z)$ denote the nonempty set of multipliers $(\lambda,\ \mu) \in \Re^m_+ \times \Re^p$ that satisfy the Karush-Kuhn-Tucker (K-K-T) conditions (3.4). For a nonnegative vector $d \in \Re^m$, we shall let $supp(d)$, called the support of d, be the subset of $\{1, ..., m\}$ consisting of the indexes i for $d_i > 0$. Denote

$$I(z) = \{i |\ A_i\Pi_C(z) = a_i,\ i = 1, ..., m\}. \tag{3.5}$$

Define the family $\mathcal{B}(z)$ of indexes of $\{1, ..., m\}$ as follows: $K \in \mathcal{B}(z)$ if and only if $supp(\lambda) \subseteq K \subseteq I(z)$ for some $(\lambda, \mu) \in \mathcal{M}(z)$ and the vectors

$$\{A_i^T, \ i \in K\} \cup \{B_j^T, \ j = 1, ..., p\} \tag{3.6}$$

are linearly independent. This family $\mathcal{B}(z)$ is nonempty because $\mathcal{M}(z)$ has an extreme point which easily yields a desired index set K with the stated properties.

Define

$$\mathcal{P}(z) = \{P \in \Re^{n \times n} | \ P = I - (A_K^T \ B^T) \left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \ B^T) \right)^{-1} \begin{pmatrix} A_K \\ B \end{pmatrix}, \ K \in \mathcal{B}(z)\}, \tag{3.7}$$

where I is the unit matrix of $\Re^{n \times n}$ and A_K is the matrix consisting of the K rows of A.

Remark 3.1. The existence of $\left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \ B^T) \right)^{-1}$ comes from the linear independence of the vectors $\{A_i^T, \ i \in K\} \cup \{B_j^T, \ j = 1, ..., p\}$. Note that for all $P \in \mathcal{P}(z)$, we have $P^T = P$, $P^2 = P$, and $\|P\| \leq 1$. These simple facts will be used later.

Combining the results of [19] and the K-K-T conditions (3.4), we get

Lemma 3.1 [5]. (i) There exists a neighborhood $N(z)$ of z such that when $y \in N(z)$, we have

$$\mathcal{B}(y) \subseteq \mathcal{B}(z) \text{ and } \mathcal{P}(y) \subseteq \mathcal{P}(z);$$

(ii) when $\mathcal{B}(y) \subseteq \mathcal{B}(z)$, $\Pi_C(y) = \Pi_C(z) + P(y - z) \quad \forall \ P \in \mathcal{P}(y)$.

Denote

$$\mathcal{W}(x) = \{W \in \Re^{n \times n} | \ W = I - P(I - V), \ P \in \mathcal{P}(x - F(x)), \ V \in \partial_Q F(x)\}. \tag{3.8}$$

From the facts that $\partial_Q F(x)$ and $\mathcal{P}(x - F(x))$ are nonempty compact sets, and the mappings $\partial_Q F(\cdot)$ and $\mathcal{P}(\cdot)$ are upper semi-continuous, we know that $\mathcal{M}(x)$ is a nonempty compact set and the mapping $\mathcal{M}(\cdot)$ is upper semi-continuous.

The Newton method for solving (1.1) can be described as following:

Given $x^0 \in \Re^n$.
Do for $k = 0, 1, ...$:
Choose $V_k \in \partial_Q F(x^k)$, $P_k \in \mathcal{P}(x^k - F(x^k))$, and compute

$$W_k := I - P_k(I - V_k) \in \mathcal{W}(x^k)$$

Solve

$$W_k s + H(x^k) = 0$$

for s^k

$$x^{k+1} = x^k + s^k. \tag{3.9}$$

From the above Newton method we know that at the k-th step one needs to solve a linear equations while in [8], one needs to solve the following nonlinear subproblem

$$x - \Pi_C[x - [F(x^k) + V_k(x - x^k)]] = 0$$

to get x^{k+1}.

Theorem 3.1. Suppose that $F : \Re^n \to \Re^n$ is locally Lipschitzian and semismmooth at x^*, C is of the form (1.10), and x^* is a solution of (1.1). If all $W_* \in \mathcal{W}(x^*)$ are nonsingular, then there exists a neighborhood N of x^* such that when the initial vector x^0 is chosen in N, then the entire sequence $\{x^k\}$ generated by (3.9) is well defined and converges to x^* Q-superlinearly.

Proof. From Lemma 3.1 we know that there exists a neighborhood N of x^* such that $\mathcal{B}(x - F(x)) \subseteq \mathcal{B}(x^* - F(x^*))$ and $\mathcal{P}(x - F(x)) \subseteq \mathcal{P}(x^* - F(x^*))$ hold for all $x \in N$. So from the assumption that all $W_* \in \mathcal{W}(x^*)$ are nonsingular and the facts that $\mathcal{W}(x^*)$ is a nonempty compact set and the mapping $\mathcal{W}(\cdot)$ is upper semi-continuous, we know that there exists a positive number $\beta > 0$ such that

$$\|W^{-1}\| \le \beta$$

for all $W \in \mathcal{W}(x)$, $x \in N$. So (3.9) is well defined for the first step.

When $x^k \in N$, $\mathcal{B}(x^k - F(x^k)) \subseteq \mathcal{B}(x^* - F(x^*))$ holds. So from (ii) of Lemma 3.1 we get

$$\Pi_C[x^k - F(x^k)] - \Pi_C[x^* - F(x^*)] = P_k[x^k - F(x^k) - (x^* - F(x^*))]. \tag{3.10}$$

From $W_k s^k + H(x^k) = 0$ we have

$$W_k(x^{k+1} - x^*) + W_k(x^* - x^k) + H(x^k) = 0.$$

Therefore,

$$\begin{aligned}
\|x^{k+1} - x^*\| &\le \beta\|H(x^k) - H(x^*) - W_k(x^k - x^*)\| \\
&= \beta\{\|[x^k - x^* - I(x^k - x^*)] - [\Pi_C[x^k - F(x^k)] - \Pi_C[x^* - F(x^*)] \\
&\quad - P_k(I - V_k)(x^k - x^*)]\|\} \qquad (3.11) \\
&= \beta\|P_k[F(x^k) - F(x^*) - V_k(x^k - x^*)]\| \\
&\le \beta\|F(x^k) - F(x^*) - V_k(x^k - x^*)\|.
\end{aligned}$$

Since F is semismooth at x^*, each of its components is semismooth at x^* also. So from (3.1) and (3.2) we know that for any $V \in \partial_Q F(x^* + h)$

$$F(x^* + h) - F(x^*) - Vh = o(\|h\|) \tag{3.12}$$

as $h \to 0$.

Combining (3.11) and (3.12), we have

$$\|x^{k+1} - x^*\| = o(\|x^k - x^*\|),$$

which completes the proof of the Q-superlinear convergence of $\{x^k\}$. $\square$

For the assumption of nonsingularity of $W_* \in \mathcal{W}(x^*)$, we have the following result.

Proposition 3.1. Suppose that $V \in \partial_Q F(x)$ is strictly copositive on the cone

$$\mathcal{C}(x;C) = \bigcup_K \{v|\ A_K v = 0,\ Bv = 0,\ K \in \mathcal{B}(x - F(x))\},$$

i.e.,

$$v^T V v > 0 \quad \forall v \in \mathcal{C}(x;C)\backslash 0, \tag{3.13}$$

then for any $P \in \mathcal{P}(x - F(x))$, the matrix

$$W := I - P(I - V) \in \mathcal{W}(x)$$

is nonsingular.

Proof. For any $P \in \mathcal{P}(x - F(x))$, there exists $K \in \mathcal{B}(x - F(x))$ such that

$$P = I - (A_K^T\ B^T)\left(\begin{pmatrix} A_K \\ B \end{pmatrix}(A_K^T\ B^T)\right)^{-1}\begin{pmatrix} A_K \\ B \end{pmatrix}.$$

Assume that v is such that

$$Wv = 0,$$

i.e.,

$$v - P(I - V)v = 0. \tag{3.14}$$

Multiplying $(PVv)^T$ in both sides of (3.14) and noting that $P^T = P$ and $P^2 = P$, we have

$$\begin{aligned} 0 &= (PVv)^T v - (PVv)^T P(I - V)v \\ &= v^T V^T P v - v^T V^T P^2 v + (PVv)^T(PVv) \\ &= v^T V^T P v - v^T V^T P v + (PVv)^T(PVv) \\ &= (PVv)^T PVv, \end{aligned}$$

i.e.,

$$PVv = 0. \tag{3.15}$$

Substituting (3.15) into (3.14) gives

$$v = Pv. \tag{3.16}$$

But

$$\begin{pmatrix} A_K \\ B \end{pmatrix} Pv = \begin{pmatrix} A_K \\ B \end{pmatrix} v - \begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \; B^T) \left(\begin{pmatrix} A_K \\ B \end{pmatrix} (A_K^T \; B^T) \right)^{-1} \begin{pmatrix} A_K \\ B \end{pmatrix} v$$

$$= \begin{pmatrix} A_K \\ B \end{pmatrix} v - \begin{pmatrix} A_K \\ B \end{pmatrix} v = 0,$$

which, and (3.16), means that

$$\begin{pmatrix} A_K \\ B \end{pmatrix} v = 0. \tag{3.17}$$

Using (3.14) and (3.15) we get

$$v^T V v = v^T V P v - v^T V P V v$$

$$= 0 - 0 = 0,$$

From this, (3.13), and (3.17), we know that

$$v = 0,$$

which shows that W is nonsingular. □

Remark 3.2. Proposition 3.1 states that if (3.13) is satisfied at x^* for all $V_* \in \partial_Q F(x^*)$, then all $W_* \in \mathcal{W}(x^*)$ are nonsingular.

Remark 3.3. When $C = \Re^n_+$ and $F \in C^2$, the nonsingularity assumption of $W_* \in \mathcal{W}(x^*)$ is equivalent to the b-regular assumption [16]. For this sake, in the following we will say that H is b-regular at x^* if all $W \in \mathcal{W}(x^*)$ are nonsingular.

Proposition 3.2. Suppose that F is locally Lipschitzian, semismooth, and b-regular at the solution point x^*. Then there exist positive constants $c_1, \; c_2 \in (0, \; \infty)$ and a neighborhood N of x^* such that

$$c_1\|x - x^*\| \leq \|H(x)\| \leq c_2\|x - x^*\|. \tag{3.18}$$

Proof. Since F is locally Lipschitzian at x^*, there exist a neighborhood N_1 of x^* and a constant $c \in (0, \; \infty)$ such that

$$\|F(x) - F(x^*)\| \leq c\|x - x^*\|.$$

Therefore,

$$\begin{aligned} \|H(x)\| &= \|H(x) - H(x^*)\| \\ &= \|x - \Pi_C[x - F(x)] - (x^* - \Pi_C[x^* - F(x^*)])\| \\ &\leq \|x - x^*\| + \|x - F(x) - (x^* - F(x^*))\| \\ &\leq (2 + c)\|x - x^*\|. \end{aligned}$$

Take $c_2 = 2 + c$, then

$$\|H(x)\| \leq c_2\|x - x^*\|. \tag{3.19}$$

According to Lemma 3.1, there exists a neighborhood $N_2(\subseteq N_1)$ of x^* such that

$$\|W^{-1}\| \leq \beta$$

and

$$\Pi_C[x - F(x)] - \Pi_C[x^* - F(x^*)] = P[x - F(x) - (x^* - F(x^*)]$$

for any $W \in \mathcal{W}(x)$, $P \in \mathcal{P}(x - F(x))$, and $x \in N_2$.

So

$$\begin{aligned} H(x) &= x - \Pi_C[x - F(x)] - (x^* - \Pi_C[x^* - F(x^*)]) \\ &= I(x - x^*) - P[x - F(x) - (x^* - F(x^*))] \\ &= [I - P(I - V)](x - x^*) + P[F(x) - F(x^*) - V(x - x^*)], \end{aligned}$$

where $P \in \mathcal{P}(x - F(x))$ and $V \in \partial_Q F(x)$.

But

$$F(x) - F(x^*) - V(x - x^*) = o(\|x - x^*\|).$$

So we can choose a neighborhood $N(\subseteq N_2)$ of x^* such that

$$\|F(x) - F(x^*) - V(x - x^*)\| \leq \frac{1}{2\beta}\|x - x^*\|.$$

Define

$$W := I - P(I - V) \in \mathcal{W}(x).$$

So for all $x \in N$,

$$\begin{aligned} \|H(x)\| &\geq \frac{1}{\beta}\|x - x^*\| - \frac{1}{2\beta}\|x - x^*\| \\ &= \frac{1}{2\beta}\|x - x^*\|. \end{aligned}$$

Take $c_1 = \frac{1}{2\beta}$, then

$$c_1\|x - x^*\| \leq \|H(x)\|. \tag{3.20}$$

So (3.19) and (3.20) make that (3.18) holds. □

Theorem 3.1 discussed the locally superlinear convergence of the Newton method established in this section. Next we will discuss a global technique for the Newton method.

Suppose that F is locally Lipschitzian and semismooth over $\Re^n$. Define

$$r(x) = \frac{1}{2}H(x)^T H(x).$$

Then according to the chain rule, we know that r is directionally differentiable and

$$r'(x;d) = H(x)^T H'(x;d), \tag{3.21}$$

where $H'(x;d) = d - \Pi'_C(x - F(x); d - F'(x;d))$. For the explicit description of $\Pi'_C(\cdot;\cdot)$, see [14]. It is easy to see that if there is a direction of d such that $r'(x;d) < 0$ and $H(x) \neq 0$, then for a given scalar $\sigma \in (0,\ 1/2)$ there exists a positive constant δ such that

$$r(x + td) \leq r(x) + \sigma t r'(x;d).$$

holds for all $t \in [0,\ \delta]$.

This, and Theorem 3.1, Proposition 3.2, stimulates us to give the following modification of the basic Newton method.

Newton Method with Line Search

Step 0. Choose an arbitrary vector $x^0 \in \Re^n$. Choose scalars α and σ with $\alpha \in (0,1)$ and $\sigma \in (0,1/2)$. $k := 0$, go to step 1.

Step 1. Choose $V_k \in \partial_Q F(x^k)$, $P_k \in \mathcal{P}(x^k - F(x^k))$, and compute

$$W_k := I - P_k(I - V_k) \in \mathcal{W}(x^k).$$

Step 2. If W_k is singular, stop; otherwise solve

$$W_k s + H(x^k) = 0$$

for s^k. If $r(x^k + s^k) \leq (1-\sigma) r(x^k)$, let $x^{k+1} = x^k + s^k$, $k := k+1$, go to step 1; otherwise, go to step 3.

Step 3. If $r'(x^k; s^k) < 0$, let $d^k = s^k$ and go to step 5; otherwise go to sep 4.

Step 4. If $r'(x^k; -s^k) < 0$, let $d^k = -s^k$ and go to step 5; otherwise, stop.

Step 5. Let $\beta^k = \alpha^{m_k}$, where m_k is the first nonnegative integer m for which

$$r(x^k + \alpha^m d^k) \leq r(x^k) + \sigma \alpha^m r'(x^k; d^k).$$

Set $x^{k+1} = x^k + \beta^k d^k$ and $k := k+1$. Go to step 1.

Remark 3.4. In the above method, the search direction d^k is obtained by solving a linear equations while in [16], d^k is obtained by solving the following equations

$$H(x^k) + H'(x^k; d) = 0,$$

which is a nonlinear and nonconvex subproblem in general.

Due to the nonsmoothness of $H(x)$, we can't expect the above algorithm to have such a global convergent property that every accumulation point of the infinite sequence $\{x^k\}$ is a solution of $H(x) = 0$. But when F has some monotonicity condition, we can combine the PC method described in § 2 and the modified Newton method to obtain a globally and superlinearly convergent method. Such a method will be discussed in § 4.

4. Safeguarded Newton Method

The Newton method with line search established in § 3 may lose global convergence, although it has locally superlinear convergence. When F is pseudomonotone over C, a practical way to get global and locally superlinear convergence is to combine the globally convergent method introduced in § 2 and the Newton method introduced in § 3. In this section we will give such a method. Suppose that F is locally Lipschitzian, semismooth over $\Re^n$, and pseudomonotone over C.

Safeguarded Newton Method

Step 0. Choose an arbitrary vector $x^0 \in \Re^n$. Choose scalars η, $\alpha, \gamma, \varepsilon_0 \in (0,1)$, $\sigma \in (0, 1/2)$, and $0 < \Delta_1 \leq \Delta_2 < 2$. $k := 0$, go to step 1.

Step 1. Choose $V_k \in \partial_Q F(x^k)$, $P_k \in \mathcal{P}(x^k - F(x^k))$, and compute

$$W_k := I - P_k(I - V_k) \in \mathcal{W}(x^k).$$

Step 2. If W_k is singular, go to step 6; otherwise solve

$$W_k s + H(x^k) = 0$$

for s^k. If

$$r(x^k + s^k) \leq (1-\sigma) r(x^k), \tag{4.1}$$

let $x^{k-1} = x^k - s^k$, $k := k - 1$, go to step 1; otherwise, go to step 3.

Step 3. If $r'(x^k; s^k) < -\varepsilon_0 r(x^k)$, let $d^k = s^k$ and go to step 5; otherwise go to sep 4.

Step 4. If $r'(x^k; -s^k) < -\varepsilon_0 r(x^k)$, let $d^k = -s^k$ and go to step 5; otherwise, go to step 6.

Step 5. (safeguarding step) Let $\beta^k = \alpha^{m_k}$, where m_k is the first nonnegative integer m such that

$$r(x^k - \alpha^m d^k) \leq r(x^k) + \sigma \alpha^m r'(x^k; d^k)$$

or

$$\alpha^m \leq \gamma$$

holds.
If $\beta^k \geq \gamma$, let $x^{k-1} = x^k - \beta_k d^k$, $k := k+1$, and go to step 1; otherwise, go to step 6.

Step 6. Set $y^0 = x^k$ and $i := 0$. Take y^0 as the initial vector and use PC method established in § 2 until to get a sequence $\{y^0, y^1, ..., y^{i(k)}\}$ such that $i(k)$ is the first positive integer i such that

$$r(y^i) \leq (1-\sigma) r(x^k).$$

Set $x^{k-1} = y^{i(k)}$ and $k := k+1$. Go to step 1.

Before giving the convergence theorem, we make several remarks.

Remark 4.1. We use the safeguarding step because H is not continuously differentiable.
Remark 4.2. The pseudomonotonicity assumption of F is used only when the Newton step fails.
Remark 4.3. Step 6 is guaranteed by Theorem 2.1.

Theorem 4.1. Let F be locally Lipschitzian and semismooth over $\Re^n$. Suppose that F is pseudomonotone over C, $C^* \neq \emptyset$, and $C_0 := \{x | \; r(x) \leq r(x^0)\}$ is bounded. Then the sequence $\{x^k\}$ generated by the above safeguarded Newton method is well defined and every accumulation point of $\{x^k\}$ is a solution of (1.1). Furthermore, if H is b-regular at an accumulation point $\bar{x}$ (i.e., all $W \in \mathcal{W}(\bar{x})$ are nonsingular), then $\{x^k\}$ converges to $\bar{x}$ Q-superlinearly.

Proof. According to the safeguarded Newton method, we have

$$\begin{aligned} r(x^{k+1}) &\le (1-\sigma\gamma\varepsilon_0)r(x^k) \\ &\le (1-\sigma\gamma\varepsilon_0)^{k+1}r(x^0). \end{aligned}$$

Therefore,

$$\lim_{k\to\infty} r(x^k) = 0. \tag{4.2}$$

From (4.2) and the boundedness of C_0, we know that $\{x^k\}$ is bounded and every accumulation point of $\{x^k\}$ is a solution of (1.1).

Furthermore, if $\bar{x}$ is an accumulation point of $\{x^k\}$ and H is b-regular at $\bar{x}$, then according to Proposition 3.2 and Theorem 3.1, when x^k is close enough to $\bar{x}$, (4.1) is satisfied and the full Newton step will be taken. So according to Theorem 3.1, the sequence $\{x^k\}$ will converge to $\bar{x}$ Q-superlinearly. □

When $F \in C^1$, by using a differentiable merit function [3], Taji, Fukushima, and Ibaraki [33] established a globally convergent Newton method for solving strongly variational inequality problems. In each step, their methods need to solve a linear variational inequality problem or a linear complementarity problem, but not a linear equations. This can be seen clearly if we take $C = \Re^n_+$. The quadratic convergence is established under the generalized strict complementarity condition, which is somewhat restrictive. It is not clear if the method of [33] can be generalized to the case that F is not differentiable but semismooth. From Proposition 3.1 and Theorem 4.1 we know that if F is strongly monotone over C and semismooth over $\Re^n$, then the iterative sequence $\{x^k\}$ will converge to the unique solution of (1.1) Q-superlinearly.

References

[1] J.P. Aubin and H. Frankowska, *Set-Valued Analysis* (Birkhaüser, Boston, 1990).

[2] F.H. Clarke, *Optimization and Nonsmooth Analysis* (John Wiley and Sons, New York, 1983).

[3] M. Fukushima, "Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems", *Mathematical Programming* 53 (1992) 99-110.

[4] J. Han and D. Sun, "Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity", in D.Z. Du, L. Qi and R.S. Womersley, eds., *Recent Advances in Nonsmooth Optimization* (World Scientific Publishing Co., New Jersy, 1995), forthcoming.

[5] J. Han and D. Sun, "Newton and quasi-Newton methods for normal maps with polyhedral set", manuscript, Institute of Applied Mathematics, Academia Sinica, Beijing, China (November 1994).

[6] P.T. Harker and J.-S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications", *Mathematical Programming* 48 (1990) 339-357.

[7] B.S. He, "A class of projection and contraction methods for monotone variational inequalities", Reports of Institute of Mathematics 94-11, Nanjing University, Nanjing, China (1994).

[8] H. Jiang and L. Qi, "Local Uniqueness and Newton-type methods for nonsmooth variational inequalities", AM 93/14, Applied Mathematics Preprint, The University of New South Wales, Australia (1993) (under revision).

[9] N.H. Josephy, "Newton's method for generalized equations", Technical Summary Report No. 1965, Mathematical Research Center, University of Wisconsin (Madison, WI, 1979).

[10] M. Kojima and S. Shindo, "Extensions of Newton and quasi-Newton methods to systems of PC^1 equations", *Journal of the Operations Research Society of Japan* 29 (1986) 352-374.

[11] B. Kummer, "Newton's method for non-differentiable functions", in J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach and K. Zimmerman, eds., *Advances in Mathematical Programming* (Academi Verlag, Berlin, 1988) 114-125.

[12] P. Marcotte and J.-P. Dussault, "A note on a globally convergent Newton method for solving monotone variational inequalities", Operations Research Letters 6 (1987) 35-42.

[13] R. Mifflin, "Semismoothness and semiconvex functions in constrained optimization", *SIAM Journal on Control and Optimization* 15 (1972) 957-972.

[14] J.-S. Pang, "Newton's method for B-differentiable equations", *Mathematics of Operations Research* 15 (1990) 311-341.

[15] J.-S. Pang, "A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems", *Mathematical Programming* 51 (1991) 101-131.

[16] J.-S. Pang and S.A. Gabriel, "NE/SQP: a robust algorithm for the nonlinear complementarity problem", Mathematical Programming 60 (1993) 295-337.

[17] J.-S. Pang and L. Qi, "Nonsmooth equations: motivation and algorithms", *SIAM Journal on Optimization* 3 (1993) 443-465.

[18] J.-S. Pang and L. Qi, "A globally convergent Newton method for convex SC^1 minimization problems", to appear in *Journal of Optimization Theory and Applications.*

[19] J.-S. Pang and Daniel Ralph, "Piecewise smoothness, local invertibility, and parametric analysis of normal maps", Manuscript, Department of Mathematics, The University of Melbourne, Parkville, Victoria, Australia (1993).

[20] L. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations", *Mathematics of Operations Research* 18 (1993) 227-244.

[21] L. Qi and H. Jiang, "A globally and superlinearly convergent trust region algorithm for convex SC^1 minimization problems and its application to stochastic programs", AMR 94/7, Applied Mathematics Report, The University of New South Wales, Australia (1994).

[22] L. Qi and J. Sun, "A nonsmooth version of Newton's method", *Mathematical Pro-*

gramming 58 (1993) 353-368.

[23] L.Qi and W. Sun, "An iterative method for the minimax problem", AMR 94/15, Applied Mathematics Report, The University of New South Wales, Australia (1994).

[24] L. Qi and R. Womersley, "An SQP algorithm for extended linear-quadratic problems in stochastic programming", AM 92/23, Applied Mathematics Preprint, The University of New South Wales, Australia (August, 1992) (under revision).

[25] S.M. Robinson, "Strongly regular generalized equations", *Mathematics of Operations Research* 5 (1980) 43-62.

[26] S.M. Robinson, "Generalized equations", in A. Bachem, M. Gröschel and B. Korte, eds., *Mathematical Programming: The State of the Art* (Springer-Verlag, Berlin, 1983) 346-367.

[27] S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity", *Mathematical Programming Study* 30 (1987) 45-66.

[28] S.M. Robinson, "Newton's methods for a class of nonsmooth functions", Industrial Engineering Working Paper, University of Wisconsin (Madison, USA, 1988).

[29] D. Sun, "A projection and contraction method for the nonlinear complementarity problem and its extensions", *Mathematica Numerica Sinica* 16 (1994) 183-194.

[30] D. Sun, "A new step-size skill for solving a class of nonlinear projection equations", *Journal of Computational Mathematics* 13 (1995) forthcoming.

[31] D. Sun, "A class of iterative methods for solving nonlinear projection equations", manuscript, Institute of Applied Mathematics, Academia Sinica, Beijing, China (July, 1994 (submitted).

[32] D. Sun and J. Han, "Newton and quasi-Newton methods for a class of nonsmooth equations and related problems", Technical Report 94/026, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, P.R. China (submitted).

[33] K. Taji, M. Fukushima and T. Ibaraki, "A globally convergent Newton method for solving strongly monotone variational inequalities", *Mathematical Programming* 58 (1993) 369-383.

[34] E.H. Zarantonello, "Projections on convex sets in Hilbert space and spectral theory", in E.H. Zarantonello, ed., *Contributions to Nonlinear Functional Analysis* (Academic Press, New York, 1971).

Appendix

Publications

[1] Defeng Sun, "Projected extragradient method for finding saddle points of general convex programming", *Journal of Qufu Normal University* 19:4 (1993) 10-17 (in Chinese).

[2] Defeng Sun and Jinde Wang, "An approximation method for stochastic programming with recourse", *Mathematica Numerica Sinica* 16 (1994) 80-92 (in Chinese). Also in *Chinese Journal of Numerical Mathematics and Applications* 16:2 (1994) 70-83.

[3] Defeng Sun, "A projection and contraction method for the nonlinear complementarity problem and its extensions", *Mathematica Numerica Sinica* 16 (1994) 183-194 (in Chinese). Also in *Chinese Journal of Numerical Mathematics and Applications* 16:3 (1994) 73-84.

[4] Defeng Sun, "An iterative method for solving variational inequality problems and complementarity problems", *Numerical Mathematics, A Journal of Chinese Universities* 16 (1994) 145-153 (in Chinese).

[5] Dachuan Xu and Defeng Sun, "A modification of successive approximation method for nonsmooth equations", *Journal of Qufu Normal University* 20:3 (1994) 14-20.

[6] Defeng Sun and Dachuan Xu, "Quasi-Newton method for nonlinear complementarity problem", in Min-I Yue, ed., *Optimization Theory and Applications* 579-584 (Xidian, 1994).

[7] Defeng Sun, "A new step-size skill for solving a class of nonlinear projection equations", *Journal of Computational Mathematics* 13 (1995) forthcoming.

[8] Jiye Han and Defeng Sun, "Superlinear convergence of approximate Newton methods for LC^1 optimization problems without strict complementarity", in D.Z. Du, L. Qi and R.S. Womersley, eds., *Recent Advances in Nonsmooth Optimization* (World Scientific Publishing Co., New Jersy, 1995), forthcoming.

[9] Defeng Sun, "On the convergence properties of a projection and contraction method", to appear in *Numerical Mathematics, A Journal of Chinese Universities* (English Series).

[10] Defeng Sun, "An improved extragradient method for linear variational inequality problems with lower and upper bounds", to appear in *Nanjing Da Xue Shu Xue Ban Nian Kan* (in Chinese).

[11] Guanghui Liu, Jiye Han, and Defeng Sun, "Global convergence of BFGS method with nonmonotone line search", submitted to *Optimization* (revised in October 1994).

[12] Jiye Han, Guanghui Liu, Defeng Sun, and Hongxia Yin, "A global convergence theorem for conjugate gradient methods for optimization and its applications", Technical Report No. 94/011, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, P.R. China (submitted).

[13] Defeng Sun, "A class of iterative methods for solving nonlinear projection equations", manuscript, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, P.R. China (July, 1994) (submitted).

[14] Defeng Sun and Jiye Han, "Newton and quasi-Newton methods for a class of nonsmooth equations and related problems", Technical Report No. 94/026, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, P.R. China (submitted).

[15] Jiye Han and Defeng Sun, "Newton and quasi-Newton methods for normal maps with polyhedral set", manuscript, Institute of Applied Mathematics, Academia Sinica, Beijing 100080, P.R. China (November 1994).