ITERATIVE METHODS OF SOLUTION FOR COMPLEMENTARITY PROBLEMS

by

Pudukkottai K. Subramanian

Computer Sciences Technical Report #607

June 1985

ITERATIVE METHODS OF SOLUTION FOR COMPLEMENTARITY PROBLEMS

by

Pudukkottai K. Subramanian

A thesis submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

1985

To the revered memory of my father

P. KRISHNAMOORTHY

for the wisdom of yesterday

and to my daughter

JYOTI

for the promise of tomorrow.

ACKNOWLEDGEMENTS

It is not often that one undertakes to go for a second doctorate. Many people have helped me realize this goal and I would like to thank all of them.

To Professor Olvi Mangasarian I would like to express my heartfelt gratitude for agreeing to take me as his student, for his guidance, advice and encouragement. He has always been most generous with his time. In many ways he has been a *Guru* in the real sense of the word: friend, teacher and guide. I shall most certainly miss our Tuesday meetings and this thesis owes more to him than to anyone else.

I wish to express my appreciation to Professor Stephen Robinson who encouraged me to go for the Ph.D. He was instrumental in my coming to Madison and has always been generous with his time, advice and help. I am thankful to Professors Robert Meyer and Tzong-Huei Shiau for being members of the examination and committee and to Professors Peter Glynn and Ken Kunen for reading drafts of the thesis.

I would like to express my gratitude to the Mathematics Research Center for providing financial support during my stay at Madison.

I wish to thank my dear friend Ravishankar for sharing with me many hours of discussion on Indian philosophy, culture and music, for sharing many of my frustrations and excitements as a graduate student, for guiding me through the intricacies of UNIX and most importantly for initiating me into TEX .

My family has always been my source of strength and support. To my mother Mrs. Vilasini Krishnamoorthy who insisted that her children excel in education and who has always been a bedrock of support, to my "other" parents Mr. & Mrs. Jerome Rest who enthusiastically supported our decision to return to school, who shared with us the pleasures and pains of going to school, for their continued love and support, who were always there when we needed them and finally to Karen, my companion for almost two decades and who had to take care of a family in addition to going to school for her own Ph.D, I can only say that I owe to you all more than I can possibly ever express.

ITERATIVE METHODS OF SOLUTION FOR COMPLEMENTARITY PROBLEMS

by Pudukkottai K. Subramanian

under the supervision of Professor Olvi Mangasarian

ABSTRACT

Many problems in optimization theory such as linear programming, quadratic programming and problems arising in diverse areas such as economic equilibria , electronic circuit simulation and free boundary problems can be formulated as complementarity problems. It is well known that where the matrices involved are large sparse matrices, the usual pivoting methods are not very efficient and sometimes even fail. This thesis is a contribution to the ongoing research in the area of iterative methods for the solution of linear and nonlinear complementarity problems.

We begin by considering complementarity problems where the operators are montone and consider their Tihonov regularizations. We obtain bounds for the solutions of the perturbed problems and in particular, estimates for the rate of growth of these solutions. In the case of linear complementarity problems with positive semidefinite matrices, these results reduce the solution of the LCP to the solution of a sequence of positive definite symmetric quadratic programs. We propose SOR (successive overrelaxation) algorithms

to solve these subproblems.

In the case of complementarity problems with nonempty interior, we show that given any preassigned feasibility tolerance our algorithm solves the problem by solving a Tihonov regularization problem for a single value of the parameter.

We consider monotone complementarity problems as fixed point problems. We prove convergence of iterative algorithms which find fixed points of carefully constructed projection maps using summability methods.

Since the solution of the nonlinear complementarity problem is equivalent to the solution of a system of nonlinear equations, we develop damped Gauss-Newton methods which under appropriate hypotheses solve this system with a local quadratic convergence rate. We present computational results which show that the use of SOR methods in conjunction with the Gauss-Newton methods is computationally effective.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. The Complementarity Problem

Let $\Re^n$ be the n-dimensional *real* linear space and let F be a mapping from $\Re^n$ to $\Re^n$. The *complementarity problem* consists in finding an $x \in \Re^n_+$ (if it exists) such that $F(x) \in \Re^n_+$ *and* $< x, F(x) > = 0$ where $\Re^n_+$ is the *non-negative orthant* of $\Re^n$, i.e., $\Re^n_+ = \{(x_1, \ldots, x_n) \ : \ x_i \geq 0)\}$. When F is an affine function and is of the form $F(x) = Mx + q$, where M is an $n \times n$ real matrix and $q \in \Re^n$, the complementarity problem is referred to as the *Linear Complementarity Problem* and is denoted by $LCP(M, q)$. When F is not affine, the complementarity problem is referred to simply as the *Nonlinear Complementarity Problem* and we shall write $NLCP(F)$ in this case. This thesis is concerned with *iterative methods* for the solution of such problems.

2. Sources of complementarity problems

The complementarity problem arises in many situations. Its primary sources are equilibrium problems (both economic and physical) as well as necessary optimality conditions for mathematical programming problems.

We give below some well known instances of complementarity problems. For proofs of our statements regarding the first two examples, we refer the reader to [Cottle & Dantzig, 1968] and for the third to [Mangasarian, 1969] .

2.1 Symmetric Dual Linear Programs

Consider the linear program (LP),

$$\text{LP:} \quad \min \; c^T x \qquad \text{subject to} \quad Ax \geq b, \quad x \geq 0$$

and its dual (DLP),

$$\text{DLP:} \quad \max \; b^T y \qquad \text{subject to} \quad A^T y \leq c, \quad y \geq 0$$

where A is an $m \times n$ matrix, and c and b are given vectors in $\Re^n$ and $\Re^m$ respectively. Solvability of either program, and hence of both, is equivalent to the solvability of the $LCP(M,q)$ where

$$M = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix} \qquad \text{and} \qquad q = \begin{pmatrix} c \\ -b \end{pmatrix}.$$

It is well known that $\bar{z} = (\bar{x}, \bar{y})$ in $\Re^n \times \Re^m$ solves the $LCP(M,q)$ if and only if $\bar{x}$ solves LP and $\bar{y}$ solves DLP.

2.2 Quadratic Program

Let D be a positive semidefinite symmetric $n \times n$ matrix and let A, b, and c be as in (2.1) above. The quadratic program (QP)

$$\text{QP:} \quad \min \; c^T x + \tfrac{1}{2} x^T D x \qquad \text{subject to} \quad Ax \geq b, \quad x \geq 0$$

is solvable if and only if the $LCP(M,q)$ with

$$M = \begin{pmatrix} D & -A^T \\ A & 0 \end{pmatrix} \quad \text{and} \quad q = \begin{pmatrix} c \\ -b \end{pmatrix}$$

is solvable. If $\bar{z} = (\bar{x},\, \bar{y})$ in $\Re^n \times \Re^m$ is the solution of $LCP(M,q)$, then $\bar{x}$ solves QP.

2.3 Convex Program

Let $f : \Re^n \to \Re$ and let $g : \Re^n \to \Re^m$ be differentiable convex functions on $\Re^n$. Consider the convex program

$$\min f(x) \quad \text{subject to} \quad g(x) \le 0,\ x \ge 0.$$

Let $L(x,u) = f(x) + u^T g(x)$ be the *Lagrangian*. The Karush- Kuhn-Tucker conditions for this program [Mangasarian, 1969] are :

$$(\text{KKT}) : \quad \begin{cases} v = \nabla_x L(x,u) \ = \nabla f(x) + u^T \nabla g(x) \ge 0 \\ y = -\nabla_u L(x,u) = -g(x) \ge 0 \\ x \ge 0, \quad u \ge 0 \\ x^T v = 0 \\ u^T y = 0 \end{cases}$$

If any of the standard constraint qualifications are satisfied, then for each solution of the convex program, the KKT conditions are satisfiable. Thus the KKT conditions are equivalent to the nonlinear complementarity problem $NLCP(F)$ with

$$z = \begin{pmatrix} x \\ u \end{pmatrix} \in \Re^{n+m}, \quad F(z) = \begin{pmatrix} \nabla_x L(x,u) \\ -\nabla_u L(x,u) \end{pmatrix}.$$

2.4 Other sources of complementarity problems

Complementarity problems occur in the numerical solution of partial differential equations (where the underlying matrices are often large and sparse) [Cottle, 1977], [Cottle, Golub, Sacher, 1978]. They also occur in the Christopherson method for solving free boundary problems in journal bearings [Cryer, 1969], in the simulation of electronic circuits [v. Bokhoven, 1980], price models in commodity future markets [J. C. Cheng, 1975], energy models, e. g., the Project Independence Evaluation System (PIES) [Hogan, 1975].

3. Methods of solution for complementarity problems

In the case of the linear complementarity problem, numerical methods of solution are essentially of two types: *direct methods* and *iterative methods*. The former are based on the process of pivoting on the elements of the underlying matrix. Typically, direct methods terminate in a finite number of steps. In contrast, iterative methods produce an infinite sequence of iterates which converge to a solution.

The best known direct methods are those of Lemke [1965] and Cottle & Dantzig [1968]. These methods are known to be efficient when the matrices are not large , are dense and have nice characteristics e.g., copositive plus, principal positive minors etc. However, for large systems (which is often the case in applications), when the matrices are sparse these methods may be ineffective in terms of storage requirements and speed. Also the sparsity of the matrices is lost after a few iterations unless special techniques are employed. Examples of large size matrices are known where the pivoting methods fail to produce a solution [Mangasarian, 1984].

Iterative methods are particularly advantageous for large scale sparse problems and can be conveniently stored. Sparsity is preserved and since no matrix inversion occurs, iterates are computed fairly easily. The real disadvantage is the large number of iterates often required to meet a termination criterion.

Iterative methods, often using a relaxation procedure, such as successive overrelaxation (SOR) methods have been proposed by several authors for the case when the underlying matrix M is symmetric. In this case, the LCP is the equivalent of the Karush-Kuhn-Tucker conditions for the quadratic program

$$\min f(x) = \frac{1}{2}\, x^T M x + q^T x \quad \text{subject to} \quad x \geq 0$$

and f serves as a descent function. Well known amongst these are the methods of Cryer [1971], Cottle et al [1978] and Mangasarian [1977]. In particular, Mangasarian's algorithm includes many existing algorithms as special cases. Ahn [1981] has shown that Mangasarian's algorithm converges for special classes of non-symmetric matrices (H-matrices).

When LCP(M,q) is solvable, it has a vertex solution. Utilizing this fact, Mangasarian [1976, 1978, 1979] studied the possibility of solving an LCP as a linear program. Shiau [1983] has proposed solving an LCP as a finite sequence of linear programs.

Methods of solution for nonlinear complementarity problems are inevitably iterative in nature. A solution of $NLCP(F)$ for F continuous is also a fixed point of the map

$$x \quad \longmapsto \quad \left(x - F(x)\right)_+ := \min\{0,\ x - F(x)\}$$

where x_+ is the projection of x on $\Re^n_+$. Thus the best known methods exploit this property. Well known amongst these are the *simpilicial division methods* due to Scarf [1967, 1973], Todd [1976], Eaves [1972].

If F is differentiable, given the point x_k , we can consider the *linearization* $\mathcal{L}_k(F)$ of F at x_k :

$$\begin{aligned} \mathcal{L}_k(F)x &= F(x_k) + \nabla F(x_k)(x - x_k) \\ &= F(x_k) - \nabla F(x_k)x_k + \nabla F(x_k)x \end{aligned}$$

where $\nabla F(x_k)$ is the Jacobian of F at x_k . Hence if we let

$$M_k := \nabla F(x_k), \qquad q_k := F(x_k) - \nabla F(x_k)x_k,$$

we get a sequence $\{x_k\}$ of iterates such that x_{k+1} solves $LCP(M_k, q_k)$. Using results from the theory of *generalized equations* as developed by Robinson [1976, 1978, 1979, 1982], Josephy [1979] has shown that under suitable hypotheses, the iterates converge locally to a solution x^* of $NLCP(F)$, thus reducing $NLCP(F)$ to a sequence of LCPs. He has used this method in particular to solve Hogan's PIES model (Hogan[1975]) using Lemke's algorithm to solve the resulting LCPs.

4. Methods proposed in this thesis

In Chapter 2 we shall consider *monotone operators* $F : \Re^n \longrightarrow \Re^n$, that is operators F such that $< F(x) - F(y),\ x - y > \geq 0$ for all x, y in $\Re^n$. When $F = Mx + q$, this is equivalent to having positive semidefinite M. Under suitable conditions, $NLCP(F_\varepsilon)$, where $F_\varepsilon = F + \varepsilon I$ is the *Tihonov regularization* of F, has a unique solution x_ε ([Brézis, 1973], [Karamardian, 1972]). If $NLCP(F)$ has a solution, then as $\varepsilon \downarrow 0$, x_ε converges to the least two-norm solution $\bar{x}$ of $NLCP(F)$, that is

$$\bar{x} = \text{argmin}\{\,\|x\| \,:\, x \text{ solves } NLCP(F)\}$$

where $\|\ \|$ is the two-norm on $\Re^n$. We shall establish growth rates for x_ε from which the above known results follow as corollaries.

The above results are used in Chapter 3 to consider $LCP(M,\ q)$ when M is positive semidefinite. We shall formulate the problem as a *dual exact penalty function* problem [Han & Mangasarian, 1983]. We propose SOR algorithms to solve $LCP(M + \varepsilon I,\ q)$ for $\varepsilon > 0$ and this reduces $LCP(M,\ q)$ to a sequence of LCPs to be solved by the above SOR procedure. In particular, when the feasible set of $LCP(M,\ q)$ has a nonempty interior, we show how $LCP(M,\ q)$ can be solved to any preassigned tolerance by solving $LCP(M + \varepsilon I,\ q)$ for a single $\varepsilon > 0$.

In Chapter 4, we consider positive semidefinite M and fixed point methods. Using carefully selected sequences α_n and ε_n and a summability matrix

B, we show that the iterates

$$x_{n+1} = \left(x_n - \alpha_n(Mx_n + \varepsilon_n x_n + q)\right)_+$$

are bounded if and only if $LCP(M,q)$ is solvable and in this case, the B–transform $B(\{x_n\})$ converges to a solution of $LCP(M,q)$. As an interesting corollary we develop a similar algorithm to solve $NLCP(F)$ when F is monotone and satisfies a *distributed Slater constraint qualification* as defined by Mangasarian and McLinden [1984]. In this setting these authors have shown that $NLCP(F)$ is solvable.

The fixed point methods of Chapter 4, by their global nature, are slow and particularly so near a solution point. Thus their utility is essentially in the generation of suitable starting points for fast converging Newton- like methods. We consider one such method in Chapter 5. It has been shown by Mangasarian [1976] that the general $NLCP(F)$ is equivalent to system of nonlinear equations. We consider a damped Gauss-Newton procedure to solve such a system. When $NLCP(F)$ has a nondegenerate solution $\bar{z}$ such that $\nabla F(\bar{z})$ has nonsingular principal minors then this algorithm, under suitable conditions, converges locally quadratically to $\bar{z}$.

The Gauss-Newton method is particularly useful in conjunction with the SOR and fixed point algorithms of the previous chapters leading to the notion of *poly algorithms*. The thesis concludes with a report on our computational experience with the above algorithms and with some suggestions for further research.

5. Notions and Notations

As indicated earlier, we shall be principally concerned with real finite dimensional spaces, and real matrices and vectors. In particular:

(i) $\Re^n$ stands for the space of of real ordered n-tuples.

(ii) All vectors are column vectors. Given a vector x, we shall denote its i^{th} component by x_i. We say $x \geq 0$ if one has $x_i \geq 0$ for all i. For a given scalar λ, we define $(\lambda)_+ = \max\{0, \lambda\}$. If $x \in \Re^n$, we write x_+ for the vector whose i^{th} component is $(x_i)_+$.

(iii) Superscripts are used to distinguish between vectors, e.g., x^1, x^2 etc.

(iv) For $x, y \in \Re^n$, x^T indicates the transpose of x, $x^T y$ their inner product. Occasionally the superscript T will be suppressed and we also use $< \, , \, >$ for the inner product to improve clarity.

(v) All matrices are indicated by upper case letters A, B, C etc. The i^{th} row of A is indicated by A_i and we write $A_{.j}$ to indicate the j^{th} column. The $(i, j)^{th}$ element of A is indicated by A_{ij}. The transpose of A is denoted by A^T. The symbol I indicates the identity matrix of appropriate dimension while e shall indicate a vector of all ones of appropriate dimension.

(vi) Real valued functions defined on subsets of $\Re^n$ are denoted by f, g, h etc., and we write ∇f and $\nabla^2 f(x)$ to indicate the gradient vector and the Hessian matrix at the point x. If F is an operator, $F : \Re^n \longrightarrow \Re^m$, we shall write ∇F to indicate the $m \times n$ Jacobian matrix at the point x. For the most part, we prefer to use upper case letters F, G, etc., to indicate operators.

(vii) For $x \in \Re^n$, $\| x \| = \{x^T x\}^{1/2}$ is the standard Euclidean norm. When the norm is induced by a positive definite symmetric matrix G, we write $\| x \|_G$ for $\{x^T Gx\}^{1/2}$.

(viii) If K is a closed convex set in $\Re^n$, given $x \in \Re^n$ we write $P_K(x)$ for the *projection* of x on K, that is

$$P_K(x) = \text{argmin}\left\{ \| z - x \| : z \in K \right\}.$$

(ix) Given $NLCP(F)$, we shall write $S(F)$ for its *feasible set* and $\bar{S}(F)$ for its *solution set* that is,

$$S(F) = \{x \in \Re^n_+ : F(x) \in \Re^n_+\}$$

$$\bar{S}(F) = \{x \in S(F) : < x, F(x) > = 0\}.$$

In the case of $LCP(M, q)$ we denote these sets by $S(M, q)$ and $\bar{S}(M, q)$ respectively.

(x) We use the notation a.b.c to refer to subsection b.c (or a displayed equation) in section b of Chapter a. Within the same chapter, the chapter number will be omitted. We indicate bibliographic references by author's name and year of publication, e.g., [Mangasarian, 1969]. All references are arranged alphabetically and chronologically for each author.

(xi) Finally, the end of the proof of an assertion is indicated by ▌.

CHAPTER 2

MONOTONE OPERATORS

AND TIHONOV REGULARIZATION

1. Introduction

In this chapter we shall be concerned with operators $F : \Re^n \longrightarrow \Re^n$ which are *monotone* and their *Tihonov regularization*, $F + \varepsilon I$ for $\varepsilon > 0$. In this case $NLCP(F + \varepsilon I)$ has a unique solution $x(\varepsilon)$ and we study the growth rate of $\|x(\varepsilon)\|$. Our prinicpal result, Theorem 5.1, sharpens some results due to Brézis [1973] which are given for a multifunction on a Hilbert space. Although our results are all couched in an $\Re^n$ setting, they are all extendable with minor modifications to the more general setting considered by Brézis.

Our main reference throughout this chapter shall be [Auslender, 1976].

2. Definition. *Let $D \subseteq \Re^n$. An operator $F: D \longrightarrow \Re^n$ is said to be monotone on D if*

$$< F(x) - F(y),\ x - y > \geq 0 \quad \forall x,\ y \in D.$$

It is said to be strictly monotone if the above inequality holds strictly. We say F is strongly monotone on D with modulus α if

$$< F(x) - F(y),\ x - y > \geq \alpha\|x - y\|^2 \quad \forall x,\ y \in D.$$

When $D = \Re^n$ we simply say monotone, strictly monotone, strongly monotone etc.

3. Examples of monotone operators.

We consider in this section examples of monotone operators. For proofs of our assertions regarding examples (3.1) and (3.2) we refer the reader to [Auslender, 1976], pages 118 and 40–41 respectively. The proof for example (3.3) may be found e.g., in [Mangasarian and McLinden, 1985].

3.1 Projection on a closed convex set.

Let C be a nonempty closed convex set in $\Re^n$. Then the projection operator P defined on $\Re^n$ by

$$P : x \longrightarrow P_C(x) := \operatorname{argmin}\{\|z - x\| : z \in C\}$$

is monotone.

3.2 Derivatives of convex functions

Let f be a convex function on the convex set $D \subseteq \Re^n$, that is, $f : D \rightarrow \Re$ and satisfies

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y) \quad \forall x, y \in D.$$

If f is finite and differentiable on D with gradient $\nabla f(x)$ then $F(x) = \nabla f(x)$ is monotone on D. If f is strongly convex on D, that is, $\exists \delta > 0$ such that for all x, y in D we have

$$f\left(\lambda x + (1-\lambda)y\right) \leq \lambda f(x) + (1-\lambda)f(y) - \delta\lambda(1-\lambda)\|x-y\|^2$$

then $F(x)$ is strongly monotone on D.

3.3 Convex Programs

Consider the convex program 1.2.3 of Chapter 1 :

$$\min f(x) \quad \text{subject to} \quad g(x) \leq 0,\ x \geq 0$$

where $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$ are differentiable and convex. The Karush-Kuhn-Tucker optimality conditions for this problem are equivlant to $NLCP(F)$,

$$z = \begin{pmatrix} x \\ u \end{pmatrix} \in \Re^{n+m} \quad F(z) = \begin{pmatrix} \nabla_x L(x,u) \\ -\nabla_u L(x,u) \end{pmatrix},$$

$L(x,u) = f(x) + u^T g(x)$ being the standard Lagrangian. Then $F(z)$ is monotone.

3.4 Affine operators

Let $F(x) = Mx + q$, M being positive semidefinite real $n \times n$ matrix, that is, $x^T M x \geq 0$ for all $x \in \Re^n$. Then $F(x)$ is monotone. If M is positive definite, that is, $\exists\, \lambda,\ \mu > 0$ such that

$$\lambda\,\|x\|^2 \leq x^T M x \leq \mu\,\|x\|^2$$

then $F(x)$ is strongly monotone.

4. Variational inequalities

We shall find it convenient in the sequel to consider complementarity problems as *variational inequalities*:

4.1 Definition. *Let $D \subseteq \Re^n$, $F: D \to \Re^n$. The variational inequality problem consists in finding $z_o \in D$, if it exists, such that*

$$< F(z_o),\ x - z_o > \geq 0 \quad \forall x \in D.$$

In this case we say that z_o solves the variational inequality

$$\text{(VI)}: \qquad < F(z),\ x - z > \geq 0 \quad \forall x \in D.$$

Although many problems can be cast as variational inequality problems, our interest in them stems from the following well known proposition (see e.g., [Karamardian, 1972]).

4.2 Proposition. *Let $F: \Re^n_+ \to \Re^n$. Then z_o solves $NLCP(F)$ if and only if z_o solves* (VI).

Proof

If z_o solves $NLCP(F)$, then $z_o \geq 0$, $F(z_o) \geq 0$ and $< z_o,\ F(z_o) > = 0$. Hence,

$$< F(z_o),\ x - z_o > = < F(z_o),\ x > \geq 0 \quad \forall x \in \Re^n_+.$$

Conversely, if z_o solves (VI), taking $x = 0$ and $x = 2z_o$ successively we get $< z_o,\ F(z_o) > = 0$. For $i = 1, \ldots, n$ take $x = z_o + e_i$ where e_i is the unit vector with 1 in the i^{th} place, 0 elsewhere then $\big(F(z_o)\big)_i \geq 0$ for $1 \leq i \leq n$. By definition, $z_o \geq 0$. ∎

4.3 Definition. *Let C be a closed convex set in $\Re^n$, and let $F{:}\,\Re^n \rightarrow \Re^n$. We say F is hemicontinuous on C if for all $x,\ y \in C$, the map*

$$\lambda \longmapsto < F\big(\lambda x + (1-\lambda)y\big),\ x - y >$$

is continuous on the interval $[0,\ 1]$.

4.4 Proposition. *Let C be a closed convex set contained in D and let $F{:}\,D \longrightarrow \Re^n$ be monotone and hemicontinuous on C. Then*

$$< F(z^\star),\ (z - z^\star) > \ \geq \ 0 \qquad \forall z \in C$$

if and only if

$$< F(z),\ (z - z^\star) > \ \geq \ 0 \qquad \forall z \in C\,. \tag{4.5}$$

Further, $Z_o = \{z^\star : z^\star \text{ solves } (4.5)\}$ is closed and convex.

Proof

See Auslender [1976, page 121] . ∎

Suppose F is affine and monotone, that is, $F(x) = Mx + q$ with M positive semidefinite. It is well known that if the feasible set

$$S(M,\ q) = \{x \ \geq \ 0 : Mx + q \ \geq 0\}$$

is nonempty, then $LCP(M,\ q)$ is solvable [Eaves, 1971]. This is no longer the case even for continuous nonlinear monotone mappings. For a counter example, see [Megiddo, 1977] (this example also appears in [Garcia, 1977]). However, if F satisfies some growth conditions defined below then $NLCP(F)$ is solvable. Alternately, if F satifies a regularity condition such as the *distributed Slater constraint qualification* [Mangasarian and McLinden, 1985], then again the result is true.

4.6 Definition. *Let $C \subseteq D$ be a nonempty closed convex set and assume $F: D \to \Re^n$. We say F is coercive (strongly coercive) if there exist $v_o \in C$, $\lambda \in \Re$ positive such that*

$$v \in C,\ \|v\| \geq \lambda \implies F(v)(v - v_o) > 0$$

(respectively,

$$v \in C,\ \|v\| \to \infty \implies \frac{F(v)(v - v_o)}{\|v - v_o\|} \to +\infty).$$

The proof of the following Theorem may be found in [Auslender, 1976].

4.7 Theorem. *Let $F : \Re^n \to \Re^n$ be a monotone operator, coercive and hemiconintuous on $\Re^n_+$. Then $NLCP(F)$ is solvable and if F is strongly coercive, then it has a unique solution.*

We now define the *Tihonov regularization* of an operator.

4.8 Definition. *Let $F : \Re^n \longrightarrow \Re^n$ and let $\varepsilon > 0$. The Tihonov regularization F_ε of F is defined by $F_\varepsilon(x) = F(x) + \varepsilon\, x$.*

If $F : \Re^n \to \Re^n$ is monotone and hemicontinuous, then F_ε is also hemicontinuous and strongly monotone with modulus of monotonicity at least ε. It is immediate that F_ε is strongly coercive. Thus we get the following useful corollary to Theorem 4.7.

4.9 Corollary. *Let $F : \Re^n \to \Re^n$ be monotone and hemicontinuous. Then $\forall \varepsilon > 0$, there exists a unique $x(\varepsilon)$ (called ε-approximant or simply approximant), which solves $NLCP(F)$.*

5. Properties of approximants

In this section we shall prove the main theorem of this chapter on the growth rate of ε-approximants.

5.1 Theorem. *Let $F : \Re^n \to \Re^n$ be a monotone operator which is hemicontinuous on $\Re^n_+$. Let $\{\varepsilon_n\}$ be a sequence of positive reals such that $\varepsilon_n \downarrow 0$. Let $F_n = F + \varepsilon_n I$ be the Tihonov regularization of F and let x_n be the unique solution of $NLCP(F_n)$. Let $m > n$ and assume that $F(0) \not\geq 0$. Then*

a) $\|x_m\| > \|x_n\|$

b) $\varepsilon_m\|x_m\| \leq \varepsilon_n\|x_n\|$

c) $\|x_m - x_n\|^2 \leq \{(\varepsilon_n - \varepsilon_m)/(\varepsilon_n + \varepsilon_m)\} \cdot \{\|x_m\|^2 - \|x_n\|^2\}$

d) $< x_m, x_n > \geq (\varepsilon_m\|x_m\|^2 + \varepsilon_n\|x_n\|^2)/(\varepsilon_m + \varepsilon_n)$

Let $\overline{S} = \{x : x \text{ solves } NLCP(F)\}$. Then

e) $\sup\{\|x_n\|\} < \infty \Longleftrightarrow x_n \longrightarrow \overline{x} = P_{\bar{S}}(0) \Longleftrightarrow \overline{S} \neq \emptyset.$

Proof

From Proposition 4.2, it follows that

$$< F_m(x_m),\ x - x_m > \geq 0 \qquad \forall x \in \Re^n_+.$$

By taking $x = x_n$,

$$< F_m(x_m),\ x_n - x_m > \geq 0. \tag{5.2}$$

Likewise,

$$< F_n(x_n),\ x_m - x_n > \geq 0,$$

which we rewrite as

$$< -F_n(x_n),\ x_n - x_m > \geq 0. \tag{5.3}$$

Adding (5.2) and (5.3) we get

$$< F_m(x_m) - F_n(x_n),\ x_n - x_m > \geq 0.$$

Hence remembering that $F_m = F + \varepsilon_m I$,

$$< F(x_m) + \varepsilon_m x_m - F(x_n) - \varepsilon_n x_n,\ x_n - x_m > \geq 0.$$

From the monotonicity of F this yields

$$< \varepsilon_m x_m - \varepsilon_n x_n,\ x_n - x_m > \geq < F(x_n) - F(x_m),\ x_n - x_m > \geq 0,$$

that is,

$$\varepsilon_m < x_m - x_n,\ x_n - x_m > + (\varepsilon_m - \varepsilon_n) < x_n,\ x_n - x_m > \geq 0.$$

By assumption $m > n$ so that $\varepsilon_m < \varepsilon_n$. We now have

$$(\varepsilon_n - \varepsilon_m) < x_n,\ x_m - x_n > \geq \varepsilon_m \|x_m - x_n\|^2. \tag{5.4}$$

By Cauchy-Schwarz inequality,

$$\frac{\varepsilon_n - \varepsilon_m}{\varepsilon_m} \|x_n\| \|x_m - x_n\| \geq \|x_m - x_n\|^2.$$

We shall prove (a) shortly so that $x_m \neq x_n$. Hence

$$\frac{\varepsilon_n - \varepsilon_m}{\varepsilon_m} \|x_n\| \geq \|x_m - x_n\| \geq \|x_m\| - \|x_n\|$$

and

$$\varepsilon_n \|x_n\| \geq \varepsilon_m \|x_m\|.$$

This proves (b).

Next we prove (c). Obviously,

$$\|x_m\|^2 = \|x_m - x_n\|^2 + \|x_n\|^2 + 2 < x_m - x_n,\ x_n >$$

so that from (5.4) we now get

$$\|x_m\|^2 \geq \|x_m - x_n\|^2 + \|x_n\|^2 + \left\{ \frac{2\varepsilon_m}{\varepsilon_n - \varepsilon_m} \right\} \|x_m - x_n\|^2.$$

Hence,

$$\|x_m\|^2 - \|x_n\|^2 \geq \left\{ \frac{\varepsilon_n + \varepsilon_m}{\varepsilon_n - \varepsilon_m} \right\} (\|x_m - x_n\|^2).$$

This establishes (c).

We now prove (a). Observe that $m > n$ implies that $x_m \neq x_n$. To see this, suppose the contrary and write $x_m = x_n = x$. Then x solves $NLCP(F_i)$ for $i = m, n$ so that

$$< F(x) + \varepsilon_m x,\ x > = 0$$

and

$$< F(x) + \varepsilon_n x,\ x > = 0$$

which imply $< \varepsilon_m x - \varepsilon_n x,\ x > = 0$. Since $\varepsilon_m < \varepsilon_n$ we must have $x = 0$. But $F_m(x) \geq 0$, so that we must have $F(0) \geq 0$ contradicting our hypothesis. Hence $\|x_m - x_n\| > 0$ and clearly (a) follows from (c).

To prove (d), we have from (c) that

$$\|x_m\|^2 - 2 < x_m,\ x_n > + \|x_n\|^2 \leq \left\{ \frac{\varepsilon_n - \varepsilon_m}{\varepsilon_n + \varepsilon_m} \right\} \left(\|x_m\|^2 - \|x_n\|^2 \right)$$

which implies that

$$2 < x_n,\ x_m > \geq \frac{2\varepsilon_m}{\varepsilon_n + \varepsilon_m} \|x_m\|^2 + \frac{2\varepsilon_n}{\varepsilon_n + \varepsilon_m} \|x_n\|^2$$

or

$$< x_m,\ x_n > \geq \frac{\varepsilon_m \|x_m\|^2 + \varepsilon_n \|x_n\|^2}{\varepsilon_m + \varepsilon_n}.$$

This proves (d).

Finally we prove (e).

We start by showing that $\{\|x_n\|\}$ bounded $\Rightarrow$ x_n converges to an element of $\bar{S}$. From (a), since $\{\|x_n\|\}$ is strictly increasing, $\sup \|x_n\| =$

$\lim \|x_n\|$. Taking $m > n$ and letting $m,\ n \to \infty$, it follows from (c) that $\{x_n\}$ is Cauchy. Hence x_n converges. Let $x_n \to \xi$. Since x_n solves $NLCP(F_n)$,

$$x_n \geq 0, \quad F_n(x_n) \geq 0, \quad < x_n,\ F_n(x_n) > = 0$$

which implies

$$\xi \geq 0, \quad F(\xi) \geq 0, \quad , < \xi,\ F(\xi) > = 0$$

so that $\xi \in \bar{S}$.

On the other hand, if $\bar{S} \neq \emptyset$, let $\bar{z}$ be any arbitrary element of $\bar{S}$. Assume that n is arbitrary but fixed. By Proposition (4.2),

$$< F_n(x_n),\ x - x_n > \geq 0 \qquad \forall x \in \Re^n_+.$$

Take $x = \bar{z}$ to get

$$< F(x_n) + \varepsilon_n x_n,\ \bar{z} - x_n > \geq 0. \tag{5.5}$$

Since $\bar{z}$ solves $NLCP(F)$, by Proposition (4.4),

$$< F(x),\ x - \bar{z} > \geq 0 \qquad \forall x \in \Re^n_+.$$

Taking $x = x_n$,

$$< F(x_n),\ x_n - \bar{z} > \geq 0. \tag{5.6}$$

From (5.5) and (5.6) we get

$$\varepsilon_n < x_n,\ \bar{z} - x_n > \geq 0 \tag{5.7}$$

so that $< x_n, \bar{z} > \geq \|x_n\|^2$. Hence $\|x_n\| \leq \|\bar{z}\|$, that is, $\sup_n \|x_n\|$ is bounded proving the converse.

It remains only to show that if $x_n \rightarrow \xi$ then $\xi = P_{\bar{S}}(0)$. But from (5.7) we have $< \xi, \bar{z} - \xi > \geq 0$ and since $\bar{z}$ was an arbitrary element of $\bar{S}$ it follows that $\xi = P_{\bar{S}}(0)$. This completes the proof. ▮

5.8 Remarks

1. Parts (a) and (e) of Theorem 5.1 are known when F is a multifunction on a Hilbert space H. For a proof using the theory of Yosida approximations, see [Brézis, 1974], who also proves a weaker form of (c).

2. We shall find Theorem 5.1 useful in the next chapter. In particular, part (e) is used in developing successive overrelaxation (SOR) algorithms for $LCP(M, q)$ when M is positive semidefinite. We shall use part (b) in finding $\varepsilon - \textit{approximate solutions}$ satisfying preassigned tolerances.

CHAPTER 3

DUAL EXACT PENALTY FORMULATION

1. Introduction

In this chapter we shall be concerned with $LCP(M,\ q)$ when the matrix M is positive semidefinite. Following [Han & Mangasarian, 1983], we first consider a dual exact penalty formulation for $LCP(M,q)$ when M is positive definite, thus reducing the resulting problem to a maximization problem over the non-negative orthant. A modification of an SOR procedure due to Mangasarian [1977] is proposed to solve this maximization problem. When M is only semidefinite, we consider the Tihonov regularization $M+\varepsilon_n I$ of M for a sequence of positive reals $\varepsilon_n \downarrow 0$ and solve the resulting sequence $\{LCP(M+\varepsilon_n I,\ q)\}$ by the above procedure.

When M is positive semidefinite and the interior $int\, S(M,\ q)$ of the feasible set $S(M,\ q)$ is nonempty, we show using Theorem 1.5.1 that for any preassigned tolerance δ, we can find approximate solutions by solving $LCP(M+\bar{\varepsilon}I,\ q)$ for a specific $\bar{\varepsilon} > 0$ depending on δ.

2. Dual exact penalty for LCP(M, q)

We begin with the following definition.

2.1 Definition. *Given the nonlinear program (NLP),*

$$\text{(NLP)}: \quad \begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & g(x) \leq 0 \end{array}$$

where $f : \Re^n \longrightarrow \Re$ *and* $g : \Re^n \longrightarrow \Re^m$, *we say that* $(\overline{x}, \overline{u}) \in \Re^n \times \Re^m$ *is a Karush-Kuhn-Tucker (KKT) point for (NLP) if*

$$\left.\begin{array}{l} \nabla f(\overline{x}) + \overline{u}g(\overline{x}) = 0 \\ g(\overline{x}) \leq 0, \ \overline{u}g(\overline{x}) = 0, \ \overline{u} \geq 0. \end{array}\right\} \tag{2.2}$$

In this case, $\overline{x}$ *is called a stationary point of (NLP).*

Let M be positive semidefinite and let us rewrite $LCP(M, q)$ as

$$\left.\begin{array}{ll} \text{minimize} & f(x) = < x, \ Mx + q > \\ \text{subject to} & Mx + q \geq 0 \\ & x \geq 0. \end{array}\right\} \tag{2.3}$$

In this case $f(x)$ and the constraints are convex so that the *Wolfe Dual* for $LCP(M, q)$ is the program [Mangasarian, 1969],

$$\left.\begin{array}{ll} \text{maximize} & L(x, u, v) \\ \text{subject to} & \nabla_x L(x, u, v) = 0 \\ & u \geq 0, \ v \geq 0 \end{array}\right\} \tag{2.4}$$

where $L(x, u) = f(x) - u^T(Mx + q) - x^T v$ is the standard Lagrangian and $\nabla_x L(x, u, v)$ its gradient with respect to x. We rewrite (2.4) fully as

$$\left.\begin{array}{ll} \text{maximize} & x(Mx + q) - u(Mx + q) - xv \\ \text{subject to} & (M + M^T)x + q - M^T u - v = 0 \\ & u \geq 0, \ v \geq 0. \end{array}\right\} \tag{2.5}$$

The *dual exact penalty function* problem associated with (2.3) is the penalty function problem associated with (2.5), namely,

$$\begin{aligned} \text{maximize} \quad & D(x,\, u,\, v) \\ \text{subject to} \quad & (u,\, v) \;\geq\; 0 \end{aligned} \tag{2.6}$$

where

$$D(x,u,v) := x(Mx+q) - u(Mx+q) - vx - \frac{\gamma}{2}\|(M+M^T)x + q - M^T u - v\|^2,$$

γ being the penalty parameter [Han and Mangasarian, 1983]. We use the term *exact penalty* expressedly (as against exterior penalty) since, as pointed out by the above mentioned authors, the optimal Lagrange multiplier associated with the equality constraints of the Wolfe dual (2.5) is zero under certain natural conditions (to be explored further later) and hence the parameter γ remains finite.

We first show that the stationary points of (2.6) are of particular interest to us.

2.7 Theorem. *Assume that M is positive semidefinite and that either (i) $\gamma = 0$ or (ii) $\gamma \neq 0$ and $1/\gamma \notin$ spectrum $(M+M^T)$. Then every stationary point $(x,\, u,\, v)$ of (2.6) solves $LCP(M,q)$.*

Proof

By definition, x is a stationary point of the program

$$\min_{u \geq 0} \quad \theta(x,\, u)$$

if for some $\xi \ \geq \ 0$, $(x,\, u,\, \xi)$ is a KKT point satisfying

$$\begin{aligned} \nabla_x L(x,\, u,\, \xi) &= 0 \\ \nabla_u L(x,\, u,\, \xi) &= 0 \\ u \geq 0,\ \xi &\geq 0 \\ \xi u &= 0 \end{aligned}$$

where $L(x,\, u,\, \xi) \ = \ \theta(x,\, u) - \xi u$. These conditions translate into

$$\nabla_x \theta(x,\, u) = 0, \quad \nabla_u \theta(x,\, u) \geq 0, \quad u \nabla_u \theta(x,\, u) = 0, \quad u \geq 0.$$

Using the above facts and noting that (2.6) is a maximization problem, we find that $(x,\, u,\, v)$ is a stationary point for (2.6) if and only if

$$\begin{aligned} \nabla_x D(x,\, u,\, v) &= 0 \\ \nabla_u D(x,\, u,\, v) &\leq 0 \\ \nabla_v D(x,\, u,\, v) &\leq 0 \\ u \nabla_u D(x,\, u,\, v) &= 0 \\ v \nabla_v D(x,\, u,\, v) &= 0 \\ u \geq 0,\ v &\geq 0. \end{aligned}$$

Hence $(x,\, u,\, v)$ is a stationary point if and only if

$$\nabla_x D = \big(I - \gamma(M + M^T)\big)\big((M + M^T)x + q - M^T u - v\big) = 0,$$

$$\nabla_u D = -(Mx + q) + \gamma M\big((M + M^T)x + q - M^T u - v\big) \leq 0,$$

$$\nabla_v D = -x + \gamma\big((M + M^T)x + q - M^T u - v\big) \leq 0,$$

$$u\nabla_u D = 0, \; v\nabla_v D = 0, \; (u, v) \geq 0.$$

By our hypothesis on γ, these conditions reduce to

$$(M + M^T)x + q - M^T u - v = 0,$$

$$-(Mx + q) \leq 0, \quad -x \leq 0,$$

$$u\big(-(Mx + q)\big) = 0, \quad v(-x) = 0,$$

$$u \geq 0, \quad v \geq 0.$$

Hence,

$$\left.\begin{array}{ll} \text{(i)} & (M + M^T)x + q = M^T u + v \\ \text{(ii)} & vx = u(Mx + q) = 0 \\ \text{(iii)} & x \geq 0, \; u \geq 0, \; v \geq 0, \; Mx + q \geq 0 \end{array}\right\} \qquad (2.8)$$

It follows from (i) that

$$Mx + q = M^T(u - x) + v.$$

Multiplying this equation by u and x successively and using (ii) we get

$$0 = u(Mx + q) = uM^T(u - x) + uv$$

$$x(Mx + q) = xM^T(u - x)$$

Subtracting and using (iii) and the fact M is positive semidefinite we finally get

$$0 \geq -x(Mx+q) - uv = (u-x)M^T(u-x) \geq 0.$$

Hence

$$x(Mx+q) = 0, \quad x \geq 0, \quad Mx+q \geq 0,$$

that is x solves $LCP(M, q)$. ∎

2.9 Lemma. *Let M be positive definite, ρ the smallest eigenvalue of $(M + M^T)$. If $\rho > 1/\gamma$ then the Hessian $\nabla^2 D$ of $D(x, u, v)$ in (2.6) is negative definite and symmetric.*

Proof

Let us define

$$z = (x,\ u,\ v)^T \quad \text{and} \quad B = (M + M^T \quad -M^T \quad -I)$$

so that $Bz = (M+M^T)x - M^T u - v$ and the expression for D now becomes

$$D(z) = x(Mx+q) - u(Mx+q) - vx - \frac{\gamma}{2} \cdot \|Bz+q\|^2.$$

Thus

$$\nabla D(z) = \begin{pmatrix} Bz+q \\ -(Mx+q) \\ -x \end{pmatrix} - \gamma B^T(Bz+q),$$

and hence

$$\nabla^2 D(z) = \begin{pmatrix} M+M^T & -M^T & -I \\ -M & 0 & 0 \\ -I & 0 & 0 \end{pmatrix} - \gamma B^T B$$

$= A - \gamma B^T B$ say. Clearly $\nabla^2 D$ is symmetric, while

$$
\begin{aligned}
& z^T D z \\
&= z^T A z - \gamma \cdot z^T B^T B z \\
&= < z,\ (Bz\ - Mx\ - x) > -\gamma \cdot \|Bz\|^2 \\
&= 2xMx - 2xM^T u - 2xv - \gamma \|Bz\|^2 \\
&= -2xMx + 2x\{(M + M^T)x - M^T u - v\} - \gamma\|Bz\|^2 \\
&\leq\ -\rho \cdot \|x\|^2 + 2\|x\| \cdot \|Bz\| - \gamma\|Bz\|^2 \\
&= -\rho\{\|x\|^2 - (2/\rho)\|x\|\,\|Bz\| + \|Bz\|^2/\rho^2\} + (\tfrac{1}{\rho} - \gamma)\|BZ\|^2 \\
&= -\rho\{\|x\| - (1/\rho)\|Bz\|\}^2 - (\gamma - \tfrac{1}{\rho})\|Bz\|^2\} \\
&< 0
\end{aligned}
$$

since $\gamma\rho\ < 1$. Hence $\nabla^2 D$ is negative definite. ∎

2.10 Theorem. *Let M be positive definite and ρ the smallest eigenvalue of $(M + M^T)$. Suppose $\gamma > 1/\rho$. Then the (unique) solution $(x(\gamma),\ v(\gamma)$ of the program*

$$
\min_{(x,v)\geq 0} \quad \{xv + \frac{\gamma}{2}\|Mx + q - v\|^2\} \tag{2.11}
$$

solves $LCP(M,\ q)$ with $v(\gamma) = Mx(\gamma) + q$.

Proof

From Theorem 2.7, every stationary point of (2.6) solves $LCP(M, q)$, which, since M is positive definite, has a unique solution $\bar{x}$. Now $D(x, u, v)$ is strictly concave (Lemma 2.9) so that (2.6) has at most one stationary point satisfying the conditions (2.8). However, (2.8) is satisfied if we take $u = x$ so that any stationary point of (2.6) *must* satisfy $u = x$ also.

Hence we can simplify (2.6) to obtain (2.11). Finally, letting $x = \overline{x}$ and $v = M\overline{x} + q$ in (2.11) we see that

$$0 \leq \min_{(x,v)\geq 0} \left\{xv + \frac{\gamma}{2}\|Mx + q - v\|^2\right\} \leq 0$$

so that $(\overline{x},\ M\overline{x} + q)$ solves (2.11) uniquely. ∎

We note that (2.11) was also considered by Eijndhoven [1985].

2.12 Corollary. *Let M be positive semidefinite and $\{\varepsilon_n\}$ a sequence of positive reals such that $\varepsilon_n \downarrow 0$. Assume that $S(M,\ q)$ is not empty. Then*

$$\min_{(x,v)\geq 0} \left\{\varepsilon_n xv + \frac{1}{2}\|(M + \varepsilon_n I)x + q - v\|^2\right\} \tag{2.13}$$

has a sequence of solutions $\{x(\varepsilon_n),\ v(\varepsilon_n)\}$ such that $x(\varepsilon_n) \to \overline{x}$ and $v(\varepsilon_n) \to M\overline{x} + q$ where $\overline{x}$ is the least two-norm solution of $LCP(M,\ q)$.

Proof

For any $\varepsilon_n > 0$, $M_n\ (:= M + \varepsilon_n I)$ is positive definite. The smallest eigenvalue of $M_n + M_n^T$ is at least $2\varepsilon_n$. Now use Theorem 2.10. The convergence of $x(\varepsilon_n)$ follows from Theorem 1.5.1 (e). ∎

2.14 Preassigned tolerance

Suppose that M is positive semidefinite with *int* $S(M,\ q)$ not empty. When an approximate solution satisfying a preassigned tolerance would suffice, it is sufficient to solve (2.13) for a single $\varepsilon > 0$. We make this precise in the following Proposition.

2.15 Theorem. *Let $\delta > 0$ be a preassigned tolerance. Assume that M is positive semidefinite and that $int\, S(M,\ q) \neq \emptyset$. Then there exists $\bar{\varepsilon} > 0$ such that $\forall \varepsilon,\quad 0 < \varepsilon < \bar{\varepsilon}$, the unique solution $x(\varepsilon)$ of $LCP(M + \varepsilon I,\ q)$ satisfies*

$$x(\varepsilon) \geq 0, \qquad \|w(\varepsilon) - w(\varepsilon)_+\| \leq \delta \quad \textit{and} \quad |< x(\varepsilon),\ w(\varepsilon) >| < \delta$$

where $w(\varepsilon) = Mx(\varepsilon) + q$.

Proof

From the last Corollary, $x(\varepsilon)$ solves (2.13) with $\varepsilon_n = \varepsilon,\ v(\varepsilon) = w(\varepsilon) + \varepsilon x(\varepsilon)$. It is known that $int\, S(M,\ q) \neq \emptyset \implies \overline{S}(M,\ q)$ is bounded. For a proof see [Mangasarian, 1982], where this is shown for the more general case when M is *copositive plus*, that is,

$$\text{(i)} \quad x \geq 0 \implies xMx \geq 0 \qquad \text{(ii)} \quad xMx = 0 \implies Mx = 0.$$

Hence $\exists K \geq 1$ such that $\|\overline{S}(M,\ q)\| \leq K$ (to find K one can use the bounds obtained by Mangasarian and McLinden [1985] by solving a single linear program if necessary). Now choose $\bar{\varepsilon} = \delta / K^2$.

For any ε, from (2.13) we have

$$x(\varepsilon)v(\varepsilon) = 0, \quad v(\varepsilon) - w(\varepsilon) = \varepsilon x(\varepsilon).$$

Hence,

$$\begin{aligned} |x(\varepsilon)w(\varepsilon)| = \varepsilon\|x(\varepsilon)\|^2 \ &\le \ K\,\varepsilon\,\|x(\varepsilon)\| \\ &\le K\,\bar{\varepsilon}\|x(\bar{\varepsilon})\| \\ &< K^2 \cdot \delta/K^2 \\ &= \delta, \end{aligned}$$

from Theorem 1.5.1 (b). Also since $w(\varepsilon)_+$ is the closest point to $w(\varepsilon)$ in $\Re^n_+$, we have

$$\begin{aligned} \|w(\varepsilon) - w(\varepsilon)_+\| \le \|w(\varepsilon) - v(\varepsilon)\| &= \varepsilon \cdot \|x(\varepsilon)\| \\ &\le \bar{\varepsilon} \cdot \|x(\bar{\varepsilon})\| \ \le \ \frac{\delta}{K^2} \cdot K \\ &< \ \delta. \end{aligned}$$

This completes our proof. ▮

3. SOR Algorithms

The solution of $LCP(M,\ q)$ when M is a positive semidefinite matrix reduces to the solution of a sequence of subproblems of the type (2.13). We consider in this section iterative procedures to solve these subproblems.

Let $\varepsilon > 0$ and write $N = M + \varepsilon I$. Let $z = (x,\ v)^T \in \Re^{2n}_+$. The objective function $\varphi(z)$ in (2.13) can then be written as

$$\varphi(z) = \varepsilon\, xv + \frac{1}{2}\|Nx + q - v\|^2.$$

We then have

$$\nabla\varphi(z) = \varepsilon \begin{pmatrix} v \\ x \end{pmatrix} + \begin{pmatrix} N^T(Nx + q - v) \\ -(Nx + q - v) \end{pmatrix}$$

and

$$\nabla^2\varphi(z) = \begin{pmatrix} N^T N & \varepsilon I - N^T \\ \varepsilon I - N & I \end{pmatrix}.$$

Hence (2.13) is simply a quadratic program with a symmetric positive definite Hessian (Lemma 2.9). Thus we need to consider iterative procedures to solve problems of the type

$$\min_{x\geq 0} \quad f(x) := \frac{1}{2}x^T M x + qx \tag{3.1}$$

where M is positive definite and symmetric. For simplicity, we assume that M is size $n \times n$ and $q \in \Re^n$.

Cryer [1971] has proposed the following successive overrelaxation algorithm (SOR) to solve (3.1)

3.2 Cryer's Algorithm

(1) Let $0 < \omega < 2$, and M positive definite symmetric.

(2) Let $x^0 \geq 0$ be any arbitrary starting point.

(3) For $k \geq 0$ given x^k, determine x^{k+1} as follows: Having

$$x_1^{k+1}, \cdots x_j^{k+1}, \quad j < n,$$

let

$$\xi^j = (x_1^{k+1}, \cdots, x_j^{k+1}, x_{j+1}^k, \cdots, x_n^k).$$

Then

$$x_{j+1}^{k+1} = \left\{x_{j+1}^k - \omega \cdot D_{jj}^{-1} \cdot (\nabla f(\xi^j)_j\right\}_+$$

where D is the diagonal of M.

Cryer has shown that the iterates $\{x^k\}$ converge to the unique solution of (3.1). This algorithm, however, is a special case of a more general class of algorithms due to Mangasarian [1977, Algorithm 2.1].

3.3 Mangasarian's Algorithm

(1) Let $x^0 \geq 0$ and M symmetric.

(2) For $k \geq 0$, having x^k determine x^{k+1} from

$$x^{k+1} = \lambda\Big\{x^k - \omega E^k\big(Mx^k + q + K^k(x^{k+1} - x^k)\big)\Big\}_+ + (1-\lambda)x^k$$

where $0 < \lambda \leq 1$, $\omega > 0$, $\{E^k\}$ and $\{K^k\}$ are bounded sequences of $n \times n$ matrices with each E^k being a positive diagonal matrix satisfying $E^k > \alpha I$ for some $\alpha > 0$ and such that for some $\gamma > 0$ we have

$$y^T\Big\{(\lambda\omega E^k)^{-1} + K^k - \frac{M}{2}\Big\}y \geq \gamma\|y\|^2 \tag{3.4}$$

for all k and for all $y \in \Re^n$.

Mangasarian has shown that $f(x^k) \geq f(x^{k+1})$ and that every accumulation point of x^k , if there exists one, solves $LCP(M, q)$. When M is positive definite, the level sets $\big\{x : f(x) \leq f(x^0)\big\}$ are compact, and in this case x^k converges to the (unique) solution of $LCP(M, q)$. It is easy to see that Cryer's algorithm follows from (3.3) by taking $\lambda = 1$, $E^k = D$, the diagonal of M and $K^k = L$ the strict lower triangular part of M with $0 < \omega < 2$.

4. Modified Mangasarian Algorithm (MSOR)

In this section we shall consider a modified form of Algorithm (3.3) providing at the same time, perhaps, a simplified proof of Mangasarian's convergence theorem.

(1) Let $x^0 \geq 0$ be arbitrary and M symmetric.

(2) For $k \geq 0$, having x^k, determine z^k from

$$z^k = \lambda \Big\{ x^k - \omega E^k \big(Mx^k + q + K^k(z^k - x^k) \big) \Big\}_+ + (1 - \lambda) x^k$$

where the sequences of matrices $\{E^k\}$ and $\{K^k\}$ satisfy all the requirements of Algorithm (3.3), $0 < \lambda \leq 1$, $\omega > 0$.

(3) Choose $x^{k+1} \geq 0$ such that $f(z^k) \geq f(x^{k+1})$.

Theorem 4.2. *Assume that $x^\star$ is an accumulation point of $\{x^k\}$. Then $x^\star$ solves $LCP(M,\ q)$.*

Proof

As in the proof of [Mangasarian, 1977, Theorm 2.1], we have

$$f(x^k) - f(x^{k+1}) \geq f(x^k) - f(z^k) \geq \gamma \|z^k - x^k\|^2 \geq 0.$$

Suppose now that $\exists \{k_j\}$ such that $x^{k_j} \to x^\star$ and $E^{k_j} \to E^\star$. It is easy to see that

$$\begin{aligned} f(x^{k_j}) - f(x^{k_{j+1}}) &\geq f(x^{k_j}) - f(x^{k_j+1}) \\ &\geq f(x^{k_j}) - f(z^{k_j}) \\ &\geq \gamma \|z^{k_j} - x^{k_j}\|^2 \end{aligned}$$

By continuity of f the left hand side goes to 0 in the limit so that $z^{k_j} - x^{k_j} \rightarrow 0$. Since K^{k_j} is bounded, this implies that

$$x^* = \lim_{j \to \infty} z^{k_j} = \lambda\big(x^* - \omega E^*(Mx^* + q)\big)_+ + (1-\lambda)x^*$$

and by [Mangasarian, 1977, Lemma 2.1], x^* solves $LCP(M,\ q)$. ∎

5. Implementation of MSOR

We shall now consider the implementation of MSOR to solve (3.1) and more generally the case when M is symmetric and positive semidefinite. Assume that the point x^k is given and define d^k, the *direction* at the point x^k to be $z^k - x^k$. A convenient choice for the point x^{k+1} is given by $x^k + \sigma_k d^k$ where

$$\sigma_k = \arg\min\big\{f(x^k + \sigma d^k) : x^k + \sigma d^k \geq 0,\ \sigma \geq 0\big\}.$$

It is easy to see that $d^k = 0 \iff x^k$ solves $LCP(M,\ q)$. Indeed, the proof of Theorem 4.2 essentially rests on the fact that the sequence d^k has a subsequence $d^{k_j} \rightarrow 0$. We assume in the following that $d^k \neq 0$.

Consider first the case when M is positive definite. Then $\|d^k\|_M \neq 0$. Define

$$\beta_k = \min_{1 \leq j \leq n} \left\{ \frac{-x_j^k}{d_j^k} : d_j^k < 0 \right\}$$

where $\beta_k = \infty$ if $d^k \geq 0$. Also the *unconstrained minimum* of $f(x^k + \sigma d^k)$ occurs at the point given by

$$0 = \frac{d}{d\sigma} f(x^k + \sigma d^k) = d^k\big(M(x^k + \sigma d^k) + q\big)$$

that is at the point

$$x^k + s_k d^k \quad \text{where} \quad s_k = \frac{-d^k(Mx^k + q)}{\|d^k\|_M^2}.$$

It is now easy to see that the stepsize σ_k for our choice of x^{k+1} is given by $\sigma_k = \min(\beta_k, s_k)$.

When M is only positive semidefinite, it is possible that $d^k M d^k = 0$, that is, $Md^k = 0$ so that $s_k = \infty$. In this case $\sigma_k = \beta_k$ provided that β_k is finite. If $\beta_k = \infty$, that is, $d^k \geq 0$ then $f(x)$ is unbounded in (3.1). Thus in this case the algorithm terminates. We note that the iterates $\{x^k\}$ need not have any accumulation point in general when M is only positive semidefinite.

Thus for purposes of implementation, MSOR can be described as follows. Let M be positive semidefinite and symmetric. Let λ, ω, $\{K^k\}$ and E^k be as in MSOR.

(1) Let $x^0 \geq 0$.

(2) For $k \geq 0$, given x^k, define

$$z^k = \lambda\Big\{x^k - \omega E^k\big(Mx^k + q + K^k(z^k - x^k)\big)\Big\}_+ + (1-\lambda)x^k,$$
$$d^k = z^k - x^k.$$

(3) If $d^k = 0$ stop; x^k is optimal.

(4) If $Md^k = 0$ and $d^k \geq 0$ stop; problem has no solution.

(5) If $d^k \not\geq 0$, let

$$\beta_k = \min\Big\{-\frac{x_j^k}{d_j^k} : d_j^k < 0\Big\}.$$

(6) If $Md^k \neq 0$, let

$$s_k = \frac{-d^k(Mx^k + q)}{d^k M d^k}$$

(7) Set $\sigma_k = \min(\beta_k,\ s_k),\quad x^{k+1} = x^k + \sigma_k d^k.$

6. Remarks

The above algorithm was implemented as follows. We chose $\lambda = 1,\quad \omega = 1.8,\quad K^k = L$ where L is the strict lower triangular part of M. Also, E^k was chosen to be E, where $E_{ii} = D_{ii}$, for $D_{ii} > 0$ else $E_{ii} = 1$. We found the algorithm to be robust and fast. For positive definite symmetric matrices of size 40×40, the problems were usually solved in about .05 seconds. As is to be expected, the algorithm is much faster than Cryer's algorithm. We shall present a more detailed report in Chapter 6.

CHAPTER 4

FIXED POINT METHODS

1. Introduction

In this chapter we shall be concerned with $NLCP(F)$ when F is monotone and with $LCP(M, q)$ when M is positive semidefinite. Since a fixed point $\overline{x}$ of the map $x \mapsto \big(x - F(x)\big)_+$ solves $NLCP(F)$ we shall consider iterative methods such that the iterates converge to such a fixed point.

We shall also consider summability methods to solve $LCP(M, q)$, when M is positive semidefinite. We shall construct a sequence of iterates which is shown to be bounded if and only if $LCP(M, q)$ is solvable. In this case, a (summability) matrix transform of these iterates is shown to converge to a solution of $LCP(M, q)$.

As an interesting corollary, we consider the case of $NLCP(F)$ where F is monotone and satisfies the *distributed Slater constraint qualification* of Mangasarian & McLinden [1985]. In this case, these authors have established that $NLCP(F)$ is solvable. We shall show that the matrix transform of the iterates constructed as above converges to a solution.

2. Fixed point methods

We begin this section with the well known notion of a *contraction mapping*.

2.1 Definition. *Let $P : D \subseteq \Re^n \longrightarrow \Re^n$. We say P is Lipschitzian with modulus $L > 0$ if*

$$\|P(x) - P(y)\| \leq L\,\|x - y\| \qquad \forall x,\ y \in D.$$

When $L \leq 1$ $(L < 1)$ we say P is non-expansive (contractive) .

The following Theorem is classical; see e.g., [Ortega and Rheinboldt, 1970, page 120].

2.2 Theorem. *(Banach's contraction mapping princple). Let P: $D \subseteq \Re^n \longrightarrow \Re^n$, D_0 a closed subset of D and such that $PD_0 = \{P(x) : x \in D_0\} \subseteq D_0$. If P is a contraction mapping on D_0 with modulus L, then P has a unique fixed point $\overline{x}$ in D_0. Further, for any point x^0 in D_0, the sequence $\{x^k\}$ where $x^{k+1} = P(x^k)$, converges to $\overline{x}$ with the following linear rate :*

$$\frac{\|x^{k+1} - \overline{x}\|}{\|x^k - \overline{x}\|} \leq L.$$

The content of the following Proposition is well known. We state it in following form for later use and furnish a proof for the sake of completeness.

2.3 Proposition. *Let F: $D \subseteq \Re^n \longrightarrow \Re^n$ be monotone and Lipschitzian with modulus L. Suppose that $\varepsilon > 0$, $\alpha > 0$ and $\varepsilon\alpha \leq 1$. Then the projection map* $\mathbb{P}$ defined by

$$\mathbb{P}(x) = \Big\{x - \alpha\big(F(x) + \varepsilon x\big)\Big\}_+, \qquad x \in D$$

is also Lipschitzian with modulus $k(\alpha) = \sqrt{(1 - \alpha\varepsilon)^2 + (\alpha L)^2}$. If $\alpha < 2\varepsilon/\sqrt{\varepsilon^2 + L^2}$, then $\mathbb{P}$ is a contraction and k attains its minimum value

$$k_{\min}(\alpha) = L/\sqrt{L^2 + \varepsilon^2} \quad for \quad \alpha = \varepsilon/L^2 + \varepsilon^2.$$

Proof

We have

$$\|\mathbb{P}(x) - \mathbb{P}(y)\|^2 = \|\{x - \alpha(F(x) + \varepsilon x)\}_+ - \{y - \alpha(F(y) + \varepsilon y)\}_+\|^2$$
$$\leq \|\{x - \alpha(F(x) + \varepsilon x)\} - \{y - \alpha(F(y) + \varepsilon y)\}\|^2$$

since projection on $\Re^n_+$ is non-expansive. Hence,

$$\|\mathbb{P}(x) - \mathbb{P}(y)\|^2 \leq \|(x-y)(1 - \varepsilon\alpha) - \alpha(F(x) - F(y))\|^2$$
$$= \|x-y\|^2 (1 - \alpha\varepsilon)^2 + \alpha^2 \|F(x) - F(y)\|^2$$
$$- 2\alpha(1 - \alpha\varepsilon)(x-y)(F(x) - F(y)).$$

Since $\alpha\varepsilon \leq 1$ and $< F(x) - F(y), x - y > \geq 0$ from the monotonicity of F,

$$\|\mathbb{P}(x) - \mathbb{P}(y)\|^2 \leq \|x-y\|^2 \{(1 - \alpha\varepsilon)^2 + (\alpha L)^2\}.$$

The other claims about $k(\alpha)$ are easy to verify. ∎

2.4 Theorem. *Let $F : \Re^n \longrightarrow \Re^n$ be monotone and Lipschitzian with modulus L. Let $\{\varepsilon_n\}$ be a sequence of positive reals, $\varepsilon_n \downarrow 0$. For $n = 1, 2, \cdots$ let*

$$\mathbb{P}_n(x) = \{x - \alpha_n(F(x) + \varepsilon_n x)\}_+$$

and for $m = 1, 2, \cdots$ and $x \in \Re^n$ let

$$\mathbb{P}_n^m(x) = \underbrace{\mathbb{P}_n \circ \cdots \circ \mathbb{P}_n}_{m \text{ times}}(x) = x(n,m).$$

Suppose further that

$$\alpha_n = \frac{\varepsilon_n}{\varepsilon_n^2 + L^2}, \quad k_n = \frac{L}{\sqrt{L^2 + \varepsilon_n^2}}, \quad \delta_n = \varepsilon_n(1 - k_n).$$

For $n = 1, 2, \cdots,$ let $\overline{x}^n$ be defined by

$$\overline{x}^n = x(n,m), \quad where \quad \|x(n,m+1) - x(n,m)\| < \delta_n.$$

Then the sequence $\{\|\overline{x}^n\|\}$ is bounded if and only if $NLCP(F)$ is solvable and in this case, $\overline{x}_n \rightarrow \overline{x}$, the least two-norm solution of $NLCP(F)$.

Proof

From Proposition 2.3, $\mathbb{P}_n$ is a contraction with modulus $k_n < 1$. By the contraction mapping principle, given any x^0,

$$\lim_{j\to\infty} \mathbb{P}_n{}^j \longrightarrow z^n, \quad \mathbb{P}_n(z^n) = z^n.$$

Note that z^n solves $NLCP(F + \varepsilon_n I)$ uniquely. Since, by definition,

$$\mathbb{P}_n(x(n,m)) = x(n,\, m+1)$$

we have

$$\delta_n > \|x(n,m+1) - x(n,m)\| \geq \|x(n,m) - z^n\| - \|x(n,m+1) - z^n\|$$

and

$$\|x(n,m+1) - z^n\| = \|\mathbb{P}_n(x(n,m) - \mathbb{P}_n(z^n)\| \leq k_n \cdot \|x(n,m) - z^n\|$$

it follows that

$$\delta_n > \|x(n,m+1) - x(n,m)\| \geq (1 - k_n)\, \|x(n,m) - z^n\|$$

and that

$$\|\overline{x}^n - z^n\| < \varepsilon_n.$$

From Theorem 1.5.1 (d) our conclusions about $\{\overline{x}^n\}$ follow. ∎

We remark that the last Theorem is a *two-step* process in the sense that for a given ε_n, the contraction $\mathbb{P}_n$ is iterated m times until $x(n,m)$ is close enough to the solution z^n of $NLCP(F + \varepsilon_n)$. One then takes a smaller ε_n and the process repeats. Our aim now is to prove convergence for an algorithm which combines both steps into a single step. We shall need the following notions from the theory of Summability.

2.5 Definition. *An infinite matrix $A = (A_{ij})$, $i,\ j = 1,\ 2, \cdots$, is said to be convergence preserving if for any sequence $\{x_n\}$, the sequence $\{y_n\}$ defined by*

$$y_n = \sum_{j=1}^{\infty} A_{nj} x_j \tag{2.6}$$

is well defined and $\lim x_n = \lim y_n$. *We call $\{y_n\}$ the A-transform of $\{x_n\}$ and write $y_n = A(\{x_n\})$.*

The following Theorem is classical. Its proof may be found for instance in [Peyerimhoff, 1969].

2.7 Theorem. *(O. Toeptlitz). An infinite matrix $A = (A_{ij})$, $i,\ j = 1,\ 2, \cdots$ is convergence preserving if and only if*

(1) $\sum_{j=1}^{\infty} |A_{ij}| = \sigma_i$ *exists,*

$\{\sigma_i\}$ *is bounded,*

(2) $\lim_i(\sum_{j=1}^{\infty} A_{ij}) = 1$,

(3) $\lim_i A_{ij} = 0$.

We are now ready to prove the principal theorem of this chapter.

2.8 Theorem. *Let M be a positive semidefinite matrix. Assume that that the sequences $\{\overline{\alpha}_n\}$ and $\{\overline{\varepsilon}_n\}$ of positive reals are such that*

$$\left.\begin{aligned} &\sum_{n=1}^{\infty} \overline{\alpha}_n \ \textit{diverges}, \\ &\sum_{n=1}^{\infty} \overline{\alpha}_n^2 \ \textit{converges}, \\ &\sum_{n=1}^{\infty} \overline{\alpha}_n \overline{\varepsilon}_n \ \textit{converges and} \\ &\overline{\varepsilon}_n \leq 1, \quad \overline{\rho}_n = \frac{\overline{\alpha}_n}{\overline{\varepsilon}_n} \downarrow 0. \end{aligned}\right\} \tag{2.9}$$

Suppose that k is the smallest positive integer satisfying

$$\sqrt{2\overline{\rho}_k} + 2\overline{\rho}_k < L := \frac{1}{1 + \|M\|} \tag{2.10}.$$

Let $B = (B_{ij})$ be the infinite matrix whose n^{th} row B_n is defined by

$$B_n = \left(\frac{\overline{\alpha}_{1+k}}{S_n}, \frac{\overline{\alpha}_{2+k}}{S_n}, \ldots, \frac{\overline{\alpha}_{n+k}}{S_n}, 0, \cdots \right)$$

where $S_n = \sum_{j=1}^{n} \overline{\alpha}_{j+k}$. Let $x^0 = 0$ and having x^n, determine x^{n+1} from

$$x^{n+1} = \left\{ (1 - \overline{\alpha}_{n+k}\overline{\varepsilon}_{n+k})x^n - \overline{\alpha}_{n+k}(Mx^n + q) \right\}_+ \tag{2.11}.$$

Let $\{y^n\}$ be the B-transform of $\{x^n\}$, that is

$$y^n = \frac{1}{S_n}\left(\sum_{j=1}^{n} \overline{\alpha}_{j+k} x^j \right) \tag{2.12}.$$

Then $\overline{S}(M, q) \neq \emptyset \iff \{x^n\}$ *is bounded. When this condition holds,*

$$y^n \longrightarrow y^* \in \overline{S}(M, q).$$

Proof

Assume that k satisfies (2.10). For notational convenience, we shall write

$$\alpha_n = \overline{\alpha}_{n+k}, \quad \varepsilon_n = \overline{\varepsilon}_{n+k} \quad \text{and} \quad \rho_n = \overline{\rho}_{n+k}.$$

Obviuosly, the sequences $\{\alpha_n\}$, $\{\varepsilon_n\}$ and $\{\rho_n\}$ also satisfy the conditions (2.13). We shall write

$$Fx = Mx + q, \; F_n x = Fx + \varepsilon_n x.$$

Thus we can write (2.11) in the form

$$x^{n+1} = \left\{(1 - \alpha_n \varepsilon_n)x^n - \alpha_n F x^n\right\}_+ \qquad (2.13).$$

We first assume that $\{x^n\}$ is bounded and show that in this case $y^n \to y^* \in \overline{S}(M, q)$ so that $\overline{S}(M, q) \neq \emptyset$.

Assume then that $\{x^n\}$ is bounded. Clearly, $\exists K_1 > 0$ and $K_2 > 0$ such that

$$\begin{aligned} \|x^n\| &\leq K_1, \\ \|F_n x^n\| &= \|Mx^n + q + \varepsilon_n x^n\| \\ &\leq (1 + \|M\|) \cdot \|x^n\| + \|q\| \\ &\leq (1 + \|M\|) \cdot K_1 + \|q\| \\ &:= K_2. \end{aligned}$$

Let $x \in \Re^n_+$ be arbitrary but fixed. Then from (2.13) we have

$$\begin{aligned} \|x^{n+1} - x\|^2 &= \|(x^n - \alpha_n(F_n x^n))_+ - x\|^2 \\ &\leq \|(x^n - x) - \alpha_n(Fx^n + \varepsilon_n x^n)\|^2 \\ &\leq \|(x^n - x)\|^2 - 2\alpha_n(Fx^n)(x^n - x) \\ &\quad - 2\alpha_n\varepsilon_n x^n(x^n - x) + \alpha_n^2 K_2^2. \end{aligned} \tag{2.14}$$

Since M is positive semidefinite we also have

$$(Fx^n)(x^n - x) \geq (Fx)(x^n - x).$$

Let

$$K_3 = \sup_n \|x^n\| \cdot \|x^n - x\|.$$

From (2.14) we now get

$$2\alpha_n(Fx)(x^n - x) \leq \|x^n - x\|^2 - \|x^{n+1} - x\|^2 + 2\alpha_n\varepsilon_n K_3 + \alpha_n^2 K_2^2.$$

Summing this from 1 to k we obtain

$$2(Fx)\sum_{n=1}^{k} \alpha_n(x^n - x) \leq \|x^1 - x\|^2 - \|x^{k+1} - x\|^2 + 2K_3 \sum_{n=1}^{k} \alpha_n\varepsilon_n + K_2^2 \sum_{n=1}^{k} \alpha_n^2.$$

Divide this last inequality by S_k and let $k \to \infty$. From the assumed properties in (2.9) of the sequences $\{\alpha_n\}$, $\{\varepsilon_n\}$ and from the definition of $\{y^n\}$, we now have

$$\liminf_k < Fx,\ x - y^k > \geq 0.$$

Since y^n is a convex combination of $x^1,\ x^2, \cdots, x^n$ it follows that $\{x^n\}$ bounded $\Rightarrow \{y^n\}$ is bounded. Hence $\{y^n\}$ has a limit point y^* for which

$$< Fx,\ x - y^* > \ \geq\ 0.$$

Since $x \in \Re^n_+$ was arbitrary, y^* solves $LCP(M,\ q)$. This completes our proof hat

$$\{x^n\} \quad \text{bounded} \quad \Longrightarrow \quad \overline{S}(M,\ q) \neq \emptyset.$$

Next we prove $y^n \longrightarrow y^*$.

Since $\overline{S}(M,\ q) \neq \emptyset$, choose $z \in \overline{S}(M,\ q)$ arbitrary but fixed. By Theorem 2.4.2, z satisfies

$$< Fx^n,\ x^n - z > \ \geq\ 0 \qquad (2.15)$$

since $x^n \geq 0$. From (2.13) and (2.15) we have

$$\begin{aligned} \|x^{n+1} - z\|^2 &\leq \|(x^n - z) - \alpha_n(Fx^n + \varepsilon_n x^n)\|^2 \\ &\leq \|x^n - z\|^2 - 2\alpha_n(Fx^n)(x^n - z) \\ &\quad - 2\alpha_n\varepsilon_n\, x^n(x^n - z) + \alpha_n^2 K_2^2 \\ &\leq \|x^n - z\|^2 + 2\alpha_n\varepsilon_n\, |x^n(x^n - z)| + \alpha_n^2 K_2^2. \end{aligned}$$

Define $\beta_n(z)$ by

$$\beta_n(z) := 2\alpha_n\varepsilon_n\, \|x^n\|\, \|x^n - z\| + \alpha_n^2\, K_2^2 \qquad (2.16)$$

and we now have

$$\|x^{n+1} - z\|^2 \ \leq\ \|x^n - z\|^2 + \beta_n(z) \qquad (2.17).$$

Let $\overline{S}$ denote $\overline{S}(M, q)$ and let

$$z^n = P_{\overline{S}}(x^n).$$

We are going to show that $\exists z^*$ such that

$$z^n \longrightarrow z^*, \quad y^n \longrightarrow z^*.$$

From (2.17) and the definition of z^n,

$$\|x^{n+1} - z^{n+1}\|^2 \leq \|x^{n+1} - z^n\|^2 \leq \|x^n - z^n\|^2 + \beta_n(z).$$

Since $\sum_n \beta_n(z)$ converges, by [Cheng, 1981, Lemma 2.2.12], we can conclude that

$$\|x^n - z^n\| \quad \text{converges.} \tag{2.18}$$

By parallelogram law, for $m > 0$,

$$\|z^{n+m} - z^n\|^2 = 2\|x^{n+m} - z^n\|^2 + 2\|x^{n+m} - z^{n+m}\|^2$$
$$- 4\|x^{n+m} - \frac{1}{2}(z^n + z^{n+m})\|^2.$$

Since $\overline{S}$ is convex, $(z^n + z^{n+m})/2 \in \overline{S}$. Also, z^{n+m} is the closest point to x^{n+m} in $\overline{S}$. Hence,

$$\|z^{n+m} - z^n\|^2 \leq 2\|x^{n+m} - z^n\|^2 - -2\|x^{n+m} - z^{n+m}\|^2. \tag{2.19}$$

Letting $z = z^n$ in (2.17) and noting that z^n is the closest point to x^n in $\overline{S}$, it follows that $\beta_n(z^n) \leq \beta_n(z)$. Now let $z = z^n$ in (2.17) and use induction to get

$$\|x^{n+m} - z^n\|^2 \leq \|x^n - z^n\|^2 + \sum_{j=n}^{n+m} \beta_j(z), \qquad m > 0.$$

Substitute this in (2.19) and we have

$$\|z^{n+m} - z^n\|^2 \leq 2\|x^n - z^n\|^2 - 2\|x^{n+m} - z^{n+m}\|^2 + 2\sum_{j=n}^{n+m} \beta_j(z). \tag{2.20}$$

¿From (2.18) and the fact $\sum_n \beta_n(z)$ converges, we have by letting $n, m \to \infty$ in (2.20) that

$$\|z^{n+m} - z^n\| \longrightarrow 0$$

so that $\{z^n\}$ is Cauchy. Since $\overline{S}$ is closed, $\exists\, z^\star \in \overline{S}$ such that $z^n \to z^\star$. We shall now show that $y^n \to z^\star$ as well.

Since $\{y^n\}$ is also bounded, let $y^\star$ be any of its limit points. Assume that the subsequence y^{n_k} converges to $y^\star$. From our proof earlier, $y^\star \in \overline{S}$. Observe that

$$z^j = P_{\overline{S}}(x^j) \Longrightarrow < x^j - z^j, y^\star - z > \leq 0. \tag{2.21}$$

Multiply (2.21) by α_j^2 and sum from $j = 1,\ 2, \cdots, n_k$ to get

$$\left\langle \sum_{j=1}^{n_k} \alpha_j(x^j - z^j),\ \sum_{j=1}^{n_k} \alpha_j(y^\star - z^j) \right\rangle \leq 0.$$

Divide the last inequality by $S_{n_k}^2$ to obtain

$$\left\langle y^{n_k} - \frac{1}{S_{n_k}} \sum_{j=1}^{n_k} \alpha_j z^j,\ y^\star - \frac{1}{S_{n_k}} \sum_{j=1}^{n_k} \alpha_j z^j \right\rangle \leq 0. \tag{2.22}$$

Notice however that

$$\xi^{n_k} := \frac{1}{S_{n_k}} \left(\sum_{j=1}^{n_k} \alpha_j z^j \right)$$

is simply a subsequence of the B-transform of $\{z^n\}$, that is of

$$\{\xi^n\} = B(\{z^n\}).$$

Since B satisfies all the conditions of Theorem 2.7, it is a convergence preserving matrix. However, $z^n \to z^*$ so that both ξ^n and ξ^{n_k} also converge to z^*. If we take limits as $k \to \infty$ in (2.22), we get

$$< y^* - z^*,\ y^* - z^* > \leq 0$$

so that $y^* = z^*$. But y^* was any *arbitrary* limit point of $\{y^n\}$. Hence $y^n \to z^*$. This completes our proof that

$$\{x^n\} \text{ bounded} \quad \Longrightarrow \quad \overline{S}(M, q) \neq \emptyset \ \text{ and } \ y^n \longrightarrow z^* \in \overline{S}(M, q).$$

We now prove the converse, that is we shall assume that $\overline{S}(M, q) \neq \emptyset$ and show that $\{x^n\}$ is bounded.

Recall from (2.10) that k satisfies

$$\sqrt{2\overline{\rho}_k} + 2\overline{\rho}_k \ < \ L \ = \frac{1}{1 + \|M\|}.$$

Hence there exists σ, $0 < \sigma < 1/2$ for which

$$\sqrt{2\overline{\rho}_k} + \overline{\rho}_k \ < \ \frac{1}{(1+\sigma)(1 + \|M\|)} \ := \ L_\sigma \quad .$$

The function $f(r)$,

$$f(r) := \frac{r}{r(1+\sigma)(1 + \|M\|) + \|q\|}$$

is strictly increasing in $[0, \infty]$, $\lim_r f(r) = L_\sigma$. Thus $\exists \bar{r} > 0$ such that for $r > \bar{r}$,

$$\sqrt{2\bar{\rho}_k} + 2\bar{\rho}_k \;<\; f(\bar{r}) \;<\; f(r).$$

Since $\bar{\rho}_n \downarrow 0$ and $\bar{\rho}_{n+k} = \rho_n$, we have for all $n > 0$ and $r > \bar{r}$,

$$\sqrt{2\rho_n} + 2\rho_n \;<\; f(\bar{r}) \;<\; f(r). \tag{2.23}$$

By assumption, $\overline{S} \neq \emptyset$. Let $z = P_{\overline{S}}(0)$, that is z is the least two-norm solution of $LCP(M, q)$. Define

$$r = \max(\bar{r}, \ \frac{1}{\sigma}\|z\|) + 1.$$

Our aim is to show that

$$\|x^n - z\| \;\leq\; r, \qquad \forall n \;\geq\; 0,$$

that is $\{x^n\}$ is bounded and this would complete our proof.

We use induction.

For $n = 0$, $\|x^0 - z\| = \|z\| \;<\; \sigma r \;<\; r$.

Suppose now that $\|x^n - z\| \;\leq\; r$. Let $\mu_n = \|x^n - z\|$. From (2.13),

$$\begin{aligned} \mu_{n+1}^2 &= \|x^{n+1} - z\|^2 \\ &\leq \|x^n - z\|^2 - 2\alpha_n\varepsilon_n\, x^n(x^n - z) \\ &\quad - 2\alpha_n(Fx^n)(x^n - z) + \alpha_n^2\|Fx^n + \varepsilon_n x^n\|^2. \end{aligned} \tag{2.24}$$

Since $z \in \overline{S}$, $(Fx^n)(x^n - z) \geq 0$. Also if $\mu_{n+1} \leq \mu_n$, we are done. So assume $\mu_{n+1} > \mu_n$. From (2.24) we thus get

$$2\alpha_n \varepsilon_n x^n(x^n - z) < \alpha_n^2 \|Fx^n + \varepsilon_n x^n\|^2,$$

that is

$$x^n(x^n - z) < \frac{\rho_n}{2}\|Fx^n + \varepsilon_n x^n\|^2. \tag{2.25}$$

Since

$$\begin{aligned}\|x^n\| &\leq \|x^n - z\| + \|z\| \\ &\leq r + \sigma r \\ &= (1+\sigma)r,\end{aligned}$$

we have

$$\begin{aligned}\|Fx^n + \varepsilon_n x^n\| &\leq \|Mx^n + q\| + \varepsilon_n \|x^n\| \\ &\leq \|M\|\,\|x^n\| + \|q\| + \|x^n\| \\ &\leq (1+\sigma)r \cdot (1 + \|M\|) + \|q\| \\ &= \frac{r}{f(r)} \\ &=: \xi \quad \text{say.}\end{aligned} \tag{2.26}$$

From (2.25) we now get

$$\begin{aligned}x^n(x^n - z) &< \frac{\rho_n}{2}\xi^2, \\ (x^n - z)(x^n - z) &< \frac{\rho_n}{2}\xi^2 - z(x^n - z) \\ &\leq \frac{\rho_n}{2}\xi^2 + \|z\| \cdot \|x^n - z\| \quad .\end{aligned}$$

Rewriting this last inequality,

$$\mu_n^2 < \frac{\rho_n}{2}\xi^2 + r\sigma\mu_n,$$

whence

$$2\mu_n^2 - 2r\sigma\mu_n - \rho_n\xi^2 < 0.$$

Since $\mu_n \geq 0$, we must have

$$\begin{aligned}\mu_n &< \frac{2r\sigma + \sqrt{4r^2\sigma^2 + 8\xi^2\rho_n}}{4} \\ &\leq \frac{r\sigma}{2} + \left(\frac{2r\sigma + 2\xi\sqrt{2\rho_n}}{4}\right)\end{aligned}$$

and finally

$$\mu_n < r\sigma + \frac{\xi}{2}\sqrt{2\rho_n}. \tag{2.27}$$

Again from the definition of x^{n+1} in (2.13),

$$\begin{aligned}\mu_{n+1} &= \|x^{n+1} - z\| \\ &\leq \|x^n - z - \alpha_n(F_n x^n)\| \\ &\leq \mu_n + \alpha_n\|F_n x^n\| \\ &\leq \mu_n + \rho_n\xi, \end{aligned} \tag{2.28}$$

where we have used (2.26) and the fact $\alpha_n < \rho_n$. If we use our estimate of μ_n from (2.27) in (2.28) we get

$$\mu_{n+1} < r\sigma + \frac{\xi}{2}\sqrt{2\rho_n} + \rho_n\xi.$$

Substituting for ξ from (2.26) and using (2.23) we finally get

$$\begin{aligned}\mu_{n+1} &< r\sigma + \frac{(\sqrt{2\rho_n} + 2\rho_n)}{2} \cdot \frac{r}{f(r)} \\ &< r\sigma + \frac{f(r)}{2} \cdot \frac{r}{f(r)} \\ &= r\sigma + \frac{r}{2} \\ &< r\end{aligned}$$

since $\sigma < 1/2$. Hence $\mu_{n+1} < r$. This completes our induction and also the proof of the Theorem. ▌

2.29 Remark

Our proof showing that $\{y^n\}$ converges by considering $z^n = P_{\overline{S}}(x^n)$ is patterned after [Baillon, 1975], who uses this technique to construct fixed points of non-expansive maps. Notice also Baillon's use of the *Cesàro matrix* C where $C_{ij} = 1/i$ for $j \leq i$, while $C_{ij} = 0$ for $j > i$.

3. Application to NLCP(F)

We shall now show that the proof of Theorem 2.8 can be used to construct a solution of $NLCP(F)$ when F is monotone and satisfies some regularity conditions. Our reference for this section is [Mangasarian and McLinden, 1985].

3.1 Definition. *Let* F: $D \subseteq \Re^n \longrightarrow \Re^n$. *We say that* F *satisfies the distributed Slater constrainst qualification (DSCQ) if there exist* p *points* $z^1, z^2, \cdots,, z^p \in D$, *nonnegative weights* $\lambda_1, \lambda_2, \cdots \lambda_p$ $\left(\sum_j \lambda_j = 1\right)$ *such that* $\hat{z} = \sum_j \lambda_j z^j \geq 0$ *and* $\hat{w} = \sum_j \lambda_j w^j > 0$ *where* $w^j = F(z^j)$.

Mangasarian and McLinden have proved the following Theorem.

3.2 Theorem. *Let* $F\colon D \subseteq \Re^n \longrightarrow \Re^n$, $\Re^n_+ \subset D$ *and suppose that* F *is monotone and continuous on* D . *Assume that* F *satisfies (DSCQ). Let*

$$\gamma > \max\big(1, \ -\hat{z}\hat{w} + \sum_{j=1}^{p} \lambda_j z^j w^j\big),$$

$$C = \{z \in \Re^n_+ \ : \ \hat{w}z \ \leq \ \hat{w}\hat{z} + \gamma\}$$

where λ_j, z^j, w^j, $\hat{z}$ *and* $\hat{w}$ *are as in (DSCQ). Then* $NLCP(F)$ *is solvable and has a solution* $z^\star$ *such that* $\hat{w}z^\star \ < \ \hat{w}\hat{z} + \gamma$.

We shall now show that the technique used in the proof of Theorem 2.9 can be used to construct a solution of $NLCP(F)$ guaranteed by Theorem 3.2.

3.3 Theorem. *Assume that* F *satisfies the hypotheses of Theorem 3.2 and let* C *be the compact convex set as defined in that Theorem. Let* $x^0 = 0$ *and given* x^n *find from* x^{n+1}

$$x^{n+1} = P_C\{x^n - \frac{F(x^n)}{n}\}.$$

Let B *be the Césaro matrix with*

$$B_n = \left(\frac{1}{S_n}, \ \frac{1}{2S_n}, \ \cdots, \ \frac{1}{nS_n}, \ 0, \ 0, \cdots\right), \quad S_n = \sum_{j=1}^{n} \frac{1}{j}$$

and let $\{y^n\} = B\big(\{x^n\}\big)$. *Then* y^n *converges to a solution of* $NLCP(F)$.

Proof

We shall give only a brief outline. Since $\{x^n\}$ and hence $\{y^n\}$ are both bounded, $\{y^n\}$ has a limit point $y^\star$. One uses the monotonicity of F to show that

$$\langle F(y^*, x - y^*\rangle \geq 0, \qquad \forall x \in C.$$

Hence y^* is a fixed point of the map $x \mapsto P_C(x - F(x))$. However, Mangasarian and McLinden show that any such fixed point satisfies $\hat{w}y^* < \hat{w}\hat{z} + \gamma$. Hence y^* solves $NLCP(F)$. One can now show that $y^n \to y^*$ by considering the projection z^n of x^n on $\bar{S}(F)$. ∎

3.4 Remark

From a computational point of view the fixed point methods in general, and those considered in this chapter in particular, are not viable methods. They are extremely slow and particularly so in the vicinity of a solution point since the step sizes taken in such a vicinity are extremely small. Their slowness in part is also due to the fact that they do not utilize special features of the matrix M in the case of $LCP(M, q)$. Their real utility is perhaps in generating good starting points for fast Newton-type algorithms (we consider one such algorithm in the next chapter). However, the SOR methods are much faster than the fixed point methods even for generation of starting points.

CHAPTER 5

GAUSS-NEWTON METHODS

1. Introduction

The SOR methods of Chapter 3 and the fixed point methods of Chapter 4 have the obvious advantage that the starting point can be any feasible point, usually the origin. The iterates are also easily calculated, thus reducing the cost of computation. Unfortunately, however, these methods are sometimes slow and require fairly large number of iterates. This is especially true of the fixed point methods which tend to slow down in the vicinity of a solution point.

In this chapter, we shall consider a class of faster algorithms, which is a version of the so called damped Gauss-Newton methods. We shall show that under fairly simple conditions these algorithms exhibit local superlinear convergence and under stronger hypotheses, local quadratic convergence. However, like all Newton-type methods, the algorithm is computationally expensive since one needs to solve a set of linear equations at each iteration and one obtains quadratic convergence only if one is close to a solution point. Nevertheless, as we shall see in the next chapter, this method used in conjunction with the slower methods has proved computationally effective.

2. Nonlinear equations and NLCP(F)

Our starting point for this chapter is the following Theorem due to Mangasarian [1976] which shows the equivalence of $NLCP(F)$ to the solution of a system of nonlinear equations.

2.1 Theorem. *Let* $F : \Re^n \longrightarrow \Re^n$, θ *a strictly increasing function such that* $\theta : \Re \to \Re$ *with* $\theta(0) = 0$. *Then* $z^\star$ *solves* $NLCP(F)$ *if and only if* $z^\star$ *solves the system of nonlinear equations* $G_i(z) = 0 \; i = 1, 2 \cdots n$ *where*

$$G_i(z) = \theta\big(|F_i(z) - z_i|\big) - \theta\big(F_i(z)\big) - \theta(z_i)$$

and $F_i(z)$ *is the* i^{th} *component of* $F(z)$.

We shall call $G(z)$ the *M-function* associated with F . Mangasarian also proved the following corollary which we shall find particularly useful later.

2.2 Corollary. *Suppose that* $z^\star$ *solves* $NLCP(F)$. *Assume further that* F *is differentiable at* $z^\star$,

(1) $z^\star$ *is nondegenerate, that is* $z^\star + F(z^\star) \; > \; 0$,

(2) $\nabla F(z^\star)$, *the Jacobian of* F *at* $z^\star$ *has nonsingular principal minors and*

(3) θ *is differentiable, strictly increasing such that*

$$\varsigma \; > \; 0 \Longrightarrow \theta'(\varsigma) + \theta'(0) \; > \; 0.$$

Then $z^\star$ *solves* $NLCP(F)$ *and the Jacobian* $\nabla G(z^\star)$ *of* G *is nonsingular.*

3. A Gauss-Newton algorithm

Let $F\colon \Re^n \longrightarrow \Re^n$ and consider $NLCP(F)$. For the rest of this chapter we shall assume that $\nabla F(z)$ exists and the function $\theta(\varsigma)$ in Theorem 2.1 is taken to be $\theta(\varsigma) = \varsigma|\varsigma|$. Let $G(z)$ be the associated M-function. Define

$$g(z) = \frac{1}{2}\|G(z)\|^2.$$

We are going to develop an algorithm to minimize g , somewhat in the spirit of least squares minimization. We note that $z^\star$ solves $G(z) = 0$ if and only if it is a global minimizer of g . Also, if $z^\star$ is a critical point of g and $\nabla G(z^\star)$ is nonsingular, then $G(z^\star) = 0$ since

$$\nabla g(z^\star) = \nabla G(z^\star)^T G(z^\star).$$

In this case, from Theorem 2.1, $z^\star$ solves $NLCP(F)$. Hence our aim is to find algorithms to find critical points of g.

Given $s \in \Re^n$, let us linearize G about s and consider

$$g_s(x) = \frac{1}{2}\|G(s) + \nabla G(s)(x-s)\|^2.$$

Then the gradient $\nabla g_s(x)$ of g_s is given by

$$\nabla g_s(x) = \nabla G(s)^T\big(G(s) + \nabla G(s)(x-s)\big).$$

Hence the Hessian $H_s(x)$ of g_s is given by

$$H_s(s) = \nabla G(s)^T \nabla G(s) =: A_s \qquad (3.1).$$

We note that A_s is positive semidefinite and symmetric.

3.2 Lemma, *Let* $\lambda > 0$. *For any* $x \in \Re^n$, *let* $A_x = \nabla G(x)^T \nabla G(x)$. *Suppose that* $\nabla g(x) \neq 0$. *Then the direction* p *given by*

$$(A_x + \lambda I)p = \nabla g(x)$$

is an ascent direction for g. *In particular,* $\exists w > 0$ *such that* $g(x - wp) < g(x)$.

Proof

There exist constants $\gamma \geq 0$ and $\nu > 0$ such that

$$\gamma \|h\|^2 \leq h^T A_x h \leq \nu \|h\|^2 \qquad \forall h \in \Re^n.$$

It follows that

$$(\gamma + \lambda)\|h\|^2 \leq h^T (A_x + \lambda I)h \leq (\nu + \lambda)\|h\|^2 \qquad \forall h \in \Re^n.$$

Since $\nabla g(x) \neq 0$, $p \neq 0$. If we take $h = p$, we get

$$p^T \nabla g(x) \geq (\gamma + \lambda)\|p\|^2 > 0.$$

It follows that $\nabla g(x) \cdot p > 0$ and hence [Ortega and Rheinboldt, 1970, 8.2.1] that p is an ascent direction for g. ∎

3.3 Damped Gauss-Newton Algorithm

The Lemma just proved leads us to consider the following algorithm.

(1) Let x^0 be given. Having x^k, determine x^{k+1} as follows :

(2) If $\nabla g(x^k) = 0$, stop.

(3) If $\nabla g(x^k) \neq 0$, define

$$A_k = \nabla G(x^k)^T \nabla G(x^k),$$

$$\lambda_k = g(x^k) \quad \text{and}$$

$$p^k = (A_k + \lambda_k I)^{-1} \nabla g(x^k).$$

Let ω_k be the largest element in the set $\Omega = \{1,\ 1/2, \cdots 1/2^n, \cdots\}$ such that $g(x^k - \omega_k p^k) < g(x^k)$. Set

$$x^{k+1} = x^k - \omega_k p^k.$$

3.4 Theorem. *Let x^0 be given and let $\{x^k\}$ the sequence determined by Algorithm 3.3. Assume that*

(3.5) $\sup_k \|\nabla G(x^k)\| < \infty$ *and*

(3.6) $\nabla g(x)$ *is Lipschitzian with modulus K on $\Re^n$.*

Then either $\{x^k\}$ terminates at a stationary point of g or else every limit point of $\{x^k\}$, if it exists, is a stationary point of g.

Proof

The first assertion is obvious. Let $0 < \delta \leq 1$. Let γ_k be the smallest eigenvalue of A_k. We claim that

$$\omega_k > \frac{(\lambda_k + \gamma_k)}{K}(1 - \delta). \tag{3.7}$$

If

$$\omega_k > \frac{2(\lambda_k + \gamma_k)(1-\delta)}{K},$$

then (3.7) holds trivially. Assume, therefore, that

$$\omega_k \leq \frac{2(\lambda_k + \gamma_k)(1-\delta)}{K}. \tag{3.8}$$

By our assumptions on $\nabla g(x)$,

$$\|\nabla g(y) - \nabla g(z)\| \leq K\|y - z\|.$$

Hence [Ortega, 1972, p 144],

$$\left| g(x^k - \omega_k p^k) - g(x^k) + \nabla g(x^k)\omega_k p^k \right| \leq \frac{K}{2} \cdot \omega_k^2 \|p^k\|^2.$$

Hence for $\nabla g(x^k) \neq 0$ (or equivalently, $p^k \neq 0$),

$$\begin{aligned}
& g(x^k) - g(x^k - \omega_k p^k) \\
& \geq \omega_k \cdot \nabla g(x^k) p^k - \frac{K}{2} \cdot \omega_k^2 \|p^k\|^2 \\
& = \omega_k \left\{ (1-\delta)\nabla g(x^k) p^k - \frac{K\omega_k \|p^k\|^2}{2} + \delta \nabla g(x^k) \cdot p^k \right\} \\
& = \omega_k \left\{ (1-\delta)\|p^k\|^2 \left[\frac{\nabla g(x^k) p^k}{\|p^k\|^2} - \frac{K\omega_k}{2(1-\delta)} \right] + \delta \nabla g(x^k) p^k \right\}.
\end{aligned} \tag{3.9}$$

Now by (3.5), $\exists \gamma > 0$ such that for all $h \in \Re^n$,

$$(\gamma_k + \lambda_k)\|h\|^2 \leq h^T(A_k + \lambda_k I)h \leq (\gamma + \lambda_k)\|h\|^2, \tag{3.10a}$$

$$\frac{1}{\gamma + \lambda_k}\|h\|^2 \leq h^T(A_k + \lambda_k I)^{-1}h \leq \frac{1}{\gamma_k + \lambda_k}\|h\|^2. \tag{3.10b}$$

Take $h = p^k$ in (3.10a) to get

$$\frac{\nabla g(x^k)p^k}{\|p^k\|^2} \geq (\lambda_k + \gamma_k).$$

Now take $h = \nabla g(x^k)$ in (3.10b) to get

$$\nabla g(x^k)p^k \geq \frac{\|\nabla g(x^k)\|^2}{(\lambda_k + \gamma)}.$$

Using (3.8) we now have that the square bracket in (3.9) is nonnegative and hence from (3.9),

$$\begin{aligned} g(x^k) - g(x^k - \omega_k p^k) &> \delta\omega_k \nabla g(x^k)p^k \\ &\geq \frac{\delta\omega_k\|\nabla g(x^k)\|^2}{(\gamma + \lambda_k)}. \end{aligned} \tag{3.11}$$

Hence (3.8) implies that $g(x^k) > g(x^k - \omega_k p^k)$. Since ω_k is the largest $\omega \in \Omega$ chosen to satisfy $g(x^k) > g(x^k - \omega p^k)$, it follows that $2\omega_k$ violates (3.8) so that (3.7) holds. This proves our claim.

Assume now that x^* is a limit point of $\{x^k\}$ and that $x^{k_j} \to x^*$. From (3.11) we have

$$\begin{aligned} g(x^{k_j}) - g(x^{k_{j+1}}) &\geq g(x^{k_j}) - g(x^{k_j+1}) \\ &> \frac{\delta\omega_{k_j}\|\nabla g(x^{k_j})\|^2}{(\lambda_{k_j} + \gamma)} \\ &\longrightarrow 0. \end{aligned} \tag{3.12}$$

If $\liminf \omega_{k_j} = 0$ then by (3.7),

$$0 = \liminf \lambda_{k_j} = \liminf g(x^{k_j}) = \lim g(x^{k_j}) = g(x^*),$$

that is $G(x^*) = 0$ and hence $\nabla g(x^*) = 0$. If, however, $\liminf \omega_{k_j} = w^* > 0$, then from (3.12) and the continuity of $\nabla g(x)$, $\|g(x^*)\| = 0$, that is $\nabla g(x^*) = 0$. This completes our proof. ▮

4. Local superlinear convergence

In this section we shall investigate situations under which conditions (3.5) and (3.6) of Theorem (3.4) can be realized. We shall then prove that under appropriate conditions, Algorithm (3.3) exhibits local superlinear convergence. We begin with the notion of a *locally Lipschitzian operator*.

4.1 Definition. *An operator* $F\colon \Re^n \longrightarrow \Re^n$ *is said to be locally Lipschitzian at a point* z *if* $\exists K > 0$ *and a neighborhood* $N = N(z)$ *of* z *such that* F *is Lipschitzian on* N *that is*

$$\|F(x) - F(y)\| \leq K\|x - y\| \qquad \forall x,\ y \in N$$

and we write $F \in L(N)$.

4.2 Lemma. *Let* $F\colon \Re^n \longrightarrow \Re^n$ *and consider* $NLCP(F)$. *Let* $G\colon \Re^n \longrightarrow \Re^n$ *be the associated M-function,* $g(x) = \|G(x)\|^2/2$. *Suppose further that* $\nabla F_i(z)$, $i = 1, 2, \cdots n$ *is locally Lipschitzian at some point* $\bar{z}$. *Then* $\exists \delta > 0$ *such that* ∇g *is Lipschitzian on* S, *where*

$$S = S(\bar{z}, \delta) = \{x : \|x - \bar{z}\| \leq \delta\}.$$

Proof

Recall that for $i = 1, 2, \cdots n$

$$G_i(z) = \{F_i(z) - z_i\}^2 - F_i(z)|F_i(z)| - z_i|z_i| \tag{4.3}$$

so that

$$\nabla G_i(z) = 2\big(F_i(z) - z_i\big)\big(\nabla F_i(z) - e_i\big) \\ - 2|F_i(z)|\nabla F_i(z) - 2|z_i|e_i. \tag{4.4}$$

Since $\nabla F_i(z) \in L(S)$, it is continuous and so are $\nabla F(z)$, $\nabla G_i(z)$ and $\nabla G(z)$. Hence all these operators are bounded on S. This implies that $F_i(z)$, $G_i(z)$ and $G(z) \in L(S)$.

Now if $A, B : D \subseteq \Re^n \rightarrow \Re^m$ and $A, B \in L(D)$, it is trivial that $A \pm B \in L(D)$. However, if A and B are also bounded on D, it follows from the identity

$$\|A(x)B(x) - A(y)B(y)\|$$
$$= \frac{1}{2}\|\big(A(x) + A(y)\big)\big(B(x) - B(y)\big) + \big(B(x) + B(y)\big)\big(A(x) - A(y)\big)\|$$

that $AB \in L(S)$. From (3.4) we now deduce that $\nabla G_i(z) \in L(S)$ and hence that $\nabla G(z) \in L(S)$. Similar considerations show that $\nabla g(z) = \nabla G(z)^T G(z) \in L(S)$. ∎

4.5 Theorem. *Let* $F\colon \Re^n \longrightarrow \Re^n$ *and consider* $NLCP(F)$. *Let* $G\colon \Re^n \longrightarrow \Re^n$ *be the associated M-function. Let* z^* *solve* $NLCP(F)$, $\nabla F_i(z)$ *be locally Lipschitzian at* z^*, $1 \leq i \leq n$. *Assume that the hypotheses of Corol-*

lary (2.2) are satisfied. Then there exists a neighborhood $N(z^*)$ such that if $x^0 \in N(z^*)$ then the sequence $\{x^k\}$ defined by Algorithm (3.3) satisfies :

(1) $x^k \in N(z^*)$, $x^k \rightarrow z^*$,

(2) *Conditions (3.5) and (3.6) of Theorem 3.4 hold and*

(3) *if K is the constant given by (3.6), γ^* the smallest eigenvalue of $\nabla G(z^*)^T \nabla G(z^*)$ and $2\gamma^* > K$ then $x^k \rightarrow z^*$ superlinearly.*

Proof

By Corollary 2.2, $\nabla G(z^*)$ is nonsingular and by Theorem 2.1, $G(z^*) = 0$. Let $\mu_1, \mu_2, \cdots \mu_n$ be the eigenvalues of $\nabla G(z^*)$, $\Sigma = \{\mu_i : Re(\mu_i) < 0\}$ where $Re(\mu)$ denotes the real part of μ. Let

$$\eta = \begin{cases} \min\{-|\mu_i|^2/Re(\mu_i) \ : \ \mu_i \in \Sigma\} & \\ +\infty, & \text{if } \Sigma = \emptyset \ . \end{cases}$$

We can find $\delta > 0$ such that for all $z \in S = S(z^*, \delta) = \{x : \|x - z^*\| \leq \delta\}$, we have $g(z) = \|G(z)\|^2/2 < \eta$ and that $\nabla g(z)$ is Lipschitzian with modulus K on S (Lemma 4.2).

Let x^k be in S. In Algorithm 3.3, $0 < \omega_k \leq 1$ by choice and since $\lambda_k = g(x^k) < \eta$, it follows that

$$0 < \omega \leq 1, \qquad 0 \leq \frac{\lambda_k}{2 - \omega_k} < \eta.$$

It is a consequence of [Ortega and Rheinboldt, 10.2.3] that z^* is a point of attraction of $\{x^k\}$ and that $x^k \rightarrow z^*$. By Lemma 4.2 and our choice of S, it is clear that (3.5) and (3.6) hold.

Assume now that $\gamma^* > K/2$. Recall that ω_k is the largest element in Ω such that $g(x^k - \omega_k p^k) < g(x^k)$. We saw in the proof of Theorem 3.4 that

$$\omega < \frac{2(\lambda_k + \gamma_k)}{K} \Longrightarrow g(x^k - \omega p^k) < g(x^k).$$

Since $\lambda_k \to 0$, it follows that $2(\lambda_k + \gamma_k)/K \to 2\gamma^* > 1$ and hence that $\omega_k \to 1$. Thus, $\lambda_k \to 0$, $\omega_k \to 1$ and by [Ortega and Rheinboldt, p124] that $x^k \to z^*$ superlinearly. ▮

5. Local quadratic convergence

Let x^0 be given. If we assume that $\nabla G(z)$ satisfies hypothesis stronger than (3.6), we can modify Algorithm 3.3 so that the resulting iterates converge locally quadratically.

5.1 Modified Gauss-Newton Algorithm

Consider the following algorithm in which the *perturbation parameters* λ_k are chosen slightly differently:

(1) Let x^0 be given. Having x^k, define x^{k+1} as follows:

(2) If $\nabla g(x^k) = 0$, stop.

(3) If $\nabla g(x^k) \neq 0$, let

$$A_k = \nabla G(x^k)^T \nabla G(x^k),$$

$$\gamma_k = \text{ smallest eigenvalue of } A_k.$$

(4) Define the λ_k by

$$\lambda_k = \begin{cases} 0, & \text{if } \gamma_k > 0; \\ g(x^k), & \text{otherwise}. \end{cases}$$

(5) Set $p^k = (A_k + \lambda_k I)^{-1} \nabla g(x^k)$.

(6) Let ω_k be the largest element in $\Omega = \{1, 1/2, \cdots\}$ such that

$g(x^k - \omega_k p^k) < g(x^k)$.

(7) Set $x^{k+1} = x^k - \omega_k p^k$.

5.2 Theorem. *Let $F: \Re^n \longrightarrow \Re^n$ and assume that x^0 is given. Let $\{x^k\}$ be the sequence of iterates of Algorithm 5.1. Assume that :*

(5.3) $0 < \liminf_k \|\nabla G(x^k)\| \leq \sup_k \|\nabla G(x^k)\| < \infty$ *and*

(5.4) $\nabla g(x)$ *is Lipschitzian with modulus K on $\Re^n$.*

Then either $\{x^k\}$ terminates at a stationary point of g, or else every limit point of $\{x^k\}$, if there exists one is a stationary point of g.

Proof

The proof is similar to that of Theorem 3.4 and so we give only an outline. Let $0 < \delta < 1$.

Let γ_k be the smallest eigenvalue of A_k. By (5.3), there exists $\gamma > 0$ such that

$$\gamma_k \|h\|^2 \leq h^T A_k h \leq \gamma \|h\|^2$$

for all $h \in \Re^n$. As in Theorm 3.4 it is easy to show that

$$\begin{aligned}
\omega_k &\leq \frac{2(\lambda_k + \gamma_k)(1-\delta)}{K} \Rightarrow g(x^k - \omega_k p^k) < g(x^k), \\
\omega_k &> \frac{(\lambda_k + \gamma_k)(1-\delta)}{K}, \\
g(x^k) - g(x^k - \omega_k p^k) &> \frac{\delta \omega_k \|\nabla g(x^k)\|^2}{(\lambda_k + \gamma)} \geq 0.
\end{aligned}$$

If $x^\star$ is a limit point of $\{x^k\}$ with $x^{k_j} \to x^\star$, one then shows that

$$\omega_{k_j} \cdot \|\nabla g(x^{k_j})\| \to 0.$$

But

$$\liminf_j \omega_{k_j} \; > \; \frac{(1-\delta)}{K} \liminf_j \gamma_{k_j} \; > \; 0$$

by (5.3) so that

$$0 = \lim \|\nabla g(x^{k_j}\| = \|\nabla g(x^\star)\|. \blacksquare$$

(Note that unlike Theorem 3.4, $\lambda_k = 0$ for all k is possible so that $\lambda_k \to 0 \not\Rightarrow \nabla g(x^\star) = 0$. This is a consequence of our insistence that for points sufficiently close to a nonsingular point, $A_k + \lambda_k I$ must in fact be identical to A_k, and hence $\lambda_k = 0$.)

Corresponding to Theorem 4.5 we have the following result.

5.5 Theorem. *Let F: $\Re^n \longrightarrow \Re^n$, G: $\Re^n \longrightarrow \Re^n$ the corresponding M-function. Assume that $z^\star$ solves $NLCP(F)$ and that the conditions of Corollary 2.2 hold. Assume further that $\nabla F_i(z)$, $1 \le i \le n$ is locally Lipschitzian at $z^\star$. Then there exists a neighborhood $N(z^\star)$ of $z^\star$ such that if $x^0 \in N(z^\star)$ then $\{x^k\}$ defined by Algorithm 5.1 satisfies :*

(1) *$x^k \in N(z^\star)$, $x^k \to z^\star$ superlinearly,*

(2) *Conditions (5.3) and (5.4) of Theorem 5.4 hold and*

(3) *if K is the constant given by (5.4), $\gamma^\star$ the smallest eigenvalue of $\nabla G(z^\star)^T \nabla G(z^\star)$ and $\gamma^\star > K/2$, then $x^k \to z^\star$ quadratically.*

Proof

By Corollary 2.2, $\nabla G(z^\star)$ is nonsingular and $G(z^\star) = 0$. Using Lemma 4.2 we can find $\delta > 0$ such that for all z in $S := S(z^\star, \delta) = \{x : \|x - z^\star\} < \delta$, $\nabla G(z)$ is nonsingular and $\sup \|\nabla G(z)\|$ is finite.

Let $x^k \in S$, $k \geq 0$. Then A_k is positive definite so that in Algorithm 5.1, $\gamma_k > 0$ and $\lambda_k = 0$. Hence $0 < \omega_k \leq 1$, $\lambda_k \equiv 0$ for all k. By [Ortega and Rheinboldt, 1970, 10.2.3], $z^\star$ is a point of attraction for $\{x^k\}$. By [Ortega and Rheinboldt, 1970, p124], $x^k \to z^\star$ superlinearly. It is easy to see that (5.3) and (5.4) hold.

Suppose now that $\gamma^\star > K/2$. Notice that since $\lambda_k = 0$, we have

$$
\begin{aligned}
x^{k+1} &= x^k - \omega_k A_k^{-1} \nabla g(x^k) \\
&= x^k - \omega_k \nabla G(x^k)^{-1} \left[\nabla G(x^k)^T\right]^{-1} \nabla G(x^k)^T G(x^k) \\
&= x^k - \omega_k \nabla G(x^k)^{-1} G(x^k).
\end{aligned}
$$

One now shows exactly as in the proof of Theorm 4.5 that $2(\lambda_k + \gamma_k)/K \to 2\gamma^\star/K > 1$. Hence by our choice, for large k, $\omega_k = 1$. Hence Algorithm 5.1 is simply the Newton process and since $\nabla G(z)$ is Lipschitzian in S (Lemma 4.2), it follws that $x^k \to z^\star$ quadratically. ∎

5.6 Remarks

1. In Algorithm 5.1, one can also take $\lambda_k = \lambda > 0$ whenever $\gamma_k = 0$ and computationally this may be preferable.
2. The robustness of the algorithm depends strongly on $\gamma^\star$. If $\gamma^\star$ is very small, then it is possible that $\omega_k \to 0$ and the algorithm fails.

3. Although the conditions of Corollary are only *sufficient* conditions, the algorithms developed in this chapter failed to solve problems in which *none* of those conditions were fulfilled indicating, perhaps, the sharpness of those conditions. Nevertheless, many problems in which the nondegeneracy condition failed were successfully solved. We shall present a fuller report in the next chapter.

CHAPTER 6

COMPUTATIONAL RESULTS

1. Introduction

In this chapter we shall report on our computational experience with the various algorithms descirbed in this thesis. The reader may recall that in Chapter 3 the solution of $LCP(M, q)$, when M is positive semidefinite, was reduced to the solution of a sequence $LCP(M + \varepsilon_n I, q)$ of LCPs for $\varepsilon_n \downarrow 0$. We suggested two SOR algorithms for the solution of these subproblems, Cryer's algorithm [Cryer, 1971] and a modification of a more general algorithm due to Mangasarian [1977]. For future reference we shall refer to these as CSOR and MSOR respectively.

In Chapter 4 we considered fixed point methods and two algorithms to solve $LCP(M, q)$, M positive semidefinite. A *two-step* algorithm was given in Theorem 4.2.4 and a *one-step* algorithm in Theorem 4.2.8. Given a sequence of $\varepsilon_n \downarrow 0$, in the two-step method one iterates until a point close to the fixed point of the map $x \mapsto \big(x - \alpha_n(Mx + q + \varepsilon_n x)\big)_+$ is obtained and then a smaller ε_n is chosen. The step sizes α_n are defined in terms of ε_n . In the one-step method both ε_n and α_n are changed at each iteration. These two methods will be referred to as FP2 and FP1 respectively.

Finally in Chapter 5 we considered a damped Gauss-Newton method. This algorithm is applicable to general nonlinear complementarity problems satisfying the requirements of Theorem 5.4.5 or Theorem 5.5.2. We shall call this method as DGN.

We shall be comparing the performance of our algorithms against the algorithm due to Lemke [1968] currently used widely. For brevity, this algorithm will be simply called LEMKE.

2. Performance of algorithms

The two relaxation algorithms CSOR and MSOR are both robust and capable of handling large scale problems. Of these two, MSOR is decidedly significantly faster and is applicable to a wide class of symmetric $LCP(M, q)$. However, both methods tend to slow down appreciably near solution points.

The fixed point methods FP2 and FP1 are both excruciatingly slow, perhaps because they do not exploit any special features of the matrix M. This is also true of the nonlinear version of FP1 (Theorem 4.3.2). Although relatively fewer iterates are needed to satisfy a weak termination tolerance (e.g.,10^{-2}), many hundreds of iterates are often needed to satisfy a stringent tolerance (e.g., 10^{-8}).

For the class of problems for which it is applicable, the damped Gauss-Newton algorithm DGN is quite robust. It was used successfully to solve Colville test Problems 1 and 2 [Colville, 1968]. Problem 1 was solved in 6 iterations (2.1 seconds). Problem 2 considered to be more difficult was solved in 13 iterations (5.1 seconds). Using Iterative Quadratic Programming (IQP) [Garcia-Palamores and Mangasarian, 1976], we solved these problems (using LEMKE to solve the quadratic subproblems) in 0.5 seconds and 4.5 seconds respectively. DGN compares favorably in comparison with LEMKE for positive definite LCPs. As we shall see later, these methods are faster than the SOR methods for medium size problems ($n \approx 35$).

3. Test Problems

All the algorithms were tested on some interesting small LCPs such as the following example due to Kostreva [1979] :

$$M = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix} \qquad q = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

The matrix M is positive semidefinite and positive definite on $\Re^n_+$. The unique solution is

$$z^* = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)^T, \qquad w^* = M\,z^* + q = (0,\ 0,\ 0)^T$$

It is known that LEMKE cycles. In fact, for $0 \le \varepsilon \le .5$, LEMKE cycles for the problem $LCP(M + \varepsilon I,\ q)$.

All the other algorithms solved the above $LCP(M,\ q)$ successfully starting from the origin and the results are given in Table 3.1.

TABLE 3.1 : KOSTREVA'S EXAMPLE

Algorithms	Avg. No. iterations	CPU time (seconds)
CSOR	6	0.008
MSOR	4	0.003
FP2	82	0.090
FP1	141	0.180
DGN	4	0.010
LEMKE	NA	NA

Next the algorithms were tested on several randomly generated test problems. Several problems of sizes n = 5, 10, 15, 25, 30, 40 respectively were generated using MATLAB, an interactive matrix language program which in turn uses LINPACK, a state of the art linear algebra programs. The problems were designed as follows :

1. Generate the matrix M with each entry being a random number uniformly distributed in the interval $[-1, 1]$. Sparsity is built in by placing nonzero elements in random positions in the matrix.
2. Find the eigenvalues of $M + M^T$, perturb diagonals of M to make M positive semidefinite.

3. For each index i, randomly choose one of x_i and w_i to be zero and the other to be a random floating number in [0, 1000].

4. Define $q = w - Mx$.

All programs were written in Fortran 77 and tested on Vax 11-780 computer under the virtual UNIX (Berkeley 4.2) operating system. All floating point computations are in double precision providing about 16 figure decimal accuracy.

Each problem was solved three times (when a given method succeeded) and an average of the CPU times taken. For each problem size, we attempted to solve 20 test problems. The tables to follow indicate the percentage of the number of problems solved, the average number of iterations taken by each method tested and the average CPU time in seconds. The fixed point methods FP2 and FP1 were tested only problem size n=5 where they took hundreds of iterations just to reach a tolerance of 10^{-2}. It was therefore decided not to attempt to solve larger size problems with these methods.

We have already mentioned that the SOR methods are initially quite fast but slow down near solution points. On the other hand, DGN works best in the vicinity of a solution point. Thus it is of interest to try a combination of the two methods, using SOR to generate a good starting point for DGN. Such algorithms are called *polyalgorithms* in the literature. We experimented with one such algorithm, using MSOR until the iterates satisfy a tolerance of 10^{-3} and using the output as the starting point for DGN. We refer to this combination as POLY-MSOR. We observe that unlike DGN, performance of ordinary LEMKE is not influenced by the starting point being near a solution point; thus polyalgorithms using LEMKE to complete the solution need *not* be effective.

Our test results seem to suggest that LEMKE is particularly robust for medium size problems, $25 \leq n \leq 35$ and also faster than SOR algorithms in each cases. However, for $n < 15$ or for $n \geq 40$, the SOR algorithms are superior to LEMKE.

It must be noted that for positive semidefinite nonsymmetric matrices M, $LCP(M,\, q)$ was solved by CSOR and MSOR using Tihonov regularization . Thus these methods essentially solve problems twice the size of M. The Tihonov parameter ε_n was chosen as 10^{-n}, the output $x(\varepsilon_n) = x_n$ at iteration n being the input for iteration $n+1$. The iterations were stopped as soon as

$$|x_n(Mx_n + q)| < \delta, \qquad \|(Mx_n + q) - (Mx_n + q)_+\| < \delta$$

for some tolerance δ. The tables to follow give the *total* number of iterations to reach optimality, that is, # iterations $= \sum N(n)$ where $N(n) =$ number of iterations to solve $LCP(M + \varepsilon_n I,\, q)$.

TABLE 3.2 : PROBLEM SIZE $n = 5$
NO. of test problems = 20

Algorithms	Percent. solved	Avg. No. iterations	CPU time (seconds)
CSOR	100	400	0.2
MSOR	100	300	0.1
FP2	60	1800	1.2
FP1	53	2400	1.8
DGN	96	15	0.7
LEMKE	100	8	0.3
POLY-MSOR	96	295	0.7

3.3 Remarks

1. CSOR and MSOR have the disadvantage of having to solve a problem of size $2n$ when M is not symmetric.

2. The problems in which DGN failed were those in which neither of the conditions of Corollary 5.2.2 were fulfilled.

3. The SOR methods were usually faster. Notice that POLY-MSOR was no more efficient than DGN. On the average 290 SOR iterations and 5 DGN iterates were needed to reach optimality.

TABLE 3.4 : PROBLEM SIZE $n = 15$
NO. of test problems = 20

Algorithms	Percent. solved	Avg. No. iterations	CPU time (seconds)
CSOR	100	1500	0.8
MSOR	100	1000	0.6
DGN	93	25	2.0
LEMKE	100	12	0.4
POLY-MSOR	93	900	1.3

3.5 Remarks

1. LEMKE was faster than all the other algorithms. The maximum number of iterations was 18.

2. DGN performed as fast as LEMKE when M was positive definite.

3. Notice that POLY-MSOR is faster than DGN. In most cases MSOR reached a tolerance of 10^{-3} in 890 iterations. The output at this stage was used as starting point for DGN and optimality reached in an additional 10 DGN iterates.

TABLE 3.6 : PROBLEM SIZE $n = 25$
NO. of test problems = 20

Algorithms	Percent. solved	Avg. No. iterations	CPU time (seconds)
CSOR	95	2600	1.35
MSOR	97	2000	0.85
DGN	60	40	3.00
LEMKE	95	25	0.60
POLY-MSOR	60	1500	1.50

3.7 Remarks

1. DGN failed to solve any problem in which both conditions of Corollary 5.2.2 failed. However it solved problems in which only one of the conditions was violated albeit being slow in such cases.

2. LEMKE was still faster than SOR methods. Our experience confirms robustness of LEMKE for positive semidefinite M. It is known, however, that for $n \approx 25$ and general M, LEMKE performs poorly [Shiau, 1983].

3. Problems solvable by DGN were also solved by MSOR and POLY-MSOR.

TABLE 3.8 : PROBLEM SIZE $n = 40$
NO. of test problems = 20

Algorithms	Percent. solved	Avg. No. iterations	CPU time (seconds)
CSOR	90	5000	2.5
MSOR	92	4200	1.5
DGN	50	70	7.0
LEMKE	95	48	6.3
POLY-MSOR	54	3000	1.0

3.9 Remarks

1. For problems with size $n > 30$, LEMKE was much slower than SOR methods. In some it took as many as 10.5 seconds while these were solved by MSOR in 1 second.

2. DGN was the slowest and failed when the nonsingular principal minor condition failed. Otherwise it was usually comparable to LEMKE.

3. POLY-MSOR was the fastest for problems solvable by DGN. It also solved two which failed with DGN. Typically 90% of the iterations were needed to reach an accuracy of 10^{-4} and the rest used in DGN.

TABLE 3.10 : PROBLEM SIZE $n = 40$, M SYMMETRIC
NO. of test problems = 20

Algorithms	Percent. solved	Avg. No. iterations	CPU time (seconds)
MSOR	98	350	0.05
LEMKE	90	50	6.50

3.11 Remarks

1. This table gives computational results for a class of $LCP(M, q)$, M randomly generated but positive semidefinite and *symmetric*. *For these problems, Tihonov regularization was not used and MSOR used directly.* Our computational results attest to the efficiency of MSOR vis-a-vis LEMKE for large sparse problems. MSOR was remarkably fast and 60% of the problems were solved within 100 iterations often in .01 sec.

2. In many problems LEMKE was terminated when the maximum number of iterations (200) or maximum time (10 seconds) were exceeded.

4. Suggestions for further research

In this section we list several directions in which the results of this thesis may be extended.

1. In Chapter 2, we considered least two-norm solutions of feasible positive semidefinite $LCP(M, q)$. We showed that for each $\varepsilon > 0$, the solution $x(\varepsilon)$ of $LCP(M+\varepsilon I, q)$ satisfies $\|x(\varepsilon)\| < \|x^*\|$ where x^* is the least two-norm solution thus providing a *lower* bound for $\| x^*\|$. It would be of interest to find an *upper* bound for $\|x^*\|$. Such bounds are known when $int\, S(M, q) \neq \emptyset$ (see e.g., [Mangasarian, 1982].
2. In Chapter 3, dual exact penalty was used to transform positive semidefinite $LCP(M, q)$ into a sequence of positive definite symmetric quadratic subprograms to be solved by SOR procedures. However, this introduces two difficulties, viz., (a) the subproblems are of dimension $2n$ and (b) the presence of the matrix $(M + \varepsilon I, q)^T(M + \varepsilon I)$ in the subproblems possibly destroying sparsity structures of M unless special procedures are implemented. The real advantage of the dual exact penalty formulation is that it provides a strictly convex descent function. Iterative procedures using perhaps the *gap function* $\varphi(x) = x^T(Mx + q)$ would be very useful.
3. The fixed point methods of Chapter 4, while mathematically pleasing are not computationally effective. It is conjectured that the rate of convergence is sublinear for FP1 while this is obvious for FP2 since

the contraction constants for the latter tend to 1. We believe that if the strong features of SOR procedures viz., (a) utilize relaxation parameters ω , (b) use of the lower triangular and diagonal matrices etc., are used in the fixed point methods they would prove faster. One would have to carefully modify the convergence proofs in this case.

4. The DGN algorithm of Chapter 5 seems to work efficiently in the presence of nondegeneracy of the solution point although this information may not be available. Although convergence cannot be guaranteed, problems in which the nondegeneracy condition fails have been solved. In using DGN one has solve the system of linear equations $(A_k + \lambda_k I)p^k = \nabla g(x^k)$ (see Algorithm 5.3.3) to find the direction p^k. Since MSOR is particularly efficient in solving such equations when the underlying matrix is positive definite and symmetric, it is expected that considerable time would be saved if these equations are solved iteratively.

5. Since $NLCP(F)$ is equivalent to the system of equations $G(z) = 0$, it would be of interest to solve this system as a sequence of linear equations by linearizing $G(z)$ at z^k and solving

$$\mathcal{L}_k(G)z = G(z^k) + (z - z^k)\nabla G(z^k) = 0$$

to find z^{k+1}. The resulting subproblems can be solved conveniently by using SOR procedures which are robust and fast. It may not be necessary to solve each subproblem fully and it is conjectured that in the vicinity of a solution when the conditions of Theorem 5.5.2 are satisfied, quadratic convergence would occur.

BIBLIOGRAPHY

Aganagic, M. (1981). On diagonal dominance in linear complementarity, *Linear Algebra and its Applications* **39**, pp 41–49.

Agmon, S. (1954). The relaxation method for linear inequualities, *Canadian J. Math.* **6**, pp 382–392.

Ahn, B. -H. (1981). Solution of nonsymmetric linear complementarity problems by iterative methods, *J. Opt. Th. Applics.* **33**, pp 175–185.

Armijo, L. (1966). Minimization of functions having Lipschitz continuous first partial derivatives, *Pacific J. Math.* **16**, pp 1–3.

Auslender, A. (1976). *Optimisation Méthods Numériques*, Masson, Paris.

Baillon, J. -B. (1975). Un théorème de type ergodique pour les contractions non linéaires dans un espace de Hilbert, *Comp. Rend. Acad. Sci. Paris* **280**, pp 1511–1514.

Berge, C. (1963). *Topological Spaces*, McMillan, New York.

Bermon, A. and Plemmons, P. (1979). *Nonnegative matrices in the mathematical sciences*, Academic Press, New York.

Bertsekas, D. P. (1975). Necessary and sufficient conditions for a penalty method to be exact, *Mathematical Programming* **9**, pp 87–99.

van Bokhoven, W. M. G. (1980). Macromodelling and simulation of mixed analog-digital networks by a piecewise linear system approach, *IEEE 1980 Circuits and Computers*, pp 361–365.

Brégman, L. M. (1965). The Method of Successive Projection for finding a common point of convex sets, *Soviet Mathematics Doklady* **6**, pp 688–692.

Brézis, H. (1973). *Operateurs maximux monotones*, North-Holland Publishing Co., Amsterdam.

Cheng. J. C. (1975). *Analysis of a quantum price model in commodity future markets and a fair salary administration system*, Ph.D. Thesis, Department of Mathematics, MIT, Cambridge, MA.

Cheng, Y. C. (1981). *Iterative methods for solving Linear Complementarity and Linear Programming problems*, Ph.D. Thesis, Department of Computer Sciences, University of Wisconsin-Madison.

Chung, S.J. and Murty, K. G. (1981). "Polynomially bounded Ellipsoid Algorithms for Convex Quadratic Programming", in *Nonlinear Programming 4* (O. L. Mangasarian, S. M. Robinson and R. R. Meyer, eds.), pp 439–485. Academic Press, London and New York.

Colville, A. R. (1968). A comparative study on nonlinear programming codes, IBM New York Scientific Report 320–2949.

Cottle, R. W. (1977). "Numerical methods for complementarity problems in Engineering and Applied science", in *Coimputing methods in Applied Sciences and Engineering 704* (R. Glowinski and J. -L. Lions, eds.), pp 37–52. Springer-Verlag, Berlin-Heidelberg-New York.

Cottle, R. W. (1980). "Some recent developments in linear complementarity theory", in *Variational Inequalities and complementarity problems* (R. W. Cottle, F. Gianessi and J. -L. Lions, eds.), John Wiley, New York.

Cottle R. W. and Dantzig, G. (1968). Complementary Pivot theory of Mahtematical Programming, *Linear Algebra and its Applications* **1**, pp 103–125.

Cottle, R. W., Golub, G. H. and Sacher, R. S. (1978). On the solution of large structured complementarity problems, *Applied Mathematics and Optimization* **4**, pp 347–363.

Cryer, C. W. (1969). The method of Christopherson for solving free boundary problems for infinite journal bearings by means of finite differences, Tech. Report 72, Computer Sciecnes Department, University of Wisconsin–Madison.

Cryer, C. W. (1971). The solution of a quadratic programming problem using systematic overrelaxation, *SIAM J. Control* **9**, pp 385–392.

Dantzig,G. B. (1963). *Linear Programming and extensions*, Princeton University Press, New Jersey.

Dantzig, G. B. and Cottle, R. W. (1967). "Positive (semi) definite programming", in *Nonlinear Programming* (J. Abadie, ed.), pp 55–73, North-Holland Publications, Amsterdam.

Eaves, B. C. (1971). The linear complementarity problem, *Management Science* **17**, pp 612–634.

Eaves, B. C. (1971). On quadratic programming, *Management Science* **17**, pp 698–711.

Eaves, B. C. (1972). Homotopies for computation of fixed points, *Mathematical Programming* **3**, pp 1–22.

Eaves, B. C. (1978). Computing stationary points, *Mathematical Programming Study* **7**, pp 1-14.

van Eijndhoven, J. T. J. The solving of the Linear Complementarity problem in circuit simulation, *preprint*.

Fiedler, M. and Pták, V. (1962). On matrices with non-positive off-diagonal elements and positive principal minors, *Czechoslovak Math. J.* **12**, pp 382–400.

Frank, M. and Wolfe, P. (1956). An algorithm for quadratic programming, *Naval Res. Logistics Quart.* **3**, pp 95–110.

Garcia, G. B. (1977). A note on the Complementarity Problem, *J. Opt. Th. Applics.* **21**, pp 529–530.

Garcia-Palomares, U. M. and Mangasarian, O. L. (1976). Superlinear convergent quasi-Newton algorithms for nonlinearly constrained optimization problems, *Mathematical Programming* **11**, pp 1–13.

Han, S. -P. and Mangasarian, O. L. (1983). A dual differentiable exact penalty function, *Mathematical Programming Study* **25**, pp 293–306.

Hogan, W. W. (1975). Energy Policy Models for Project Independence, *Comput. and Ops. Res.* **2**, pp 251–271.

Josephy, N. H. (1979). *Newton's method for generalized equations and the PIES energy model*, Ph.D. Thesis, Department of Industrial Engineering, University of Wisconsin-Madison.

Karamardian, S. (1972). The Complementarity Problem, *Mathematical Programming* **2**, pp 107–129.

Keller, H. B. (1965). On the solution of singular and semidefinite linear systems by iteration, *SIAM J. Num. Anal.* **2**, pp 281–290.

Kostreva, M. M. (1979). Cycling in linear complementarity problems, *Mathematical Programming* **16**, pp127–130

Lemke, C. E. (1965). Bimatrix equilibrium points and mathematical programming, *Management Science* **11**, pp 681–689.

Lemke, C. E. (1968). "On Complementarity pivot theory", in *Mathematics of the Decision Sciences* (G. B. Dantzig and A. F. Veinott, Jr., eds.), American Mathematical Society, Providence, Rhode Island.

Lemke, C. E. (1980). "A survey of complementarity theory", in *Variational inequalities and complementarity problems* (R. W. Cottle, F. Gianessi and J. -L. Lions eds.), John Wiley & Sons, New York.

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

Mangasarian, O. L. (1976). Equivalence of the complementarity problem to a system of nonlinear equations, *SIAM J. Applied Math.* **31**, pp 89–92.

Mangasarian, O. L. (1976). Linear complementarity problems solvable by a single linear program, *Mathematical Programming* **10**, pp 263–270.

Mangasarian, O. L. (1977). Solution of symmetric linear complementarity problems by iterative methods, *J. Opt. Th. Applics.* **22**, pp 465–485.

Mangasarian, O. L. (1978). Characterizations of linear complementariy problems solvable as linear programs, *Mathematical Programming Study* **7**, pp 74–87.

Mangasarian, O. L. (1979). Simplified characterizations of linear complementarity problems solvable as linear programs, *Mathematics of Operations Research* **4**, pp 268–273.

Mangasarian, O. L.. (1980). Locally unique solutions of quadratic programs, linear and nonlinear complementarity problems, *Mathematical Programming* **19**, pp 200–212.

Mangasarian, O. L. (1982). Characterizations of bounded solutions of linear complementarity problems, *Mathematical Programming Study* **19**, pp 153–166.

Mangasarian, O. L. (1984). Normal Solutions of linear programs, *Mathematical Programming Study* **22**, pp 206–216.

Mangasarian, O. L. (1985). Simple computable bounds for solutions of linear complementarity problems and linear programs, *Mathematical Programming Study* (to appear)

Mangasarian, O. L. and McLinden, L. (1985). Simple bounds for solutions of monotone complementarity problems and convex programs, *Mathematical Programming* **32**, pp 32–40.

Megiddo, N. (1977). A monotone complementarity problem with feasible solutions but no complementarity solutions, *Mathematical Programming* **12**, pp 131–132.

Moré, J. J. (1974). Classes of functions and feasibility conditions in nonlinear complementarity problems, *Mathematical Programming* **6**, pp 32–338.

Murty, K. G. (1972). On the number of solutions to the complementarity problems and spanning properties of complementarity cones, *Linear Algebra and its Applications* **5**, pp 65–108.

Ortega, J. M. (1972). *Numerical Analysis, a second course*, Academic Press, New York.

Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*, Academic Press, New York.

Pang, J. -S. (1979). A new characterization of real H-matrices with positive diagonals, *Linear Algebra and its Applications* **25**, pp 162–167.

Pang, J. -S. (1981). "The Implicit Complementarity Problem", in *Nonlinear Programming 4* (O. L. Mangasarian, R. R. Meyer and S. M. Robinson, eds.), Academic Press, New York.

Pang, J. -S. (1982). On the convergence of a basic iterative method for the implicit complementarity problem, *J. Op. Th. Applics.* **17**, pp 149–162.

Peyerimhoff, A. (1969). *Lectures on Summability*, Lecture Notes in Mathematics, Springer-Verlag, Berlin-Heidelberg-New York.

Robinson, S. M. (1975). Stability Theory for Systems of Inequalities, Part I: Linear Systems, *SIAM J. Num. Anal.* **12**, pp 754–769.

Robinson, S. M. (1976a). An Implicit-Function Theorem for Generalized Variational Inequalities, Technical Summary Report No. 1672, Mathematics Research Center, University of Wisconsin-Madison.

Robinson, S. M. (1976b). Stability Theory for Systems of Inequalities, Part II: Differentiable Nonlinear Systems, *SIAM J. Num. Anal.* **13**, pp 497–513.

Robinson, S. M. (1976d). Regularity and Stability for Convex Multivalued Functions, *Mathematics of Operations Research* **1**, pp 130–143.

Robinson, S. M. (1979). Generalized equations and their solutions, Part I: Basic Theory, *Mathematical Programming Study* **10**, pp 128–141.

Robinson, S. M. (1980). Strongly regualr generalized equations, *Mathematics of Operations Research* **5**, pp 43–62.

Robinson, S. M. (1982). Generalized equations and their solutions, Part II: Applications to Nonlinear Programming, *Mathematical Programming Study* **19**, pp 200–221.

Robinson, S. M. (1982). "Generalized Equations" in Mathematical Programming: The State of the Art, Bonn 1982 (A. Bachem, M. Grötschel and B. Korte, eds.), Springer-Verlag, Berlin.

Scarf, H. (1967). The approximation of fixed points of a continuous mapping, *SIAM J. Applied Math.* **15**, pp 1328–1343.

Scarf, H. (1973). *The computation of Economic Equilibria*, Yale University Press, New Haven, Connecticut.

Shiau, T. -H. (1983). *Iterative linear programming for linear complementarity and related problems*, Ph.D. Thesis, Computer Sciences Department, University of Wisconsin-Madison.

Todd, M. (1976). On triangulations for computing fixed points, *Mathematical Programming* **10**, pp 322–346.

Varga, R. S. (1976). *Matrix Iterative Analysis*, Prentice Hall, Inc., Englewood Cliffs, New Jersey.

Zangwill, W. I. (1969). *Nonlinear Programming: A unified approach*, Prentice Hall, Inc., Engelwood Cliffs, New Jersey.