# CONVERGENCE RATE ANALYSES OF OPTIMIZATION ALGORITHMS FOR COMPUTER-AIDED DESIGN

by

Edward Joseph Wiest

Memorandum No. UCB/ERL M90/94

16 October 1990

COVER PAGE

# CONVERGENCE RATE ANALYSES OF OPTIMIZATION ALGORITHMS FOR COMPUTER-AIDED DESIGN

by

Edward Joseph Wiest

Memorandum No. UCB/ERL M90/94

16 October 1990

**ELECTRONICS RESEARCH LABORATORY**

College of Engineering
University of California, Berkeley
94720

# CONVERGENCE RATE ANALYSES OF OPTIMIZATION ALGORITHMS FOR COMPUTER-AIDED DESIGN

by

Edward Joseph Wiest

Memorandum No. UCB/ERL M90/94

16 October 1990

**ELECTRONICS RESEARCH LABORATORY**

College of Engineering
University of California, Berkeley
94720

TITLE PAGE

## ACKNOWLEDGEMENTS

I would like to express my gratitude to my thesis advisor, Professor Elijah Polak, for supervising my research. I have learned much about computational mathematics (and research in general) from him. Much of the research on which this thesis is based was performed in collaboration with him. He proposed the variable metric technique which is analyzed in Chapter 3, spotted numerous errors in my proofs, and, of course, provided guidance throughout the process. Chapters 2, 3 and 4 derive from papers which were co-authored by him.

I would like to thank the members of my thesis committee, Professors Pravin Varaiya and Shmuel Oren, for reading my thesis and Professors Ilan Adler and Seth Sanders for attending my qualifying exam.

I would like to thank my co-workers, Tom Yang, Limin He, Ted Baker and Ywh-Pyng Harn, for their help. Joe Higgins deserves special thanks for his ready answers to my optimization questions and his patient answers to my UNIX questions. I would also like to thank Cormac Conroy and Mike Jackson for many interesting discussions and for describing to me what electrical engineers do.

I would like to thank my parents for their support and encouragement throughout my education. My fond thanks goes to Caroline McCall and Beate Lohser for their emotional support and friendship.

Finally, I would like to thank my high school science teacher, Elizabeth Crippen, for encouraging me to study mathematics independently.

This research was supported in part by the National Science Foundation (grant ECS-8713334), the Air Force Office of Scientific Research (contract AFOSR-86-0116), the State of California MICRO Program (grant 532410-19900) and by a Howard Hughes Doctoral Fellowship (Hughes Aircraft Company). I would like to acknowledge these organizations for their assistance.

## Table of Contents

# CHAPTER 1

# INTRODUCTION

## 1.1 BACKGROUND

Many engineering design problems can be formulated as nonsmooth optimization problems of the form,

$$\min_{x \in \mathbb{R}^n} \{ \psi(x) \mid f^j(x) \le 0, \forall\, j \in \mathbf{p} \} \, , \tag{1.1.1}$$

where $\psi(x) \triangleq \max_{k \in \mathbf{q}} e^k(x)$, the functions $f^k:\mathbb{R}^n \to \mathbb{R}$ and $e^k:\mathbb{R}^n \to \mathbb{R}$ are continuously differentiable, and where $\mathbf{p}$ and $\mathbf{q}$ denote the sets $\{ 1, \ldots, p \}$ and $\{ 1, \ldots, q \}$, respectively. The components of the vector $x$ represent design variables, the functions $e^k(\cdot)$ represent competing measures of cost and performance which are to be minimized, and each constraint $f^j(x) \le 0$, $j \in \mathbf{p}$, represents a design specification which must be satisfied [Nye.1, Pol.4, Wuu.1].[1] The max function $\psi(\cdot)$ provides a way of simultaneously considering competing measures of performance. The function $\psi(\cdot)$ is nondifferentiable, and, although the points at which $\psi(\cdot)$ is nondifferentiable form a set of measure zero, the nondifferentiability may cause ordinary nonlinear programming algorithms to converge to a nonstationary point when applied to (1.1.1) [Pol.4].[2] As a consequence, algorithms which exploit the special structure of the constrained minimax problem (1.1.1) are needed.

---

[1]More general problems can be specified. For example, if one wishes to design a feedback controller for a linear time-invariant plant so that the maximum deviation of the closed-loop impulse response from some desired response is minimized, then the max in the function $\psi(\cdot)$ will include a continuum of functions, not just a finite number. In this case, $\psi(\cdot)$ will have the form $\max_{y \in Y} \phi(x, y)$ where $Y$ is a compact subset of $\mathbb{R}^m$. See [Pol.4] for a discussion of such *semi-infinite* problems. Another example of a more general problem is the computation of a continuous-time optimal control. In such a problem, the design vector is an element of an infinite-dimensional space. Algorithms for these more general kinds of optimization problems are based on progressive discretization [He.1], that is, the general problem is reduced to partial solution of a sequence of problems of the form (1.1.1) which are obtained by discretizing the more general problem. Hence, algorithms developed for (1.1.1) can be incorporated into algorithms for the more general problem. The paper [Cha.2] discusses ways of modifying algorithms of the type presented here to handle equality constraints.

[2]Problem (1.1.1) may be transcribed into an equivalent differentiable nonlinear programming problem, $\min \{ w \mid e^k(x) - w \le 0, k \in \mathbf{q}, f^j(x) \le 0, j \in \mathbf{p} \}$. However, this transcription is not recommended, because nonlinear programming algorithms converge more slowly on the transcription than minimax algorithms designed specifically for (1.1.1).

In many engineering design problems, there are a few "hard" constraints, the violation of which, by however small an amount, is unacceptable [Nye.1]. An example of such a constraint in the design of a controller for a linear system is one which constrains the poles of the closed-loop system to lie in the left half of the complex plane. If this constraint is violated, the system is unstable. Furthermore, some of the performance functions $e^k(\cdot)$ may not be defined outside of the feasible region. For example, the value of performance functions involving integrals of the impulse response may not be well-defined when the system is unstable. Other hard constraints ensure the physical realizability of the design. For example, the length of a beam in a seismic-resistant structure cannot be negative. For these reasons, it is important that, once an algorithm for the solution of (1.1.1) constructs an estimate of the solution $x_i$ which is feasible, all subsequent iterates are feasible, i.e., that

$$f^j(x_l) \le 0 \ , \quad \forall j \in \mathbf{p}, \tag{1.1.2a}$$

for all $l > i$. It is also desirable that, once a feasible iterate is computed, the value of the objective function be decreased at each iteration, i.e., that

$$f^0(x_{l+1}) < f^0(x_l) \ , \tag{1.1.2b}$$

for all $l > i$. We will refer to an algorithm which satisfies both (1.1.3a) and (1.1.3b) as a *feasible descent* algorithm.

A number of feasible descent algorithms have been developed, e.g., [Hua.1-2, Mey.1, Mif.1, Pir.1, Pol.4, Pol.7, Top.1, Her.1, Pan.1-2, Zou.1],[3] but, for the most part, they are slow. This is a serious impediment to the use of the algorithms for several reasons. Obviously, for design situations in which an optimization involves the investment of several days of computing time, the use of a faster algorithm means a considerable savings in time and money. More commonly, however, the solution of a serious optimization problem arising in design involves some hours of computing time. The benefit of reducing computing time from several hours to a fraction of an

[3]Most of the algorithms referred to are intended for special case where $\psi(x)$ is differentiable.

hour is that the optimization becomes an interactive part of the design process. It is then convenient for a designer to initiate the optimization, review the results, reformulate the optimization problem (for example, by changing the topology of a circuit under design) and rerun the optimization.

Speed of convergence is important for yet another reason. Few problems have a single local solution; more often there are several local minima and a number of nonminimum stationary points. (The latter are analogous to the inflection points and maxima of a single-variable function). The algorithms we describe are local; they are able to locate a nearby stationary point, but do not, in general, find the global minimizer. Multistart methods are a successful way of using a local algorithm to locate a global minimizer [Rin.1-2]. In a multistart method, a local optimization algorithm is initiated with a variety of starting points. The best local minimizer found in this search is taken as an approximation to the global minimizer. Fast algorithms make it possible to try a large number of starting points in a reasonable amount of time. Conversely, without fast algorithms, it may be impractical to perform a global search on problems which involve time-consuming function evaluations.

The aim of our research was to produce faster algorithms for the solution of (1.1.1). However, to simplify our analysis, we consider separately two simpler problems, the unconstrained minimax problem,

$$\min_{x \in \mathbf{R}^n} \psi(x) , \tag{1.1.3a}$$

and the smooth, inequality-constrained nonlinear programming problem,

$$\min_{x \in \mathbf{R}^n} \{ f^0(x) \mid f^j(x) \leq 0, \forall j \in \mathbf{p} \} . \tag{1.1.3b}$$

We present and analyze iterative algorithms for the solution of both of these problems. We demonstrate the reliability of the algorithms presented by proving that they converge only to points satisfying necessary conditions for optimality.

The efficiency of the algorithms is evaluated by two methods. We performed numerical experiments, in which we ran our algorithms on a variety of problems and compared their performance with other algorithms of similar type. Tests of the performance of an algorithm on a few problems, however, are an unreliable indicator of the general behavior of an algorithm. The small number of problems on which it is practical to test an algorithm constitute only a sparse sampling of the whole "space" of optimization problems. For this reason, we invest most of our effort in this dissertation in deriving the asymptotic rates of convergence of our algorithms as they approach solutions. Such results characterize the terminal behavior of algorithms on a large proportion of the possible problems.[4] While it would have been preferable to characterize the speed of convergence throughout the computation, rather than just in a neighborhood of the solution, such results are difficult to obtain (see [Nem.1] for examples). Also, it is not clear whether they are a better indicator of "average" performance than asymptotic rates.

## 1.2 ALGORITHMS

We will be concerned with iterative optimization algorithms for the solution of (1.1.3a) and (1.1.3b) which are based on a fixed procedure for improving upon any given estimate of the solution. By repeating the procedure over and over, an iterative algorithm constructs from an initial guess at the solution a sequence of ever-improving approximations $\{ x_i \}_{i \in \mathbb{N}}$. Most algorithms construct the iterate $x_{i+1}$ from the current iterate $x_i$ by computing a *search direction* $h_i \in \mathbb{R}^n$, choosing a *step length* $\lambda_i \in \mathbb{R}$, and setting $x_{i+1} = x_i + \lambda_i h_i$.

One class of algorithms for solving the minimax problem (1.1.3a) obtains a search direction $h_i$ at each iteration by solving the subproblem,

$$\min_{h \in \mathbb{R}^n} \max_{k \in I_i} e^k(x_i) + \langle \nabla e^k(x_i), h \rangle + \tfrac{1}{2}\gamma \| h \|^2 . \tag{1.1.4}$$

[4] We make the assumption that the algorithm which requires the fewest iterations to solve a problem is the most efficient in terms of CPU time. We justify this assumption as follows. The evaluation of some of the functions appearing in engineering design problems often involves simulation or requires the numerical solution of differential equations. The evaluation of these functions and their gradients consumes so much time that the overhead involved in computing the next iterate from this information is negligible.

Algorithms in this class are characterized by the quantities $I_i \subset \mathbf{q}$, $\gamma > 0$, and a rule for computing the step size. Algorithms of this type were proposed first by Pshenichnyi [Psh.1-2] and later, independently, by Pironneau and Polak [Pir.1]. (Hence, we will call such algorithms PPP algorithms.) A linear rate of convergence was established for a PPP algorithm by Pshenichnyi under fairly strong assumptions. Under different, but equally strong, convexity assumptions, Polak and Wiest [Pol.2-3] showed that the sequence of function values $\{ \psi(x_i) \}_{i \in \mathbb{N}}$ constructed by a PPP algorithm converges Q-linearly to the minimum, i.e., there exists $\delta \in (0, 1)$ such that

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \leq \delta\,, \tag{1.1.5}$$

where $\hat{\psi}$ is the minimum value of $\psi(x)$. The quantity on the left-hand side of (1.1.5) is sometimes referred to as the *convergence ratio* of the sequence $\{ \psi(x_i) \}_{i \in \mathbb{N}}$ [Lue.1]. In Chapter 2, we show that these assumptions can be significantly relaxed. The relaxed assumptions are only slightly stronger than the second-order sufficiency conditions for $\hat{x}$ to be a local minimizer of $\psi(\cdot)$. The convergence rate theory developed in Chapter 2 is used to prove convergence rate results in Chapters 3, 4 and 5.

Minimax problems having a special structure arise in the design of feedback compensators and in the computation of open-loop optimal controls for a linear discrete-time system. (See the appendix to Chapter 3.) In this class of problems, each function $e^k(x)$ is the composition of convex function with a linear function, i.e., $e^k(x) = g^k(A_k x)$ for each $k \in \mathbf{q}$, where $g^k : \mathbb{R}^{m_k} \to \mathbb{R}$ and $A_k \in \mathbb{R}^{m_k \times n}$. Hence, (1.1.3a) becomes

$$\min_{x \in \mathbb{R}^n} \max_{k \in \mathbf{q}} g^k(A_k x)\,. \tag{1.1.6}$$

If the intersection of the null spaces of the matrices $A_k$ is nontrivial, then the minimax problem will not have any isolated minimizers. Instead, for any minimizer $\hat{x}$, every point in the affine space $\hat{x} + \bigcap_{k \in \mathbf{q}} Null(A_k)$ is a local minimizer. (If only some of the functions $e^k(\cdot)$ are active at a minimizer, then the minimizing set may be neither isolated nor a whole affine space.) Of course, no minimizer of such a problem satisfies the second-order sufficiency conditions for optimality. Yet, when a PPP minimax algorithm is applied to such a problem, linear convergence of the function values to the minimum is observed. In the latter part of Chapter 2, we prove that linear convergence is attained by two PPP algorithms on minimax problems of this kind.

The convergence ratio bound derived in Chapter 2 depends in part upon the conditioning of the Hessian matrix of a Lagrangian-type function, $\sum_{k \in \mathbf{q}} \hat{\mu}^k \nabla^2 e^k(\hat{x})$, where $\hat{\mu}$ is a vector of optimal multipliers associated with the solution $\hat{x}$. In particular, when the smallest positive eigenvalue of this matrix is very small, the convergence ratio bound is nearly one. The speed of convergence is then quite slow. In Chapter 3, we explore the effect which rescaling the domain space has on the speed with which PPP algorithms solve composite minimax problems of the form (1.1.6). Suppose that the change of variables $x = Qy$ is made (where $Q$ is an invertible, symmetric matrix), so that (1.1.6) becomes

$$\min_{x \in \mathbf{R}^n} \max_{k \in \mathbf{q}} g^k(A_k Q y) . \tag{1.1.7}$$

Then the Lagrangian Hessian matrix becomes $Q^T \sum_{k \in \mathbf{q}} \hat{\mu}^k \nabla^2 e^k(\hat{x}) Q$.[5] The choice $Q \triangleq \left[ \left( \sum_{k \in \mathbf{q}} \hat{\mu}^k \nabla^2 e^k(\hat{x}) \right)^{1/2} \right]^{\dagger}$ (where $X^{\dagger}$ denotes the pseudoinverse of the matrix $X$) yields a Hessian matrix with a smallest positive eigenvalue of one.[6] The result is an improvement in the convergence ratio bound derived for the PPP algorithm in Chapter 2.

Unfortunately, the matrix $Q$ cannot be computed in advance, since the vector $\hat{x}$ and the optimal multipliers $\hat{\mu}$ are unknown until the problem has been solved. This situation, in which there exists a transformation of the domain space which would greatly increase the speed of an

[5] The vector of optimal multipliers is invariant under such a domain transformation.

[6] The matrices $\nabla^2 e^k(\hat{x})$ must be positive semi-definite in order for $Q$ to be real. We augment the zero eigenvalues of $Q$ as defined in order to make it invertible.

algorithm in solving a problem but which is unknown, occurs elsewhere in optimization. For example, the convergence of the method of steepest descent on the problem $\min_{x \in \mathbb{R}^n} f(x)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is a uniformly convex, twice Lipschitz continuously differentiable function is, in general, linear. However, the method of steepest descent converges superlinearly on the transformed problem, $\min_{y \in \mathbb{R}^n} f(\nabla^2 f(\hat{x})^{-1/2} y)$, where $\hat{x}$ is the solution. The usual approach in such a situation is to rescale the domain at each iteration of the basic optimization algorithm with an estimate of the desired, but unknown, transformation matrix $\nabla^2 f(\hat{x})^{-1/2}$. Such a method is known as a *variable metric* method. If the sequence of matrices defining the transformations actually used converges to the matrix defining the desired transformation, then an increase in the speed of convergence is obtained. Referring to the example of smooth unconstrained optimization, rescaling the domain at iteration $i$ using the transformation defined by $Q_i \triangleq \nabla^2 f(x_i)^{-1/2}$ and then applying one iteration of steepest descent yields an algorithm which is equivalent to Newton's method with an exact line search. The latter method is superlinearly convergent.

Han proposed such a variable metric technique for the minimax problem (1.1.3a). The method is equivalent to rescaling the domain at iteration $i$ using the transformation defined by $Q_i \triangleq \left[ \left( \sum_{k \in \mathbf{q}} \mu_i^k \nabla^2 e^k(x_i) \right)^{1/2} \right]^{\dagger}$, and applying one iteration of a PPP algorithm. Han showed that his algorithm is convergent, but it does not converge superlinearly due to the Maratos effect.[7] In Chapter 3, we use the bound on the convergence ratio of sequences constructed by PPP algorithms which was derived in Chapter 2 to gauge the effect of incorporating a variable metric into a PPP algorithm. Our analysis shows that the use of this variable metric decreases the convergence ratio bound. Numerical results show that there is a corresponding improvement in the actual convergence ratios. The variable metric method also improves the performance of a non-PPP algorithm.

---

[7]The Maratos effect [Mar.1] is a phenomenon observed in the behavior of nonlinear programming algorithms, but it can be observed in the behavior of minimax algorithms as well. The Maratos effect occurs when the insistence of the step size rule on a decrease in $\psi(x)$ prevents a step size of one from being taken, even near a solution. As a result, the convergence of an algorithm which would be superlinearly convergent if unity step sizes were taken is degraded to linear.

Rather than compute the second-derivatives, $\nabla^2 e^k(x_i)$, Han considered estimating them using quasi-Newton updates. We take another approach to avoid computing these second-derivatives. Note that

$$\sum_{k \in \mathbf{q}} \hat{\mu}^k \nabla^2 e^k(\hat{x}) = \sum_{k \in \mathbf{q}} \hat{\mu}^k A_k^T \nabla^2 g^k(\hat{x}) A_k . \tag{1.1.8}$$

Since the functions $g^j(\cdot)$ encountered in control system design are generally well-conditioned, most of the benefit of the original rescaling can be obtained by using transformations defined by $Q_i \triangleq (\sum_{k \in \mathbf{q}} \mu_i^k A_j^T A_j)^{1/2 \dagger}$.

In the second half of the dissertation, we turn our attention to the smooth, inequality-constrained optimization problem (1.1.3b). Algorithms for the solution of (1.1.3b) can be divided, with some overlap, into several types. In penalty methods, Lagrangian methods and methods of centers, the constrained problem is transcribed into a sequence of unconstrained problems. The most widely used type of algorithm is based upon successive approximation to the optimality conditions of (1.1.3b). The sequential quadratic programming method of Wilson is an early example of such a method [Wil.1].[8] At each iteration of this algorithm, a search direction is obtained by solving a linear complementarity problem[9] which approximates the Kuhn-Tucker necessary conditions for (1.1.3b).

A fourth type of algorithm, which has been much less thoroughly investigated, is based upon successive approximation to (1.1.3b) itself. In such methods, a search direction is obtained at each iteration by solving a natural approximation to (1.1.3b) in which each function $f^j(\cdot)$ is replaced by an approximation. Sequential linear programming [Gri.1] and Pshenichnyi's method of linearization [Psh.1] are first-order algorithms of this type. In [Pol.10], algorithms of this type were proposed in which each function $f^j(\cdot)$ is replaced by the quadratic approximation

[8]There has been a great deal of research aimed at constructing second-order methods of this type which are both globally and superlinearly convergent. Algorithms which are based upon successive approximations to optimality conditions and which, of the available second-order information, make use only of the Lagrangian Hessian may fail to achieve their potential superlinear rate of convergence due to truncation of the step size [Mar.1]. Additional evaluations of the constraints are needed to prevent this [May.1, Fle.1, Fuk.1, Gab.1, Col.1-2, Pan.1-2].

$f^j(x) + \langle \nabla f^j(x), h \rangle + \frac{1}{2}\langle h, H^j h \rangle$, for some matrix $H^j \in \mathbb{R}^{n \times n}$. The resulting subproblem is a quadratic program with quadratic constraints, which we will call a *generalized quadratic program* (GQP):

$$\min_{h \in \mathbb{R}^n} \{ f^0(x) + \langle \nabla f^j(x), h \rangle + \frac{1}{2}\langle h, H^j h \rangle \mid f^j(x) + \langle \nabla f^j(x), h \rangle + \frac{1}{2}\langle h, H^j h \rangle \leq 0 , \quad \forall j \in \mathbf{p} \} . \quad (1.1.9)$$

This dissertation presents the first thorough analyses of convergence and rate of convergence of *implementable* GQP-based algorithms. While there has been some theoretical analysis of GQP-based algorithms [Pol.10, Pan.3-4], the algorithms considered were conceptual, that is, they assumed that the GQP subproblem is solved exactly. These algorithms were not implemented (to our knowledge) because no finite step procedures for solving problem (1.1.3b) were known [Pol.10, Pan.4]. We resolve this difficulty by approximating the solution to the subproblem using an active set method. By determining the set of constraints which are active at the solution to the GQP subproblem, the inequality-constrained subproblem (1.1.9) is reduced to an equality-constrained problem. The optimality conditions for the equality-constrained problem constitute a system of polynomial equations (unlike the optimality conditions for (1.1.9), which constitute a nonlinear complementarity problem) to which a root-finding method may be applied. We obtain both a good estimate of the active set and a good starting point for a root-finding method by computing the Polak-Mayne-Trahan search direction [Pol.7, Pir.1].

In Chapter 4, we present a first-order GQP-based algorithm, in which each $H^j$ is taken to be a multiple of the identity. The search direction at each feasible point is a feasible descent direction. We show that the algorithm converges linearly with a convergence ratio bound that is smaller than that obtained for the Polak-Mayne-Trahan algorithm. Numerical results show the algorithm to be generally superior to the Polak-Mayne-Trahan method and comparable to the first-order feasible descent method of Herskovits [Her.1].

---

[9] This linear complementarity problem is equivalent to a quadratic program, and it is often presented and analyzed in that form.

In Chapter 5, we treat a class of second-order GQP-based algorithms, in which the matrix $H^j$ is an estimate of the Hessian matrix $\nabla^2 f^j(\hat{x})$. Second-order algorithms discard curvature information when they combine estimates of the individual Hessian matrices to form a Lagrangian Hessian matrix. The algorithms presented in this chapter are the first to fully use the full-second order information. The algorithms converge globally and superlinearly.

## 1.3 DISSERTATION OUTLINE

In Chapter 2, we prove that two versions of a minimax algorithm converge linearly under assumptions little stronger than second-order sufficiency conditions. For a class of composite minimax problems which do not satisfy these assumptions, we prove that the algorithms converge linearly under a strict complementarity assumption.

In Chapter 3, we present a variable metric method which can be applied to any first-order minimax algorithm to improve the speed of convergence on a class of composite minimax problems. We prove that the technique improves the convergence ratio bound obtained for the minimax algorithms of Chapter 2. Numerical experiments are presented which show that the technique improves the overall speed of convergence of both the minimax algorithm in Chapter 2 and a barrier-function type minimax algorithm for which no convergence rate theory exists.

In Chapter 4, we present a first-order GQP-based algorithm for problem (1.1.3b). We show that the algorithm converges globally and linearly. Numerical results are presented which show that the algorithm is superior to the Polak-Mayne-Trahan method of feasible directions, and is comparable to the first-order feasible descent algorithm of Herskovits [Her.1].

In Chapter 5, we present a class of second-order GQP-based algorithms for problem (1.1.3b). We show that the algorithms converge globally and superlinearly. The rate of convergence obtained ranges from superlinear to 3/2 depending upon the degree of accuracy of the Hessian approximations. Numerical experiments with one algorithm from the class show it to be competitive with the superlinearly convergent, feasible descent algorithm of [Pan.1].

# CHAPTER 2

# ON THE RATE OF CONVERGENCE OF TWO MINIMAX ALGORITHMS

## 2.1 INTRODUCTION

We are concerned in this chapter with algorithms for solving minimax problems of the form

$$\min_{x \in \mathbb{R}^n} \max_{j \in \mathbf{p}} f^j(x), \tag{2.1.1a}$$

where $\mathbf{p} \triangleq \{ 1, 2, ..., p \}$ and each $f^j : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable.

Most minimax algorithms have been shown to converge locally or globally under various conditions. However, the literature dealing with their *rate* of convergence is rather fragmentary (see, e.g., [Psh.1, Pol.1-3, Sho.1, Kiw.1, Pev.1, Mad.1, Dau.1]. In this chapter, we establish the rate of convergence of two versions of a minimax algorithm which was proposed first by Pshenichnyi [Psh.1-2] (who calls it the *method of linearizations*) and later by Pironneau and Polak [Pir.1] as a subprocedure in an implementation of the Huard method of centers[1] [Hua.1].

We will briefly review the literature dealing with the rate of convergence of first-order minimax algorithms for solving problem (2.1.1a). For this, we need to define the function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$\psi(x) = \max_{j \in \mathbf{p}} f^j(x) . \tag{2.1.1b}$$

First, since problem (2.1.1a) can be transcribed into the equivalent constrained form,

$$\min \{ w \mid f^j(x) - w \leq 0, \ j \in \mathbf{p} \} , \tag{2.1.2}$$

it can also be solved by first-order nonlinear programming algorithms.[2] For example, it can be

[1] One of the definitions of the "center" of a set described by inequalities, given by Huard [Hua.1], is in terms of a minimax subproblem of the form (2.1.1a). Consequently, every implementation of the Huard method of centers (e.g. - [Pir.1, Hua.2]) incorporates a minimax algorithm as a subprocedure. This fact was not widely recognized, and some of these imbedded minimax algorithms were later rediscovered independently.

[2] The transcription of (2.1.1a) into (2.1.2) is not recommended, because nonlinear programming algorithms converge more slowly on (2.1.2) than minimax algorithms designed specifically for (2.1.1a).

solved by the Pironneau-Polak method of centers [Pir.1] which, as shown by Chaney [Cha.1], converges linearly on (2.1.2) whenever a strengthened second-order sufficiency condition is satisfied.

Subgradient and bundle methods designed for the more general problem of minimizing locally Lipschitz functions can be used for minimizing the function $\psi(\cdot)$ [Kiw.1]. In Polyak [Pol.1] and Shor [Sho.1], we find proofs that several subgradient methods converge linearly when $\psi(\cdot)$ is strongly convex.

Next, there are several algorithms which were designed specifically for solving minimax problems of the form (2.1.1a). One of the oldest is that of Demyanov [Dem.1-2], which computes $\delta$-approximations to the minimum value of $\psi(\cdot)$ with $\delta > 0$. It computes search directions by solving a linear program defined by the linearizations of the $\delta$-active functions $f^j(\cdot)$ (c.f., the Zoutendijk method of feasible directions [Zou.1], and employs an Armijo-like step size rule. It was shown by Pevny [Pev.1] that, when $\psi(\cdot)$ is strongly convex, the Demyanov algorithm converges linearly in function value.[3] Madsen et al [Mad.1] propose a trust region algorithm for the linearly constrained minimax problem in which a linear program is solved at each iteration. When the solution $\hat{x}$ of (2.1.1a) is a "vertex" solution (also called a *Chebyshev point* or a *Haar point*), the algorithm in [Mad.1] converges quadratically. However, when $\hat{x}$ is not a vertex solution, the rate of convergence of this algorithm is unknown.

The minimax algorithms which we will discuss in this chapter belong to a family conforming to the following algorithm model, which uses the Pshenichnyi-Pironneau-Polak (PPP) search direction subprocedure [Psh.1, Pir.1]:

**PPP Algorithm Model**

**Data:** $x_0 \in \mathbb{R}^n$ ; $\gamma > 0$.

[3] Pevnyi also shows that, if the functions fail to be strongly convex but are convex with bounded level sets, convergence to a $\delta$-optimal value is arithmetic.

**Step 1:** Given $x_i$, compute the search direction,

$$h_i \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in I_i} f^j(x_i) + \langle \nabla f^j(x_i), h \rangle + \tfrac{1}{2}\gamma \| h \|^2 . \tag{2.1.3}$$

**Step 2:** Compute the step size $\lambda_i$.

**Step 3:** Set $x_{i+1} = x_i + \lambda_i h_i$, replace $i$ by $i + 1$ and go to Step 1. □

Algorithms in this family are specified by the quantities $I_i \subset \mathbf{p}$, $\gamma > 0$, and a rule for computing the step size $\lambda_i$. Thus, in [Psh.1], we find a minimax algorithm in the PPP family with $I_i \triangleq \{ j \in \mathbf{p} \mid f^j(x_i) \geq \psi(x_i) - \delta \}$ (with $\delta > 0$), $\gamma = 1$, and the *constant* step size $\lambda_i \equiv \lambda$. It is shown in [Psh.1] that the resulting algorithm converges linearly, provided that the initial point is sufficiently close to $\hat{x}$. The proof assumes that $\lambda$ is sufficiently small, and that strict complementary slackness, affine independence of the gradients of the active functions and second-order sufficiency conditions hold at $\hat{x}$. It is also shown in [Psh.1] that, if $\lambda = 1$ and $\hat{x}$ is a "vertex" solution, then the *local* algorithm converges quadratically.

In [Psh.1], we also find a PPP minimax algorithm which uses the step size rule

$$\lambda_i = \max_{k \in \mathbb{N} \cup \{0\}} \{ 2^{-k} \mid \psi(x_i + 2^{-k} h_i) - \psi(x_i) \leq -2^{-k} \alpha \| h_i \|^2 \} , \quad \alpha \in (0, 1) , \tag{2.1.4}$$

where $\mathbb{N}$ is the set of all nonnegative integers. It was shown in [Dau.1] that, if (2.1.1a) has a "vertex" solution $\hat{x}$, then the step size in the above algorithm eventually becomes unity. It therefore follows from [Psh.1], that if a sequence $\{x_i\}_{i=0}^{\infty}$, constructed by the PPP algorithm using (2.1.4), converges to a "vertex" solution $\hat{x}$, then it converges quadratically.

In [Pol.2], $I_i = \mathbf{p}$, $\gamma > 0$ and an Armijo step size rule [Arm.1] similar to (2.1.4) is used, while, in [Pol.3], $I_i = \mathbf{p}$, $\gamma = 1$ and an exact minimizing line search is used to determine step size. It was shown in these papers that both of these PPP methods converge linearly under the assumption that the functions $f^j(\cdot)$ are strongly convex.

In Sections 3 and 4 of this chapter, we show that the PPP algorithms, considered in [Pol.2-3], converge linearly under a slightly strengthened form of the standard second-order sufficiency condition. This condition is considerably weaker than the strong convexity assumption used in [Pol.2-3]. Furthermore, unlike in [Psh.1], we assume neither strict complementary slackness nor affine independence of the gradients of the active functions.

In Section 5, we consider the *composite* minimax problem,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \mathbf{p}} g^j(A_j x) , \tag{2.1.5}$$

in which each continuously differentiable function $g^j : \mathbb{R}^{l_j} \to \mathbb{R}$ is composed with a different linear function $A_j : \mathbb{R}^n \to \mathbb{R}^{l_j}$. Minimax problems of this form arise in the design of feedback compensators and of discrete-time optimal controls. We show that, despite the fact that the solution set is generally nonunique,[4] the PPP algorithm described in [Pol.2] converges linearly, under somewhat more stringent assumptions than for the general case.

## 2.2 THE PPP MINIMAX ALGORITHM WITH EXACT LINE SEARCH

In this section, we will consider the algorithm which results when the step size $\lambda_i$ in the PPP Algorithm Model is computed by exact minimization along the search direction. To simplify notation, we define

$$\phi^j(h \mid x) \triangleq f^j(x) + \langle \nabla f^j(x), h \rangle + \tfrac{1}{2}\gamma \|h\|^2 . \tag{2.2.1}$$

**Algorithm 2.2.1 (PPP-ELS):** (see Algorithm 5.2 and Corollary 5.1 in [Pol.4])

**Data:** $x_0 \in \mathbb{R}^n$ ; $\gamma > 0$.

**Step 0:** Set $i = 0$.

**Step 1:** Compute the search direction,[5]

[4] In fact, the solution set must contain a translation of the intersection of the null spaces of the matrices $A_j$.

[5] For the convenience of the proofs to follow, we subtract the term $\psi(x_i)$ from the minimand in (2.1.3), so as to make the value $\theta(x_i)$ of the search direction finding problem less than or equal to zero. This has no effect on the resulting search direction.

$$h(x_i) \triangleq \arg \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{p}} \phi^j(h \mid x_i) - \psi(x_i). \tag{2.2.2}$$

**Step 2:** Compute the minimizing step size, $\lambda_i = \arg \min_{\lambda \in \mathbb{R}} \psi(x_i + \lambda h(x_i))$.

**Step 3:** Set $x_{i+1} = x_i + \lambda_i h(x_i)$, replace $i$ by $i+1$ and go to Step 1. □

Let the standard unit simplex be denoted by $\Sigma_p \triangleq \{ \mu \in \mathbb{R}^p \mid \mu^j \geq 0 , \sum_{j \in \mathbf{p}} \mu^j = 1 \}$. Then the search direction finding problem (2.2.2) can be transformed as follows:

$$\theta(x) \triangleq \min_{h \in R^n} \left[ \max_{j \in \mathbf{p}} \phi^j(h \mid x) - \psi(x) \right]$$

$$= \min_{h \in R^n} \left[ \max_{\mu \in \Sigma_p} \sum_{j \in \mathbf{p}} \mu^j \phi(h \mid x) - \psi(x) \right]. \tag{2.2.3}$$

Next, by Theorem 2.7.1, the max and min in (2.2.3) can be interchanged, and hence we obtain that

$$\theta(x) = \max_{\mu \in \Sigma_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \mu^j \phi^j(h \mid x) - \psi(x)$$

$$= \max_{\mu \in \Sigma_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \mu^j \left( f^j(x) + \langle \nabla f^j(x), h \rangle - \psi(x) \right) + \tfrac{1}{2}\gamma \| h \|^2 . \tag{2.2.4}$$

The solution $\mu$ to (2.2.4) is not always unique, and hence we define the solution set

$$U(x) \triangleq \arg \max_{\mu \in \Sigma_p} \left[ \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \mu^j \left( f^j(x) + \langle \nabla f^j(x), h \rangle - \psi(x) \right) + \tfrac{1}{2}\gamma \| h \|^2 \right]. \tag{2.2.5a}$$

By solving the inner minimization problem in (2.2.5a), we see that $U(x)$ is the solution set to a positive semi-definite quadratic program,

$$U(x) = \arg \max_{\mu \in \Sigma_p} \sum_{j \in \mathbf{p}} \mu^j \left( f^j(x) - \psi(x) \right) - \frac{1}{2\gamma} \Big\| \sum_{j \in \mathbf{p}} \mu^j \nabla f^j(x) \Big\|^2 . \tag{2.2.5b}$$

Several methods exist for solving such problems (see, for example, [Gil.1, Hig.1, Kiw.2-3, Rus.1, von.1]).

As a consequence of the extended von Neumann Minimax Theorem (Theorem 2.7.1), we have that, for any $\bar{\mu} \in U(x)$,

$$\sum_{j \in \mathbf{p}} \bar{\mu}^j \phi^j(h(x) \mid x) = \max_{\mu \in \Sigma_p} \sum_{j \in \mathbf{p}} \mu^j \phi^j(h(x) \mid x)$$

$$= \min_{h \in \mathbb{R}^n} \max_{\mu \in \Sigma_p} \sum_{j \in \mathbf{p}} \mu^j \phi^j(h \mid x)$$

$$= \max_{\mu \in \Sigma_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \mu^j \phi^j(h \mid x)$$

$$= \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \bar{\mu}^j \phi^j(h \mid x) . \tag{2.2.6}$$

Hence, any multiplier vector $\bar{\mu} \in U(x)$ yields the solution,

$$h(x) = \arg \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \bar{\mu}^j \phi^j(h \mid x) \tag{2.2.7}$$

to the search direction finding problem (2.2.2) (for $x = x_i$), which is unique since the function $\max_{j \in \mathbf{p}} \phi^j(\cdot \mid x)$ is strictly convex.

Next we recall the following necessary optimality condition for problem (2.1.1a).

**Theorem 2.2.1: [Cla.1, Dem.3, Joh.1, Pol.4]** *If $\hat{x} \in \mathbb{R}^n$ is a solution to problem (2.1.1a), then there exists a vector of multipliers $\hat{\mu} \in \Sigma_p$ such that*

$$\sum_{j \in \mathbf{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0 , \tag{2.2.8a}$$

$$\sum_{j \in \mathbf{p}} \hat{\mu}^j [fx^j(\hat{x}) - \psi(\hat{x})] = 0 . \tag{2.2.8b}$$

When the functions $f^j(\cdot)$ are convex, equations (2.2.8a-b) are also a sufficient condition for optimality. We denote the minimum value for problem (2.1.1a) by $\hat{\psi} \triangleq \min_{x \in \mathbb{R}^n} \psi(x)$ and the set of minimizers by $\hat{G} \triangleq \arg\min_{x \in \mathbb{R}^n} \psi(x)$. For any $\hat{x} \in \hat{G}$, the set of multiplier vectors $\hat{\mu} \in \Sigma_p$ which satisfy equations (2.2.8a-b) together with $\hat{x}$ is exactly $U(\hat{x})$.

**Theorem 2.2.2: [Pol.4]** *Suppose that the functions $f^j(\cdot)$ in problem (2.1.1a) have continuous derivatives. If $\bar{x}$ is an accumulation point of a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 2.2.1, then $\bar{x}$ satisfies the optimality condition (2.2.8.a-b).*

## 2.3 RATE OF CONVERGENCE OF THE PPP-ELS ALGORITHM

We now proceed to prove that the sequence of values, $\{ \psi(x_i) \}_{i=0}^{\infty}$, constructed by Algorithm 2.2.1 converges linearly to the minimum value under weaker assumptions than those used in [Pol.2-3]. Our proof draws on ideas which appeared in the proofs of linear convergence of the Pironneau-Polak algorithm for inequality-constrained minimization in [Pir.1, Cha.1]. We make the following assumptions. Let $F^j(x)$ denote the second derivative matrix of $f^j(x)$ for each $j \in \mathbf{p}$.

**Hypothesis 2.3.1:** *We will assume that*

(i) *the functions $f^j(\cdot)$ are twice continuously differentiable,*

(ii) *there exists $T \in \mathbb{R}$ such that the set $S \triangleq \{ x \in \mathbb{R}^n \mid \psi(x) \le T \}$ is bounded and such that there is a single point $\hat{x} \in S$ which satisfies the necessary conditions* (2.2.8a-b),

(iii) *for some $M' < \infty$*, all $x \in \mathbb{R}^n$ *and all $j \in \mathbf{p}$, $\| F^j(x) \|_2 < M'$.*

For any stationary point $\hat{x}$, we define

$$J(\hat{x}) \triangleq \{ j \in \mathbf{p} \mid \exists \mu \in U(\hat{x}) : \mu^j > 0 \} . \tag{2.3.1}$$

**Hypothesis 2.3.2:** *Let $\hat{x}$ be as defined in Hypothesis 2.3.1, let $B$ denote the null space of the matrix with columns $\{ \nabla f^j(\hat{x}) \}_{j \in J(\hat{x})}$. We will assume that there exists $m' > 0$ such that, for all $\hat{\mu} \in U(\hat{x})$,*

$$m' \| h \|^2 < \langle h, \left[ \sum_{j \in \mathbf{p}} \hat{\mu}^j F^j(\hat{x}) \right] h \rangle \quad \forall h \in B . \tag{2.3.2}$$

Hypothesis 2.3.2 and equations (2.2.8a-b) together constitute a strengthening of the standard second-order sufficiency conditions for $\hat{x}$ to be a local minimizer of $\psi(\cdot)$ [Wom.1]. Note that, while (2.3.2) must hold for *all* multiplier vectors in $U(\hat{x})$, the subspace $B$ over which the inequality must hold may be quite small, because all of the multiplier vectors in $U(\hat{x})$ are used to determine the set $J(\hat{x})$. $\square$

The proof of linear convergence requires several technical lemmas involving the following quantities. With $\hat{x}$ as in Hypothesis 2.3.1 (ii) and $B$ as in Hypothesis 2.3.2, let $P : \mathbb{R}^n \to \mathbb{R}^n$ denote the projection operator with range equal to $B$, and let $P^{\perp}$ be the projection operator with range equal to $B^{\perp}$. Let

$$m \triangleq \min\{ m', \gamma \} . \tag{2.3.3}$$

For any $y \in \mathbb{R}^n$ and $\mu \in \Sigma_p$, we define

$$R(y,\mu) \triangleq \tfrac{1}{2} m\, I - \int_0^1 (1-s) \sum_{j \in \mathbf{p}} \mu^j F^j(\hat{x} + (1-s)y)\,ds . \tag{2.3.4}$$

The function $R(\cdot,\cdot)$ is continuous, and, by Hypothesis 2.3.2, for any $\hat{\mu} \in U(\hat{x})$, $R(0,\hat{\mu})$ is negative definite on the subspace $B$.

We will use the notation $z_i \to Z$ to indicate the convergence of the sequence $\{ z_i \}_{i=0}^{\infty} \subset \mathbb{R}^n$ to the set $Z \subset \mathbb{R}^n$, *i.e.*, the fact that $\lim_{i \to \infty} \min_{y \in Z} \| z_i - y \| = 0$. The following two results are established in the Appendix.

**Lemma 2.3.1:** *If Hypotheses 2.3.1 and 2.3.2 hold, and $\hat{x}$ is defined as in Hypothesis 2.3.1, then there exists $K > 0$ such that*

$$\limsup_{\substack{y \to 0 \\ \mu \to U(\hat{x})}} \frac{\langle y, R(y,\mu) y \rangle}{\| y \| \, \| P^{\perp} y \|} < K . \tag{2.3.5}$$

$\square$

**Lemma 2.3.2:** *If Hypotheses 2.3.1 and 2.3.2 hold, and $\hat{x}$* is defined as in Hypothesis 2.3.1, then

$$\lim_{x \to \hat{x}} \frac{\| x - \hat{x} \| \, \| P^{\perp}(x - \hat{x}) \|}{\psi(x) - \hat{\psi}} = 0 . \tag{2.3.6}$$

$\square$

We now relate the potential decrease in the function $\psi(\cdot)$ at $x$ to the decrease predicted by $\theta(x)$.

**Lemma 2.3.3:** *If Hypotheses 2.3.1 and 2.3.2 hold, then*

$$\limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \frac{\theta(x)}{\psi(x) - \hat{\psi}} \le -\frac{m}{\gamma}. \tag{2.3.7}$$

**Proof:** Referring to (2.2.4) and (2.2.5a-b), we see that, for any $\bar{\mu} \in U(x)$,

$$\theta(x) = \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \bar{\mu}^j \phi^j(h \mid x) - \psi(x). \tag{2.3.8}$$

Let $s \triangleq m / \gamma$. By the definition of $m$ above, $s \le 1$. Substituting $h = s(\hat{x} - x)$ in (2.3.8) and using the definition of $\phi^j(\cdot \mid \cdot)$ in (2.1.2), we obtain that

$$\begin{aligned}
\theta(x) &\le \sum_{j \in \mathbf{p}} \bar{\mu}^j \phi^j(s(\hat{x} - x) \mid x) - \psi(x) \\
&= \sum_{j \in \mathbf{p}} \bar{\mu}^j \left[ f^j(x) - \psi(x) + \langle \nabla f^j(x), s(\hat{x} - x) \rangle + \tfrac{1}{2}\gamma s^2 \|\hat{x} - x\|^2 \right] \\
&\le s \left\{ \sum_{j \in \mathbf{p}} \bar{\mu}^j f^j(x) + \langle \sum_{j \in \mathbf{p}} \bar{\mu}^j \nabla f^j(x), \hat{x} - x \rangle + \tfrac{1}{2} m \|\hat{x} - x\|^2 - \psi(x) \right\},
\end{aligned} \tag{2.3.9}$$

since $s \in (0, 1]$ and $f^j(x) \le \psi(x)$. Adding and subtracting the term $\langle x - \hat{x}, \left[ \int_0^1 (1-t) \sum_{j \in \mathbf{p}} \mu^j F^j(\hat{x} + (1-t)(x - \hat{x}))dt \right](x - \hat{x}) \rangle$ to the right hand side of (2.3.9), we find that

$$\begin{aligned}
\theta(x) \le s \Bigg\{ &\sum_{j \in \mathbf{p}} \bar{\mu}^j f^j(x) + \langle \sum_{j \in \mathbf{p}} \bar{\mu}^j \nabla f^j(x), \hat{x} - x \rangle \\
&+ \langle x - \hat{x}, \left[ \int_0^1 (1-t) \sum_{j \in \mathbf{p}} \bar{\mu}^j F^j(\hat{x} + (1-t)(x - \hat{x}))dt \right](x - \hat{x}) \rangle \\
&- \psi(x) + \langle x - \hat{x}, R(x - \hat{x}, \bar{\mu})(x - \hat{x}) \rangle \Bigg\}
\end{aligned} \tag{2.3.10}$$

The first three terms in the right hand side of (2.3.10) constitute the second-order Taylor expansion of $\sum_{j \in \mathbf{p}} \bar{\mu}^j f^j(\hat{x})$. Hence,

$$\theta(x) \le s\left\{\sum_{j \in \mathbf{p}} \bar{\mu}^j f^j(\hat{x}) - \psi(x) + \langle x - \hat{x}, R(x-\hat{x}, \bar{\mu})(x-\hat{x})\rangle\right\}$$

$$\le s\left\{\psi(\hat{x}) - \psi(x) + \langle x - \hat{x}, R(x-\hat{x}, \bar{\mu})(x-\hat{x})\rangle\right\}. \tag{2.3.11}$$

Dividing both sides of (2.3.11) by $\psi(x) - \hat{\psi}$, we get

$$\frac{\theta(x)}{\psi(x) - \psi(\hat{x})} \le s\left\{-1 + \frac{\langle x - \hat{x}, R(x-\hat{x}, \bar{\mu})(x-\hat{x})\rangle}{\psi(x) - \psi(\hat{x})}\right\}. \tag{2.3.12}$$

By Lemma 2.3.1,

$$\limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \max_{\mu \in U(x)} \frac{\langle x - \hat{x}, R(x-\hat{x}, \mu)(x-\hat{x})\rangle}{\|x - \hat{x}\| \, \|P^{\perp}(x-\hat{x})\|} < K, \tag{2.3.13}$$

and, by Lemma 2.3.2,

$$\limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \frac{\|x - \hat{x}\| \, \|P^{\perp}(x-\hat{x})\|}{\psi(x) - \hat{\psi}} = 0. \tag{2.3.14}$$

Taking the lim sup of (2.3.12) as $x \to \hat{x}$ and using (2.3.13), (2.3.14) and

$$s = \frac{m}{\gamma}, \tag{2.3.15}$$

yields (2.3.7). □

We combine Lemma 2.3.3 with a relation between the decrease predicted by $\theta(x)$ and the actual decrease obtained at $x$ in the direction $h(x)$ using an exact line search. Let

$$M \triangleq \max\{M', \gamma\}. \tag{2.3.16}$$

**Lemma 2.3.4:** *If Hypotheses 2.3.1 and 2.3.2 hold, then*

$$\limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \leq 1 - \frac{\min\{m', \gamma\}}{\max\{M', \gamma\}} . \tag{2.3.17}$$

**Proof:** Since by Hypothesis 2.3.1(ii), $\hat{x}$ is the only point in the level set $S$ satisfying the necessary condition for optimality (2.2.8a-b), it follows that $\psi(\hat{x}) = \hat{\psi}$ and $\psi(x) > \hat{\psi}$ for all $x \in S \setminus \hat{x}$. Since $\theta(x)$ is zero if and only if the necessary conditions (2.2.8a-b) are met at $x$, $\theta(x) < 0$ for all $x \in S \setminus \hat{x}$.

The second derivative bound of Hypothesis 2.3.1(iii) implies that for each $f^j(\cdot)$,

$$f^j(y + z) - f^j(y) - \langle \nabla f^j(y), z \rangle \leq \tfrac{1}{2} M \|z\|^2 , \quad \forall y , \; z \in \mathbb{R}^n . \tag{2.3.18}$$

Thus, for any $\bar{\lambda} \in (0, 1]$ and $x \in S \setminus \hat{x}$,

$$\begin{aligned} \min_{\lambda \in \mathbb{R}} \psi(x + \lambda h(x)) - \psi(x) &\leq \psi(x + \bar{\lambda} h(x)) - \psi(x) \\ &\leq \max_{j \in \mathbf{p}} f^j(x) - \psi(x) + \langle \nabla f^j(x), \bar{\lambda} h(x) \rangle + \tfrac{1}{2} M \|\bar{\lambda} h(x)\|^2 , \\ &\leq \bar{\lambda} \left[ \max_{j \in \mathbf{p}} f^j(x) - \psi(x) + \langle \nabla f^j(x), h(x) \rangle + \tfrac{1}{2} \bar{\lambda} M \|h(x)\|^2 \right] . \end{aligned} \tag{2.3.19}$$

Setting $\bar{\lambda} = \gamma / M$ and using the definition of $M$,

$$\begin{aligned} \min_{\lambda \in \mathbb{R}} \psi(x + \lambda h(x)) - \psi(x) &\leq \bar{\lambda} \left[ \max_{j \in \mathbf{p}} f^j(x) - \psi(x) + \langle \nabla f^j(x), h(x) \rangle + \tfrac{1}{2} \gamma \|h(x)\|^2 \right] \\ &= \bar{\lambda} \theta(x) . \end{aligned} \tag{2.3.20}$$

Since $\theta(x) < 0$ for all $x \neq \hat{x}$,

$$\min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \psi(x)}{\theta(x)} \geq \bar{\lambda} = \frac{\gamma}{M} . \tag{2.3.21}$$

Applying inequality (2.3.21) and Lemma 2.3.3 to the left hand side of (2.3.17), we obtain

$$\limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \psi(x)}{\psi(x) - \hat{\psi}} = \limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \psi(x)}{\theta(x)} \; \frac{\theta(x)}{\psi(x) - \hat{\psi}}$$

$$\le \frac{\gamma}{M} \limsup_{\substack{x \to \hat{x} \\ x \ne \hat{x}}} \frac{\theta(x)}{\psi(x) - \hat{\psi}}$$

$$\le \frac{\gamma}{M} \left[ \frac{-m}{\gamma} \right]$$

$$= -\frac{m}{M} = -\frac{\min\{m', \gamma\}}{\max\{M', \gamma\}} . \tag{2.3.22}$$

The second step holds because $\theta(x) < 0$ and $\psi(x) > \hat{\psi}$. Adding 1 to both sides yields the desired result. □

**Theorem 2.3.2:** *If Hypotheses 2.3.1 and 2.3.2 hold and Algorithm 2.2.1 generates a sequence* $\{ x_i \}_{i=0}^{\infty}$, *starting from a point* $x_0 \in S$, *then* (a) $x_i \to \hat{x}$ *as* $i \to \infty$, *and* (b) *either* $x_i = \hat{x}$ *for all large* $i$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \frac{\min\{m', \gamma\}}{\max\{M', \gamma\}} . \tag{2.3.23}$$

**Proof:** (a) The sequence $\{ x_i \}_{i=0}^{\infty}$ lies in the compact set $S$, and hence it converges to the set of its accumulation points. By Theorem 2.2.2, each accumulation point must satisfy the necessary conditions (2.2.8a-b). Since, by Hypothesis 2.3.1(ii), only $\hat{x} \in S$ satisfies (2.2.8a-b), the sequence converges to $\hat{x}$.

(b) Follows from (a) and Lemma 2.3.5. □

Following Luenberger [Lue.1], we refer to the quantity $\limsup_{i \to \infty} (\psi(x_{i+1}) - \hat{\psi}) / (\psi(x_i) - \hat{\psi})$ as the *convergence ratio* of the sequence $\{ \psi(x_i) \}_{i=0}^{\infty}$. The right-hand side of inequality (2.3.23) bounds the convergence ratio of any sequence constructed by Algorithm 2.2.1 in solving any problem in the class defined by Hypotheses 2.3.1 and 2.3.2.

**Remark 2.3.1:** The functions $\phi^j(\cdot \mid \cdot)$, could have been defined in (2.2.1) using different values of $\gamma$ for each $j \in \mathbf{p}$. This would have had two effects. First, the search direction finding problem would have been considerably more difficult to solve. Second, the convergence ratio bound in the right-hand side of inequality (2.3.23) could turn out to be larger; certainly it would not be smaller. However, if individual bounds on the $\|F^j(x)\|$ are known, one may be able to establish lower bounds by using different $\gamma_j$. □

**Remark 2.3.2:** From the definition of $m$ and $M$ in (2.3.3) and (2.3.16), the ratio $m/M$ appearing in the convergence ratio bound is independent of $\gamma$ provided that $\gamma \in [m', M']$. However, for $\gamma$ outside this range, $m/M$ is smaller and the convergence ratio bound is greater. The following example shows that this dependence of the convergence ratio bound on $\gamma$ is not an artifact of our proof technique, but that it reflects the dependence of the actual convergence ratios on $\gamma$. We applied Algorithm 2.4.1 (see Section 4) to the problem of minimizing the maximum of $f^1(x) \triangleq -6x_0 + 4(x_0^2 + x_1^2)$ and $f^2(x) \triangleq x_0 + ½(x_0^2 + x_1^2)$ using a variety of values for $\gamma$. For this problem $m' = 2$ and $M' = 8$. We started the algorithm from the point $(1, 1)$, and used $\gamma \in \{2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^4, 2^5, 2^6\}$. Figure 2.1 displays both the convergence ratio bounds computed from the right-hand side of (2.4.2) and the convergence ratios which were observed. □

## 2.4 RATE OF CONVERGENCE OF THE PPP-ARMIJO ALGORITHM

The step size rule used in Algorithm 2.2.1 calls for the exact minimization of a function of a single variable. In practice, we use a step size rule which can be executed in a finite number of steps. A suitable replacement for Step 2 in Algorithm 2.2.1 is the following generalization [Pol.4] of the Armijo rule for differentiable functions [Arm.1],

**Step 2':** Compute a step size, $\lambda_i = \beta^{k_i}$, where $k_i \in \mathbb{Z}$ is any integer such that

$$\psi(x_i + \beta^{k_i} h_i) - \psi(x_i) \le \alpha \beta^{k_i} \theta(x_i) , \tag{2.4.1a}$$

and

$$\psi(x_i + \beta^{k-1} h_i) - \psi(x_i) > \alpha\beta^{k-1}\theta(x_i)\,, \tag{2.4.1b}$$

with fixed parameters $\alpha, \beta \in (0, 1)$. We will call the the algorithm resulting from the replacement of Step 2 in Algorithm 2.2.1 by Step 2' Algorithm 2.4.1. The convergence result, Theorem 2.2.2, holds for Algorithm 2.4.1 [Pol.4]. We show that a rate of convergence result very similar to Theorem 2.3.2 holds as well.

**Theorem 2.4.1:** *If Hypotheses 2.3.1 and 2.3.2 hold and Algorithm 2.4.1 generates a sequence* $\{ x_i \}_{i=0}^{\infty}$*, starting from a point* $x_0 \in S$*, then* (a) $x_i \to \hat{x}$*, as* $i \to \infty$*, and* (b) *either* $x_i = \hat{x}$ *for all large* $i$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \alpha\beta \frac{\min\{ m', \gamma \}}{\max\{ M', \gamma \}}\,. \tag{2.4.2}$$

**Proof:** (a) The sequence $\{ x_i \}_{i=0}^{\infty}$ is contained in the compact set $S$, and hence it converges to the set of its accumulation points. Referring to [Pol.4] and using the fact that the functions $f^j(\cdot)$ are continuously differentiable, we conclude that any accumulation point must satisfy the necessary conditions (2.2.8a-b). Since the only point in $S$ satisfying these conditions is $\hat{x}$, the sequence must converge to $\hat{x}$.

(b) We obtain a bound on the decrease in $\psi(\cdot)$ obtained at each iteration, assuming that the sequence does not terminate in a finite number of steps at $\hat{x}$. The second derivative bounds again imply relation (2.3.18), and so, for all $i \in \mathbb{N}$ and $k \ge 0$,

$$\begin{aligned} \psi(x_i + \beta^k h_i) - \psi(x_i) &= \max_{j \in \mathbf{p}} f^j(x_i + \beta^k h_i) - \psi(x_i) \\ &\le \max_{j \in \mathbf{p}} f^j(x_i) + \langle \nabla f^j(x_i), \beta^k h_i \rangle - \psi(x_i) + ½ M \beta^{2k} \| h_i \|^2 \\ &\le \beta^k \left[ \max_{j \in \mathbf{p}} f^j(x_i) + \langle \nabla f^j(x_i), h_i \rangle - \psi(x_i) + ½ M \beta^k \| h_i \|^2 \right], \end{aligned} \tag{2.4.3}$$

because $\beta^k \le 1$ and $f^j(x) \le \psi(x)$. Therefore, if $\beta^k \le \gamma / M$,

$$\psi(x_i+\beta^k h_i)-\psi(x_i) \le \beta^k \left[\max_{j\in \mathbf{p}} f^{\,j}(x_i)+\langle \nabla f^{\,j}(x_i), h_i\rangle - \psi(x_i) + \tfrac{1}{2}\gamma \|h_i\|^2\right]$$
$$= \beta^k\theta(x_i) < \alpha\beta^k\theta(x_i) < 0\,. \tag{2.4.4}$$

It follows from (2.4.1a-b) that $\lambda_i \ge \beta\gamma / M$ and hence that

$$\psi(x_{i+1})-\psi(x_i) \le \alpha\lambda_i\theta(x_i) \le \frac{\alpha\beta\gamma}{M}\,\theta(x_i)\,. \tag{2.4.5}$$

Combining inequality (2.4.5) with Lemma 2.3.3 yields the desired result. □

## 2.5 RATE OF CONVERGENCE ON COMPOSITE MINIMAX PROBLEMS

Next we will establish the rate of convergence of Algorithms 2.2.1 and 2.4.1 on a class of composite minimax problems of the form

$$\min_{x\in \mathbb{R}^n}\max_{j\in \mathbf{p}} g^j(A_j x)\,, \tag{2.5.1}$$

where $g^j : \mathbb{R}^{l_j} \to \mathbb{R}$ is continuously differentiable and $A_j$ is an $l_j \times n$ real matrix. We note that (2.5.1) is a problem of the form (2.1.1a), with the functions $f^{\,j}(\cdot)$ defined by $f^{\,j} \triangleq g^j \cdot A_j$. In conformity with the previous sections, we will use the notation $\psi(x) = \max_{j\in \mathbf{p}} g^j(A_j x)$. We note that when the null spaces of the matrices $A_j$ have a nontrivial intersection, which we will call their *common null space*, problem (2.5.1) does *not* have a unique minimum and therefore does not satisfy Hypothesis 2.3.1(ii). In this case, problem (2.5.1) may also fail to satisfy the convexity requirement of Hypothesis 2.3.2. To see this, note that for problem (2.5.1), the second derivative of the Lagrangian at a minimizer $\hat{x}$ has the form,

$$\sum_{j\in \mathbf{p}} \hat{\mu}^j A_j^T G^j(A_j\hat{x})A_j\,, \tag{2.5.2a}$$

where $G^j(\cdot)$ denotes the second derivative matrix of $g^j(\cdot)$. Continuing to denote by $B$ the null space of the matrix with columns $\{\nabla f^{\,j}(\hat{x})\}_{j\in J(\hat{x})}$, we find that the second derivative matrix will only be positive *semi*-definite on the subspace $B$. However, we have observed in computational experiments that linear convergence of the values $\{ \psi(x_i) \}_{i=0}^{\infty}$ constructed by PPP-ELS and PPP-Armijo is not lost in this circumstance. In this section we will derive a bound on the rate of convergence of Algorithms 2.2.1 and 2.4.1 under the assumption that the Lagrangian Hessian is positive definite only on the orthogonal complement of the common null space of the matrices $A_j$.

Recall that, if $\hat{x} \in \hat{G}$, $U(\hat{x})$ is the set of $\hat{\mu} \in \Sigma_p$ which, together with $\hat{x}$, satisfy (2.2.8a-b). By analogy with nonlinear programming, we shall say that *strict complementary slackness* holds at $(\hat{x}, \hat{\mu})$, where $\hat{x} \in \hat{G}$ and $\hat{\mu} \in U(\hat{x})$, if

$$\hat{\mu}^j > 0 \text{ if and only if } f^j(\hat{x}) = \psi(\hat{x}), \tag{2.5.2b}$$

for $j \in \mathbf{p}$. If strict complementary slackness holds at $(\hat{x}, \hat{\mu})$ for all $\hat{\mu} \in U(\hat{x})$, then the set $U(\hat{x})$ is a singleton, the vectors $\{ A_j^T \nabla g^j(A_j \hat{x}) \}_{j \in J(\hat{x})}$ are affinely independent, and $J(\hat{x}) = \{ j \in \mathbf{p} \mid f^j(\hat{x}) = \psi(\hat{x}) \}$.

The following definition will be used throughout the dissertation. A set-valued function $S : \mathbb{R}^n \to 2^Y$ (where $Y \subset \mathbb{R}^m$) is *upper semicontinuous* in the sense of Clarke [Cla.1] if, for every $x \in \mathbb{R}^n$ and every open set $O \supset S(x)$, there exists a neighborhood $W$ of $x$ such that $S(W) \subset O$. An immediate corollary of this definition is that, if $S(x)$ is a singleton on any open set, then $S(\cdot)$ defines a continuous single-valued function on that set. Like continuity of single-valued functions, upper semicontinuity has a sequence characterization. If $S(\cdot)$ is upper semicontinuous and $S(x)$ is compact for all $x$, then $x_i \to \bar{x}$, $y_i \in S(x_i)$ and $y_i \to \bar{y}$ imply that $\bar{y} \in S(\bar{x})$.

**Proposition 2.5.1:** *Suppose that the functions $g^j(\cdot)$ are strictly convex and that strict complementary slackness holds at $(\hat{x}, \hat{\mu})$ for every $\hat{\mu} \in U(\hat{x})$ and every $\hat{x} \in \hat{G}$. Then, (a) there is a unique $\hat{\mu}$ such that $U(\hat{G}) = \{ \hat{\mu} \}$, and (b) the set $\hat{J} \triangleq J(\hat{x})$ is independent of $\hat{x}$ for all $\hat{x} \in \hat{G}$.*

**Proof:** We show (b) first. Suppose that $\mu_1, \mu_2 \in U(\hat{x})$ for some $\hat{x} \in \hat{G}$, and that $\mu_1 \neq \mu_2$. Let $t$

and $j_0$ be defined by

$$t \triangleq \min_{j \in \mathbf{p}} \{ \mu_1^j / (\mu_1^j - \mu_2^j) \mid \mu_1^j > \mu_2^j \} > 0 \,, \tag{2.5.3a}$$

$$j_0 \triangleq \arg\min_{j \in \mathbf{p}} \{ \mu_1^j / (\mu_1^j - \mu_2^j) \mid \mu_1^j > \mu_2^j \} \,. \tag{2.5.3b}$$

Then $\mu_t \triangleq \mu_1 + t(\mu_2 - \mu_1) \in \Sigma_p$ satisfies (2.2.8a-b) with $\hat{x}$, and hence $\mu_t \in U(\hat{x})$. By construction, $\mu_t^{j_0} = 0$. Hence, it follows from the strict complementary slackness assumption that $f^{j_0}(\hat{x}) < \psi(\hat{x})$. However, $\mu_1^{j_0} > 0$, and hence, again by strict complementary slackness, $f^{j_0}(\hat{x}) = \psi(\hat{x})$. This contradiction shows that $U(\hat{x})$ is a singleton for each $\hat{x} \in \hat{G}$.

Suppose that $j_1 \in J(\hat{x}')$ but $j_1 \notin J(\hat{x}'')$ for distinct points $\hat{x}'$, $\hat{x}'' \in \hat{G}$. Then $g^{j_1}(A_{j_1}\hat{x}'') < \psi(\hat{x}'')$. Let $\hat{x}_t \triangleq t\hat{x}' + (1-t)\hat{x}''$. Then $\hat{x}_t \in \hat{G}$ for all $t \in [0, 1]$, and, by the convexity of $g^{j_1}(\cdot)$, $g^{j_1}(A_{j_1}\hat{x}_t) < \psi(\hat{x}'') = \psi(\hat{x}_t)$ for all $t \in (0, 1)$. It follows from (i) above that $U(\hat{x}_t) = \{ \mu_t \}$, a singleton, and from (2.2.8b) that $\mu_t^{j_1} = 0$ for all $t \in (0, 1)$. Now, by the Maximum Theorem in [Ber.1], $U(\cdot)$ is an upper semicontinuous, compact-valued set-valued map. Since $U(\hat{x}') = \{ \hat{\mu}' \}$, a singleton, $U(\cdot)$ is continuous at $\hat{x}'$. Hence $\mu_t \to \hat{\mu}'$ as $t \to 1$, which implies that $\hat{\mu}'^{j_1} = 0$. Since this contradicts the assumption that $j_1 \in J(\hat{x}')$, we conclude that (b) holds.

Now we prove (a). Suppose that $\hat{x}', \hat{x}'' \in \hat{G}$. From (b), $g^j(A_j(\hat{x}' + t(\hat{x}'' - \hat{x}')))$ is constant for all $t \in [0, 1]$ and all $j \in \hat{J}$. Since each $g^j(\cdot)$ is strictly convex, we conclude that $A_j(\hat{x}' - \hat{x}'') = 0$ for each $j \in \hat{J}$. Therefore, for all $j \in \hat{J}$, $A_j^T \nabla g^j(A_j\hat{x}') = A_j^T \nabla g^j(A_j\hat{x}'')$ and hence any $\hat{\mu}$ satisfying (2.2.8a-b) with $\hat{x}'$ satisfies (2.2.8a-b) with $\hat{x}''$. This and the fact that $U(\hat{G})$ is a singleton imply (a). □

**Proposition 2.5.2:** *If the functions $g^j(\cdot)$ are uniformly convex, then there exists a neighborhood, $W$, of $\hat{G}$ such that, for all $x \in W$, $\mu^j = 0$ for all $\mu \in U(x)$ and $j \notin J(\hat{x})$.*

**Proof:** First note that the uniform convexity of the functions $g^j(\cdot)$ implies that $\hat{G}$ is nonempty. Since $h(x)$ is the solution of the primal problem (2.2.2), it satisfies the optimality conditions (2.2.8a-b) with the functions $f^j(\cdot)$ replaced by $\phi^j(\cdot \mid x)$. Every $\mu \in U(x)$ satisfies equations (2.2.8a-b) together with $h(x)$, and hence the second of those equations yields

$$\sum_{j \in \mathbf{p}} \mu^j \left[\phi^j(h(x) \mid x) - \psi(x) - \theta(x)\right] = 0 . \tag{2.5.4}$$

We show that $h(x) \to 0$ and $\theta(x) \to 0$ as $x \to \hat{G}$. If not, then there exists a sequence $\{ x_i \}_{i \in \mathbb{N}}$ and $\delta > 0$ such that $x_i \to \hat{G}$ and $\theta(x_i) < -\delta$ for all $i \in \mathbb{N}$. (Note that $\theta(x) = 0$ implies that $h(x) = 0$.) Let $\mathrm{A}^T = [A_1^T, ..., A_p^T]$. Now, since each $g^j(\cdot)$ is uniformly convex, the set $\{ \mathrm{A}x_i \}_{i \in \mathbb{N}}$ is bounded. Hence, there exists a bounded sequence $\{ x'_i \}_{i \in \mathbb{N}}$ such that $\mathrm{A}x'_i = \mathrm{A}x_i$ for all $i \in \mathbb{N}$. Since the sequence $\{ x'_i \}_{i \in \mathbb{N}}$ is bounded, there exists a point $\bar{x}$ and a subsequence $K \subset \mathbb{N}$, such that $x'_i \to \bar{x}$ as $i \to \infty, i \in K$. Since $\lim_{i \to \infty} \psi(x'_i) = \lim_{i \to \infty} \psi(x_i) = \hat{\psi}$, $\psi(\bar{x}) = \hat{\psi}$, and hence $\bar{x} \in \hat{G}$. By Proposition 5.5 in [Pol.4], this implies that $\theta(\bar{x}) = 0$ and $h(\bar{x}) = 0$. Since $\theta(\cdot)$ and $h(\cdot)$ are continuous, therefore, $\theta(x'_i) \to 0$ and $h(x'_i) \to 0$ as $i \to \infty, i \in K$. But $\theta(x'_i) = \theta(x_i) < -\delta$ for all $i \in \mathbb{N}$. This contradiction proves that $h(x) \to 0$ and $\theta(x) \to 0$ as $x \to \hat{G}$.

Therefore, $\phi^j(h(x) \mid x) \to g^j(A_j\hat{x})$ as $x \to \hat{x} \in \hat{G}$ for all $j$, implying that

$$\phi^j(h(x) \mid x) - \psi(x) - \theta(x) < 0 \tag{2.5.5}$$

for every $j \notin J(\hat{x})$ in some neighborhood, $W$, of $\hat{G}$. It follows from (2.5.4) and (2.5.5) that, for all $x \in W$, $\mu^j = 0$ for all $j \notin J(\hat{x})$ for all $\mu \in U(x)$. □

We now proceed to show that Algorithms 2.2.1 and 2.4.1 converge linearly on some problems of the form (2.5.1) which do not satisfy the assumptions of Theorem 2.3.2. Letting $j_1 < ... < j_b$ be the indices comprising $\hat{J}$, with $\hat{J}$ defined as in Proposition 2.5.1, we define $\hat{A}^T \triangleq [A_{j_1}^T, ..., A_{j_b}^T]$. First we will show that the tail of a sequence $\{ x_i \}_{i=0}^{\infty}$ generated either by Algorithm 2.2.1 or by Algorithm 2.4.1 is contained in a translation of the range of $\hat{A}^T$. We will then show that the sequence corresponds to that constructed by the corresponding algorithm on the following restriction of problem (2.5.1) to a translation of the range of $\hat{A}^T$:

$$\min_{y \in \mathbb{R}^a} \psi(\hat{x} + Zy)\,, \tag{2.5.6}$$

where $a \triangleq rank(\hat{A}^T)$, and $Z$ is a matrix, the columns of which form an orthonormal basis for $Range(\hat{A}^T)$. Finally, we will show that the restricted problem (2.5.6) satisfies the assumptions of Theorem 2.3.2. We will use the notation $\sigma^+[X]$ to denote the minimum *positive* eigenvalue of any symmetric, positive semi-definite matrix $X$.

**Theorem 2.5.1:** *Suppose that*

(i) *the functions $g^j(\cdot)$ are twice continuously differentiable,*

(ii) *there exist constants $0 < l' \le L'$ such that, for all $j \in \mathbf{p}$,*

$$l' \|h\|^2 < \langle h, G^j(z)h \rangle \le L' \|h\|^2\,, \quad \forall\, h, z \in \mathbb{R}^{l_j}\,, \tag{2.5.7a}$$

(iii) *strict complementary slackness holds at $(\hat{x}, \hat{\mu})$ for all $\hat{\mu} \in U(\hat{x})$ and all $\hat{x} \in \hat{G}$,*[6]

(iv) *Let $l \le l'$ and $L \ge L'$ be such that*

$$l\, \sigma^+[\sum_{j \in \mathbf{p}} \hat{\mu}^j A_j^T A_j] < \gamma < L \max_{j \in \mathbf{p}} \|Z^T A_j^T A_j Z\|\,, \tag{2.5.7b}$$

*where $\hat{\mu}$ is the sole member of $U(\hat{G})$. Under these assumptions:*

(a) *For any $\hat{x} \in \hat{G}$,*

$$\limsup_{\substack{x \to \hat{x} \\ x \in \hat{x} + Range(Z) \\ x \ne \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \le 1 - \frac{l}{L} \frac{\sigma^+[\sum_{j \in \mathbf{p}} \hat{\mu}^j A_j^T A_j]}{\max_{k \in \mathbf{p}} \|Z^T A_k^T A_k Z\|}\,. \tag{2.5.7c}$$

---

[6]The assumption of strict complementary slackness is necessary only if the matrices $A_j$ have different null spaces. For example, the linear convergence result holds without this assumption if the matrices $A_j$ are identical.

(b) *If Algorithm 2.2.1 constructs a sequence* $\{x_i\}_{i=0}^{\infty}$ *in solving problem (2.5.1), then the sequence converges to* $\hat{x}$ *for some* $\hat{x} \in \hat{G}$, *and either* $x_i = \hat{x}$ *for all large* $i$ *or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \frac{l}{L} \frac{\sigma^+[\sum_{j \in \mathbf{p}} \hat{\mu}^j A_j^T A_j]}{\max_{k \in \mathbf{p}} \| Z^T A_k^T A_k Z \|} . \tag{2.5.7d}$$

**Proof:** (a) To prove this part, we will (i) show that it is sufficient to consider the restriction of problem (2.5.1) to an affine space, (ii) verify that Hypotheses 2.3.1 and 2.3.2 hold for the restricted problem, and (iii) apply Lemma 2.3.4.

First note that the uniform convexity of the functions $g^j(\cdot)$ implies that $\hat{G}$ is nonempty. It follows from Proposition 2.5.2, that there exists a neighborhood $W \supset \hat{G}$ such that $\mu^j = 0$ for all $j \notin \hat{J}$ and $\mu \in U(W)$, and from (2.2.7), that $h(x) = \sum_{j \in \underline{p}} \mu^j A_j^T \nabla g^j(A_j x_i)$ for any $\mu \in U(x)$. Hence, for all $x \in W$,

$$h(x) \in Range(\hat{A}^T) = Range(Z) , \tag{2.5.8}$$

by the definition of $Z$ above. Let us fix $\hat{x} \in \hat{G}$, and suppose that $x \in W$. If $x \in \hat{x} + Range(Z)$, then $x + \lambda h(x) \in \hat{x} + Range(Z)$. This suggests that we consider the restriction of problem (2.5.1) to $Range(Z)$, viz.,

$$\min_{y \in \mathbb{R}^a} \psi_r(y) , \tag{2.5.9a}$$

where

$$\psi_r(y) \triangleq \psi(\hat{x} + Zy) , \tag{2.5.9b}$$

so that $\psi_r(y) = \max_{j \in \mathbf{p}} f_r^j(y)$, with $f_r^j(y) \triangleq f^j(\hat{x} + Zy)$. The search direction $d(y)$ constructed by Algorithm 2.2.1 at a point $y \in \mathbb{R}^a$ for problem (2.5.9) is given by

$$d(y) \triangleq \arg \min_{d \in \mathbb{R}^a} \max_{j \in \mathbf{p}} g^j(A_j(\hat{x} + Zy)) + \langle Z^T A_j^T \nabla g^j(A_j(\hat{x} + Zy)), d \rangle + \tfrac{1}{2}\gamma \| d \|^2$$

$$= \arg \min_{d \in \mathbb{R}^a} \max_{j \in \underline{p}} g^j(A_j(\hat{x} + Zy)) + \langle A_j^T \nabla g^j(A_j(\hat{x} + Zy)), Zd \rangle + \tfrac{1}{2}\gamma \| Zd \|^2$$

$$= \arg \min_{d \in \mathbb{R}^a} \max_{j \in \underline{p}} \phi^j(Zd \mid \hat{x} + Zy) \,, \tag{2.5.10}$$

since $Z^T Z = I_a$ and $\phi^j(h \mid x) \triangleq g^j(A_j x) + \langle A_j^T \nabla g^j(x), h \rangle + \tfrac{1}{2}\gamma \| h \|^2$. By (2.5.8), $h(\hat{x} + Zy) \in Range(Z)$. Hence, referring to (2.5.10), we see that

$$h(\hat{x} + Zy) = \arg \min_{h \in Range(Z)} \max_{j \in \mathbf{p}} \phi^j(h \mid x + Zy)$$

$$= Zd(y) \,. \tag{2.5.11}$$

Also, for $y$ such that $\hat{x} + Zy \in W$,

$$\arg \min_{\lambda \in \mathbb{R}} \psi(\hat{x} + Z(y + \lambda d(y))) = \arg \min_{\lambda \in \mathbb{R}} \psi(\hat{x} + Zy + \lambda Zd(y))$$

$$= \arg \min_{\lambda \in \mathbb{R}} \psi(\hat{x} + Zy + \lambda h(\hat{x} + Zy)) \,. \tag{2.5.12a}$$

We conclude from (2.5.11) and (2.5.12a) that

$$\limsup_{\substack{x \to \hat{x} \\ x \in \hat{x} + Range(Z) \\ x \neq \hat{x}}} \min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x)) - \hat{\psi}}{\psi(x) - \hat{\psi}} = \limsup_{\substack{y \to 0 \\ y \neq 0}} \min_{\lambda \in \mathbb{R}} \frac{\psi(\hat{x} + Z(y + \lambda d(y)) - \hat{\psi}}{\psi(\hat{x} + Zy) - \hat{\psi}} \,. \tag{2.5.12b}$$

Hence, provided problem (2.5.9) satisfies Hypothesis 2.3.1 and Hypothesis 2.3.2, we can establish (2.5.7c) by applying Lemma 2.3.4 to problem (2.5.9) to obtain an upper bound on the right hand side of (2.5.12b).

We now verify that Hypothesis 2.3.1 is satisfied by problem (2.5.9). (i) The functions $g^j(A_j(\hat{x} + Zy))$ are twice continuously differentiable in $y$ by assumption (i) of this theorem.

(ii) Let $S_r \triangleq \{ y \in \mathbb{R}^a \mid \psi(\hat{x} + Zy) \leq T \}$, with $T > \psi(\hat{G})$. Since the functions $g^j(\cdot)$ are uniformly convex by assumption (ii) of this theorem, the set $\hat{A}\, ZS_r$ is bounded. Since

$Range(Z) = Range(\hat{A}^T)$ and $Null(Z) = \{0\}$, $Null(\hat{A}Z) = \{0\}$.[7] Hence $S_r$ is bounded, and, since closed, it is compact.

The point $y = 0$ satisfies the necessary conditions for optimality (2.2.8a-b) for problem (2.5.9), since $\hat{x}$ satisfies the necessary conditions for problem (2.1.1a). Showing that 0 is the *only* point in $S_r$ which satisfies the necessary conditions (2.2.8a-b) for problem (2.5.9) is slightly involved. Let $\hat{G}_r$ denote the minimizing set for problem (2.5.9). Now suppose that there is a $y' \in \hat{G}_r$ such that $\hat{x} + Zy' \notin \hat{G}$. Then $\psi(\hat{x} + Zy') > \psi(\hat{x} + Z0)$, which contradicts the assumption that $y' \in \hat{G}_r$. Therefore, $\hat{x} + Z\hat{G}_r \subset \hat{G}$.

Now consider the set of multipliers,

$$U_r(y) \triangleq \left\{ \mu \in \Sigma_p \;\middle|\; \begin{array}{l} \sum_{j \in \mathbf{p}} \mu^j \nabla f_r^{\,j}(y) = 0 \\ \sum_{j \in \mathbf{p}} \mu^j (f_r^{\,j}(y) - \psi_r(y)) = 0 \end{array} \right\}, \tag{2.5.13}$$

which, together with $y$, satisfy the the optimality conditions (2.2.8a-b), when the functions functions $f^{\,j}(\cdot)$ are replaced by the functions $f_r^{\,j}(\cdot)$. We show that $U_r(\hat{G}_r) = \{\hat{\mu}\}$, where $\hat{\mu}$ is as defined in Proposition 2.5.1(a). For any $\bar{y} \in \hat{G}_r$, we have $\hat{x} + Z\bar{y} \in \hat{G}$ by the previous paragraph, and hence $f_r^{\,j}(\bar{y}) < \psi_r(\bar{y})$ for all $j \notin \hat{J}$. Consequently, $\bar{\mu}^j = 0$ for all $j \notin \hat{J}$ and for any $\bar{\mu} \in U_r(\bar{y})$. Therefore,

$$\sum_{j \in \mathbf{p}} \bar{\mu}^j A_j^T \nabla g^j(A_j(\hat{x} + Zy)) \in Range(\hat{A}^T) = Range(Z)\,. \tag{2.5.13}$$

For any $\bar{\mu} \in U_r(\bar{y})$, it follows from (2.5.13) that $\sum_{j \in \mathbf{p}} \bar{\mu}^j Z^T A_j^T \nabla g^j(A_j(\hat{x} + Z\bar{y})) = 0$, hence, making use of (2.5.13), we conclude that $\sum_{j \in \mathbf{p}} \bar{\mu}^j A_j^T \nabla g^j(A_j(\hat{x} + Z\bar{y})) = 0$. Hence, $\bar{\mu}$ together with $\hat{x} + Z\bar{y}$ satisfy the necessary conditions (2.2.8a-b) for the original problem (2.5.1). Thus,

[7] Suppose that $\hat{A}Zy = 0$. Since $Null(\hat{A}) \cap Range(\hat{A}^T) = \{0\}$, $Zy = 0$. But then $Null(Z) = \{0\}$ implies that $y = 0$.

$U_r(\hat{G}_r) \subset U(\hat{x}+Z\hat{G}_r)$.

Since, $\hat{x}+Z\hat{G}_r \subset \hat{G}$ by the previous paragraph, this implies that $U_r(\hat{G}_r) = \{ \hat{\mu} \}$, where $\hat{\mu}$ is the only member of $U(\hat{G})$.

Suppose that $\bar{y} \in S_r$ satisfies the optimality conditions (2.2.8a-b) for problem (2.5.9). Since $\psi_r(y)$ is convex in $y$, these necessary conditions are sufficient for optimality, and, furthermore, the entire line segment between $\bar{y}$ and 0, $[\bar{y}, 0]$, lies in $\hat{G}_r$. Since $U_r([\bar{y}, 0]) = \{ \hat{\mu} \}$ and $\hat{\mu}^j > 0$ for all $j \in \hat{J}$, $g^j(A_j(\hat{x} + Zy)) = \psi(\hat{x} + Zy) = \hat{\psi}$ for all $y \in [\bar{y}, 0]$ and all $j \in \hat{J}$. Because the functions $g^j(\cdot)$ are strictly convex, it follows that $A_jZ\bar{y} = A_jZ0 = 0$ for all $j \in \hat{J}$, and hence that $\bar{y} \in Null(\hat{A}Z)$. As mentioned above, $Null(\hat{A}Z) = \{ 0 \}$, implying that $\bar{y} = 0$. Therefore, the necessary conditions are satisfied at the unique point $0 \in S_r$ and Hypothesis 2.3.1(ii) holds.

(iii) It follows from assumption (ii) of this theorem that for all $y \in \mathbb{R}^a$, $\|F_r^j(y)\| \le L \max_{j \in \underline{p}} \|Z^T A_j^T A_j Z\|$, where $F_r^j(\cdot)$ denotes the second derivative matrix of $f_r^j(\cdot)$.

Now we verify that Hypothesis 2.3.2 holds. Letting $\underline{\sigma}[X]$ denote the minimum eigenvalue of any real symmetric matrix $X$, we obtain that

$$\begin{aligned}\underline{\sigma}\left[\sum_{j \in \mathbf{p}} \hat{\mu}^j F_r^j(0)\right] &= \underline{\sigma}\left[\sum_{j \in \mathbf{p}} \hat{\mu}^j Z^T A_j^T G^j(A_j(\hat{x} + Z0))A_j Z\right] \\ &\ge l\, \underline{\sigma}\left[\sum_{j \in \mathbf{p}} \hat{\mu}^j Z^T A_j^T A_j Z\right] \\ &= l\, \sigma^+\left[\sum_{j \in \mathbf{p}} \hat{\mu}^j A_j^T A_j\right], \end{aligned} \tag{2.5.14}$$

since the columns of $Z$ span $Range(\hat{A}^T) = Null(\hat{A})^\perp$ and $Null(\sum_{j \in \mathbf{p}} \hat{\mu}^j A_j^T A_j) = Null(\hat{A})$.

Letting the left-hand and right-hand sides of the double inequality (2.5.7b) correspond to $m$ and $M$ respectively, we can apply Lemma 2.3.4 to problem (2.5.9) to obtain

$$\limsup_{\substack{y \to 0 \\ y \neq 0}} \min_{\lambda \in \mathbb{R}} \frac{\psi(\hat{x} + Z(y + \lambda d(y))) - \hat{\psi}}{\psi(\hat{x} + Zy) - \hat{\psi}}$$

$$\leq 1 - \frac{l}{L} \frac{\sigma^+[\sum_{j \in p} \hat{\mu}^j A_j^T A_j]}{\max_{k \in p} \| Z^T A_k^T A_k Z \|}, \tag{2.5.15}$$

which, combined with (2.5.12b), gives part (a).

To show (b), we first show that $x_i \to \hat{x}$ as $i \to \infty$ for some $\hat{x} \in \hat{G}$, and then apply part (a) of this theorem. Let $\bar{A}^T \triangleq [A_1^T, \ldots, A_p^T]$. From (2.2.7), every $h_i$, constructed by Algorithm 2.2.1, is of the form $\sum_{j \in p} A_j^T z_j$, with $z_j \in \mathbb{R}^{l_j}$. Thus, the sequence $\{ x_i \}_{i=0}^{\infty}$ is contained in the closed and convex set $Q \triangleq \{ x_0 + Range(\bar{A}^T) \} \cap \{ x \in \mathbb{R}^n \mid \psi(x) \leq \psi(x_0) \}$. Suppose that $Q$ is unbounded. Then, since $Q$ is convex, there exists a nonzero $u \in Range(\bar{A}^T)$, such that, with $x_t \triangleq x_0 + tu$, $\psi(x_t) \leq \psi(x_0)$ for all $t \geq 0$. If $A_{j_0} u \neq 0$ for some $j_0 \in \underline{p}$, then the uniform convexity of $g^j(\cdot)$, which follows from assumption (ii) of this theorem, implies that $\lim_{t \to \infty} \psi(x_t) = +\infty$. Since this contradicts our assumption that $\psi(x_t) \leq \psi(x_0)$, we must have that $\bar{A}u = 0$. Hence $u \in Range(\bar{A}^T) \cap Null(\bar{A}) = \{ 0 \}$, which contradicts the assumption that $u \neq 0$. Therefore, the set $Q$ is bounded, and hence compact. Consequently, the sequence $\{ x_i \}_{i=0}^{\infty}$ must have an accumulation point, $\hat{x}$. From Corollary 5.1 and Proposition 5.5 in [Pol.4], any accumulation point $\hat{x}'$ of a sequence generated by Algorithm 2.2.1 must satisfy the first-order necessary conditions for optimality (2.2.8a-b). Since $\psi(\cdot)$ is convex, this implies that $\hat{x} \in \hat{G}$. Since $Q$ is compact, it follows that $x_i \to \hat{G}$ as $i \to \infty$.

Since $x_i \to \hat{G}$ as $i \to \infty$, there exists $i_0 \in \mathbb{N}$ such that $x_i \in W$ for all $i > i_0$. Hence, $\{ x_i \}_{i=i_0}^{\infty} \subset x_{i_0} + Range(\hat{A}^T) = x_{i_0} + Range(Z)$. Since the functions $g^j(\cdot)$ are uniformly convex, $\hat{G} \cap (x_{i_0} + Range(\hat{A}^T))$ is a singleton. Hence, the sequence $\{ x_i \}_{i=0}^{\infty}$ converges to $\hat{x} = \hat{G} \cap (x_{i_0} + Range(\hat{A}^T))$.

Inequality (2.5.7d) follows directly from convergence of the sequence to $\hat{x}$ and part (a). □

The corresponding result for Algorithm 2.4.1 can be obtained by following the steps used in Section 4 and above.

**Theorem 2.5.2:** *Suppose that the assumptions of Theorem 2.5.1 hold. If Algorithm 2.4.1 constructs a sequence $\{x_i\}_{i=0}^{\infty}$ in solving problem (2.5.1), then, the sequence converges to $\hat{x}$ for some $\hat{x} \in \hat{G}$, and either $x_i = \hat{x}$ for all large $i$ or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \le 1 - \alpha\beta \frac{l}{L} \frac{\sigma^{+}[\sum_{j \in \mathbf{p}} \hat{\mu}^j A_j^T A_j]}{\max\limits_{k \in \mathbf{p}} \| Z^T A_k^T A_k Z \|} . \tag{2.5.17}$$

## 2.6 CONCLUSIONS

We have shown that sequences $\{ \psi(x_i) \}_{i=0}^{\infty}$ generated by two PPP minimax algorithms converge linearly to the minimum value under weaker conditions than those assumed in previous analyses of the rate of convergence of PPP algorithms [Psh.1, Dau.1, Pol.2-3]. Although composite minimax problems which have nonunique, nonisolated minimizers do not satisfy the second-order sufficiency conditions Hypothesis 2.3.2, we were able to show that these PPP algorithms converge linearly on such problems provided that strong convexity and strict complementary slackness conditions are satisfied.

PPP algorithms can be generalized in a straightforward way to solve *semi-infinite* composite minimax problems [Pol.4] which arise in control system design,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \mathbf{p}} \max_{y_j \in Y_j} \phi^j(A_j x, y_j) , \tag{2.6.1}$$

where the sets $Y_j \subset \mathbb{R}^{s_j}$ are compact, and the functions $\phi^j : \mathbb{R}^{l_j} \times \mathbb{R}^{s_j} \to \mathbb{R}$, $j \in \mathbf{p}$ and $\nabla_1 \phi^j(\cdot,\cdot)$ are continuous. As before, each $A_j$ is an $l_j \times n$ matrix. Under assumptions analogous to those of Theorem 2.5.1, it can be shown that the semi-infinite versions of the PPP algorithms, considered in this chapter, also converge linearly (see [Pol.2]).

## 2.7 APPENDIX

The following extension to the von Neumann Minimax Theorem is quoted from [Pol.4].

**Theorem 2.7.1 [Pol.4]:** Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuous function such that $f(x, y)$ is convex in $x$ and concave in $y$ and let $Y$ be a compact, convex set in $\mathbb{R}^m$. Suppose also that $f(x, y) \to \infty$ as $\|x\| \to \infty$, uniformly in $y \in Y$. Then,

$$\min_{x \in \mathbb{R}^n} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in \mathbb{R}^n} f(x, y). \tag{2.7.1a}$$

□

**Proof of Lemma 2.3.1:** For any $y \in \mathbb{R}^n$, $y = Py + P^{\perp}y$. Hence, since $R(\cdot, \cdot)$ is continuous and $R(0, \hat{\mu})$ is negative definite on $B$ for any $\hat{\mu} \in U(\hat{x})$ by Hypothesis 2.3.2,

$$\begin{aligned} \langle y, R(y, \mu)y \rangle &= \langle Py + P^{\perp}y, R(y, \mu)(Py + P^{\perp}y) \rangle \\ &= \langle Py, R(y, \mu)Py \rangle + \langle P^{\perp}y, R(y, \mu)(P^{\perp}y + 2Py) \rangle \\ &\leq \langle P^{\perp}y, R(y, \mu)(P^{\perp}y + 2Py) \rangle, \end{aligned} \tag{2.7.1b}$$

for $\mu$ near $U(\hat{x})$ and $y$ small. Using the Schwarz inequality and the fact that $\|P^{\perp}y + 2Py\| \leq 2\|y\|$,

$$\begin{aligned} \langle y, R(y, \mu)y \rangle &\leq \|R(y, \mu)\| \, \|P^{\perp}y\| \, \|P^{pp}dy + 2Py\| \\ &\leq 2\|R(y, \mu)\| \, \|P^{\perp}y\| \, \|y\| \\ &\leq 3 \max_{\mu \in U(\hat{x})} \|R(0, \mu)\| \, \|P^{\perp}y\| \, \|y\|, \end{aligned} \tag{2.7.1c}$$

for $\mu$ near $U(\hat{x})$ and $y$ small, since $\|R(\cdot, \cdot)\|$ is continuous. □

**Proof of Lemma 2.3.2:** Using Taylor's Theorem, we obtain that for any $x \in \mathbb{R}^n$,

$$\psi(x) - \psi(\hat{x}) \geq \max_{j \in J(\hat{x})} f^j(x) - \psi(\hat{x})$$

$$= \max_{j \in J(\hat{x})} f^j(\hat{x}) + \langle \nabla f^j(\hat{x}), x - \hat{x} \rangle$$

$$+ \langle x - \hat{x}\,,\, \left[\int_0^1 (1-s)F^j(\hat{x} + s(x-\hat{x}))ds\right](x-\hat{x})\,\rangle - \psi(\hat{x})\,. \tag{2.7.2}$$

Since $f^j(\hat{x}) = \psi(\hat{x})$ for all $j \in J(\hat{x})$, it follows from Hypothesis 2.3.1(iii) that

$$\psi(x) - \psi(\hat{x}) \geq \max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), x - \hat{x}\,\rangle + \langle x - \hat{x}\,,\, \left[\int_0^1 (1-s)F^j(\hat{x} + s(x-\hat{x}))ds\right](x-\hat{x})\,\rangle$$

$$\geq \max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), x - \hat{x}\,\rangle - M\|x - \hat{x}\|^2\,. \tag{2.7.3}$$

Since $\langle \nabla f^j(\hat{x}), P(x-\hat{x})\rangle = 0$ for all $j \in J(\hat{x})$ and $h = Ph + P^{\perp}h$ for any $h \in \mathbb{R}^n$,

$$\max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), x - \hat{x}\,\rangle = \max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}),\, P^{\perp}(x-\hat{x})\rangle. \tag{2.7.4}$$

We will to show that there exists an $\eta > 0$ such that

$$\max_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), P^{\perp}(x-\hat{x})\rangle \geq \eta\|P^{\perp}(x-\hat{x})\|\,. \tag{2.7.5}$$

If not, then there exists a nonzero $\bar{u} \in B^{\perp}$ such that $\max\limits_{j \in J(\hat{x})} \langle \nabla f^j(\hat{x}), \bar{u}\,\rangle \leq 0$. Since $U(\hat{x})$ is convex, there exists a $\hat{\mu} \in U(\hat{x})$ such that $\hat{\mu}^j > 0$ for all $j \in J(\hat{x})$. By (2.2.8b), $\hat{\mu}^j = 0$ for $j \notin J(\hat{x})$. Therefore, by (2.2.8a),

$$\sum_{j \in J(\hat{x})} \hat{\mu}^j \langle \nabla f^j(\hat{x}), \bar{u}\,\rangle = \langle \sum_{j \in \mathbf{p}} \hat{\mu}^j \nabla f^j(\hat{x}), \bar{u}\,\rangle = \langle 0, \bar{u}\,\rangle = 0\,, \tag{2.7.6}$$

Equation (2.7.6) states that a convex combination of the nonpositive numbers, $\{\,\langle \nabla f^j(\hat{x}), \bar{u}\,\rangle\}_{j \in J(\hat{x})}$, with nonzero coefficients, $\{\hat{\mu}^j\}_{j \in J(\hat{x})}$, is zero. Hence $\langle \nabla f^j(\hat{x}), \bar{u}\,\rangle = 0$ for all $j \in J(\hat{x})$. But then $\bar{u} \in B \cap B^{\perp} = \{0\}$, contradicting the assumption that $\bar{u} \neq 0$. Hence, let $\eta > 0$ be such that (2.7.5) holds.

Substituting (2.7.4) into (2.7.5) and (2.7.5) into (2.7.3) yields

$$\psi(x) - \psi(\hat{x}) \geq \eta\|P^{\perp}(x-\hat{x})\| - M\|x - \hat{x}\|^2\,, \tag{2.7.7}$$

for $x$ in some neighborhood of $\hat{x}$.

Now we derive *another* lower bound on $\psi(x) - \psi(\hat{x})$. For any $\hat{\mu} \in U(\hat{x})$, using Taylor's Theorem and the fact that $\sum_{j \in \underline{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0$,

$$\psi(x) - \psi(\hat{x}) \geq \sum_{j \in \underline{p}} \hat{\mu}^j f^j(x) - \psi(\hat{x})$$

$$= \langle x - \hat{x}, \left[\int_0^1 (1-s) \sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x} + s(x - \hat{x}))ds\right](x - \hat{x})\rangle$$

$$= \langle P(x - \hat{x}), \left[\int_0^1 (1-s) \sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x} + s(x - \hat{x}))ds\right] P(x - \hat{x})\rangle$$

$$+ \langle P^{\perp}(x - \hat{x}), \left[\int_0^1 (1-s) \sum_{j \in \underline{p}} \hat{\mu}^j F^j(\hat{x} + s(x - \hat{x}))ds\right](2P(x - \hat{x}) + P^{\perp}(x - \hat{x}))\rangle. \quad (2.7.8)$$

Making use of Hypothesis 2.3.1(iii) and Hypothesis 2.3.2, (2.7.8) leads to

$$\psi(x) - \psi(\hat{x}) \geq \tfrac{1}{2}m \|P(x - \hat{x})\|^2 - \tfrac{1}{2}M \|P^{\perp}(x - \hat{x})\| \, \|2P(x - \hat{x}) + P^{\perp}(x - \hat{x})\|$$

$$\geq \tfrac{1}{2}m \|P(x - \hat{x})\|^2 - M \|P^{\perp}(x - \hat{x})\| \, \|x - \hat{x}\|, \quad (2.7.9)$$

for $x$ in a neighborhood of $\hat{x}$.

Combining (2.7.7) with (2.7.9) and dividing by $\|P^{\perp}(x - \hat{x})\| \|x - \hat{x}\|$ yields

$$\frac{\psi(x) - \psi(\hat{x})}{\|P^{\perp}(x - \hat{x})\| \|x - \hat{x}\|} \geq \max\left\{ \frac{\tfrac{1}{2}m \|P(x - \hat{x})\|^2}{\|P^{\perp}(x - \hat{x})\| \|x - \hat{x}\|} - M, \ \frac{\eta}{\|x - \hat{x}\|} - M \frac{\|x - \hat{x}\|}{\|P^{\perp}(x - \hat{x})\|} \right\} \quad (2.7.10)$$

for $x$ in a neighborhood of $\hat{x}$. Using the fact that $\|x\| \leq \|P^{\perp}x\| + \|Px\|$, and defining $r(x) = \|P^{\perp}(x - \hat{x})\| / \|P(x - \hat{x})\|$, we obtain that

$$\frac{\psi(x) - \psi(\hat{x})}{\|P^{\perp}(x - \hat{x})\| \|x - \hat{x}\|} \geq \max\left\{ \frac{\tfrac{1}{2}m}{r(x)^2 + r(x)} - M, \ \frac{\eta}{\|x - \hat{x}\|} - M(\frac{1}{r(x)} + 1) \right\}, \quad (2.7.11)$$

We use (2.7.11) to show that

$$\liminf_{x \to \hat{x}} \frac{\psi(x) - \psi(\hat{x})}{\|P^{\perp}(x - \hat{x})\| \, \|x - \hat{x}\|} = \infty \,, \tag{2.7.12}$$

which is equivalent to (2.3.6). Thus, given any integer $k > 0$, there exists a real number $r > 0$ such that the first term in the max in (2.7.11) is greater than $k$ if $r(x) \le r$. For $x$ such that $r(x) > r$, the second term term in the max is greater than $\eta / \|x - \hat{x}\| - M(1/r + 1)$. Hence, there exists a neighborhood, $W_k$, of $\hat{x}$ such that the max in (2.7.1171) exceeds $k$ for all $x \in W_k$, and, therefore, (2.3.6) holds. □

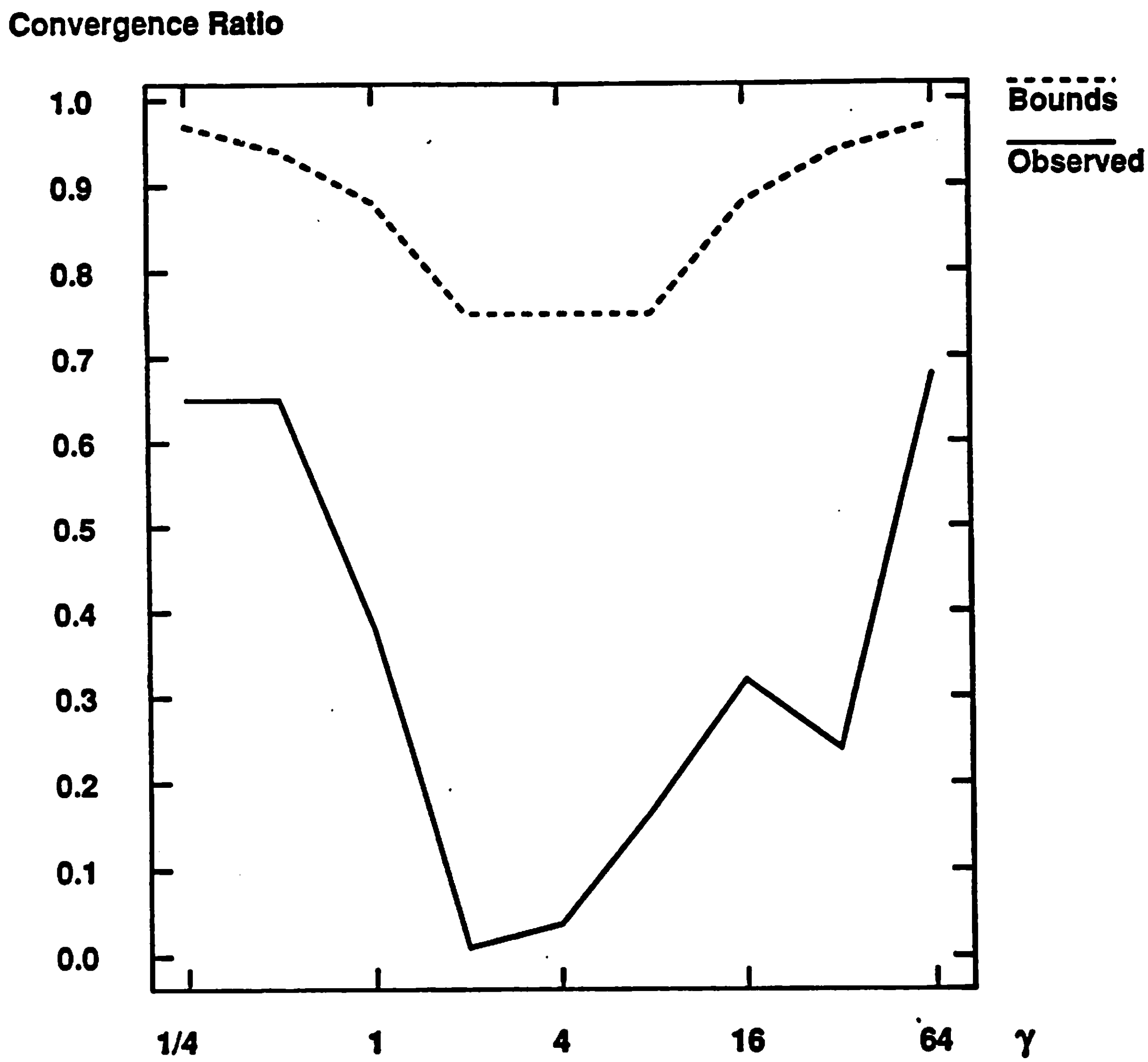


Figure 2.1: Convergence Ratio vs. Gamma

# CHAPTER 3

# A VARIABLE METRIC TECHNIQUE FOR A CLASS OF COMPOSITE MINIMAX PROBLEMS

## 3.1 INTRODUCTION

The term *variable metric method* is commonly used to describe a number of algorithms, such as those discussed in [Den.1, Byr.1], which emulate the behavior of the Newton method. The term can be applied, however, to *any* optimization algorithm which uses a sequence of linear transformations of the variables to convert the original optimization problem into a sequence of equivalent problems, to each of which it applies one iteration of a "standard" method and uses a transformed result as a starting point for the iteration on the next problem. Variable metric methods are effective when there is a linear transformation which transforms an optimization problem into a better conditioned form. Since the desired transformation is not known a priori, an approximating sequence of transformations is constructed as the computation progresses.

In the past, variable metric techniques were used as a means of improving the conditioning of an optimization problem with respect with respect to a particular algorithm. For example, the Armijo-Newton method [Gol.1] can be viewed as a combination of a sequence of transformations with the Armijo gradient method. Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{3.1.1a}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex and twice Lipschitz continuously differentiable. Given an estimate $x_i$ of the solution $\hat{x}$, at iteration $i$, the Armijo-Newton method uses the transformation $x = F(x_i)^{-1/2} y$ to construct the equivalent problem

$$\min_{y \in \mathbb{R}^n} f(F(x_i)^{-1/2} y), \tag{3.1.1b}$$

to which it applies one iteration of the Armijo gradient method. Thus, (i) it computes the search direction[1] $d_i = -\nabla_y f(F(x_i)^{-1/2} y_i)$, in the new coordinates, (ii) then it transforms this search direction back to the original space by the formula $h_i = F(x_i)^{-1/2} d_i = -F(x_i)^{-1} \nabla f(x_i)$, and (iii) computes the step size $\lambda_i$, which is unity near the solution, using a suitably transformed Armijo step size rule (see [Gol.1]). Setting $x_{i+1} = x_i + \lambda_i h_i$ completes the construction of the next iterate. As is well known, the result is a quadratically convergent algorithm. Similarly, it should be obvious that variable metric methods such as those in [Dix.1] can be viewed as combinations of sequential linear transformation techniques with the method of steepest descent, which uses an exact line search.

The above discussion shows that certain sequential linear transformations are effective in conjunction with two gradient methods. Referring to [Nem.1], we see that sequential linear transformations can be effective with *any* first-order method. It is shown in [Nem.1] that, for problem (3.1.1a), the number of iterations required by any first-order method to reduce an initial cost-error, $f(x_0) - \min_{x \in \mathbb{R}^n} f(x)$, by a factor $\kappa \in (0, 1)$ (for $x_0$ near a solution $\hat{x}$) is bounded from below by $O(\sqrt{K} \log(1/\kappa))$, where $K$ is the condition number of the Hessian $F(\hat{x})$. For problem (3.1.1a), the domain transformation $x = F(\hat{x})^{-1/2} y$ produces the equivalent problem $\min_{y \in \mathbb{R}^n} f(F(\hat{x})^{-1/2} y)$, which has an associated condition number of 1. Hence the bound on the required number of iterations is reduced from $O(\sqrt{K} \log(1/\kappa))$ to $O(\log(1/\kappa))$. For example, the method of steepest descent, which converges only linearly on problem (3.1.1a), converges superlinearly on the transformed problem [Lue.1]. Since the point $\hat{x}$ is not known a priori, a variable metric method attempting to implement this reconditioning must use a sequence of linear transformations approximating $x = F(\hat{x})^{-1/2} y$.

[1] It is interesting to observe that the Newton search direction $h_i$ is the solution of the problem

$$\min_{h \in \mathbb{R}^n} \langle \nabla f(x_i), h \rangle + \tfrac{1}{2} \|h\|^2_{F(x_i)},$$

which has the form of the Armijo gradient search direction finding problem, except that the norm (corresponding to $F(x_i) = I$) is replaced by a *new metric* at each iteration. This fact influenced the naming of the *variable metric* methods which emulate Newton's method.

Experience with solving feedback compensator design problems [Pol.5], as well as optimal control problems involving flexible structures [Bak.1], has shown that they can be very badly conditioned with respect to first-order minimax algorithms, such as those described in [Pol.4]. These problems have the form of a *composite minimax problem,*

$$\min_{x \in R^n} \max_{j \in \mathbf{p}} g^{j}(A_j x) \,, \tag{3.1.2}$$

where each $A_j$ is an $l_j \times n$ matrix, each function $g^{j} : \mathbb{R}^{l_j} \to \mathbb{R}$ is continuously differentiable, and $\mathbf{p}$ denotes the set of integers $\{ 1, ..., p \}$. Since the outer functions, $g^{j}(\cdot)$, encountered in control system design are usually very well conditioned, the ill-conditioning appears to be caused by the matrices $A_j$.

In this chapter, we present a sequential linear transformation technique which is intended to mitigate the ill-conditioning caused by the matrices $A_j$. The technique is similar to one used implicitly by Han in [Han.1]. Our technique was inspired by the observation that when all the functions $g^{j}(\cdot)$ in (3.1.2) are convex, any solution $\hat{x}$ to (3.1.2) is an unconstrained minimizer of the corresponding Lagrangian, $l(x, \hat{\mu}) \triangleq \sum_{j \in \underline{p}} \hat{\mu}^{j} g^{j}(A_j x)$, where the $\hat{\mu}^{j}$ are optimal multipliers. Although the Hessian of this Lagrangian is usually singular, a restriction to a suitable subspace can be used to recondition the problem $\min_{x \in \mathbb{R}^n} l(x, \hat{\mu})$. Since, in many engineering applications, only the matrices $A_j$ cause ill-conditioning and since second order derivatives of the $g^{j}(\cdot)$ can be very costly to compute, we replace the Hessians of the $g^{j}(\cdot)$ by identity matrices and use linear transformations to improve the conditioning of approximations to the matrix $\sum_{j \in \mathbf{p}} \hat{\mu}_j A_j^T A_j$. The resulting sequential linear transformation technique can be used in conjunction with any first-order minimax algorithm which produces estimates of the optimal multipliers. Our variable metric technique is developed in Section 2. In Sections 3 and 4, we present theoretical results which show that our variable metric technique can improve the speed of convergence of the Pshenichnyi method [Psh.1]. (Han proves only a global convergence of his related algorithm [Han.1].) In Section 5, we present numerical experiments which show that our variable metric technique reconditions problems with respect to both the Pshenichnyi method and a new interior point method [Pol.6].

## 3.2 DEVELOPMENT OF THE VARIABLE METRIC TECHNIQUE

We begin by providing a heuristic rationale for our method. Consider the *general* minimax problem,

$$\min_{x \in \mathbb{R}^n} \max_{j \in \mathbf{p}} f^j(x), \tag{3.2.1}$$

where the functions $f^j(\cdot)$ are twice continuously differentiable. We will denote the standard unit simplex in $\mathbb{R}^n$ by $\Sigma_p \triangleq \{ \mu \in \mathbb{R}^p \mid \sum_{j \in \mathbf{p}} \mu^j = 1 \ , \ \mu \geq 0 \}$, and the second derivative matrix of $f^j(\cdot)$ by $F^j(\cdot)$. We can associate with problem (3.2.1) the Lagrangian $l : \mathbb{R}^n \times \Sigma_p \to \mathbb{R}$, defined by

$$l(x, \mu) = \sum_{j \in \mathbf{p}} \mu^j f^j(x) . \tag{3.2.2}$$

We recall the following result.

**Theorem 3.2.1:** **[Cla.1, Dem.3, Joh.1, Pol.4]** *If* $\hat{x}$ is a solution to (3.2.1), then there exists a $\hat{\mu} \in \Sigma_p$ such that

$$\nabla_x l(\hat{x}, \hat{\mu}) = \sum_{j \in \mathbf{p}} \hat{\mu}^j \nabla f^j(\hat{x}) = 0, \tag{3.2.3a}$$

$$\sum_{j \in \mathbf{p}} \hat{\mu}^j [f^j(\hat{x}) - \max_{j \in \mathbf{p}} f^j(\hat{x})] = 0 . \tag{3.2.3b}$$

Now suppose that the functions $f^j(\cdot)$ are strictly convex. Then it follows from (3.2.3a) and the convexity of $l(\cdot, \hat{\mu})$ that, if $\hat{x}$ is a solution to (3.2.1), then it must also be a minimizer of the function $\phi_{\hat{\mu}}(\cdot) \triangleq l(\cdot, \hat{\mu})$. Now, as we have seen in the Section 1, the conditioning of the problem $\min_{x \in \mathbb{R}^n} \phi_{\hat{\mu}}(x)$ can be improved by a linear domain transformation based on the Hessian of $\phi(\cdot)$. Our method originated in the conjecture that this transformation would also improve the conditioning of problem (3.2.1). Han used this matrix as the basis for a variable metric method for problem (3.2.1) in [Han.1].

We now return to problem (3.1.2). The Lagrangian for problem (3.1.2) is given by $l(x, \mu) = \sum_{j \in \mathbf{p}} \mu^j g^j(A_j x)$. Hence its Hessian with respect to $x$ is given by

$$L(x, \mu) = \sum_{j \in \mathbf{p}} \mu^j A_j^T G^j(A_j x) A_j \,, \tag{3.2.4}$$

where $G^j(\cdot)$ denotes the second derivative matrix of $g^j(\cdot)$. In many engineering optimization problems, such as those mentioned in the introduction, the functions $g^j(\cdot)$ do not contribute to the ill-conditioning of the matrix $L(\hat{x}, \hat{\mu})$, at a solution. Furthermore, their Hessians may be very difficult to compute. Hence we propose to replace the Hessian matrices $G^j(A_j x_i)$ in (3.2.4) by $l_j \times l_j$ identity matrices. Thus, for any $\mu \in \Sigma_p$, let

$$R(\mu) \triangleq \sum_{j \in \mathbf{p}} \mu^j A_j^T A_j \,. \tag{3.2.5}$$

We will show that a sequential transformation method based on the matrix $R(\mu)$ can compensate for the ill-conditioning introduced by the matrices $A_j$.

To ensure that a sequential domain transformation method does not destroy the convergence properties of the algorithm which it uses, there must be both an upper bound and a strictly positive lower bound on the eigenvalues of the domain transformation matrices. However, the minimum positive eigenvalue of $R(\mu_i)$ may decrease to zero as $\mu_i \to \hat{\mu}$. Hence, we propose to modify $R(\mu_i)$ by augmenting the small eigenvalues in its spectral decomposition, as follows. For any $\mu \in \Sigma_p$, let $\lambda_1(\mu) \geq \lambda_2(\mu) \geq \ldots \geq \lambda_n(\mu)$ be the eigenvalues of $R(\mu)$. Let $U_\mu$ be any real unitary matrix such that $R(\mu) = U_\mu \operatorname{diag}(\lambda_1(\mu), \ldots, \lambda_n(\mu)) U_\mu^T$, and let $\tilde{\lambda}_j(\mu) \triangleq \max\{\lambda_j(\mu), \varepsilon\}$, where $\varepsilon > 0$ is a small fixed number. Then we define

$$Q(\mu) \triangleq U_\mu \operatorname{diag}(\tilde{\lambda}_1(\mu), \ldots, \tilde{\lambda}_n(\mu)) U_\mu^T \,. \tag{3.2.6}$$

**Proposition 3.2.1:** *The matrix-valued function $Q(\cdot)$ is well defined and continuous in $\mu$.*

**Proof:** We begin by showing that $Q(\mu)$ is well defined even though the selection of the eigenvector matrix $U_\mu$ is not unique when $R(\mu)$ has multiple eigenvalues. Letting $\lambda_0 \triangleq \infty$, we define

$$D(\mu) \triangleq \{\, j \in \underline{n} \mid \lambda_{j-1}(\mu) > \lambda_j(\mu) = \lambda_{j+1}(\mu) \cdots = \lambda_{j+m_j(\mu)-1}(\mu) > \lambda_{j+m_j(\mu)}(\mu) \,\}\ , \tag{3.2.7}$$

so that $\{\lambda_j(\mu)\}_{j \in D(\mu)}$ is the set of distinct eigenvalues of $R(\mu)$, with multiplicities $m_j(\mu)$. Next, let $u_j$ denote the $j-th$ column of $U_\mu$, $j \in \underline{n}$. Then,

$$R(\mu) = \sum_{j \in D(\mu)} \lambda_j(\mu) \left[ \sum_{k \in \underline{m_j(\mu)}} u_{j+k-1}\, u_{j+k-1}^T \right] \tag{3.2.8}$$

is a spectral decomposition of $R(\mu)$. The matrix $\sum_{k \in \underline{m_j(\mu)}} u_{j+k-1} u_{j+k-1}^T$ represents a projection operator which projects onto the eigenspace corresponding to $\lambda_j(\mu)$, and hence it is independent of the selection of $U_\mu$. Since

$$Q(\mu) \triangleq U_\mu \operatorname{diag}(\tilde{\lambda}_1(\mu), \ldots, \tilde{\lambda}_n(\mu))\, U_\mu^T = \sum_{j \in D(\mu)} \tilde{\lambda}_j(\mu) \left[ \sum_{k \in \underline{m_j(\mu)}} u_{j+k-1} u_{j+k-1}^T \right], \tag{3.2.9}$$

we see that it, too, is independent of the selection of $U_\mu$.

Now, suppose that the sequence $\{\mu_i\}_{i=0}^{\infty} \subset \Sigma_p$ converges to some $\bar{\mu} \in \Sigma_p$ as $i \to \infty$. For each $\mu_i$, let $U_i = [\, u_{1,i}, \ldots, u_{n,i} \,]$ be a unitary matrix of eigenvectors of $R(\mu_i)$, so that

$$Q(\mu_i) = \sum_{j \in D(\mu_i)} \tilde{\lambda}_j(\mu_i) \left[ \sum_{k \in \underline{m_j(\mu_i)}} u_{j+k-1,i}\, u_{j+k-1,i}^T \right]. \tag{3.2.10}$$

The sequences $\{ Q(\mu_i) \}_{i=0}^{\infty}$ and $\{ U_i \}_{i=0}^{\infty}$ are bounded, since the eigenvalues are continuous and the matrices $U_i$ are unitary. Therefore, there exists an infinite subset $K \subset \mathbb{N}$, and matrices $\bar{Q}$ and $\bar{U} = [\bar{u}_1, \ldots, \bar{u}_n]$, such that $Q(\mu_i) \to \bar{Q}$ and $U_i \to \bar{U}$ as $i \xrightarrow{K} \infty$. Because $U_i U_i^T = I$, for all $i \in \mathbb{N}$, $\bar{U}\,\bar{U}^T = I$. Since $[R(\mu_i) - \lambda_j(\mu_i) I\,] u_{j,i} = 0$, for $j \in \underline{n}$, and since the eigenvalues, $\lambda_j(\cdot)$, are continuous, $[R(\bar{\mu}) - \lambda_j(\bar{\mu}) I\,] \bar{u}_j = 0$ for $j \in \underline{n}$. Thus, $\bar{U}$ is a unitary matrix of eigenvectors for $R(\bar{\mu})$. The matrix $Q(\mu_i)$ can also be written in the form

$$Q(\mu_i) = \sum_{j \in D(\bar{\mu})} \left[ \sum_{k \in m_j(\bar{\mu})} \tilde{\lambda}_{j+k-1}(\mu_i) u_{j+k-1,i} \, u_{j+k-1,i}^T \right]. \tag{3.2.11}$$

Taking limits in (3.2.11) as $i \to \infty$, $i \in K$, yields

$$\bar{Q} = \sum_{j \in D(\bar{\mu})} \left[ \sum_{k \in m_j(\bar{\mu})} \tilde{\lambda}_{j+k-1}(\bar{\mu}) \, \bar{u}_{j+k-1} \, \bar{u}_{j+k-1}^T \right]$$
$$= \sum_{j \in D(\bar{\mu})} \tilde{\lambda}_j(\bar{\mu}) \left[ \sum_{k \in m_j(\bar{\mu})} \bar{u}_{j+k-1} \bar{u}_{j+k-1}^T \right] = Q(\bar{\mu}). \tag{3.2.12}$$

Since the sequence $\{ Q(\mu_i) \}_{i=0}^{\infty}$ is bounded and any accumulation point of this sequence equals $Q(\bar{\mu})$, it follows that $\lim_{i \to \infty} Q(\mu_i) = Q(\bar{\mu})$, and hence $Q(\cdot)$ is continuous. □

We now provide an algorithm model which shows how to combine our variable metric technique with any one-step, first-order minimax algorithm which produces multiplier estimates. To simplify notation, we rewrite problem (3.1.2) as

$$\min_{x \in \mathbb{R}^n} \psi(x), \tag{3.2.13}$$

where

$$\psi(x) \triangleq \max_{j \in \mathbf{p}} g^j(A_j x). \tag{3.2.14}$$

Now consider any first-order minimax algorithm for solving (3.2.13) which generates estimates of the optimal multipliers at each iteration. We can associate with the algorithm a $\psi$-dependent, set-valued iteration map $M_\psi : \mathbb{R}^n \to \mathbb{R}^n \times 2^{\Sigma_p}$ such that, if $\{ (x_i, \mu_i) \}_{i=1}^{\infty}$ is a sequence generated by the algorithm on the problem $\min_{x \in \mathbb{R}^n} \psi(x)$, then

$$(x_{i+1}, \mu_{i+1}) \in M_\psi(x_i), \tag{3.2.15}$$

for all $i \in \mathbb{N}$.

For any $\nu \in \Sigma_p$, let $S(\nu) \triangleq Q(\nu)^{-1/2}$. Then the function $\psi(S(\nu)y)$ can be written in the alternative form $(\psi \cdot S)(y)$, which leads to the notation $M_{\psi \cdot S(\nu)}$ for the iteration map defined for the problem transformed by $S(\nu)$. Hence a variable metric algorithm for solving problem (3.1.2), based on the the iteration map $M_\psi$ and the transformation matrix $S(\nu)$ has the form

**Variable Metric Algorithm Model 3.2.1:**

**Data:** $x_0 \in \mathbb{R}^n$, $\mu_{-1} \in \Sigma_p$, $i = 0$.

**Step 1:** Set $y_i = S(\mu_i)^{-1} x_i$ ,

**Step 2:** Compute $(y_{i+1}, \mu_{i+1}) \in M_{\psi \circ S(\mu_i)}(y_i)$ ,

**Step 3:** Set $x_{i+1} = S(\mu_i) y_{i+1}$ .

**Step 4:** Replace $i$ by $i+1$ and go to Step 1. □

Note that the multiplier vectors do not require transformation because, for any invertible matrix $S$, $(\hat{x}, \hat{\mu}) \in \mathbb{R}^n \times \Sigma_p$ satisfies equations (3.2.3a-b) with respect to problem (3.1.2) if and only if $(S^{-1}\hat{x}, \hat{\mu})$ satisfies these equations with respect to problem $\min_{y \in \mathbb{R}^n} \psi(Sy)$.

## 3.3 RATE OF CONVERGENCE OF THE PSHENICHNYI ALGORITHM

We will now summarize a number of results, established in [Pol.4] and Chapter 2, for a version of the Pshenichnyi minimax algorithm [Psh.1] which uses an exact minimizing line search. When applied to problem (3.1.2), with $\psi(\cdot)$ defined in (3.2.14), this algorithm has the following form:

**Algorithm 3.3.1** : (see Algorithm 5.2 and Corollary 5.1 in [Pol.4])

**Data:** $x_0$; $\gamma > 0$.

**Step 0:** Set $i = 0$.

**Step 1:** Compute the search direction

$$h_i = \arg \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{p}} g^j(A_j x_i) + \langle A_j^T \nabla g^j(A_j x_i), h \rangle + \tfrac{1}{2}\gamma \| h \|^2 . \tag{3.3.1}$$

**Step 2:** Compute a minimizing step size, $\lambda_i \in \arg\min_{\lambda \in \mathbb{R}} \psi(x_i + \lambda h_i)$ .

**Step 3:** Set $x_{i+1} = x_i + \lambda_i h_i$, replace $i$ by $i+1$ and go to Step 1. □

**Theorem 3.3.1:** [Pol.4] *If the functions $g^j(\cdot)$ in problem (3.1.2) are continuously differentiable, then any accumulation point $\bar{x}$ of a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 3.3.1 satisfies the first-order optimality conditions (3.2.3a-b).*

To show that the algorithm converges linearly, we need to introduce more restrictive assumptions. Let the set of minimizers for problem (3.1.2) be denoted by $\hat{G} \triangleq \arg\min_{x \in \mathbb{R}^n} \psi(x)$. By analogy with nonlinear programming convention, we say that *strict complementary slackness* holds at $(\hat{x}, \hat{\mu})$, where $\hat{x} \in \hat{G}$, $\hat{\mu} \in \Sigma_p$ and the pair satisfies (3.2.3a-b), if we have $\hat{\mu}^j > 0$ if and only if $g^j(A_j\hat{x}) = \psi(\hat{x})$.

Hypothesis 3.3.1: *Suppose that*

(i)fI the functions $g^j(\cdot)$ *are twice continuously differentiable,*

(ii) *there exist* $0 < l \leq L < \infty$ *such that, for all* $j \in \underline{p}$,

$$l \|h\|^2 \leq \langle h, G^j(z) h \rangle \leq L \|h\|^2, \quad \forall\, z, h \in \mathbb{R}^{l_j}, \tag{3.3.2}$$

(iii) *strict complementary slackness holds for all* $(\hat{x}, \hat{\mu})$ *where* $\hat{x} \in G$, $\hat{\mu} \in \Sigma_p$ *and* $(\hat{x}, \hat{\mu})$ *satisfy (3.2.3a-b).*

It follows from Hypothesis 3.3.1 that (i) for any $\hat{x} \in \hat{G}$ there is a unique optimal multiplier $\hat{\mu} \in \Sigma_p$ satisfying equations (3.2.3a-b), (ii) the set of optimal multipliers, associated with the set of optimal solutions, $\hat{G}$, is a singleton, $\{\hat{\mu}\}$, and (iii) the set of indices of functions maximal at $\hat{x}$, $\hat{J} \triangleq \{ j \in \mathbf{p} \mid g^j(A_j\hat{x}) = \psi(\hat{x}) \}$, is independent of $\hat{x} \in \hat{G}$ (see Proposition 2.5.1).

Let $j_1 < \dots < j_b$ be the indices comprising $\hat{J}$, then we define the matrix $\hat{A}^T \triangleq [A_{j_1}^T, \dots, A_{j_b}^T]$. Let $a \triangleq Rank(\hat{A}^T)$ and let $Z$ be an $n \times a$ matrix, the columns of which form an orthonormal basis for $Range(\hat{A}^T)$. Then we have the following result, established in Chapter 2 and [Wie.1].

**Theorem 3.3.2:** **[Wie.1]** *Suppose that Hypothesis 3.3.1 holds with respect to problem (3.1.2) and, in addition,*

(iv) *$l$ and $L$ are chosen so that the scaling parameter, $\gamma$, in Algorithm 3.3.1 satisfies*

$$l\ \sigma^{+}[\sum_{j \in \mathbf{p}} \hat{\mu}^{j} A_j^T A_j] < \gamma < L \max_{j \in \mathbf{p}} \| Z^T A_j^T A_j Z \| , \tag{3.3.3}$$

*where $\sigma^{+}[X]$ denotes the minimum positive eigenvalue of the symmetric matrix $X$. If Algorithm 3.3.1 constructs an infinite sequence $\{x_i\}_{i=0}^{\infty}$, then,* (a) *$x_i \to \hat{x}$ as $i \to \infty$ with $\hat{x} \in \hat{G}$, and* (b) *either there exists an $i_0 \in \mathbf{N}$ such that $x_i = \hat{x}$ for all $i \geq i_0$ or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \leq \rho , \tag{3.3.4}$$

*where*

$$\rho \triangleq 1 - \frac{l}{L} \frac{\sigma^{+}[\sum_{j \in \mathbf{p}} \hat{\mu}^{j} A_j^T A_j]}{\max_{j \in \mathbf{p}} \| Z^T A_j^T A_j Z \|} . \tag{3.3.5}$$

Following [Lue.1], we refer to the quantity $\limsup_{i \to \infty} (\psi(x_{i+1}) - \hat{\psi}) / (\psi(x_i) - \hat{\psi})$ as the *convergence ratio* of the sequence $\{ \psi(x_i) \}_{i=0}^{\infty}$. The quantity $\rho$ in (3.3.4) bounds the convergence ratio of any sequence constructed by Algorithm 3.3.1 in solving any problem in the class defined by (3.1.2) and the assumptions stated.

## 3.4 RATE OF CONVERGENCE OF VARIABLE-METRIC-PSHENICHNYI ALGORITHM

We will refer to the algorithm obtained by inserting the iteration map of Algorithm 3.3.1 into the Algorithm Model 3.2.1 as the Variable-Metric-Pshenichnyi Algorithm. We will now show that the Variable-Metric-Pshenichnyi Algorithm converges faster on problems of the form (3.1.2) than Algorithm 3.3.1.

For the transformed problem

$$\min_{x \in \mathbb{R}^n} \psi(S(\nu)y)\,, \tag{3.4.1}$$

given a point $y = S(\nu)^{-1}x$ and a $\nu \in \Sigma^p$, the search direction computation (3.3.1) has the form

$$d(y,\nu) \triangleq \arg\min_{d \in \mathbb{R}^n} \max_{j \in \mathbf{p}} g^j(A_jS(\nu)y) + \langle (A_jS(\nu))^T \nabla g^j(A_jS(\nu)y), d \rangle + ½\gamma \| d \|^2 . \tag{3.4.2}$$

The result can be transformed back to the original space using the formula

$$h(x,\nu) \triangleq S(\nu)d(y,\nu)\,. \tag{3.4.3}$$

Equivalently, $h(x,\mu)$ can be computed directly using the variable metric defined by $S(\nu)$ as follows:

$$h(x,\nu) = \arg\min_{h \in R^n} \max_{j \in \mathbf{p}} g^j(A_jx) + \langle A_j^T \nabla g^j(A_jx), h \rangle - \psi(x) + ½\gamma \langle h, Q(\nu)h \rangle . \tag{3.4.4}$$

Since the max function in (3.4.4) is strictly convex in $h$, $h(x,\nu)$ is unique, and hence it also follows that $h(\cdot,\cdot)$ is continuous.

Problem (3.4.4) can be solved by converting it to dual form by the same argument used in Section 2 of Chapter 2. Let $\theta(x,\nu)$ denote the minimum value in (3.4.4). Then for any $x \in \mathbb{R}^n$ and $\nu \in \Sigma_p$, the search direction problem can be written in the following equivalent forms:

$$\theta(x,\nu) \triangleq \min_{h \in R^n} \max_{j \in \mathbf{p}} g^j(A_jx) + \langle A_j^T \nabla g^j(A_jx), h \rangle - \psi(x) + ½\gamma \langle h, Q(\nu)h \rangle$$

$$= \min_{h \in R^n} \max_{\mu \in \Sigma_p} \sum_{j \in \mathbf{p}} \mu^j [g^j(A_jx) + \langle A_j^T \nabla g^j(A_jx), h \rangle - \psi(x)] + ½\gamma \langle h, Q(\nu)h \rangle . \tag{3.4.5}$$

By Theorem 2.7.1, the max and min in (3.4.5) can be interchanged. Hence,

$$\theta(x,\nu) = \max_{\mu \in \Sigma_p} \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p}} \mu^j [g^j(A_jx) + \langle A_j^T \nabla g^j(A_jx), h \rangle - \psi(x)] + ½\gamma \langle h, Q(\nu)h \rangle$$

$$= \max_{\mu \in \Sigma_p} \sum_{j \in \mathbf{p}} \mu^j [g^j(A_jx) - \psi(x)] - \frac{1}{2\gamma} \| \sum_{j \in \mathbf{p}} \mu^j A_j^T \nabla g^j(A_jx) \|^2_{Q(\nu)^{-1}}\,, \tag{3.4.6}$$

where the last expression is obtained by solving the inner minimization problem[2]. Since the

[2] Several methods exist (see, for example, [Gil.1, von.1, Kiw.2-3, Rus.1, Hig.1]) for solving the positive semi-definite quadratic

solution to (3.4.6) is usually not unique, we denote the solution set by

$$U(x,\nu) \triangleq \arg \max_{\mu \in \Sigma_p} \sum_{j \in p} \mu^j [g^j(A_j x) - \psi(x)] - \frac{1}{2\gamma} \| \sum_{j \in p} \mu^j A_j^T \nabla g^j(A_j x) \|^2_{Q(\nu)^{-1}} . \tag{3.4.7}$$

The set-valued function $U(\cdot,\cdot)$ has the following properties: (i) it is upper semi-continuous in the sense of Clarke [Cla.1] and compact-valued; (ii) for any minimizer $\hat{x}$ of (3.1.1a) and any $\nu \in \Sigma_p$, $U(\hat{x},\nu)$ is the set of multiplier vectors which, together with $\hat{x}$, satisfy equations (3.2.3a) and (3.2.3b), (iii) under Hypothesis 3.3.1, $U(\hat{G},\Sigma_p) = \{ \hat{\mu} \}$, a singleton, (iv) any multiplier vector $\mu \in U(x,\nu)$ yields the *unique* solution to the primal problem (3.4.4), according to the formula

$$h(x,\nu) = -\frac{1}{\gamma} Q(\nu)^{-1} \sum_{j \in \underline{p}} \mu^j A_j^T \nabla g^j(A_j x) . \tag{3.4.8}$$

Steps 2 and 3 of the Variable Metric Algorithm 3.2.1, using the iteration map of Algorithm 3.3.1, can also be performed in the original space without affecting the sequence of iterates produced. We therefore present it in this form to simplify proofs.

Algorithm 3.4.1:

Data: $x_0$; $\gamma > 0$, $\mu_{-1} \in \Sigma_p$, $\varepsilon > 0$, $i = 0$.

Step 1: Compute the multiplier vector, $\mu_i \in U(x_i, \mu_{i-1})$.

Step 2: Compute $h_i = h(x_i, \mu_{i-1})$ using (3.4.8).

Step 3: Compute the minimizing step size, $\lambda_i = \arg\min_{\lambda \in \mathbb{R}} \psi(x_i + \lambda h_i)$.

Step 4: Set $x_{i+1} = x_i + \lambda_i h_i$, replace $i$ by $i+1$ and go to Step 1. □

We will now establish several properties of Algorithm 3.4.1.

Theorem 3.4.1: *If the functions $g^j(\cdot)$ in problem (3.1.2) are continuously differentiable, then any accumulation point $\hat{x}$ of a sequence $\{ x_i \}_{i=0}^{\infty}$ generated by Algorithm 3.4.1 satisfies the necessary conditions (3.2.3a-b).*

program (3.4.6).

**Proof:** This follows from the proof of convergence for Algorithm 3.1 in [Pol.4] and the fact that the scaling matrices $S(\mu_{i-1})$ are uniformly bounded, *i.e.* - that for all $\nu \in \Sigma_p$,

$$(\max_{j \in \mathbf{p}} \|A_j^T A_j\|)^{-\frac{1}{2}} \|h\|^2 \leq \|S(\nu)h\|^2 \leq \varepsilon^{-\frac{1}{2}} \|h\|^2 . \tag{3.4.9}$$

□

Next we will show, under assumptions of convexity and complementary slackness, that the sequence of iterates, $\{x_i\}_{i=0}^{\infty}$, constructed by Algorithm 3.4.1 converges to the solution set $\hat{G}$ and that the corresponding sequence of multiplier vectors, $\{\mu_i\}_{i=-1}^{\infty}$ converges to $\hat{\mu}$, the unique optimal multiplier associated with the solution set $\hat{G}$. We will use the notation $z_i \rightarrow Z$ to represent the convergence of a sequence $\{ z_i \}_{i=0}^{\infty} \subset \mathbb{R}^n$ to a set $Z \subset \mathbb{R}^n$, *i.e.* - $\lim_{i \rightarrow \infty} \min_{y \in Z} \|z_i - y\| = 0$.

**Theorem 3.4.2:** *Suppose that Hypothesis 3.3.1 holds and that Algorithm 3.4.1 generates sequences of iterates* $\{ x_i \}_{i=0}^{\infty}$ *and of multiplier vectors* $\{ \mu_i \}_{i=0}^{\infty}$. *Then,*

(a) *there exists an open set* $W \supset \hat{G}$ *such that* $\mu^j = 0$ *for every* $j \notin \hat{J}$ *and all* $\mu \in U(W, \Sigma_p)$,

(b) *there exists* $\hat{x} \in \hat{G}$ *such that* $x_i \rightarrow \hat{x}$ *as* $i \rightarrow \infty$,

(c) *there exists* $i_0 \in \mathbb{N}$ such that $x_i \in \hat{x} + Range(\hat{A}^T)$ for all $i \geq i_0$,

(d) $\mu_i \rightarrow \hat{\mu}$ *as* $i \rightarrow \infty$.

**Proof:** (a) Since $h(\cdot, \cdot)$ defined in (3.4.3) and $\theta(\cdot, \cdot)$ defined in (3.4.5) are uniformly continuous in $(x, \nu)$ on compact sets in $\mathbb{R}^n \times \Sigma_p$, and since both functions are zero on the set $\hat{G} \times \Sigma_p$, (a) follows from the same argument as Proposition 2.5.2.

(b) Let $A^T \triangleq [A_1^T, \ldots, A_p^T]$. Equation (3.4.8) and the fact that $Range(A^T)$ is invariant under $S(\nu)$ for all $\nu \in \Sigma_p$ imply that the sequence of search directions $\{ h_i \}_{i=0}^{\infty}$ is contained in the range of $A^T$. Therefore, the sequence of iterates $\{ x_i \}_{i=0}^{\infty}$ is contained in the set

$$V \triangleq (x_0 + Range(A^T)) \cap \{ x \in \mathbb{R}^n \mid \psi(x) \leq \psi(x_0) \} . \tag{3.4.10}$$

The set $V$ is compact by the same argument as in the proof of Theorem 2.5.1, and therefore the set $\{ x_i \}_{i=0}^{\infty}$ converges to the set of its accumulation points. By Theorem 3.4.1, these must satisfy the optimality condition (3.2.3a-b). Since $\psi(\cdot)$ is convex, these necessary conditions are sufficient for optimality, implying that $x_i \to \hat{G}$ .

From part (a) of this theorem, $h(x, \nu) \in Range(\hat{A}^T)$ for all $x \in W$ and all $\nu \in \Sigma_p$. Because $x_i \to \hat{G}$, there exists $i_0 \in \mathbb{N}$ such that $x_i \in W$ for all $i > i_0$. Hence, $\{ x_i \}_{i=i_0}^{\infty} \subset x_{i_0} + Range(\hat{A}^T)$, and $x_i \to (x_{i_0} + Range(\hat{A}^T)) \cap \hat{G}$ .

We show that this limit set is a singleton. Suppose $x_1, x_2 \in (x_{i_0} + Range(\hat{A}^T)) \cap \hat{G}$ . Then, since $\psi(\cdot)$ is convex, the entire line segment between $x_1$ and $x_2$, $[x_1, x_2]$, is contained in this set. Now, $U([x_1, x_2], \Sigma_p) = \{ \hat{\mu} \}$ and $\hat{\mu}^j > 0$ for all $j \in \hat{J}$ . Hence, $g^j(A_j x) = \psi(x) = \hat{\psi}$ for all $x \in [x_1, x_2]$ and all $j \in \hat{J}$ , by equation (3.2.3b). Since the functions $g^j(\cdot)$ are strictly convex, this implies that $A_j(x_1 - x_2) = 0$ for all $j \in \hat{J}$ . Since $x_1 - x_2 \in Range(\hat{A}^T)$, this implies that $x_1 - x_2 \in Range(\hat{A}^T) \cap Null(\hat{A}) = \{ 0 \}$, *i.e.* - that $x_1 = x_2$. Thus, $\hat{G} \cap (x_{i_0} + Range(\hat{A}^T)) = \{ \hat{x} \}$ for some $\hat{x}$ .

(c) From the proof of (b), $x_i \in x_{i_0} + Range(\hat{A}^T) = \hat{x} + Range(\hat{A}^T)$ , for all $i \geq i_0$.

(d) The set-valued map $U(\cdot, \cdot)$ defined in (3.4.7) is upper semicontinuous in the sense of Clarke [Cla.1] and compact-valued, uniformly on compact sets in $\mathbb{R}^n \times \Sigma_p$. Since $x_i \to \hat{x} \in \hat{G}$ by (b) and $U(\hat{G}, \Sigma_p) = \{ \hat{\mu} \}$ , this implies that $\mu_i \to \hat{\mu}$. □

We define the function $\rho : \mathbb{R}^{n \times n} \to \mathbb{R}$ by

$$\rho(S) \triangleq 1 - \frac{l}{L} \frac{\sigma^+[\sum_{j \in p} \hat{\mu}^j S^T A_j^T A_j S]}{\max_{j \in \mathbf{p}} \| Z^T S^T A_j^T A_j S Z \|} . \tag{3.4.11}$$

Note that $\rho$ in equation (3.3.5) equals $\rho(I)$, where $I$ is the $n \times n$ identity matrix.

**Theorem 3.4.3:** *Suppose that Hypothesis 3.3.1 holds and, in addition,* (iv) *$l$ and $L$ are chosen so that the scaling parameter, $\gamma$, satisfies*

$$l\, \sigma^+[\sum_{j \in \mathbf{p}} \hat{\mu}^j S(\hat{\mu}) A_j^T A_j S(\hat{\mu})] < \gamma < L \max_{j \in \mathbf{p}} \| Z^T S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) Z \| . \tag{3.4.12a}$$

*If $\{ x_i \}_{i=0}^{\infty}$ is an infinite sequence generated by Algorithm 3.4.1, then, either there exists an $i_0 \in \mathbb{N}$ and $\hat{x} \in \hat{G}$ such that $x_i = \hat{x}$ for all $i \geq i_0$, or*

$$\limsup_{i \to \infty} \frac{\psi(x_{i+1}) - \hat{\psi}}{\psi(x_i) - \hat{\psi}} \leq \rho(S(\hat{\mu})) . \tag{3.4.12b}$$

**Proof:** By Theorem 3.4.1(b), the sequence of iterates has a limit point $\hat{x} \in \hat{G}$. Assume that $x_i \neq \hat{x}$ for all $i \in \mathbb{N}$. Hypothesis 3.3.1 and assumption (iv) of this theorem ensure that the assumptions of Theorem 3.3.2 are met for the transformed problem,

$$\min_{y \in \mathbb{R}^n} \psi(S(\hat{\mu}) y) . \tag{3.4.13}$$

Since the range of $\hat{A}^T$ is invariant under $S(\hat{\mu})$, the columns of $Z$ form a basis for the range of $S(\hat{\mu})\hat{A}^T$. This fact and assumption (iv) of this theorem imply that assumption (iv) of Theorem 3.3.2 holds with respect to problem (3.4.13). This and Hypothesis 3.3.1 ensure that the assumptions of Theorem 3.3.2 are satisfied with respect to problem (3.4.13). The following result, slightly stronger than Theorem 3.3.2, but valid under the same assumptions, is stated in Theorem 2.5.1:

$$\limsup_{\substack{y \to \hat{y} \\ y \neq \hat{y} \\ y \in Range(Z)}} \min_{\lambda \in \mathbb{R}} \frac{\psi(S(\hat{\mu})(y + \lambda d(y, \hat{\mu}))) - \hat{\psi}}{\psi(S(\hat{\mu})y) - \hat{\psi}} \le \rho(S(\hat{\mu})), \tag{3.4.14}$$

where $\hat{y} \triangleq S(\hat{\mu})^{-1}\hat{x}$ for an arbitrary $\hat{x} \in \hat{G}$. Using the substitution $y = S(\hat{\mu})^{-1}x$ and the fact that $h(S(\hat{\mu})y, \hat{\mu}) = S(\hat{\mu})d(y, \hat{\mu})$, (3.4.14) can be rephrased as follows. For any $\delta > 0$, there exists a set $V' \subset \hat{x} + Range(Z)$, which is open in the affine space $\hat{x} + Range(Z)$, such that

$$\min_{\lambda \in \mathbb{R}} \frac{\psi(x + \lambda h(x, \hat{\mu})) - \hat{\psi}}{\psi(x) - \hat{\psi}} \le (1 + \delta)\rho(S(\hat{\mu})), \tag{3.4.15}$$

for all $x \in V'$, $x \neq \hat{x}$. Since $\psi(\cdot)$ is strongly convex, the min over $\mathbb{R}$ in (3.4.15) and in Step 3 of Algorithm 3.4.1 can be replaced by a min over a closed interval $C$. With this modification, the left hand side of (3.4.15) is continuous in $(x, \mu)$, since $h(\cdot, \cdot)$ is continuous. This implies that there exists a neighborhood of $\hat{\mu}$, $D \subset \Sigma_p$, such that

$$\min_{\lambda \in C} \frac{\psi(x + \lambda h(x, \mu)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \le (1 + 2\delta)\rho(S(\hat{\mu})), \tag{3.4.16}$$

for all $x \in V'$ and $\mu \in D$. Of course, since $\delta$ was arbitrary,

$$\limsup_{\substack{x \to \hat{x} \\ x \neq \hat{x} \\ x \in \hat{x} + Range(Z)}} \min_{\lambda \in C} \frac{\psi(x + \lambda h(x, \mu)) - \hat{\psi}}{\psi(x) - \hat{\psi}} \le \rho(S(\hat{\mu})). \tag{3.4.17}$$

By Theorem 3.4.2(c), $x_i \in \hat{x} + Range(\hat{A}^T) = \hat{x} + Range(Z)$ for large $i$. Then, since $x_i \to \hat{x}$, $\mu_i \to \hat{\mu}$ and $\psi(x_i + \lambda_i h_i) = \min_{\lambda \in C} \psi(x_i + \lambda h_i)$, (3.4.12b) holds. □

The following comparison of the two convergence ratio bounds, $\rho$ given by (3.3.5) and $\rho(S(\hat{\mu}))$ given by (3.4.11), suggests that our variable metric technique results in a faster algorithm than the original Pshenichnyi Algorithm 3.3.1.

**Proposition 3.4.1:** *If* $\sigma^+[R(\hat{\mu})] > \varepsilon$, *then* $\rho(I) \geq \rho(S(\hat{\mu}))$.

**Proof:** Consider a spectral decomposition $R(\hat{\mu}) = U \Lambda U^T$, where $U$ is unitary, $\Lambda \triangleq diag(\lambda_1(\hat{\mu}), \ldots, \lambda_n(\hat{\mu}))$ and $\tilde{\Lambda} \triangleq diag(\tilde{\lambda}_1(\hat{\mu}), \ldots, \tilde{\lambda}_n(\hat{\mu}))$. We have that

$$\sum_{j \in \mathbf{p}} \hat{\mu}^j S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) = S(\hat{\mu})^T R(\hat{\mu}) S(\hat{\mu}) \tag{3.4.18}$$

$$= U \tilde{\Lambda}^{-1/2} U^T (U \Lambda U^T) U \tilde{\Lambda}^{-1/2} U^T$$

$$= U \tilde{\Lambda}^{-1} \Lambda U^T .$$

Since $\sigma^+[R(\hat{\mu})] > \varepsilon$, we have that, for each $j \in \underline{n}$, either $\tilde{\lambda}_j(\hat{\mu}) = \lambda_j(\hat{\mu})$ or $\lambda_j(\hat{\mu}) = 0$. Hence, $\sum_{j \in \mathbf{p}} \hat{\mu}^j S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) = U\, diag(1, \ldots, 1, 0, \ldots, 0)\, U$, which implies that $\sigma^+[\sum_{j \in \mathbf{p}} \hat{\mu}^j S(\hat{\mu})^T A_j^T A_j S(\hat{\mu})] = 1$, and

$$\rho(S(\hat{\mu})) \triangleq 1 - \frac{l}{L} \frac{1}{\max_{j \in \mathbf{p}} \| Z^T S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) Z \|} . \tag{3.4.19}$$

Now,

$$\| Z^T S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) Z \| = \max_{y \in \mathbb{R}^s} \frac{\langle y, Z^T S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) Z y \rangle}{\langle y, y \rangle}$$

$$= \max_{y \in \mathbb{R}^s} \frac{\| A_j S(\hat{\mu}) Z y \|^2}{\| y \|^2}$$

$$= \max_{y \in \mathbb{R}^s} \frac{\| A_j S(\hat{\mu}) Z y \|^2}{\| Z y \|^2} , \tag{3.4.20}$$

since the orthonormality of the columns of $Z$ implies that $\|Zy\| = \|y\|$ for the Euclidean norm. Making the substitution $z = S(\hat{\mu}) Z y$ yields

$$\| Z^T S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) Z \| = \max_{z \in Range(Z)} \frac{\| A_j z \|^2}{\| S(\hat{\mu})^{-1} z \|^2}$$

$$= \max_{z \in Range(Z)} \frac{\|A_j z\|^2}{\|Q(\hat{\mu})^{1/2} z\|^2}$$

$$= \max_{z \in Range(Z)} \frac{\|A_j z\|^2}{\langle z, Q(\hat{\mu}) z \rangle}. \tag{3.4.21}$$

Substituting (3.4.21) into (3.4.19) yields

$$\rho(S(\hat{\mu})) \triangleq 1 - \frac{l}{L} \min \left\{ \frac{\langle z, Q(\hat{\mu}) z \rangle}{\|A_j z\|^2} \;\middle|\; j \in \mathbf{p},\ z \in Range(Z) \right\}. \tag{3.4.22}$$

By inspection, $\rho(S(\hat{\mu}))$ is never greater than

$$\rho(I) = 1 - \frac{l}{L} \min \left\{ \frac{\langle z, Q(\hat{\mu}) z \rangle}{\|A_j y\|^2} \;\middle|\; j \in \mathbf{p},\ y, z \in Range(Z),\ \|y\| = \|z\| = 1 \right\}. \tag{3.4.23}$$

□

The difference between $\rho(I)$ and $\rho(S(\hat{\mu}))$ can be quite significant, as the following example shows.

**Example 3.4.1:** Suppose that a minimax problem involves two scaling matrices,

$$\begin{bmatrix} 1 & 0 \\ 0 & 10^{-2} \end{bmatrix}, \quad \begin{bmatrix} 10^{-2} & 0 \\ 0 & 10 \end{bmatrix}, \tag{3.4.24}$$

and that $\mu^1 = \mu^2 = 1/2$ and $l = L = 1$. The rate constant for the unscaled Algorithm 3.3.1 is $\rho(I) = 0.995$, whereas, under rescaling, it is $\rho(S(\hat{\mu})) = 0.5$. This suggests that $\lceil \log 10^{-1} / \log \rho(I) \rceil = 460$ iterations of the Pshenichnyi Algorithm 3.3.1 would be required to achieve a ten-fold reduction in $\psi(x) - \hat{\psi}$ near the solution, while only $\lceil \log 10^{-1} / \log \rho(S(\hat{\mu})) \rceil = 4$ iterations of the Variable-Metric Pshenichnyi Algorithm 3.4.1 would be required. □

## 3.5 NUMERICAL EXPERIMENTS

We performed a number of numerical experiments to evaluate the overall behavior of the variable metric technique. We compare the performance of the Pshenichnyi Algorithm 3.3.1 with that of the Variable-Metric-Pshenichnyi Algorithm 3.4.1 and with Han's method [Han.1], which performs domain rescaling by replacing the matrix $R(\mu_i)$ by $\sum_{j \in \mathbf{p}} \mu_i^j A_j^T G^j(x_i) A_j$. In addition, we compared the performance of the ***barrier function minimax algorithm*** in [Pol.6] with a corresponding variable-metric-barrier-function method which we constructed in accordance with the Variable Metric Algorithm Model 3.2.1. The barrier function method is based on the penalty function

$$\sum_{j \in \mathbf{p}} \frac{1}{\alpha - g^j(A_j x)} , \tag{3.5.1}$$

which is differentiable at all $x$ for which $\psi(x) < \alpha$. An iteration of the barrier method involves an indefinite number of inner cycles, each of which requires the evaluation of all functions and first order derivatives. The rate of convergence of this algorithm has not been established and hence we can only evaluate the effect of our sequential transformation technique on it through numerical experiments. The five algorithms were applied to the two problems below. An Armijo-like step size rule [Arm.1, Pol.4],

$$\lambda_i = \max \{ \beta^k \tilde{\lambda}_i \mid \psi(x_i + \tilde{\lambda}_i \beta^k h_i) - \psi(x_i) \le \alpha \tilde{\lambda}_i \beta^k \theta(x_i, \mu_i) \} , \tag{3.5.2}$$

with $\alpha, \beta \in (0, 1)$, was substituted for the exact minimizing line search in Algorithms 3.3.1 and 3.4.1, since problem (3.5.2) can be solved in a finite number of steps. Quadratic interpolation was used to determine a trial step size $\tilde{\lambda}_i$. In all of the experiments, the algorithm parameters were set to $\alpha = 0.7$, $\beta = 0.9$, $\gamma = 1.0$, and $\varepsilon = 10^{-10}$ (in the definition of the matrices $Q(\mu)$). Since in engineering applications, gradients and Hessians are frequently computed using finite differences, the evaluation counts in Tables 3.1 and 3.2 are tabulated as though the gradients and Hessians of the functions $g^j(\cdot)$ were evaluated by finite differences. The evaluation of a single function $g^j(z)$ incurs one function evaluation, and the gradient $\nabla g^j(z)$ incurs an additional $l_j$ evaluations. Thus, the total number of function evaluations required to obtain the information to compute a search direction for the Pshenichnyi and Variable-Metric-Pshenichnyi Algorithms is $\sum_{j \in \mathbf{p}} (l_j + 1)$. The evaluation of Hessians for use by the Han algorithm incurs an additional $\frac{1}{2}(l_j^2 + 1)$ evaluations per function $g^j(\cdot)$.

**Problem 3.5.1:** Consider the simple problem $\min_{x \in \mathbb{R}^4} \max \{ g^1(A_1 x), g^2(A_2 x) \}$, where

$$g^1(y) = y_1^2 + y_2^2 + (y_3 - 1)^2 - 1 , \tag{3.5.3a}$$

$$g^2(y) = y_1^2 + y_2^2 + (y_3 + 1)^2 - 1 , \tag{3.5.3b}$$

and the matrices $A_1$, $A_2$ are given by

$$A_1 = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10^{-1} & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 10^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} . \tag{3.5.3c}$$

An initial point of $(10^{-3}, 0, 10, 0)$ was used. The minimum value of 0 is achieved on the subspace spanned by the vector $(0, 0, 0, 1)$. Table 3.1 shows the work required for the five algorithms to achieve two given levels of accuracy in the value of $\hat{\psi}$. The units of work listed are number of iterations, the number of function evaluations and the CPU time. Figures 3.1 and 3.2 plot the function values $\{ \psi(x_i) \}$ against the number of function evaluations for the Pshenichnyi and Variable-Metric-Pshenichnyi Algorithms. □

**Problem 3.5.2:** [Hoa.1] Consider the problem of designing a controller for the feedback system in Figure 3.3 with plant,

$$P(s) = \frac{1}{(s+2)^2(s+3)} \begin{bmatrix} s^2+8s+10 & 3s^2+7s+4 \\ 2s+2 & 3s^2+9s+8 \end{bmatrix} . \tag{3.5.4}$$

Since the plant is stable, we can parametrize the controller by $C(x) = (I - R(x,s)P(s))^{-1} R(x,s)$ where $R(x,s)$ is a $2 \times 2$ matrix of rational polynomials in the complex variable $s$, which are bounded and analytic for $Re(s) \leq 0$. We chose to shape frequency

domain tracking error by solving the problem

$$\min_{x} \tfrac{1}{2} \max_{\omega \in \Omega} \| H_{e_1 u}(j\omega,\ R(x, j\omega)) \|_F^2 \ , \tag{3.5.5}$$

where $\Omega$ consists of six frequency points, equally spaced on a logarithmic scale, $\{\ 0.010, 0.029, 0.080, 0.240, 0.693, 2.0\ \}$, and $\|\cdot\|_F$ denotes the Frobenius norm. For this system, $H_{e_1 u}(x, s) = I - P(s)R(x, s)$. We used the following first order expansion of $R(x, s)$,

$$R(x, s) = \begin{bmatrix} x1 & x2 \\ x3 & x4 \end{bmatrix} \frac{1}{(s+10)} + \begin{bmatrix} x5 & x6 \\ x7 & x8 \end{bmatrix} . \tag{3.5.6}$$

The initial point $x_0 = (0, 0, 0, 0, 1, 0, 0, 1)$ was chosen, and our computations converged to the minimum value of 0.0255085 at the point

$$\hat{x} = \Big[ -80.308718709, -4.4337113582, 84.132574000, -31.534025985,$$
$$9.2348949849, -0.0051528236, -8.9338039187, 4.8550280952 \Big]$$

The work required for the algorithms to achieve two given levels of accuracy is recorded in Table 3.2. The values of $\psi(\cdot)$ are plotted versus the number of function evaluations for the Pshenichnyi and Variable-Metric-Pshenichnyi Algorithms in Figure 3.4. □

Theorems 3.3.1 and 3.4.2 apply under the same assumptions to versions of Algorithms 3.3.1 and 3.4.1 employing an Armijo-like step size rule, except that the convergence ratio bounds are given by

$$\rho \triangleq 1 - \alpha\beta \frac{l}{L} \frac{\sigma^+[\sum_{j \in p} \hat{\mu}^j A_j^T A_j]}{\max_{j \in p} \| Z^T A_j^T A_j Z \|} , \tag{3.5.7}$$

and

$$\rho(S(\hat{\mu})) \triangleq 1 - \alpha\beta \frac{l}{L} \frac{\sigma^+[\sum_{j \in p} \hat{\mu}^j S(\hat{\mu})^T A_j^T A_j S(\hat{\mu})]}{\max_{j \in p} \| Z^T S(\hat{\mu})^T A_j^T A_j S(\hat{\mu}) Z \|} . \tag{3.5.8}$$

Table 3.3 presents the actual convergence ratios of the sequences constructed by the algorithms under comparison on Problems 3.5.1 and 3.5.2, as well as the convergence ratio bounds derived above. Table 3.3 shows that the variable metric technique reduces both quantities. The reduction in computational effort corresponding to the decrease in the convergence ratios of the observed sequences is evident from Tables 3.1 and 3.2. The reduction in effort entailed by even the slight reductions in the convergence ratio bounds is also large. To show this, we have included in Table 3.3 the number of iterations which the algorithm convergence ratio bounds suggest would be required to reduce $\psi(x) - \hat{\psi}$ by a factor of 10 near a solution, *i.e.* - $\lceil \log 0.1 / \log \rho \rceil$.

If a variable metric $Q_H(\mu)$ is based on $R_H(\mu) \triangleq \sum_{j \in \mathbf{p}} \mu^j A_j^T G^j(A_j x) A_j$, rather than $R(\mu)$ as in Section 2, and if $\sigma^+[R_H(\hat{\mu})] > \varepsilon$, the search direction of Algorithm 3.4.1 coincides with that of Han's algorithm near $\hat{G}$. A result similar to Theorem 3.4.3 holds for this algorithm with

$$\rho_H = 1 - \min \left\{ \frac{\langle z, Q_H(\hat{\mu}) z \rangle}{\langle z, A_j^T G^j(A_j \hat{x}) A_j z \rangle} \;\middle|\; j \in \mathbf{p},\ z \in Range(Z) \right\}. \tag{3.5.9}$$

In general, $\rho_H > 0$, suggesting that only linear convergence is achieved despite the use of second order information. This is born out by the strictly positive convergence ratios observed for versions of the Han algorithm using an exact minimizing line search. While a sequence $\{x_i\}$ constructed by the Han algorithm with a fixed step size of 1 converges superlinearly to a minimizer, some iterations may produce an *increase* in $\psi(\cdot)$. It is likely that a *descent* algorithm based on Han's search direction will not be superlinearly convergent without the use of devices analogous to the feasibility enhancing corrections of some algorithms for nonlinear programming (see, for example, [May.1, Pan.1]).

## 3.6 CONCLUSIONS

We have introduced a variable metric technique which substantially mitigates the ill-conditioning produced in the composite minimax problem by the $A_j$ matrices. The technique does not require the evaluation of second derivatives and can be used as described in Algorithm Model 3.2.1 to speed the convergence of any first-order minimax algorithm which produces estimates of the optimal multipliers. We have analyzed the effect of the technique on the rate of convergence of the Pshenichnyi minimax algorithm. An upper bound on the convergence ratio was obtained for the variable metric version of the algorithm which can be considerably smaller than for the unscaled version. Numerical experiments verify the improvement suggested by the decrease in the convergence ratio bounds. The variable metric technique yielded a dramatic acceleration in convergence. The experiments also confirmed that the technique can speed convergence of another minimax algorithm. The variable metric technique can be applied without modification to a version of the Pshenchinyi algorithm for solving minimax problems involving semi-infinite composite max functions [Pol.4] of the form $\max_{y \in Y} \phi(x, y)$, where $Y \subset \mathbb{R}^r$ is a compact, but infinite set. The convergence rate analysis extends to this case as well.

## 3.7 APPENDIX

**Affine Parametrizations in Control:** We give two important examples of affine parametrization in control system design. The first is in the the design of a compensator according to $H^\infty$ criteria (see [Pol.5]), and the second is the construction of a discrete time control which keeps a trajectory within prescribed bounds. The resulting optimization problems involve composite functions with the inner function affine and the outer function possibly nondifferentiable.

Consider the feedback system $S(P, C)$ shown in Figure 3.3 where the plant, $P$, has a stabilizable, detectable state space representation. Let $\mathrm{U} \triangleq \{ s \in \mathbb{C} \mid \mathrm{Re}(s) > -\alpha_{\mathrm{U}} \}$, with $\alpha_{\mathrm{U}} \geq 0$, and let $\mathbb{R}_{\mathrm{U}}(s)$ be the set of rational functions that are bounded and analytic in U. If $N_r D_r^{-1}$ and $D_l^{-1} N_l$ are right and left coprime factorizations of $P$, with entries in $\mathbb{R}_{\mathrm{U}}(s)$, and $U_r$, $V_r$, $N_l$, $U_l$, $V_l$, with entries in $\mathbb{R}_{\mathrm{U}}(s)$, satisfying the Bezout identities [Des.1, You.1], then all compensators which ensure internal stability of the feedback system have the form $C(R) = (V_r - RN_l)^{-1}(U_r + RD_l)$ for some $R$ with entries in $\mathbb{R}_U(s)$. The transfer function of the closed loop system with such a compensator representation is given by

$$H(R) = \begin{bmatrix} -N_rRD_l+V_lD_l & N_rRN_l-N_rV_r & N_rRD_l-V_lD_l & N_rRD_l-V_lD_l \\ D_rRD_l+D_rU_r & -D_rRN_l+D_rV_r & -D_rRD_l-D_rU_r & -D_rRD_l-D_rU_r \\ N_rRD_l+N_rU_r & -N_rRN_l+N_rV_r & -N_rRD_l+V_lD_l & -N_rRD_l-N_rU_r \end{bmatrix}, \tag{3.7.1}$$

Referring to (3.7.1), we see that every transfer function in the feedback loop $S(P, C)$ is an *affine function* of the parameter $R$. Hence, referring to [Pol.5], we see that performance in command tracking, disturbance rejection, plant saturation avoidance and stability robustness can be expressed by inequalities of the form

$$\max_{\omega \in [0, \bar{\omega}]} \{ \bar{\sigma}[(G_l R\, G_r - F)(j\omega)] - b_f(\omega) \} \leq 0, \tag{3.7.2}$$

where $G_l$, $R$, $G_r$, $F$ are matrices with entries in $\mathbb{R}_U(s)$ and $b_f(\cdot)$ is a positive, bounded, lower semicontinuous function and $\bar{\omega} > 0$ is large.

A time domain criterion such as performance in following a given trajectory can be expressed in the form

$$\max_{t \in [0, t_0]} g(\mathbf{L}^{-1}\{(G_l R\, G_r - F)(s)\hat{u}(s)\}(t)) \leq 0, \tag{3.7.3}$$

where $g : \mathbb{R}^k \to \mathbb{R}$ is a differentiable, convex function, $e_k$ denotes the $k$-$th$ unit vector, and $G_l$, $R$, $G_r$, $F$ have elements in $\mathbb{R}_U(s)$.

The selection of an optimal compensator can be formulated in several ways, but, since it may not be clear *a priori* that the design requirements are consistent, it may be desirable to choose a compensator which minimizes the maximum violation of several such performance inequalities.

Computationally, one cannot deal with elements of $\mathbb{R}_U(s)^{n_i \times n_o}$. Hence in [Pol.5, Sal.1], the parameter $R \in \mathbb{R}_U(s)^{n_i \times n_o}$ of the compensator $C(R)$ was parametrized in terms of a vector $x \in \mathbb{R}^{n n_i n_o}$, with $n = 1, 2, 3, \cdots$, as follows. Define the matrices $X_i \in \mathbb{R}^{n_i \times n_o}$, $i = 1, 2, \ldots, n$, by filling them in order, row-wise, with the components of $x$, i.e.,

$$[X_i]_{k,l} \triangleq [x]_{(i-1)n_i n_o + (k-1)n_o + l}\ ,\ k \in \underline{n_i}\ ,\ l \in \underline{n_0}\,. \tag{3.7.4}$$

where for any $k \in \mathbb{N}$, $\underline{k}$ denotes the set $\{ 1, \ldots, k \}$. Let $p \in \mathbb{R}_+$, then [Pol.5, Sal.1] define $R_n : \mathbb{R}^{n n_i n_o} \to \mathbb{R}_U(s)^{n_i \times n_o}$ by

$$R_n(x)(s) \triangleq \sum_{i=1}^{n} X_i \left(\frac{s - p + \alpha_U}{s + p + \alpha_U}\right)^{i-1} . \tag{3.7.5}$$

It was shown in [Pol.5, Sal.1] that any $R$ with entries in $\mathbb{R}_U(s)$ can be approximated with arbitrary precision by an $R_n(x)$ for some $x$ and $n$. Thus, we can expect to obtain solutions reasonably close to the optimum over $\mathbb{R}_U(s)^{n_i \times n_o}$ if we choose $n$ large enough. Note that, with this parametrization, the compensator parameter, $R_n(x)$, is a *linear* function of the design variable vector, $x$.

Substituting (3.7.5) into (3.7.1), and then (3.7.1) into the functions in the performance inequalities (3.7.2) or (3.7.3), yields a function of the form, $\max_{y \in Y} \phi(Ax, y)$, where $Y$ is a compact, convex set in $\mathbb{R}^t$ for some $t$, $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear, and $\phi : \mathbb{R}^m \times Y \to \mathbb{R}$ is twice continuously differentiable in $x$ and continuous in $y$. (The maximum singular value function appearing in (3.7.2), $\bar{\sigma}(B)$, is not differentiable in $B$, but it can be written in the equivalent form $\max \{ \langle u , Bv \rangle \mid \|u\| = 1 ,\ \|v\| = 1 \}$.) It was shown in [Pol.4, Sal.1] that the functions in (3.7.2) and (3.7.3) are convex in the design parameter $R$. Thus, the optimal compensator design problem becomes a minimax problem involving convex, nondifferentiable functions which are composed with a linear function and has the form:

$$\mathbf{P1}: \quad \min_{x \in \mathbb{R}^{n n_i n_o}} \ \max_{k \in \mathbf{p}} \ \max_{y_k \in Y_k} \ \phi_k(A_k x, y_k). \tag{3.7.6}$$

Next consider a discrete time, time-varying linear system,

$$x_{i+1} = F_i x_i + G_i u_i , \quad x_0 = \bar{x} , \quad i = 1, 2, ..., N , \tag{3.7.7}$$

where $x_i \in \mathbb{R}^n$ is the state at time $i$ and $\mathbf{u} \in \mathbb{R}^{m \times N}$ is the control sequence. We assume that the control is required to be bounded for all time, i.e., $u_i \in U$, where $U \subset \mathbb{R}^m$ is compact. Suppose that we wish to find a feasible control that, as much as possible, keeps the trajectory within a prescribed tube, defined at $i$ by $\|x - \hat{x}_i\| \le b_i$. Then we obtain the minimax problem

$$\textbf{P2a}: \quad \min_{\mathbf{u} \in \mathbf{U}} \max_{i \in \underline{N}} \{ \|x_i - \hat{x}_i\| - b_i \} , \tag{3.7.8a}$$

where $\mathbf{U} \triangleq \{ \mathbf{u} \in \mathbb{R}^{m \times N} \mid u_i \in U , \; i = 1, 2, ..., N \}$. But $x_i = B_i \bar{x} + A_i \mathbf{u} \triangleq \prod_{j=0}^{i-1} F_j \bar{x} + \sum_{k=0}^{i-1} \prod_{j=i-1-k}^{i-1} F_j G_k u_k$. Hence (3.7.8a) is seen to be of the form

$$\textbf{P2b}: \quad \min_{\mathbf{u} \in \mathbf{U}} \max_{i \in \underline{N}} f^{\,i}(A_i \mathbf{u}) . \tag{3.7.8b}$$

| Algorithm | $\psi_i \le \hat{\psi} + 10^{-2}$ | | | $\psi_i \le \hat{\psi} + 10^{-4}$ | | |
|---|---|---|---|---|---|---|
| | Iterations | Function evaluations | Time (sec.) | Iterations | Function evaluations | Time (sec.) |
| Pshenichnyi | 291 | 5,246 | 11.6 | 397 | 7,154 | 15.9 |
| VM-Pshenichnyi | 4 | 80 | 0.3 | 6 | 116 | 0.4 |
| Han | 3 | 98 | 0.2 | 5 | 152 | 0.3 |
| Barrier | 45 | 6,806 | 9.4 | 48 | 16,640 | 22.1 |
| VM-Barrier | 40 | 2,276 | 4.8 | 43 | 2,812 | 5.7 |

**Table 3.1: Numerical Results for Problem 3.5.1.**

| Algorithm | $\psi_i \le \hat{\psi} + 10^{-2}$ | | | $\psi_i \le \hat{\psi} + 10^{-4}$ | | |
|---|---|---|---|---|---|---|
| | Iterations | Function evaluations | Time (sec.) | Iterations | Function evaluations | Time (sec.) |
| Pshenichnyi | 11628 | 976806 | 2317.9 | 11976 | 100603 | 2391.9 |
| VM-Pshenichnyi | 4 | 390 | 2.0 | 6 | 558 | 2.8 |
| Han | 4 | 1350 | 2.2 | 6 | 1902 | 3.1 |
| Barrier | 15 | 2,314,548 | 2,788.0 | 21 | 10,904,772 | 13,008.5 |
| VM-Barrier | 4 | 1422 | 7.1 | 11 | 4962 | 13.2 |

**Table 3.2: Numerical results for Problem 3.5.2.**

| Algorithm | Problem 5.1 | | | Problem 5.2 | | |
|---|---|---|---|---|---|---|
| | Convergence ratio | Convergence ratio bound | Iterations | Convergence ratio | Convergence ratio bound | Iterations |
| Pshenichnyi | .83 | .999979 | 109,646 | .9969 | .999994 | 383,763 |
| VM-Pshenichnyi | .67 | .697697 | 7 | .0805 | .937000 | 36 |
| Han | .0840 | .697697 | 7 | .0805 | .937000 | 36 |

**Table 3.3: Convergence ratios and bounds for Problems 3.5.1 and 3.5.2.**

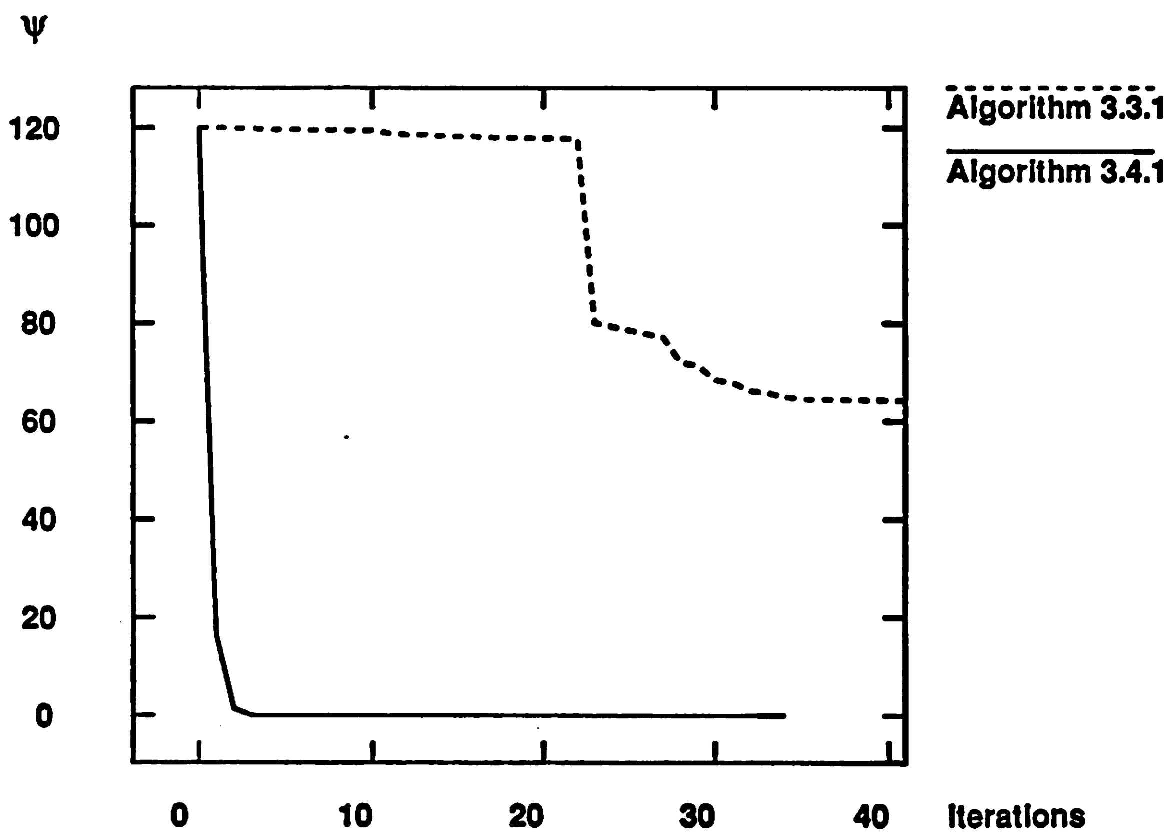


Figure 3.1: Problem 3.5.1

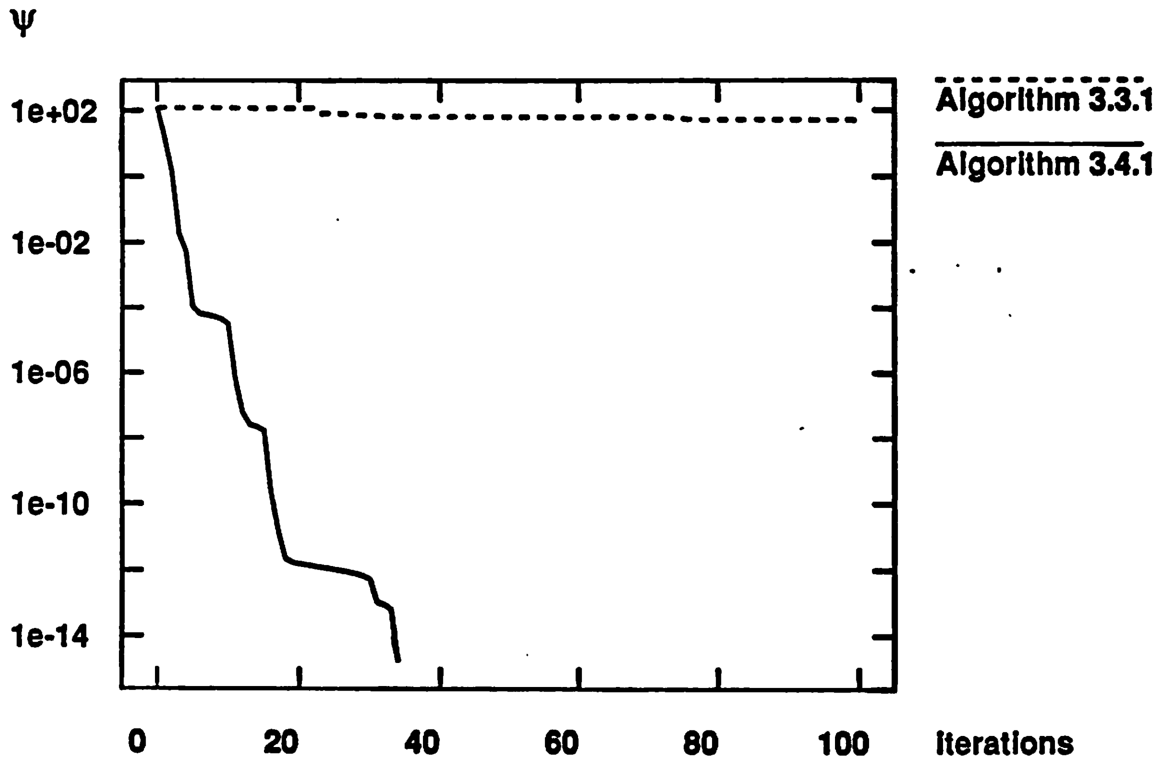


Figure 3.2: Problem 3.5.1 (log scale)

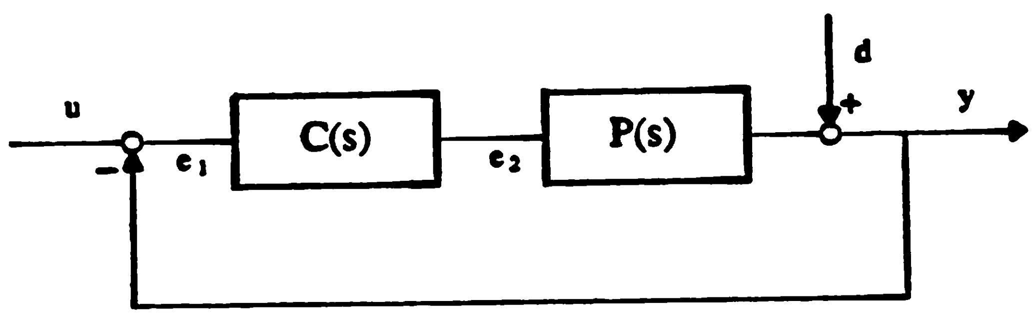


Figure 3.3: A Feedback System

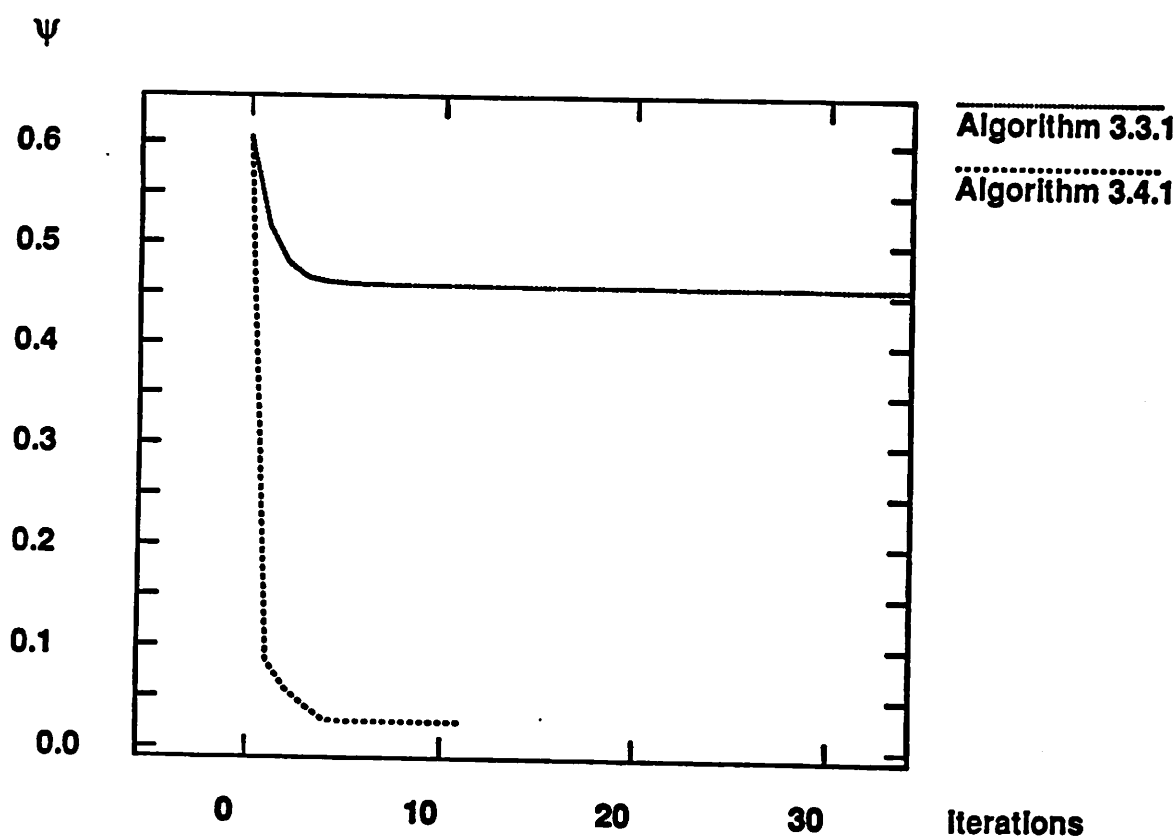


Figure 3.4: Problem 3.5.2

# CHAPTER 4

# A GENERALIZED QUADRATIC PROGRAMMING-BASED METHOD FOR INEQUALITY-CONSTRAINED OPTIMIZATION

## 4.1 INTRODUCTION

We now consider the inequality-constrained nonlinear programming problem,

$$\text{ICP} \qquad \min_{x \in \mathbb{R}^n} \{ f^0(x) \mid f^j(x) \le 0 \,,\ \forall j \in \mathbf{p} \} \,, \tag{4.1.1}$$

where p denotes the set of natural numbers $\{ 1, \dots, p \}$ and the functions $f^j : \mathbb{R}^n \to \mathbb{R}$, $j \in \mathbf{p} \cup \{ 0 \}$, are continuously differentiable. In [Pol.10], algorithms were proposed for the solution of problem (4.1.1) which obtain a search direction at each iteration by solving a natural approximation to (4.1.1) in which each function $f^j(\cdot)$ is replaced by the quadratic approximation $f^j(x) + \langle \nabla f^j(x), h \rangle + ½\langle h , H_j h \rangle$, for some $H_j \in \mathbb{R}^{n \times n}$. The resulting subproblem is a quadratic program with quadratic constraints, which we will call a *generalized quadratic program* (GQP):

$$\min_{h \in \mathbb{R}^n} \{ f^0(x) + \langle \nabla f^j(x), h \rangle + ½\langle h , H_j h \rangle \mid$$

$$f^j(x) + \langle \nabla f^j(x), h \rangle + ½\langle h , H_j h \rangle \le 0 \quad \forall j \in \mathbf{p} \} \,. \tag{4.1.2}$$

The use of GQP subproblems in algorithms for the solution of (4.1.1) offers some potential advantages over the use of quadratic programs. Information about the curvature of individual constraints can be incorporated directly into the constraints of the GQP subproblem. If the matrices $H_j$ are positive definite and the current iterate is feasible, the resulting search direction is a feasible descent direction. This chapter presents the first thorough analysis of convergence and rate of convergence of an *implementable* GQP-based algorithm.

There has been some theoretical analysis of GQP-based algorithms. The convergence of conceptual phase II[1] algorithms is treated in [Pol.10]. Rates of convergence are obtained for GQP-based minimax algorithms in [Pol.9, 11] under assumptions of uniform convexity. It is shown in [Pan.3] that, on uniformly convex problems, the norms of the search directions constructed by a conceptual GQP-based algorithm converge superlinearly to zero as the iterates approach a solution.[2]

The GQP-based algorithms proposed in [Pol.10, Pan.3-4] were conceptual, that is, they assumed that the GQP subproblem is solved exactly. These algorithms were not implemented (to our knowledge) because no finite step procedures for solving problem (4.1.2) were known [Pol.10, Pan.4]. Furthermore, (4.1.2) may not have feasible solutions if $x$ is infeasible for (4.1.1). In this chapter, we resolve these difficulties for the case of first-order information, where each $H_j$ is taken to be a multiple of the identity.

Our GQP-based method approximates the solution to (4.1.2) by adding a correction to the search direction of the Polak-Mayne-Trahan algorithm [Pol.7, Pir.1]. The approximation is exact under certain conditions, and requires the solution of only one quadratic program and a projection operation. The method uses the Polak-Mayne-Trahan search direction when no solution to (4.1.2) exists.

Because we set each $H_j$ in (4.1.2) to a multiple of the identity, the search direction at each feasible point is a feasible descent direction. Hence, once the algorithm constructs a feasible point, the inequalities

$$f^j(x_i) \le 0 \quad \forall j \in \mathbf{p} \quad \text{and} \quad f^0(x_{i+1}) < f^0(x_i)\,, \tag{4.1.3}$$

hold for all subsequent iterates. This property is important in engineering design problems for which function evaluations are extremely costly and for which designs failing to satisfy

[1]Phase I in the solution of (4.1.1) refers to the computation of a feasible point. Phase II refers to the improvement of this solution by a feasible, descent algorithm, which decreases the the value of the objective function while maintaining feasibility.

[2]Quadratic constraints have also appeared in the subproblems of trust region algorithms [More.1]. However, in these algorithms, they function to limit the search direction, rather than to represent the constraints of the problem.

specifications are unacceptable [Nye.1]. Other first-order algorithms satisfying these requirements include [Hua.1-2, Mey.1, Mif.1, Pir.1, Pol.4, Pol.7, Top.1, Her.1].

We compare the efficiency of our GQP-based algorithm with that of the Polak-Mayne-Trahan algorithm, because the GQP-based algorithm can be viewed as a modification of the Polak-Mayne-Trahan algorithm and because the Polak-Mayne-Trahan algorithm satisfies (4.1.3) and has been shown to converge linearly in Phase II [Pir.1, Cha.1] under convexity assumptions. We show that the GQP-based algorithm converges linearly with a smaller bound on the *convergence ratio* of the sequence of cost values[3] than that obtained for the Polak-Mayne-Trahan algorithm. Numerical experiments also show the new algorithm to be generally superior to the Polak-Mayne-Trahan algorithm, and competitive with the feasible descent algorithm of [Her.1].

The GQP-based algorithm presented in this chapter accepts infeasible starting points, and a linear rate of convergence obtains even if the sequence of iterates approaches feasibility only asymptotically.

In Section 3, convergence and rate of convergence results are obtained for a local, conceptual GQP-based algorithm. In Section 4, an implementation of the local, conceptual algorithm is developed. In Section 5, the convergence and rate of convergence results are obtained for the stabilized, implementable algorithm, and the results of numerical experiments are presented in Section 6. The properties of the Polak-Mayne-Trahan algorithm are reviewed in the next section.

## 4.2 THE POLAK-MAYNE-TRAHAN ALGORITHM

The Polak-Mayne-Trahan (PMT) algorithm [Pol.7] is a phase I - phase II extension of the Pironneau-Polak algorithm [Pir.1], which, in turn, is an implementation of Huard's method of centers [Hua.1-2]. The PMT algorithm is one of very few first-order phase I - phase II methods for which the rate of convergence is known (see also [Pol.8]), and its computational behavior is

[3]Recall that the *convergence ratio* of a sequence $\{x_i\}_{i \in N}$ which converges to $\hat{x}$ is defined as $\limsup_{i \to \infty} (f^0(x_{i+1}) - f^0(\hat{x})) / (f^0(x_i) - f^0(\hat{x}))$.

quite competitive in this class. We will use the PMT algorithm as a benchmark for evaluating the new algorithm. The PMT algorithm solves problems of the form

$$\min \{ f^0(x) \mid f^j(x) \le 0,\ j \in \mathbf{p} \} , \tag{4.2.1a}$$

under the assumption that the functions $f^j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable and the constraint qualification that the function $\max_{j \in \mathbf{p}} f^j(x)$ not have any stationary points outside the interior of the feasible set.

We will use the following definitions. We denote the set of natural numbers $\{ 1, ..., p \}$ by $\mathbf{p}$, and the set $\{ 0, 1, ..., p \}$ by $\mathbf{p} \cup 0$. The $p$ smooth constraints $f^j(x) \le 0$ in (4.1.1) can be combined into a single nonsmooth constraint $\psi(x) \le 0$, where $\psi(x) \triangleq \max_{j \in \mathbf{p}} f^j(x)$. Constraint violation is indicated by the values of the function $\psi_+(x) \triangleq \max \{ \psi(x), 0 \}$. Finally, we define first-order convex approximations to the functions $f^j(\cdot)$ at $x$ by

$$\tilde{f}^j(h \mid x) \triangleq \begin{cases} f^j(x) + \langle \nabla f^j(x), h \rangle + \tfrac{1}{2}\gamma \| h \|^2 & \text{if } j \in \mathbf{p} \\ \langle \nabla f^0(x), h \rangle + \tfrac{1}{2}\gamma \| h \|^2 & \text{if } j = 0 , \end{cases} \tag{4.2.1b}$$

for some fixed $\gamma > 0$. Note that $\tilde{f}^0(0 \mid x) = 0$ and $\tilde{f}^j(0 \mid x) = f^j(x)$ for all $j \in \mathbf{p}$.

**Algorithm 4.2.1**

**Data:** $x_0$; $\alpha, \beta \in (0, 1)$; $\gamma > 0$; $i = 0$.

**Step 1:** Compute the *search direction*,

$$h(x_i) \triangleq \operatorname*{argmin}_{h \in \mathbb{R}^n} \max_{j \in \mathbf{p} \cup 0} \tilde{f}^j(h \mid x_i) , \tag{4.2.2a}$$

and evaluate the *optimality function*,

$$\theta(x_i) \triangleq \max_{j \in \mathbf{p} \cup 0} \tilde{f}^j(h(x_i) \mid x_i) - \psi_+(x_i) . \tag{4.2.2b}$$

**Step 2:** If $\psi(x_i) \le 0$, set

$$\lambda_i = \max \{ \beta^k \mid f^0(x_i + \beta^k h(x_i)) - f^0(x_i) \le \alpha \beta^k \theta(x_i) \text{ and } \psi(x_i + \beta^k h(x_i)) \le 0 \} , \tag{4.2.2c}$$

else set

$$\lambda_i = \max \{ \beta^k \mid \psi(x_i + \beta^k h(x_i)) - \psi(x_i) \le \alpha\beta^k \theta(x_i) \} \ . \tag{4.2.2d}$$

**Step 3:** Set $x_{i+1} = x_i + \lambda_i h(x_i)$.

**Step 4:** Replace $i$ by $i+1$, and go to Step 1.

□

Step 2 ensures that, once a sequence generated by Algorithm 4.2.1 has entered the feasible region $X \triangleq \{ x \in \mathbb{R}^n \mid f^j(x) \le 0 \, \forall j \in \mathbf{p} \}$, it can never leave it. Referring to [Pol.4] we see that the search direction vector $h(x_i)$ can be computed in two steps. First one solves the dual of (4.2.2a), i.e., the positive semi-definite quadratic program

$$\max_{\mu \in \Sigma_{p+1}} \min_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p} \cup 0} \mu^j \tilde{f}^j(h \mid x) - \psi_+(x)$$

$$= \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \underline{p}} \mu^j f^j(x) - \psi_+(x) - \tfrac{1}{2}\gamma^{-1} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2 , \tag{4.2.3}$$

for any solution $\mu(x_i)$. We denote the set of solutions to (4.2.3) by $U_{\mathrm{PP}}(x) \triangleq \operatorname{argmax}_{\mu \in \Sigma_{p+1}} \sum_{j \in \underline{p}} \mu^j f^j(x) - \psi_+(x) - \tfrac{1}{2}\gamma^{-1} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2$. This can be done using one of several methods [Gil.1, von.1, Hig.1, Kiw.2-3, Rus.1]. The unique primal solution, $h(x_i)$, is then given by

$$h(x) = \operatorname*{argmin}_{h \in \mathbb{R}^n} \sum_{j \in \mathbf{p} \cup 0} \mu^j(x) \tilde{f}^j(h \mid x) = \frac{1}{\gamma} \sum_{j \in \mathbf{p} \cup 0} \mu^j(x) \nabla f^j(x) \, . \tag{4.2.4a}$$

From (4.2.3), we can write

$$\theta(x) = \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \underline{p}} \mu^j f^j(x) - \psi_+(x) - \tfrac{1}{2}\gamma^{-1} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2 . \tag{4.2.4b}$$

The following theorem summarizes the properties of the optimality function $\theta : \mathbb{R}^n \to \mathbb{R}$, the search direction function $h : \mathbb{R}^n \to \mathbb{R}^n$ used in the above algorithm.

**Theorem 4.2.1: [Pol.4]**

(a) *If $\hat{x}$ is a local minimizer for problem (4.1.1), then* $\theta(\hat{x}) = 0$.

(b) *For any $\bar{x} \in \mathbb{R}^n$, $\theta(\bar{x}) = 0$ if and only if there exists $\hat{\mu} \in \Sigma_{p+1}$ such that*

$$\sum_{j \in \mathbf{p} \cup 0} \hat{\mu}^j \nabla f^j(\bar{x}) = 0 , \tag{4.2.5a}$$

$$\sum_{j \in \mathbf{p}} \hat{\mu}^j f^j(\bar{x}) = \psi_+(\bar{x}) . \tag{4.2.5b}$$

(c) *Both* $\theta(\cdot)$ *and* $h(\cdot)$ *are continuous.*

Note that if $\bar{x}$ satisfies (4.2.5a-b) for some $\mu \in \Sigma_{p+1}$, and $\psi_+(\bar{x}) > 0$, then $\mu^0 = 0$, and hence $\bar{x}$ satisfies the standard first order condition for a local minimizer of $\psi(\cdot)$. If $\hat{x}$ is a local minimizer of (4.1.1), then $U_{\mathbf{pp}}(\hat{x})$ is the set of Fritz John multiplier vectors which, together with $\hat{x}$, satisfy (4.2.5a-b).

**Theorem 4.2.2:** [Pol.4] *If $\bar{x}$ is an accumulation point of a sequence $\{ x_i \}_{i=0}^{\infty}$ constructed by Algorithm 4.2.1 in solving (4.1.1), then* $\theta(\bar{x}) = 0$. *Furthermore, if, for all $x \in \mathbb{R}^n$ such that* $\psi(x) \geq 0$, $0 \notin \partial\psi(x)$ *(where $\partial\psi(x)$ denotes the generalized gradient of $\psi(\cdot)$ at $x$* [Cla.1]*), then* $\psi(\bar{x}) \leq 0$.

It was first shown in [Pir.1] that an algorithm based on the search direction rule (4.2.2a) converges linearly under convexity assumptions. Chaney [Cha.1] later established linear convergence under slightly weaker assumptions. The following theorem is a variant of Chaney's result, accounting for the fact that Algorithm 4.2.1 uses an Armijo-type line search [Arm.1] rather than an exact minimizing line search as in [Cha.1, Pir.1]. Let

$$F^j(x) \triangleq \partial^2 f^j(x)/\partial x^2 , \tag{4.2.6}$$

and $\underline{\mu}^0 \triangleq \min \{ \mu^0 \mid \mu \in U_{\mathbf{pp}}(\hat{x}) \}$.

**Theorem 4.2.3:** *Suppose that*

(i) *the functions* $f^j(\cdot)$, $j \in \mathbf{p} \cup 0$ *are twice continuously differentiable,*

(ii) *the set* $L \triangleq \{ x \in \mathbb{R}^n \mid \psi(x) \le \psi_+(x_0) \}$ *is bounded, and the necessary conditions* (4.2.5a-b) *are satisfied at a single point,* $\hat{x} \in X$, *at which the Mangasarian-Fromovitz constraint qualification holds (i.e. - there exist* $\bar{h} \in \mathbb{R}^n$ *and* $\delta > 0$ *such that* $\langle \nabla f^j(\hat{x}), \bar{h} \rangle < -\delta$ *for each* $j \in \mathbf{p}$ *such that* $f^j(\hat{x}) = 0$*),*

(iii) *for* $\hat{x}$ *as above, and with*

$$\hat{J} \triangleq \bigcup \{ J(\mu) \mid \mu \in U_{PP}(\hat{x}) \} , \tag{4.2.7a}$$

*where for any* $\mu \in \Sigma_{p+1}$, $J(\mu) \triangleq \{ j \in \mathbf{p} \mid \mu^j > 0 \}$, *there exists* $m \in (0, \gamma)$ *such that*

$$m \| h \|^2 < \langle h , \left[ \sum_{j \in \mathbf{p} \cup 0} \mu^j F^j(\hat{x}) \right] h \rangle , \tag{4.2.7b}$$

*for every* $\mu \in U_{PP}(\hat{x})$ *and for every nonzero* $h \in H$, *where*

$$H \triangleq \{ h \mid \langle \nabla f^j(\hat{x}), h \rangle = 0, \ \forall j \in \hat{J} \ \} . \tag{4.2.7c}$$

*If Algorithm 4.2.1 constructs a sequence* $\{ x_i \}_{i=0}^{\infty}$ *in solving problem (4.1.1), then* (a) $x_i \to \hat{x}$ *as* $i \to \infty$, *and* (b) *if* $\psi(x_i) \le 0$ *for any* $i \in \mathbb{N}$, *then*

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \le 1 - \alpha\beta\underline{\mu}^0 \frac{m}{M} , \tag{4.2.7d}$$

*for any* $M > \max_{j \in \mathbf{p} \cup 0} \{ \| F^j(\hat{x}) \|, \gamma \}$.

Inequality (4.2.7d) then gives an upper bound on the convergence ratio of sequences constructed by Algorithm 4.2.1.

## 4.3 A CONCEPTUAL GQP-BASED ALGORITHM

We begin by considering a conceptual, local algorithm for solving (4.1.1) which computes a search direction at $x_i$ by solving the *generalized quadratic program*,

$$\mathrm{GQP}(x): \min_{h \in \mathbb{R}^n} \{ \tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0 \quad \forall j \in \mathbf{p} \} \ , \tag{4.3.1a}$$

with $x = x_i$.

**Local Algorithm 4.3.1:**

**Data:** $x_0$; $\beta \in (0,1)$; $\gamma > 0$; $i = 0$.

**Step 1:** Compute the search direction,

$$h_i = h_{\mathrm{GQP}}(x_i) \triangleq \operatorname*{argmin}_{h \in \mathbb{R}^n} \{ \tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0 \quad \forall j \in \mathbf{p} \} \ . \tag{4.3.1b}$$

**Step 2:** Compute the step size,

$$\lambda_i = \max \{ \beta^k \mid f^0(x_i + \beta^k h_i) - f^0(x_i) \le \beta^k \tilde{f}^0(h_i \mid x_i),$$

$$\psi_+(x_i + \beta^k h_i) - \psi_+(x_i) \le \beta^k [\max_{j \in \mathbf{p}} \{ \tilde{f}^j(h_i \mid x_i), 0 \} - \psi_+(x_i)] \} \tag{4.3.1c}$$

**Step 3:** Set $x_{i+1} = x_i + \lambda_i h_i$.

**Step 4:** Replace $i$ by $i+1$, and go to Step 1.

□

Lemma 4.3.1: *Suppose that assumptions* (i)-(iii) *of Theorem 4.2.3 hold, and let* $\hat{x}$ *be as defined in assumption* (ii) *of Theorem 4.2.3. Then there exists a neighborhood* $V$ *of* $\hat{x}$ *such that* GQP($x$) *has a continuous solution,* $h_{\mathrm{GQP}}(x)$, *for all* $x \in V$.

Proof: Suppose that $x \in \mathbb{R}^n$ is such that there exists an $h' \in \mathbb{R}^n$ satisfying $\tilde{f}^j(h' \mid x) < 0$ for all $j \in \mathbf{p}$. Then the set-valued map $G(x) \triangleq \{ h \in \mathbb{R}^n \mid \tilde{f}^j(h \mid x) \le 0, \ \forall \ j \in \mathbf{p} \}$ is upper semicontinuous at $x$. $G(x)$ is compact since the functions $\tilde{f}^j(\cdot)$ are uniformly convex. Hence, by the Maximum Theorem [Ber.1], the set of solutions to GQP($x$) is a compact-valued, upper semicontinuous set-valued map at $x$. Since GQP($x$) is a strictly convex program, its solution set is a singleton, $\{ h_{\mathrm{GQP}}(x) \}$. Therefore, the solution, $h_{\mathrm{GQP}}(x)$, to GQP($x$) is continuous at any point $x$ at which GQP($x$) is strictly feasible.

By assumption (ii) of Theorem 4.2.3, there exist $\bar{h} \in \mathbb{R}^n$ and $\delta > 0$ such that $\langle \nabla f^j(\hat{x}), \bar{h} \rangle < -\delta$ for each $j \in J(\hat{x})$. Therefore, there exist $t > 0$ and a neighborhood, $V$, of $\hat{x}$ such that $\tilde{f}^j(t\bar{h} \mid x) < 0$ for all $x \in V$ and $j \in \mathbf{p}$. Therefore, GQP($x$) is strictly feasible for all $x \in V$. In light of the previous paragraph, $h_{\mathrm{GQP}}(x)$ exists and is continuous in $V$.

□

For any $x \in \mathbb{R}^n$ such that GQP($x$) has a solution, we will denote the set of *Fritz John* multiplier vectors associated with the unique solution, $h_{\mathrm{GQP}}(x)$ by

$$U_{\mathrm{GQP}}(x) \triangleq \{ \mu \in \Sigma_{p+1} \mid \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla \tilde{f}^j(h_{\mathrm{GQP}}(x) \mid x) = 0,$$

$$\sum_{j \in \mathbf{p}} \mu^j \tilde{f}^j(h_{\mathrm{GQP}}(x) \mid x) = 0 \} . \tag{4.3.1d}$$

Consider the $l_\infty$ penalty function, $p_\varepsilon(x) \triangleq \varepsilon f^0(x) + \psi_+(x)$, where $\varepsilon > 0$. The proofs below exploit the correspondence between minimizers of the constrained problem (4.1.1) and those of the minimax problem,

$$\min_{x \in \mathbb{R}^n} p_\varepsilon(x) . \tag{4.3.2a}$$

As is shown in the following lemma, the solution to (4.1.1) is also a strict local minimizer of $p_\varepsilon(\cdot)$ for sufficiently small $\varepsilon$. Let

$$d_\varepsilon(x) \triangleq \operatorname*{argmin}_{h \in \mathbb{R}^n} \{ \varepsilon \tilde{f}^0(h \mid x) + \max \{ 0, \tilde{f}^j(h \mid x) \} \} , \tag{4.3.2b}$$

and let

$$\theta_\varepsilon(x) \triangleq \varepsilon \tilde{f}^0(d_\varepsilon(x) \mid x) + \max \{ 0, \tilde{f}^j(d_\varepsilon(x) \mid x) \} - \psi_+(x) . \tag{4.3.2c}$$

Recall that $\underline{\mu}^0 \triangleq \min \{ \mu^0 \mid \mu \in U_{\mathrm{PP}}(\hat{x}) \}$.

**Lemma 4.3.2:** *Suppose that assumptions* (i)-(iii) *of Theorem 4.2.3 hold, let* $\hat{x}$ *be as defined in assumption* (ii) *of Theorem 4.2.3, and let* $V$ *be as defined in Lemma 4.3.1. Then, for any* $\varepsilon \in (0, \underline{\mu}^0/(1-\underline{\mu}^0))$, *there exists a neighborhood,* $W_\varepsilon \subset V$, *of* $\hat{x}$, *such that, for all* $x \in W_\varepsilon$, (a) $p_\varepsilon(x) \geq p_\varepsilon(\hat{x}) + \frac{1}{4}m\|x - \hat{x}\|^2$, *and* (b) $d_\varepsilon(x) = h_{GQP}(x)$.

**Proof:** (a) Assumptions (i)-(iii) of Theorem 4.2.3 ensure that the point $\hat{x}$ satisfies the standard second-order sufficiency conditions for problem (4.1.1) [McC.1]. In fact, they ensure that $\hat{x}$ satisfies these conditions for the problem,

$$\min_{x \in \mathbb{R}^n} \{ f^0(x) - \tfrac{1}{4}m\|x - \hat{x}\|^2 \mid f^j(x) - \tfrac{1}{4}m\|x - \hat{x}\|^2 \leq 0 \} . \tag{4.3.2d}$$

It follows from Theorem 4.6 of [Han.2] (see Theorem 4.8.1 of the Appendix for a restatement), therefore, that $\hat{x}$ is a strict local minimizer of $p_\varepsilon(\cdot) - \frac{1}{4}m(\varepsilon+1)\|\cdot - \hat{x}\|^2$, provided that $1/\varepsilon > \sum_{j \in \mathbf{p}} u^j$ for some Kuhn-Tucker multiplier vector for the problem (4.3.2d), $u \in \mathbb{R}^p$, associated with $\hat{x}$. Since the Kuhn-Tucker multiplier vectors for (4.3.2d) associated with $\hat{x}$ are the same as those of (4.1.1), we can construct a Kuhn-Tucker multiplier vector for (4.3.2d) from any Fritz John multiplier vector, $\mu \in U_{PP}(\hat{x})$, as follows:

$$u_\mu = (\mu^1, \ldots, \mu^p)/\mu^0 , \tag{4.3.2e}$$

because the Mangasarian-Fromovitz constraint qualification (assumption (ii) of Theorem 4.2.3) ensures that $\mu^0 \geq \underline{\mu}^0 > 0$. Hence, if $1/\varepsilon > \|u_\mu\|_1 = (1-\underline{\mu}^0)/\underline{\mu}^0$, then $\hat{x}$ is a strict local minimizer of $p_\varepsilon(\cdot) - \frac{1}{4}m(\varepsilon+1)\|\cdot - \hat{x}\|^2$. This implies that $p_\varepsilon(\hat{x}) \leq p_\varepsilon(x) - \frac{1}{4}m(\varepsilon+1)\|x - \hat{x}\|^2$ for $x$ in some neighborhood of $\hat{x}$.

(b) We recall that by Lemma 4.3.1, the solution $h_{GQP}(x)$ to GQP$(x)$ exists for all $x$ in a neighborhood $V$ of $\hat{x}$. We will now prove that, for any $\varepsilon < \underline{\mu}^0/(1-\underline{\mu}^0)$, $d_\varepsilon(x) = h_{GQP}(x)$ for all $x$ in a neighborhood of $\hat{x}$.

We first show that, for $x$ near $\hat{x}$, the norm of some Kuhn-Tucker multiplier vector associated with the solution to GQP($x$) is bounded from above by $(1-\underline{\mu}^0)/\underline{\mu}^0$. We denote the set of Kuhn-Tucker multiplier vectors for GQP($x$) by

$$KT_{\mathrm{GQP}}(x) \triangleq \{ u \in \mathbb{R}^p_+ \mid \nabla \tilde{f}^0(h_{\mathrm{GQP}}(x) \mid x) + \sum_{j \in \mathbf{p}} u^j \nabla \tilde{f}^j(h_{\mathrm{GQP}}(x) \mid x) = 0,$$

$$\sum_{j \in \mathbf{p}} \mu^j \tilde{f}^j(h_{\mathrm{GQP}}(x) \mid x) = 0 \} \ , \qquad (4.3.3)$$

for $x \in V$. Since $h_{\mathrm{GQP}}(\hat{x}) = 0$ and $\psi_+(\hat{x}) = 0$, an inspection of (4.3.1d) reveals that $U_{\mathrm{GQP}}(\hat{x}) = U_{\mathrm{PP}}(\hat{x})$. By assumption (ii) of Theorem 4.2.3, $\underline{\mu}^0 > 0$. Since $U_{\mathrm{GQP}}(\cdot)$ is an upper semicontinuous, compact-valued set-valued map at $\hat{x}$, there exists, for any $\delta \in (0, \underline{\mu})$, a neighborhood, $W_\delta \subset V$, of $\hat{x}$, such that $\mu^0 > \underline{\mu}^0 - \delta$ for every $\mu \in U_{\mathrm{GQP}}(W_\delta)$. Now, every Fritz John multiplier vector $\mu \in U_{\mathrm{GQP}}(x)$ corresponds to a Kuhn-Tucker multiplier vector, $u_\mu \triangleq (\mu^1/\mu^0, \ldots, \mu^p/\mu^0) \in KT_{\mathrm{GQP}}(x)$. For such Kuhn-Tucker multiplier vectors, $\|u_\mu\|_1 = (1-\mu^0)/\mu^0 < (1-\underline{\mu}^0+\delta)/(\underline{\mu}^0-\delta)$ for every $\mu \in U_{\mathrm{GQP}}(W_\delta)$.

Because (i) for any $\delta \in (0, \underline{\mu})$, there exists a neighborhood $W_\delta$ of $\hat{x}$ such that $\min\{ \|u\|_1 \mid u \in KT_{\mathrm{GQP}}(x) \} < (1-\underline{\mu}^0+\delta)/(\underline{\mu}^0-\delta)$ for $x \in W_\delta$ (from the previous paragraph), (ii) $\max_{j \in \mathbf{p}} \tilde{f}^j(h' \mid x) < 0$ for $x \in V$ and some $h' \in \mathbb{R}^n$ (from the proof of Lemma 4.3.1), and (iii) problem GQP(x) is a convex program, we can apply Theorem 4.9 of [Han.2] to conclude that, for $\varepsilon < (\hat{\mu}^0-\delta)/(1-\hat{\mu}^0+\delta)$, $h_{\mathrm{GQP}}(x)$ is the unique minimizer of the convex function $\min_{d \in \mathbb{R}^n} \varepsilon \tilde{f}^0(d \mid x) + \max\{0, \tilde{f}^j(d \mid x)\}$ for all $x \in W_\delta$. (See Theorem 4.8.2 of the Appendix for a restatement of Theorem 4.9 of [Han.2].) Hence, $h_{\mathrm{GQP}}(x) = d_\varepsilon(x)$ for all $x \in W_\delta$. Since $\delta$ was arbitrary, such a neighborhood exists for any $\varepsilon < \underline{\mu}^0/(1-\underline{\mu}^0)$.

□

Theorem 4.3.1: *Suppose that assumptions* (i)-(iii) *of Theorem 4.2.3 hold, and let* $\hat{x}$ *be as defined in assumption* (ii) *of Theorem 4.2.3. Then, for any neighborhood,* $W$, *of* $\hat{x}$, *there exists a neigh-*

*borhood*

*$V_w \subset W$ of $\hat{x}$ such that, if $x_0 \in V_w$, the sequence $\{x_i\}_{i \in \mathbf{N}}$ constructed by Algorithm 4.3.1 remains in $V_w$ and converges to $\hat{x}$.*

**Proof:** Let $A(\cdot)$ denote the *iteration map* of Algorithm 4.3.1. The function $A(\cdot)$ maps one iterate into the next, i.e., $x_{i+1} = A(x_i)$. The sequence $\{x_i\}_{i \in \mathbf{N}}$ will remain in a set $V_w$ if the set $V_w$ is invariant under $A(\cdot)$, i.e., $A(V_w) \subset V_w$. We now show that such a neighborhood $V_w \subset W$ of $\hat{x}$ exists.

Let $\varepsilon < \underline{\mu}^0 / (1 - \underline{\mu}^0)$ be arbitrary. By Lemma 4.3.2(a), there exists a neighborhood $W_\varepsilon$ of $\hat{x}$ such that $p_\varepsilon(x) \geq p_\varepsilon(\hat{x}) + \frac{1}{4}m' \|x - \hat{x}\|^2$ for $x \in W_\varepsilon$. For small enough $\delta > 0$, therefore, the set $V_w \triangleq \{x \in W_\varepsilon \mid p_\varepsilon(x) < p_\varepsilon(\hat{x}) + \delta\}$ is contained in $W$. By the continuity of $p_\varepsilon(\cdot)$, the set $V_w$ is a neighborhood of $\hat{x}$.

By Step 2 of Algorithm 4.3.1, with $x_1 = A(x_0)$ for any $x_0 \in V$,

$$p_\varepsilon(x_1) - p_\varepsilon(x_0) = \varepsilon[f^0(x_1) - f^0(x_0)] + [\psi_+(x_1) - \psi_+(x_0)]$$

$$\leq \lambda_0[\varepsilon\tilde{f}^0(h_{GQP}(x_0) \mid x_0) + \max_{j \in \mathbf{p}} \{\tilde{f}^j(h_{GQP}(x_0) \mid x_0), 0\} - \psi_+(x_0)]. \tag{4.3.4a}$$

By Lemma 4.3.2(b), $h_{GQP}(x_0) = h_\varepsilon(x_0)$ for $x_0 \in V_w$, and, hence,

$$p_\varepsilon(x_1) - p_\varepsilon(x_0) \leq \lambda_0[\varepsilon\tilde{f}^0(h_\varepsilon(x_0) \mid x_0) + \max_{j \in \mathbf{p}} \{\tilde{f}^j(h_\varepsilon(x_0) \mid x_0), 0\} - \psi_+(x_0)]$$

$$= \lambda_0\theta_\varepsilon(x_0) \leq 0. \tag{4.3.4b}$$

Therefore, $p_\varepsilon(x_1) \leq p_\varepsilon(x_0) \leq p_\varepsilon(\hat{x}) + \delta$, implying that $A(V_w) \subset V_w$.

Now we show that only $\hat{x}$ can be an accumulation point of the sequence $\{x_i\}_{i \in \mathbf{N}}$ constructed by Algorithm 4.3.1, from an $x_0 \in V_w$. Suppose that $\{x_i\}_{i \in K}$ converges to $\bar{x} \in V_w$, where $K \subset \mathbf{N}$ and $\bar{x} \neq \hat{x}$. Since, by assumption (ii) of Theorem 4.2.3, $\hat{x}$ is the only stationary point for (4.1.1) in $V_{\bar{w}}$, $\bar{x}$ cannot be stationary for problem (4.3.2a). By Lemma 4.3.1, $\tilde{f}^{0}(h_{GQP}(x) \mid x))$ is continuous in $V_{\bar{w}}$, and therefore there exist $\delta > 0$ and a neighborhood, $W' \subset V_{\bar{w}}$, of $\bar{x}$ such that

$$\theta_{\varepsilon}(x) < -\delta\,, \tag{4.3.4c}$$

for all $x \in W'$. Clearly, there exists an $i_0 \in K$, such that $x_i \in W'$ for all $i > i_0$, $i \in K$. Let $M' < \infty$ be such that $\|F^{j}(x)\| \le M'$ for all $x \in W$. Then,

$$f^{j}(x_i + \lambda h_{GQP}(x_i)) \le f^{j}(x_i) + \lambda \langle \nabla f^{j}(x_i), h_{GQP}(x_i) \rangle + \tfrac{1}{2} M' \lambda^2 \|h_{GQP}(x_i)\|^2\,, \tag{4.3.4d}$$

for all $i \in K$, $i > i_0$. Hence, for $j \in \mathbf{p}$ and $\lambda \le 1$,

$$f^{j}(x_i + \lambda h_{GQP}(x_i)) - \psi_{+}(x_i) \le \lambda \{ f^{j}(x_i) + \langle \nabla f^{j}(x_i), h_{GQP}(x_i) \rangle + \tfrac{1}{2} M' \lambda \|h_{GQP}(x_i)\|^2 - \psi_{+}(x_i) \}, \tag{4.3.4e}$$

since $\psi_{+}(x_i) \ge f^{j}(x_i)$. For $\lambda \le \gamma / M'$, then

$$f^{j}(x_i + \lambda h_{GQP}(x_i)) - \psi_{+}(x_i) \le \lambda \{ f^{j}(x_i) + \langle \nabla f^{j}(x_i), h_{GQP}(x_i) \rangle + \tfrac{1}{2} \gamma \|h_{GQP}(x_i)\|^2 - \psi_{+}(x_i) \}$$

$$= \lambda \{ \tilde{f}^{j}(h_{GQP}(x_i) \mid x_i) - \psi_{+}(x_i) \}\ . \tag{4.3.4f}$$

Taking the maximum over $j \in \mathbf{p}$,

$$\psi_{+}(x_i + \lambda h_{GQP}(x_i)) - \psi_{+}(x_i) \le \lambda \left\{ \max_{j \in \mathbf{p}} \{ 0, \tilde{f}^{\,j}(h_{GQP}(x_i) \mid x_i) \} - \psi_{+}(x_i) \right\}, \tag{4.3.4g}$$

for all $\lambda \in (0, \gamma / M']$ and $i > i_0$, $i \in K$. Setting $j = 0$ in (4.3.4d),

$$f^{0}(x_i + \lambda h_{GQP}(x_i)) - f^{0}(x_i) \le \lambda \{ \langle \nabla f^{j}(x_i), h_{GQP}(x_i) \rangle + \tfrac{1}{2} \gamma \|h_{GQP}(x_i)\|^2 \}$$

$$= \lambda \tilde{f}^{0}(h_{GQP}(x_i) \mid x_i)\,, \tag{4.3.4h}$$

for $\lambda \le \gamma / M'$ and $i \in K$, $i > i_0$. Inequalities (4.3.4g) and (4.3.4h) and Step 2 of Algorithm 4.3.1 imply that $\lambda_i \ge \beta\gamma / M'$. From Step 2 of Algorithm 4.3.1 and the fact that $h_{GQP}(x_i) = h_{\varepsilon}(x_i)$ for $i > i_0$,

$$p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i) = p_\varepsilon(x_i + \lambda_i h_{GQP}(x_i)) - p_\varepsilon(x_i)$$

$$\leq \lambda_i \left\{ \varepsilon \bar{f}^0(h_{GQP}(x_i) \mid x_i) + \max_{j \in p} \{ 0, \bar{f}^j(h_{GQP}(x_i) \mid x_i) \} - \psi_+(x_i) \right\}$$

$$\leq \lambda_i \left\{ \varepsilon \bar{f}^0(h_\varepsilon(x_i) \mid x_i) + \max_{j \in p} \{ 0, \bar{f}^j(h_\varepsilon(x_i) \mid x_i) \} - \psi_+(x_i) \right\}$$

$$= \lambda_i \theta_\varepsilon(x_i) . \tag{4.3.4i}$$

Then for $i \in K$, $i > i_0$,

$$p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i) = p \leq \frac{\beta\gamma}{M}\theta_\varepsilon(x_i) \leq -\frac{\beta\gamma\delta}{M} . \tag{4.3.4j}$$

Now $p_\varepsilon(x_{i+1}) \leq p_\varepsilon(x_i)$ for *all* $i > i_0$ by (4.3.4i). Hence (4.3.4j) implies that $p_\varepsilon(x_i) \to -\infty$ as $i \to \infty$. However, this is impossible, since $\{ x_i \}_{i \in \mathbf{N}}$ is contained in the bounded set $V_{\bar{w}}$. Therefore, $\bar{x} \neq \hat{x}$ cannot be an accumulation point for the sequence.

Since $V_{\bar{w}}$ is compact, the sequence $\{ x_i \}_{i \in \mathbf{N}} \subset V_{\bar{w}}$ must converge to the set of its accumulation points. We have shown that $\hat{x}$ can be the only accumulation point for the sequence. Therefore, the sequence converges to $\hat{x}$.

□

Let $\bar{\mu}^0 \triangleq \max \{ \mu^0 \mid \mu \in U_{PP}(\hat{x}) \}$.

**Theorem 4.3.2:** *Suppose that assumptions* (i)-(iii) *of Theorem 4.2.3 hold with $\hat{x}$ as defined there, that $x_0 \in V_{\bar{w}}$, with any $V_{\bar{w}}$ as defined in Theorem 4.3.1, and that Algorithm 4.3.1 constructs a sequence $\{ x_i \}_{i=0}^{\infty}$ in solving (4.1.1) starting from a point $x_0 \in V_{\bar{w}}$. Then,* (a) *for any* $\varepsilon < \underline{\mu}^0 / (1 - \underline{\mu}^0)$,

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(\hat{x})}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \leq 1 - \beta\frac{m}{M} \min \{ \frac{\varepsilon}{\bar{\mu}^0(1+\varepsilon)}, 1 \} , \tag{4.3.5a}$$

*and* (b) *if* $\psi(x_{i_0}) \leq 0$ *for any* $i_0 \in \mathbf{N}$,

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \le 1 - \beta(\underline{\mu}^0/\bar{\mu}^0)\frac{m}{M} . \tag{4.3.5b}$$

Unless $\hat{x}$ is also an *unconstrained* minimizer of $f^0(\cdot)$ (in which case, $\bar{\mu}^0 = 1$), the bound in (4.3.5b) on the convergence ratios of sequences constructed by Algorithm 4.3.1 is smaller than the bound in (4.2.7d) for sequences constructed by Algorithm 4.2.1,

$$1 - \alpha\beta(\underline{\mu}^0/\bar{\mu}^0)\frac{m}{M}\underline{\mu}^0 < 1 - \underline{\mu}^0\frac{m}{M} .$$

**Proof:** (a) Let positive $\varepsilon \in (0, \underline{\mu}^0 / (1 - \underline{\mu}^0))$ be arbitrary. The proof of Theorem 4.3.1 gives us a relation between the decrease in the penalty function $p_\varepsilon(x)$ at iteration $i$ and the decrease predicted by $\theta_\varepsilon(x)$,

$$p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i) \le \frac{\beta\gamma}{M}\,\theta_\varepsilon(x_i) . \tag{4.3.6}$$

for large $i$. Hence,

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le \frac{\beta\gamma}{M} \limsup_{i \to \infty} \frac{\theta_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} . \tag{4.3.7}$$

To complete our proof, we will make use of Theorem 2.3.3. This result provides an upper bound on the right-hand side of (4.3.7). For this purpose, we will show that the assumptions of Theorem 2.3.3 hold. Assumptions (i) and (ii) of Theorem 4.2.3 ensure that assumptions (i) and (ii) of Theorem 2.3.3 hold with respect to the minimax problem (4.3.2a) at $\hat{x}$. Next we turn to assumption (iii) of Theorem 2.3.3.

We associate with the minimax problem (4.3.2a) the set of multiplier vectors $U_\varepsilon(\hat{x})$ consisting of those $\mu \in \Sigma_{p+1}$ such that

$$\mu^0\varepsilon\nabla f^0(\hat{x}) + \sum_{j \in \mathbf{p}} \mu^j \{ \varepsilon\nabla f^0(\hat{x}) + \nabla f^j(\hat{x}) \} = 0 , \tag{4.3.8a}$$

$$\mu^0 \varepsilon f^0(\hat{x}) + \sum_{j \in \underline{p}} \mu^j \{ \varepsilon f^0(\hat{x}) + f^j(\hat{x}) \} = p_\varepsilon(\hat{x}) . \tag{4.3.8b}$$

The sets $U_\varepsilon(\hat{x})$ and $U_{PP}(\hat{x})$ are related as follows. Since $\psi_+(\hat{x}) = 0$, (4.3.8a-b) can be rewritten as

$$\varepsilon \nabla f^0(\hat{x}) + \sum_{j \in \mathbf{p}} \mu^j \nabla f^j(\hat{x}) = 0 , \tag{4.3.9a}$$

$$\sum_{j \in \mathbf{p}} \mu^j f^j(\hat{x}) = 0 . \tag{4.3.9b}$$

Then, since $1 - \mu^0 = \sum_{j \in \mathbf{p}} \mu^j$, $(\varepsilon, \mu^1, \ldots, \mu^p) / (\varepsilon + 1 - \mu^0) \in U_{PP}(\hat{x})$ for any $\mu \in U_\varepsilon(\hat{x})$. It follows from assumption (iii) of Theorem 4.2.3, that, with $H$ as defined in Theorem 4.2.3,

$$m \|h\|^2 < \langle h , \left[ \frac{\varepsilon}{\varepsilon + 1 - \mu^0} F^0(\hat{x}) + \sum_{j \in \mathbf{p}} \frac{\mu^j}{\varepsilon + 1 - \mu^0} F^j(\hat{x}) \right] h \rangle, \quad \forall\, h \in H,\ h \neq 0, \tag{4.3.10a}$$

for any $\mu \in U_\varepsilon(\hat{x})$. Hence for any $\mu \in U_\varepsilon(\hat{x})$,

$$m_\varepsilon \|h\|^2 < \langle h , \left[ \mu^0 \varepsilon F^0(\hat{x}) + \sum_{j \in \mathbf{p}} \mu^j \{ \varepsilon F^0(\hat{x}) + F^j(\hat{x}) \} \right] h \rangle \quad \forall\, h \in H ,\ h \neq 0 , \tag{4.3.10b}$$

where $m_\varepsilon \triangleq \min \{ m (\varepsilon + 1 - \mu^0) \mid \mu \in U_\varepsilon(\hat{x}) \} = m (\varepsilon + 1 - \max \{ \mu^0 \mid \mu \in U_\varepsilon(\hat{x}) \} )$. Hence, assumption (iii) of Theorem 2.3.3 is satisfied at $\hat{x}$ for the minimax problem (4.5.1), and it therefore follows from Theorem 2.3.3 (and the fact that

$$\theta_\varepsilon(x) \le \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{p} \cup 0} g^j(x) + \langle \nabla g^j(x), h \rangle + \tfrac{1}{2}(1 + \varepsilon)\gamma \|h\|^2 - p_\varepsilon(x) ,$$

where $g^j(x) \triangleq \varepsilon f^0(x) + f^j(x)$ for $j \in \mathbf{p}$ and $g^0(x) \triangleq \varepsilon f^0(x)$) that

$$\limsup_{i \to \infty} \frac{\theta_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le - \frac{\min \{ m_\varepsilon, (1 + \varepsilon)\gamma \}}{(1 + \varepsilon)\gamma} . \tag{4.3.11}$$

Combining (4.3.11) with (4.3.7) yields

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le - \frac{\beta\gamma}{M} \frac{\min \{ m_\varepsilon, (1 + \varepsilon)\gamma \}}{(1 + \varepsilon)\gamma}$$

$$= -\beta \frac{\min \{ m_\varepsilon / (1+\varepsilon), \gamma \}}{M} . \tag{4.3.12}$$

Next consider any $\mu \in U_\varepsilon(\hat{x})$. As mentioned above, $(\varepsilon, \mu^1, \ldots, \mu^p) / (\varepsilon + 1 - \mu^0) \in U_{PP}(\hat{x})$. Recall that $\bar{\mu}^0 \triangleq \max \{ \mu^0 | \mu \in U_{PP}(\hat{x}) \}$. Then

$$\frac{\varepsilon}{\varepsilon + 1 - \mu^0} \le \bar{\mu}^0 , \tag{4.3.13}$$

and hence

$$m_\varepsilon = m(\varepsilon + 1 - \max \{ \mu^0 | \mu \in U_\varepsilon(\hat{x}) \}) \ge m \frac{\varepsilon}{\bar{\mu}^0} . \tag{4.3.14}$$

Substituting (4.3.14) into (4.3.12) yields

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(x_i)}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le -\frac{\beta}{M} \min \{ m\varepsilon / (\bar{\mu}^0(1+\varepsilon)), \gamma \} \le -\beta \frac{m}{M} \min \{ \frac{\varepsilon}{\bar{\mu}^0(1+\varepsilon)}, 1 \} \tag{4.3.15}$$

Adding 1 to each side of the inequality in (4.3.15), we obtain (4.3.5a).

(b) Using the fact that $p_\varepsilon(x_i) = f^0(x_i)$ for $i > i_0$,

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \le 1 - \beta \frac{m}{M} \min \{ \frac{\varepsilon}{\bar{\mu}^0(1+\varepsilon)}, 1 \} . \tag{4.3.16}$$

Since $\varepsilon < \underline{\mu}^0 / (1 - \underline{\mu}^0)$ is arbitrary, (4.3.5b) holds.

□

## 4.4 GLOBALIZATION AND IMPLEMENTATION OF THE GQP SUBPROCEDURE

There are two issues associated with the use of the problem

$$\text{GQP}(x): \quad \min \{ \tilde{f}^0(h \mid x) | \tilde{f}^j(h \mid x) \le 0, \ \forall \ j \in \mathbf{p} \} , \tag{4.4.1}$$

as a search direction subprocedure that must be resolved. The first is the issue of globalization. When $x$ is not feasible for (4.1.1) and is far from a solution to (4.1.1), GQP($x$) may not have any feasible solutions. The second is the issue of implementation. Unlike the search direction problem (4.2.2a) of Algorithm 4.2.1, GPQ($x$) cannot be transformed into a quadratic program to be solved by known methods. We must find an efficient method for solving it in a neighborhood of any solution $\hat{x}$ of (4.1.1), where, by Lemma 4.3.1, GQP($x$) is known to have a solution.

We will develop the globalized, implementable search direction subprocedure in three steps. First, we will show that GPQ($x$) is equivalent to a problem GQ$\tilde{\text{P}}$($x$) with *linear* equality constraints and a *single* quadratic inequality constraint, determined by the constraints active at the solution to GPQ($x$). Second, we will use the PMT search direction subprocedure to predict which constraints are active at the solution. This will allow us to construct a problem with linear equality constraints and a single quadratic inequality constraint which approximates GQ$\tilde{\text{P}}$($x$). We will show that, when the approximating problem has a solution, it can be easily obtained from the PMT search direction vector $h(x)$. Third, we will incorporate these observations in a search direction subprocedure which reverts to the PMT search direction when the approximating problem has no solution.

Because, the PMT search direction subprocedure correctly predicts the constraints active at the solution to GQP($x$) when $x$ is near a solution to (4.1.1) at which strict complimentary slackness holds, the globalized, implementable search direction subprocedure leads to a phase I - phase II algorithm which has the same robustness properties as the PMT algorithm and the same rate of convergence as the *conceptual* Algorithm 4.3.1.

Thus, we begin by developing an equivalent statement for GQP($x$). For any $x \in \mathbf{R}^n$ and set $J \subset \mathbf{p}$, we define the problem

$$\text{P}(x,J): \quad \min_{h \in \mathbf{R}^n} \{ \tilde{f}^0(h \mid x) \mid \tilde{f}^{j_0}(h \mid x) \le 0,\ \tilde{f}^j(h \mid x) = \tilde{f}^{j_0}(h \mid x), \forall j \in J \setminus j_0 \}, \tag{4.4.2a}$$

where $j_0 \in J$ is arbitrary. A brief inspection of (4.4.2a) reveals that the problem P($x$, $J$) is independent of the selection of $j_0 \in J$. We will denote the solution to P($x$, $J$) by $d(x, J)$.

Since the functions $\tilde{f}^j(\cdot \mid x)$ all have the same quadratic term, $\frac{1}{2}\gamma \| h \|^2$, the equality constraints in (4.4.2a) are linear. Hence, problem (4.4.2a) requires the minimization of a quadratic function subject to *linear* equality constraints and a *single* positive-definite quadratic inequality constraint. A subproblem of this form appears in trust region methods, and efficient methods for solving it have been developed [Mor.1]. However, because $\tilde{f}^0(\cdot \mid x)$ and $\tilde{f}^{j_0}(\cdot \mid x)$ have the *same* quadratic term, a simpler technique can be used to solve (4.4.2a) for our choice of $J$ (see Proposition 4.4.2).

Assuming that (4.4.1) is feasible, we define the *active constraint index set* by

$$J_{\mathrm{GQP}}(x) \triangleq \{\, j \in \mathbf{p} \mid \tilde{f}^j(h_{\mathrm{GQP}}(x) \mid x) = 0 \,\} \,. \tag{4.4.2b}$$

(The set $J_{\mathrm{GQP}}(x)$ may be empty.) A small amount of reflection confirms that the problem $\mathrm{GQP}(x)$ is equivalent to the problem $\mathrm{P}(x, J_{\mathrm{GQP}}(x))$. (Problem $\mathrm{P}(x, J_{\mathrm{GQP}}(x))$ is what we referred to above as $\mathrm{GQ\tilde{P}}(x)$.) Hence, when the set $J_{\mathrm{GQP}}(x)$ is known, the problem $\mathrm{GQP}(x)$ is relatively easy to solve. Next, for any $\mu \in \Sigma_{p+1}$, let

$$J(\mu) \triangleq \{\, j \in \mathbf{p} \mid \mu^j > 0 \,\} \tag{4.4.3a}$$

and let $\mu_{\mathrm{PP}}(x)$ be any selection from $U_{\mathrm{PP}}(x)$. In the following propositions, we will prove that the use of

$$J_{\mathrm{PP}}(x) \triangleq J(\mu_{\mathrm{PP}}(x)) \,, \tag{4.4.3b}$$

as an estimate of $J_{\mathrm{GQP}}(x)$ has several desirable consequences.

The following proposition shows that $d(x, J_{\mathrm{PP}}(x))$ can be obtained rather easily from $h(x)$. Recall that, for any $x \in \mathbb{R}^n$ such that $\mathrm{GQP}(x)$ has a solution, we denote the set of Fritz John multiplier vectors associated with the solution by $U_{\mathrm{GQP}}(x)$ (see (4.3.1d)).

Proposition 4.4.2: *Suppose that problem $\mathrm{P}(x, J_{\mathrm{PP}}(x))$ has a solution $d(x, J_{\mathrm{PP}}(x))$. Let $j_0 \in J_{\mathrm{PP}}(x)$ be arbitrary, let $G_x$ be a matrix with columns $\nabla f^j(h(x) \mid x) - \nabla f^{j_0}(h(x) \mid x)$, $j \in J_{\mathrm{PP}}(x) \setminus j_0$, let $N_x$ be a matrix whose columns form an orthonormal basis for the null space* *of $G_x^T$, and let $P_x \triangleq N_x N_x^T$ be the orthogonal projection operator whose range is the null space of $G_x^T$. Then there exists a $\tau \in \mathbb{R}$ such that*

$$d(x, J_{PP}(x)) = h(x) + \tau P_x \nabla \tilde{f}^0(h(x) \mid x). \tag{4.4.4}$$

**Proof:** First, we rewrite $P(x, J_{PP}(x))$ in the form

$$\min\{\tilde{f}^0(h \mid x) \mid \tilde{f}^{j_0}(h \mid x) \le 0,\ g_x + G_x^T h = 0\}, \tag{4.4.5a}$$

where $j_0 \in J_{PP}(x)$ is arbitrary, $g_x$ is the vector with elements $f^j(h(x) \mid x) - f^{j_0}(h(x) \mid x)$, $j \in J_{PP}(x) \setminus j_0$. Since $g_x + G_x^T h(x) = 0$, it follows that if we set $h = h(x) + \delta h$ in (4.4.5a), then we must have $G_x^T \delta h = 0$, which implies that $\delta h = N_x y$ for some $y$. Hence, By substituting $\delta h = N_x y$ into (4.4.5a), the equality constraint in (4.4.5a) can be eliminated. Upon expansion of the functions $\tilde{f}^j(\cdot \mid x)$ around $h(x)$, (4.4.5a) becomes

$$\min\{\tilde{f}^0(h(x) \mid x) + \langle \nabla \tilde{f}^0(h(x) \mid x), N_x y \rangle + \tfrac{1}{2}\gamma \| N_x y \|^2 \mid$$

$$\tilde{f}^{j_0}(h(x) \mid x) + \langle \nabla \tilde{f}^{j_0}(h(x) \mid x), N_x y \rangle + \tfrac{1}{2}\gamma \| N_x y \|^2 \le 0\}. \tag{4.4.5b}$$

If $J_{PP}(x) = \emptyset$, then $\mu_{PP}^0 = 1$, then $\nabla \tilde{f}^0(h(x) \mid x) = 0$ and the optimal solution to (4.4.5a) is $\delta h(x) = 0$. Now suppose that $\nabla \tilde{f}^0(h(x) \mid x) \ne 0$. This implies that $\mu_{PP}^0(x) < 1$ and that $J_{PP}(x) \ne \emptyset$. Then the solution $\delta h(x)$ for problem (4.4.5b) satisfies the first-order condition

$$N_x^T \Big[ \mu^0 \nabla \tilde{f}^0(h(x) \mid x) + (1 - \mu^0) \nabla \tilde{f}^{j_0}(h(x) \mid x) + \gamma \delta h(x) \Big] = 0, \tag{4.4.5c}$$

for some $\mu^0 \in [0, 1]$. Since $N_x N_x^T \delta h(x) = P_x \delta h(x) = \delta h(x)$, we obtain from (4.4.5c) that

$$\delta h(x) = \gamma^{-1} \Big[ \mu^0 P_x \nabla \tilde{f}^0(h(x) \mid x) + (1 - \mu^0) P_x \nabla \tilde{f}^{j_0}(h(x) \mid x) \Big]. \tag{4.4.5d}$$

Now, $h(x)$, the solution to (4.2.2a), satisfies the optimality condition $\sum_{j \in p \cup 0} \mu_{PP}^j(x) \nabla \tilde{f}^j(h(x) \mid x) = 0$. Rearranging this equation (and dropping the dependence of $\mu_{PP}$ on $x$) yields

$$0 = \mu^0_{PP} \nabla \tilde{f}^{\,0}(h(x) \mid x) + (1 - \mu^0_{PP}) \nabla \tilde{f}^{\,j^*}(h(x) \mid x)$$
$$+ \sum_{j \in \mathbf{p}} \mu^j_{PP} \left[ \nabla \tilde{f}^{\,j}(h(x) \mid x) - \nabla \tilde{f}^{\,j^*}(h(x) \mid x) \right]. \tag{4.4.5e}$$

Applying $P_x$ to both sides of (4.4.5e), we conclude that

$$0 = \mu^0_{PP} P_x \nabla \tilde{f}^{\,0}(h(x) \mid x) + (1 - \mu^0_{PP}) P_x \nabla \tilde{f}^{\,j^*}(h(x) \mid x), \tag{4.4.5f}$$

since $P_x(\nabla \tilde{f}^{\,j}(h(x) \mid x) - \nabla \tilde{f}^{\,j^*}(h(x) \mid x)) = 0$ for all $j \in \mathbf{p}$ by the definition of $P_x$. Since $\mu^0_{PP} < 1$,

$$P_x \nabla \tilde{f}^{\,j^*}(h(x) \mid x) = -\frac{\mu^0_{PP}}{1 - \mu^0_{PP}} P_x \nabla \tilde{f}^{\,0}(h(x) \mid x). \tag{4.4.5g}$$

Substituting (4.4.5g) into (4.4.5d) yields

$$\delta h(x) = \gamma^{-1} \left[ \mu^0 - (1 - \mu^0) \frac{\mu^0_{PP}}{1 - \mu^0_{PP}} \right] P_x \nabla \tilde{f}^{\,0}(h(x) \mid x). \tag{4.4.5h}$$

□

The search direction $d(x, J_{PP(x)})$ may not be a feasible solution for GQP($x$). The following subprocedure returns the Polak-Trahan-Mayne search direction in this case.

**Search Direction Subprocedure 4.4.1:**

**Step 1:** Compute the Polak-Trahan-Mayne search direction $h(x)$ and identify the set $J_{PP}(x)$.

**Step 2:** Compute the step, $\Delta h(x) = P_x \nabla \tilde{f}^{\,0}(h(x) \mid x)$.

**Step 3:** Compute $\tau \in \mathbb{R}$ by solving

$$\min \{ \tilde{f}^{\,0}(h(x) + \tau \Delta h(x) \mid x) \mid \tilde{f}^{\,j}(h(x) + \tau \Delta h(x) \mid x) \le 0 \ \ \forall\, j \in \mathbf{p} \}. \tag{4.4.6}$$

(If problem (4.4.6) is infeasible, set $\tau = 0$.)

**Step 4:** Set $d(x) = h(x) + \tau \Delta h(x)$.

□

The minimization in Step 3 can be performed very quickly since it involves only quadratic functions of a single variable. Note that $\Delta h(x)$ of the Search Direction Subprocedure 4.4.1 is equal to $\tau\delta h(x)$, with $\delta h(x)$ as defined in the proof of Proposition 4.4.2. The following proposition summarizes the useful properties of $d(x)$.

We now prove that, if $h(x)$ is feasible for GQP($x$), then $d(x)$ is a feasible direction promising as much decrease in the objective as $h(x)$. If $h(x)$ is not feasible for GQP($x$), then $d(x)$ provides as much improvement in the constraint violation as $h(x)$.

**Proposition 4.4.3:**

(a) *If* $\tilde{f}^j(h(x) \mid x) \le 0$ *for each* $j \in \underline{p}$, *then* $\tilde{f}^0(d(x) \mid x) \le \tilde{f}^0(h(x) \mid x)$ *and* $\tilde{f}^j(d(x) \mid x) \le 0$ *for each* $j \in \underline{p}$.

(b) *If* $\max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) > 0$, *then* $\max_{j \in \mathbf{p}} \tilde{f}^j(d(x) \mid x) \le \max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x)$.

(c) *If* GQP($x$) *is feasible and* $J_{\mathrm{PP}}(x) = J_{\mathrm{GQP}}(x)$, *then* $d(x)$ *solves* GQP($x$).

Lemma 4.5.1 shows that the assumptions of Proposition 4.4.3(c) hold in a neighborhood of a solution $\hat{x}$ to (4.1.1), provided that strict complementary slackness holds at $\hat{x}$.

Proof: (a) This follows from the fact $\tau = 0$ is feasible for the single-variable minimization in Step 4.

(b) If problem (4.4.6) is feasible, then

$$\max_{j \in \mathbf{p}} \tilde{f}^j(d(x) \mid x) = 0 \le \max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) . \tag{4.4.7}$$

If problem (4.4.6) is infeasible, $d(x) = h(x)$.

(c) Since $J_{\mathrm{PP}}(x) = J_{\mathrm{GQP}}(x)$, $d(x, J_{\mathrm{PP}}(x))$ solves GQP($x$). We show that Algorithm 4.4.1 computes $d(x, J_{\mathrm{PP}}(x))$. Since $d(x, J_{\mathrm{PP}}(x))$ minimizes $\tilde{f}^0(\cdot \mid x)$ over $\{ h \in \mathbb{R}^n \mid \tilde{f}^j(h \mid x) \le 0, j \in \mathbf{p} \}$,

$$\tilde{f}^0(d(x, J_{\mathrm{PP}}(x))) = \min_{h \in \mathbb{R}^n} \{ \tilde{f}^0(h \mid x) \mid \tilde{f}^j(h \mid x) \le 0, j \in \mathbf{p} \}$$

$$\le \min_{\tau \in \mathbb{R}} \{ \tilde{f}^0(h(x) + \tau\Delta h(x) \mid x) \mid \tilde{f}^j(h(x) + \tau\Delta h(x) \mid x) \le 0, j \in \mathbf{p} \} .$$

Since $d(x, J_{PP}(x))$ can be expressed as $h(x) + \tau_0 \Delta h(x) \mid x)$ for some $\tau_0 \in \mathbb{R}$, problem (4.4.6) is feasible and has the solution $\tau_0$. Therefore, $d(x) = d(x, J_{PP}(x))$. □

## 4.5 A STABILIZED IMPLEMENTABLE GQP-BASED ALGORITHM

We replace Step 2 of Algorithm 4.3.1 with the Search Direction Subprocedure 4.4.1 to obtain a global phase I - phase II method, and we establish its convergence properties.

**Algorithm 4.5.1:**

**Data:** $x_0$; $\beta \in (0, 1)$; $\gamma > 0$; $i = 0$.

**Step 1:** Compute a search direction $d_i = d(x_i)$ by means of Search Direction Subprocedure 4.4.1.

**Step 2:** Compute a step size,

$$\lambda_i = \max_{k \in \mathbb{N}} \{ \beta^k \mid \; f^0(x_i + \beta^k d_i) - f^0(x_i) \le \beta^k \tilde{f}^0(d_i \mid x_i),$$

$$\psi_+(x_i + \beta^k d_i) - \psi_+(x_i) \le \beta^k [\max_{j \in \mathbf{p}} \{ \tilde{f}^j(d_i \mid x_i), 0 \} - \psi_+(x_i)] \} \; . \qquad (4.5.1)$$

**Step 3:** Set $x_{i+1} = x_i + \lambda_i d_i$.

**Step 4:** Replace $i$ by $i+1$, and go to Step 1.

□

The three cases listed in Theorem 4.5.1 are exhaustive. In case (b), $\theta(\bar{x}) = 0$ implies that $0 \in \partial\psi(\bar{x})$, where $\partial\psi(x)$ denotes the generalized gradient of $\psi(\cdot)$ at $x$. This case is normally ruled out by assumption. The convergence result obtained for Algorithm 4.5.1 is slightly weaker than that obtained for Algorithm 4.2.1 in Theorem 4.2.2. In case (c), where Algorithm 4.5.1 constructs a sequence which remains infeasible but has feasible accumulation points, not all of the accumulation points are guaranteed to be stationary points of problem (4.1.1).

**Theorem 4.5.1:** *Suppose that the functions $f^j(\cdot)$ in (4.1.1) have continuous derivatives, that Algorithm 4.5.1 constructs a sequence $\{ x_i \}_{i=0}^{\infty}$ in solving (4.1.1), and that $\bar{x}$ is an accumulation point of the sequence.*

(a) *If there exists an* $i_0 \in \mathbb{N}$ *such that* $\psi(x_{i_0}) \le 0$, *then* $\theta(\bar{x}) = 0$.

(b) *If* $\psi(x_i) > 0$ *for all* $i \in \mathbb{N}$ *and* $\psi(\bar{x}) > 0$, *then* $\theta(\bar{x}) = 0$.

(c) *If* $\psi(x_i) > 0$ *for all* $i \in \mathbb{N}$ *and* $\psi(\bar{x}) = 0$, *then* $\liminf_{i \to \infty} |\theta(x_i)| = 0$.

**Proof:** First we derive bounds on $\|d(x)\|$ for use in the proof of parts (a) and (b). Suppose that the subsequence $\{x_i\}_{i \in K}$ converges to $\bar{x}$, for some subset $K \subset \mathbb{N}$, and that $\theta(\bar{x}) \ne 0$. By Theorem 4.2.1, $\theta(\cdot)$ is continuous, and, by (4.4.2b), $\theta(x) \le 0$ for all $x \in \mathbb{R}^n$. Therefore there exists a $\delta > 0$ and a neighborhood, $W_0$, of $\bar{x}$ such that

$$\theta(x) = \max_{j \in \mathbf{p} \cup 0} \{ \tilde{f}^j(h(x) \mid x) \} - \psi_+(x) < -\delta , \tag{4.5.2a}$$

for all $x \in W_0$. We use this fact and Proposition 4.4.3 to show that $\|d(x)\| > 0$ for all $x$ in a neighborhood of $\bar{x}$.

Suppose that $\psi(\bar{x}) \le 0$. In view of (4.5.2a), there exists a neighborhood, $W_1 \subset W_0$, of $\bar{x}$, such that $\psi(x) < \frac{1}{2}\delta$ for all $x \in W_1$. Then, for $x \in W_1$,

$$\max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) \le \theta(x) + \psi_+(x) \le -\tfrac{1}{2}\delta < 0 . \tag{4.5.2b}$$

From Proposition 4.4.3(a), we have that

$$\tilde{f}^0(d(x) \mid x) - \psi_+(x) \le \tilde{f}^0(h(x) \mid x) - \psi_+(x) \le \theta(x) < -\delta , \tag{4.5.2c}$$

for all $x \in W_0$. Since $\psi(x) < \frac{1}{2}\delta$, if follows from (4.5.2b) and (4.5.2c) that $\tilde{f}^0(d(x) \mid x) < -\frac{1}{2}\delta$ for all $x \in W_1$. Hence, since $\tilde{f}^0(0 \mid x) = 0$, and since $\tilde{f}^0(h \mid x)$ is continuous in $h$, uniformly in $x$, there exists $b' > 0$ such that $\|d(x)\| > b'$ for all $x \in W_1$.

Now suppose that $\psi(\bar{x}) > 0$. We proceed in a manner similar to that in the previous paragraph. There exists a neighborhood, $W_2 \subset W_0$, of $\bar{x}$, such that $\psi(x) > \frac{1}{2}\psi(\bar{x})$ for each $x \in W_2$. For each $x \in W_2$, either $\max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) > 0$, or else $\max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) \le 0$. In the former case, it follows from Proposition 4.4.3(b) that

$$\max_{j \in \mathbf{p}} \tilde{f}^j(d(x) \mid x) - \psi_+(x) \le \max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) - \psi_+(x) < -\delta . \tag{4.5.2d}$$

In the latter case, it follows from Proposition 4.4.3(b) that

$$\max_{j \in \mathbf{p}} \tilde{f}^j(d(x) \mid x) - \psi_+(x) < 0 - \psi_+(x) = -\tfrac{1}{2}\psi(\bar{x}) , \tag{4.5.2e}$$

for all $x \in W_2$. Therefore, for all $x \in W_2$, $\max_{j \in \mathbf{p}} \tilde{f}^j(d(x) \mid x) - \psi_+(x) < -\min\{ \delta, \tfrac{1}{2}\psi(\bar{x}) \}$. Hence, since $\max_{j \in \mathbf{p}} \tilde{f}^j(0 \mid x) - \psi_+(x) = 0$ for $x \in W_2$, and since the function $\max_{j \in \mathbf{p}} \tilde{f}^j(h \mid x)$ is continuous in $h$, uniformly in $x$, there exists $b \in (0, b')$ such that $\|d(x)\| > b$ for all $x \in W_2$.

Because the functions $\tilde{f}^k(\cdot \mid x)$ are strongly convex in $h$, uniformly in $x$, $\|d(x)\|$ is also bounded from above in $W_2$. Because $\|d(x)\|$ is bounded on $W_2$ and the gradients $\nabla f^j(\cdot)$ are continuous, there exist $\bar{\lambda} > 0$ and a neighborhood, $W_3$, of $\bar{x}$, such that $\|\int_0^1 \nabla f^j(x + s\lambda d(x))ds - \nabla f^j(x)\| < \tfrac{1}{2}\gamma b$ for all $x \in W_3$, $\lambda \in [0, \bar{\lambda}]$ and $j \in \mathbf{p} \cup 0$. (We assume without loss of generality that $W_3 \subset W_1$ if $\psi(\bar{x}) \le 0$ and that $W_3 \subset W_2$ if $\psi(\bar{x}) > 0$.)

(a) Suppose that $\psi(x_{i_0}) \le 0$ for some $i_0 \in \mathbb{N}$. (This implies that $\psi(x_i) \le 0$ for all $i \ge i_0$ and that $\psi(\bar{x}) \le 0$.) Then there exists an $i_1 \ge i_0$ such that $x_i \in W_3$ for all $i \ge i_1$, $i \in K$. For $i > i_1$, $i \in K$ and $\lambda \in (0, \bar{\lambda}]$,

$$\begin{aligned} f^0(x_i + \lambda d_i) - f^0(x_i) &= \langle \nabla f^0(x_i), \lambda d_i \rangle + \langle \textstyle\int_0^1 [\nabla f^0(x_i + s\lambda d_i) - \nabla f^0(x_i)]ds , \lambda d_i \rangle \\ &\le \lambda \{ \langle \nabla f^0(x_i), d_i \rangle + \|d_i\| \, \| \textstyle\int_0^1 [\nabla f^j(x_i + s\lambda d_i) - \nabla f^j(x_i)]ds \| \} \\ &\le \lambda \{ \langle \nabla f^0(x_i), d_i \rangle + \tfrac{1}{2}\gamma b \|d_i\| \} \\ &\le \lambda \{ \langle \nabla f^0(x_i), d_i \rangle + \tfrac{1}{2}\gamma \|d_i\|^2 \} = \lambda \tilde{f}^0(d_i \mid x_i) . \end{aligned} \tag{4.5.3a}$$

Similarly, for $\lambda \in (0, \bar{\lambda}]$, $i > i_1$, $i \in K$, and $j \in \mathbf{p}$,

$$f^j(x_i + \lambda d_i) \le \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \|d_i\| \, \| \textstyle\int_0^1 [\nabla f^j(x_i + s\lambda d_i) - \nabla f^j(x_i)]ds \| \}$$

$$\leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \tfrac{1}{2}\gamma b \| d_i \| \} \leq \lambda \tilde{f}^j(d_i \mid x_i) . \qquad (4.5.3b)$$

Taking the maximum over $j \in \mathbf{p}$, and using the fact that $\psi_+(x_i) = 0$, we obtain from (4.5.3b) that

$$\psi_+(x_i + \lambda d_i) - \psi_+(x_i) \leq \lambda[\max_{j \in \mathbf{p}} \{ \tilde{f}^j(d_i \mid x_i), 0 \} - \psi_+(x_i)] , \qquad (4.5.3c)$$

for $i > i_1$, $i \in K$ and $\lambda \in (0, \bar{\lambda}]$. It follows from (4.5.3a), (4.5.3c) and Step 2 of Algorithm 4.5.1 that $\lambda_i > \beta\bar{\lambda}$ for $i > i_1$, $i \in K$. By Proposition 4.4.3(a), $\tilde{f}^0(d_i \mid x_i) \leq \theta(x_i)$ for $i > i_1$, $i \in K$, and hence

$$f^0(x_i + \lambda_i d_i) - f^0(x_i) \leq \lambda_i \tilde{f}^0(d_i \mid x_i) \leq \lambda_i \theta(x_i) \leq -\tfrac{1}{2}\beta\bar{\lambda}\delta , \qquad (4.5.3d)$$

for $i > i_1$, $i \in K$.

However, this is impossible, since $f^0(x_i)$ is monotone decreasing for $i \geq i_1$ and $f^0(x_i) \xrightarrow{K} f^0(\bar{x})$, as $i \xrightarrow{K} \infty$. Thus, the necessary condition (4.2.5a-b), must be satisfied at $\bar{x}$ in this case.

(b) Now suppose that $\psi(x_i) > 0$ for all $i \in \mathbb{N}$ and that $\psi(\bar{x}) > 0$. Then there exists an $i_1 \in \mathbb{N}$ such that $x_i \in W_3$ for all $i \geq i_1$, $i \in K$. For any $x \in \mathbb{R}^n$ such that $\tilde{f}^j(h(x) \mid x) \leq 0$ for each $j \in \underline{p}$, $\tilde{f}^j(d(x) \mid x) \leq 0$ for each $j \in \mathbf{p}$ by Proposition 4.4.3(a). For any $x \in \mathbb{R}^n$ such that $\max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) > 0$, $\max_{j \in \mathbf{p}} \tilde{f}^j(d(x) \mid x) \leq \max_{j \in \mathbf{p}} \tilde{f}^j(h(x) \mid x) \leq 0$ by Proposition 4.4.3(b). Therefore, $\tilde{f}^j(d(x) \mid x) \leq \psi(x)$ for all $j \in \mathbf{p}$ and $x \in \mathbb{R}^n$. Hence,

$$f^j(x_i + \lambda d_i) - \psi_+(x_i) = f^j(x_i) + \langle \nabla f^j(x_i), \lambda d_i \rangle + \langle \int_0^1 [\nabla f^j(x_i + s\lambda d_i) - \nabla f^j(x_i)] ds , \lambda h(x_i) \rangle - \psi_+(x_i)$$

$$\leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \tfrac{1}{2}\gamma b \| d_i \| - \psi_+(x_i) \}$$

$$\leq \lambda \{ f^j(x_i) + \langle \nabla f^j(x_i), d_i \rangle + \tfrac{1}{2}\gamma \| d_i \|^2 - \psi_+(x_i) \}$$

$$= \lambda \{ \tilde{f}^{j}(d_i \mid x_i) - \psi_+(x_i) \} , \tag{4.5.4a}$$

for all $i > i_1, i \in K, \lambda \in (0, \bar{\lambda}]$, and $j \in \mathbf{p}$. Taking the maximum over $j \in \mathbf{p}$, and using the fact that $0 - \psi_+(x_i) \le \lambda[\max_{j \in \mathbf{p}} \tilde{f}^{j}(d_i \mid x_i) - \psi_+(x_i)]$,

$$\psi_+(x_i + \lambda d_i) - \psi_+(x_i) \le \lambda[\max_{j \in \mathbf{p}} \{ \tilde{f}^{j}(d_i \mid x_i), 0 \} - \psi_+(x_i)] . \tag{4.5.4b}$$

Similarly,

$$f^{0}(x_i + \lambda d_i) - f^{0}(x_i) \le \lambda \tilde{f}^{0}(d_i \mid x_i) , \tag{4.5.4c}$$

for all $i > i_1, i \in K, \lambda \in (0, \bar{\lambda}]$, and $j \in \mathbf{p}$. It follows from (4.5.4b), (4.5.4c) and Step 2 of Algorithm 4.5.1 that $\lambda_i > \beta\bar{\lambda}$ for $i > i_1, i \in K$.

From Proposition 4.4.3(b), if $\max_{k \in \mathbf{p}} \tilde{f}^{k}(h(x_i) \mid x_i) > 0$,

$$\max_{k \in \mathbf{p}} \tilde{f}^{k}(d_i \mid x_i) - \psi_+(x_i) \le \max_{k \in \mathbf{p}} \tilde{f}^{k}(h_i \mid x_i) - \psi_+(x_i) \le \theta(x_i) \le -\delta . \tag{4.5.4d}$$

Otherwise, $\max_{k \in \mathbf{p}} \tilde{f}^{k}(h(x_i) \mid x_i) \le 0$, which, together with Proposition 4.4.3(a), implies that

$$\max_{k \in \mathbf{p}} \tilde{f}^{k}(d_i \mid x_i) - \psi_+(x_i) \le 0 - \psi_+(x_i) . \tag{4.5.4e}$$

There exists $i_2 > i_1$ such that $\psi_+(x_i) > \tfrac{1}{2}\psi_+(\bar{x})$ for $i > i_2, i \in K$. Substituting (4.5.4d) and (4.5.4e) into (4.5.4f),

$$\psi_+(x_i + \lambda d_i) - \psi_+(x_i) \le -\lambda_i \min \{ \psi_+(x_i), \delta \} \le -\beta\bar{\lambda} \min \{ \tfrac{1}{2}\psi_+(\bar{x}), \delta \} , \tag{4.5.4f}$$

for $i > i_1, i \in K$.

Since $\psi(x_i)$ is monotone decreasing, (4.5.4f) implies that $\psi(x_i) \to -\infty$ as $i \to \infty$. However, this is impossible, since $\psi(x_i) \xrightarrow{K} \psi(\bar{x})$ as $i \xrightarrow{K} \infty$, Therefore, the necessary condition (4.2.5a-b) must be satisfied at $\bar{x}$.

(c) Now suppose that $\psi(x_i) > 0$ for all $i \in \mathbb{N}$ and that $\psi(\bar{x}) = 0$. In this case, we do not show that $\theta(\bar{x}) = 0$, but merely that $\liminf_{i \to \infty} |\theta(x_i)| = 0$.

To obtain a contradiction, suppose that $\liminf_{i\to\infty}\theta(x_i) < -\delta' < 0$. Then there exists $i_1 \in \mathbb{N}$ such that $\theta(x_i) < -\delta'$ for all $i > i_1$. By Proposition 4.4.3(a-b),

$$\max_{j \in \mathbf{p}} \tilde{f}^j(d_i \mid x_i) \le \max\{0, \max_{j\in\mathbf{p}} \tilde{f}^j(h(x_i) \mid x_i)\}$$

$$\le \max\{0, \theta(x_i) + \psi_+(x_i)\}\ . \tag{4.5.5a}$$

Hence

$$\max_{j\in\mathbf{p}} \tilde{f}^j(d_i \mid x_i) - \psi_+(x_i) \le \max\{-\psi_+(x_i), \theta(x_i)\} \le \max\{-\psi_+(x_i), -\delta'\} < 0 \tag{4.5.5b}$$

for all $i > i_1$. This implies that $\psi_+(x_i)$ is monotone decreasing, and, since $\psi(\bar{x}) = 0$, the sequence $\{\psi_+(x_i)\}_{i\in\mathbb{N}}$ converges to 0. Therefore, there exists $i_2 > i_1$ such that $\psi_+(x_i) < \frac{1}{2}\delta'$ for all $i > i_2$. Hence,

$$\max_{j\in\mathbf{p}} \tilde{f}^j(h(x_i) \mid x_i) \le \theta(x_i) + \psi_+(x_i) \le -\delta' + \tfrac{1}{2}\delta' < 0\ , \tag{4.5.5c}$$

for all $i > i_2$. From Proposition 4.4.3(a), then,

$$\tilde{f}^0(d_i \mid x_i) \le \tilde{f}^0(h(x_i) \mid x_i) \le \theta(x_i) + \psi_+(x_i) \le -\tfrac{1}{2}\delta'\ , \tag{4.5.5d}$$

for all $i > i_2$. This implies that $f^0(x_i)$ is monotone decreasing for for $i > i_2$.

Now we use the fact that $\bar{x}$ is an accumulation point of the sequence $\{x_i\}_{i\in\mathbb{N}}$. It follows from an argument similar to the ones used in parts (a) and (b) that there exists $\underline{\lambda} > 0$ such that $\lambda_i > \underline{\lambda}$ for all $i > i_2$, $i \in K$. Combining this fact with (4.5.5d) and Step 2 of Algorithm 4.5.1,

$$f^0(x_{i+1}) - f^0(x_i) \le -\tfrac{1}{2}\delta'\underline{\lambda}\ , \tag{4.5.5e}$$

for $i > i_1$, $i \in K$. Since $f^0(x_i)$ is monotonically decreasing, (4.5.5e) implies that $f^0(x_i) \to -\infty$ as $i \to \infty$. This is impossible, however, since $f^0(x_i) \to f^0(\bar{x})$ as $i \xrightarrow{K} \infty$. The contradiction proves that $\liminf_{i\to\infty}|\theta(x_i)| = 0$.

□

Recall the definitions of $J_{GQP}(x)$ and $J_{PP}(x)$ in (4.4.2b) and (4.4.3b) respectively, and that $h_{GQP}(x)$ denotes the solution to GQP$(x)$.

**Lemma 4.5.1:** *Suppose that assumptions* (i)-(iii) *of Theorem 4.2.3 hold, and that* (iv) *strict complementary slackness holds at the solution,* $\hat{x}$, *of (4.1.1), ( i.e. - for every* $\mu \in U(\hat{x})$ *and* $j \in \mathbf{p}$, $\mu^j > 0$ if and only if $f^j(\hat{x}) = 0$*). Then, there exist a neighborhood,* $V''$, *of* $\hat{x}$, $h' \in \mathbb{R}^n$ *and* $\delta > 0$ *such that, for all* $x \in V''$, (a) $J_{PP}(x) = J_{GQP}(x)$, *and* (b) $d(x) = h_{GQP}(x)$.

**Proof:** First we observe that assumption (ii) of Theorem 4.2.3 implies that $\psi(\hat{x}) \le 0$. Assumption (iv), above, implies that $U_{PP}(\hat{x})$ is a singleton $\{\hat{\mu}\}$ for some $\hat{\mu} \in \Sigma_{p+1}$, and hence that $\hat{J} = J(\hat{\mu}) = \{ j \in \mathbf{p} \mid f^j(\hat{x}) = 0 \}$. Let $V$ be as defined in Lemma 4.3.1.

(a) Because (i) $U_{PP}(\hat{x}) = \{\hat{\mu}\}$, (ii) $U_{PP}(\cdot)$ is an upper semicontinuous, compact-valued set-valued map, and (iii) $\hat{\mu}^j > 0$ for all $j \in \hat{J}$, there exists a neighborhood $W_0 \subset V$ of $\hat{x}$ such that $\mu^j > 0$ for every $j \in \hat{J}$ and $\mu \in U_{PP}(W_0)$. From the definition of $J_{PP}(x)$ in (4.4.3b), $J_{PP}(x) \supset \hat{J}$ for all $x \in W_0$. Now we show that $J_{PP} \subset \hat{J}$. By strict complementary slackness, $f^j(\hat{x}) < 0$ for every $j \notin \hat{J}$. Since $h(\hat{x}) = 0$ and $h(\cdot)$ is continuous [Pol.4], there exists a neighborhood, $W_1 \subset W_0$, of $\hat{x}$ such that $\tilde{f}^j(h(x) \mid x) - \psi_+(x) < 0$ for all $j \notin \hat{J}$ and $x \in W_1$. It follows from the definition of $U_{PP}(x)$ that $\mu^j = 0$ for every $j \notin \hat{J}$ and every $\mu \in U_{PP}(W_1)$. Hence $j \notin \hat{J}$ implies $j \notin J_{PP}(W_1)$. Therefore, $J_{PP}(x) = \hat{J}$ for every $x \in W_1$.

By a similar argument, we show that $J_{GQP}(x) = \hat{J}$ for all $x$ contained in a neighborhood of $\hat{x}$. (i) Since $h_{GQP}(\hat{x}) = 0$ and $\psi_+(\hat{x}) = 0$, an inspection of (4.3.1d) reveals that $U_{GQP}(\hat{x}) = U_{PP}(\hat{x}) = \{\hat{\mu}\}$. (ii) Lemma 4.3.1 implies that $h_{GQP}(x)$ is continuous in $W_1$, and hence $U_{GQP}(x)$ is an upper semicontinuous, compact-valued set-valued map. (iii) For all $j \in \hat{J}$, $\hat{\mu}^j > 0$. Hence, there exists a neighborhood, $W'_0 \subset V$, of $\hat{x}$ such that $\mu^j > 0$ for every $j \in \hat{J}$ and $\mu \in U_{GQP}(W'_0)$. From the definition of $U_{GQP}(x)$ in (4.3.1d), this implies that $\tilde{f}(h_{GQP}(x) \mid x) = 0$ for $j \in \hat{J}$ and $x \in W'_0$. Hence, by the definition of $J_{GQP}(x)$ in (4.4.2b)

$J_{GQP}(x) \supset \hat{J}$ for every $x \in W'_0$. Now we show that $J_{GQP}(x) \subset \hat{J}$. By strict complementary slackness, $f^j(\hat{x}) < 0$ for every $j \notin \hat{J}$. Since $h_{GQP}(\hat{x}) = 0$ and $h_{GQP}(\cdot)$ is continuous, there exists a neighborhood $W'_1 \subset W'_0$ of $\hat{x}$ such that $\tilde{f}^j(h_{GQP}(x) \mid x) < 0$ for every $j \notin \hat{J}$ and $x \in W'_1$. From the definition of $U_{GQP}(x)$, $\mu^j = 0$ for every $j \notin \hat{J}$ and every $\mu \in U_{GQP}(W'_1)$. Hence $j \notin \hat{J}$ implies that $j \notin J_{GQP}(W'_1)$. Therefore, $J_{GQP}(x) = \hat{J}$ for every $x \in W'_1$. Statement (a) holds with $V'' = W_1 \cap W'_1$.

(b) This follows from (a) and Proposition 4.4.3(c).

□

The following theorem asserts that, under an additional strict complementarity assumption, the implementable Algorithm 4.5.1 has the same asymptotic rate of convergence as Local Algorithm 4.3.1. Without the strict complementarity assumption, the bound on the convergence ratio which can be obtained for Algorithm 4.5.1 is the same as that obtained for Algorithm 4.2.1 in Theorem 4.2.3. However, an improved bound is not obtained for Algorithm 4.2.1 under this additional assumption. Under the strict complementarity assumption, $U_{PP}(\hat{x}) = \{ \hat{\mu} \}$ for some $\hat{\mu} \in \Sigma_{p+1}$ and hence $\underline{\mu}^0 = \bar{\mu}^0 = \hat{\mu}^0$.

**Theorem 4.5.2:** *Suppose that assumptions* (i)-(iii) *of Theorem 4.2.3 hold, that* (iv) *strict complementary slackness holds at* $(\hat{x}, \hat{\mu})$ *for every* $\hat{\mu} \in U_{PP}(\hat{x})$, *(i.e. - for every* $j \in \underline{p}$, $\hat{\mu}^j > 0$ if and only if $f^j(\hat{x}) = 0$*), and that Algorithm 4.5.1 constructs a sequence* $\{ x_i \}_{i=0}^{\infty}$ *in solving (4.1.1). Then,* (a) $x_i \to \hat{x}$ *as* $i \to \infty$, (b) *for any* $\varepsilon < \hat{\mu}^0 / (1 - \hat{\mu}^0)$,

$$\limsup_{i \to \infty} \frac{p_\varepsilon(x_{i+1}) - p_\varepsilon(\hat{x})}{p_\varepsilon(x_i) - p_\varepsilon(\hat{x})} \le 1 - \beta \frac{m}{M} \min \left\{ \frac{\varepsilon}{\hat{\mu}^0(1+\varepsilon)}, 1 \right\} , \tag{4.5.6a}$$

*and* (c) *if* $\psi(x_{i_0}) \le 0$ *for any* $i_0 \in \mathbb{N}$,

$$\limsup_{i \to \infty} \frac{f^0(x_{i+1}) - f^0(\hat{x})}{f^0(x_i) - f^0(\hat{x})} \le 1 - \beta \frac{m}{M} . \tag{4.5.6b}$$

Proof: (a) The sequence lies in the bounded set $L$ defined in assumption (ii) of Theorem 4.2.3, and hence it converges to the set of its accumulation points. By Theorem 4.5.1, $\liminf_{i\to\infty}|\theta(x_i)| = 0$.

We prove that $\hat{x}$ must be an accumulation point. Suppose not. Then there exists a neighborhood $W$ of $\hat{x}$ such that $\{x_i\}_{i\in\mathbb{N}} \subset L \setminus W$. By assumption (ii) of Theorem 4.2.3, there is no point in $L \setminus W$ which satisfies (4.2.5a-b). Since $L \setminus W$ is compact, this and Theorem 4.2.1(b) imply that $\inf\{\,\theta(x) \mid x \in L \setminus W\,\} > 0$. But this contradicts the fact that $\liminf_{i\to\infty}|\theta(x_i)| = 0$. Hence $\hat{x}$ must be an accumulation point.

Let $V''$ be as defined in Lemma 4.5.1. The iteration maps (see the proof of Theorem 4.3.1) of Algorithm 4.3.1 and 4.5.1 coincide for $x \in V''$. By Theorem 4.3.1, there exists a neighborhood $V''' \subset V''$ of $\hat{x}$ such that if the sequence $\{x_i\}_{i\in\mathbb{N}}$ enters $V'''$, it remains in $V'''$ and converges to $\hat{x}$. Since $\hat{x}$ is an accumulation point of the sequence, it must enter $V'''$. Hence, the sequence converges to $\hat{x}$.

(b) and (c) Since, by (a), $\{x_i\}_{i\in\mathbb{N}}$ converges to $\hat{x}$ and the iteration map of Algorithm 4.5.1 coincides with that of Algorithm 4.3.1 in the neighborhood $V''$ of $\hat{x}$, the results of Theorem 4.3.2 hold. Since $U_{PP}(\hat{x}) = \{\,\hat{\mu}\,\}$, $\underline{\mu}^0 = \bar{\mu}^0 = \hat{\mu}^0$.

□

## 4.6 NUMERICAL EXPERIMENTS

Algorithm 4.5.1 was compared with Algorithm 4.2.1 and the feasible descent algorithm in [Her.1] (which also satisfies (4.1.3)) on several well-known inequality-constrained problems. Table 4.1 summarizes the performances of the three algorithms on these problems. The results for the algorithm of [Her.1] are quoted from that paper.

The algorithm parameters for both Algorithms 4.2.1 and 4.5.1 were set at $\alpha = 0.9$, $\beta = 0.9$, $\gamma = 1.0$ in the experiments. To reduce the number of trial step sizes tested in the Armijo step rule, quadratic interpolation was used at each iteration of both algorithms to determine the initial trial step size.

The Rosen-Suzuki problem is problem 43 in [Hoc.1]. See Figure 4.1 for a comparison of the performance of Algorithms 4.2.1 and 4.5.1. (The y-axis label "Cost Error" of the figures refers to the quantity, $f^0(x_i) - f^0(\hat{x})$). Colville's Test Problems One and Two are problems 86 and 117, respectively, in [Hoc.1].

**Kuhn-Tucker Problem [Con.1]:** This problem has a unique minimizer at which neither the Kuhn-Tucker constraint qualification nor the Mangasarian-Fromovitz constraint qualification holds. It serves as a test of algorithm robustness. The minimum value of $-1$ occurs at $\hat{x} = (0, 1)$. Both algorithms converged to the solution from the feasible initial point $x_0 = (0.25, 0.25)$. However, Algorithm 4.2.1 converged sublinearly, while Algorithm 4.5.1 converged linearly. See Figure 4.2. □

**Circular-Quadratic Problem:** In this problem, the function approximations are exact for $\gamma = 1$, that is, $\tilde{f}^j(h \mid x) = f^j(x + h)$ for $j \in \mathbf{p} \cup 0$.

$$\min \left\{ \tfrac{1}{2}(x_1^2 + (x_2+4)^2) \mid \tfrac{1}{2}((x_1+1)^2 + x_1^2) - 2 \le 0 \,,\ \tfrac{1}{2}((x_1-1)^2 + x_2^2) - 2 \le 0 \right\} . \quad (4.6.3)$$

The minimum value of 4.5 occurs at $\hat{x} = (0, -1)$; the feasible initial point $x_0 = (1, 1)$ was used. □

**Infeasible Problem:** This simple problem was constructed to demonstrate the behavior of the algorithms when the constraints cannot be satisfied.

$$\min \left\{ -x_1 \mid (x_1+10)^2 + x_2^2 \le 0 \,, (x_1-10)^2 + x_2^2 \le 0 \right\} . \quad (4.6.2)$$

The minimum value of 1 occurs at the origin. Both Algorithms 4.2.1 and 4.5.1 converged to the solution from the initial point $x_0 = (-10, -20)$. □

## 4.7 CONCLUSIONS

We obtained a bound on the convergence ratio of sequences $\{ f^0(x_i) \}_{i \in \mathbf{N}}$ constructed by Algorithm 4.5.1 which is smaller than that obtained for Algorithm 4.2.1. On all of the standard problems on which they were tested, Algorithm 4.5.1 far surpassed the performance of Algorithm 4.2.1 and was competitive with the first-order feasible descent algorithm of [Her.1]. Search Direction Subprocedure 4.4.1 was developed as a method for approximating the solution to the GQP subproblem. The above facts show that the subprocedure can profitably be viewed as a speed-enhancing correction to the method of centers search direction (4.2.2a).

## 4.8 APPENDIX

The following two theorems are special cases of Theorems 4.6 and 4.9 of [Han.2], used in the proof of Lemma 4.3.2.

**Theorem 4.8.1:** [Han.2] *Consider the problem*

$$\min_{x \in \mathbb{R}^n} \{ g^0(x) \mid g^j(x) \le 0, \ \forall j \in \mathbf{p} \}, \tag{4.8.1}$$

*and suppose that the functions $g^j(\cdot)$ are twice continuously differentiable.*

*If $\bar{x} \in \mathbb{R}^n$, together with a Kuhn-Tucker multiplier vector $\bar{u} \in \mathbb{R}^n_+$, satisfies the standard second-order sufficiency conditions [McC.1], then, for any $\varepsilon < 1 / \|\bar{u}\|_1$, $\bar{x}$ is a strict local minimizer of the function $\varepsilon g^0(\cdot) + \max_{j \in \mathbf{p}} g^j(\cdot)$.*

**Theorem 4.8.2:** *[Han.2] Consider the problem (4.8.1) and suppose that (i) the functions $g^j(\cdot)$ are convex and continuously differentiable, and (ii) there exists $x' \in \mathbb{R}^n$ such that $g^j(x') < 0$ for all $j \in \mathbf{p}$.*

*If $\bar{x} \in \mathbb{R}^n$, together with a Kuhn-Tucker multiplier vector $\bar{u} \in \mathbb{R}^n_+$, satisfies the standard second-order sufficiency conditions [McC.1], then, for any $\varepsilon < 1 / \|\bar{u}\|_1$, $\bar{x}$ is a global minimizer of the function $\varepsilon g^0(\cdot) + \max_{j \in \mathbf{p}} g^j(\cdot)$.*

| Problem | Algorithm | NF | NG | NDF | NDG | FV |
|---|---|---|---|---|---|---|
| Rosen-Suzuki | [Her.1] | 7 | 27 | 7 | 21 | -43.81453 |
| | Algorithm 2.1 | 66 | 198 | 33 | 99 | -43.83851 |
| | Algorithm 5.1 | 6 | 18 | 3 | 9 | -43.82342 |
| | [Her.1] | 15 | 54 | 15 | 45 | -43.99907 |
| | Algorithm 2.1 | 132 | 396 | 66 | 198 | -43.99912 |
| | Algorithm 5.1 | 20 | 60 | 10 | 30 | -43.99927 |
| Colville #1 | [Her.1] | 6 | 60 | 6 | 60 | -32.03453 |
| | Algorithm 2.1 | 265 | 2650 | 127 | 1270 | -32.06142 |
| | Algorithm 5.1 | 12 | 120 | 6 | 60 | -32.21449 |
| | [Her.1] | 9 | 90 | 9 | 90 | -32.34851 |
| | Algorithm 2.1 | 884 | 8840 | 436 | 4360 | -32.34851 |
| | Algorithm 5.1 | 32 | 320 | 16 | 160 | -32.34865 |
| Colville #2 | [Her.1] | 36 | 190 | 36 | 180 | 32.81567 |
| | Algorithm 2.1 | 1840 | 9200 | 872 | 4360 | 32.81530 |
| | Algorithm 5.1 | 526 | 2630 | 246 | 1230 | 32.66952 |
| | [Her.1] | 53 | 320 | 53 | 265 | 32.34897 |
| | Algorithm 5.1 | 1741 | 8705 | 324 | 1620 | 32.34906 |
| Kuhn-Tucker | Algorithm 2.1 | 92 | 184 | 46 | 92 | -0.9009127 |
| | Algorithm 5.1 | 45 | 90 | 6 | 12 | -0.92223418 |
| | Algorithm 2.1 | 6116 | 12232 | 3058 | 6116 | -0.9900006 |
| | Algorithm 5.1 | 110 | 220 | 15 | 30 | -0.9905035 |
| Circular-Quadratic | Algorithm 2.1 | 10 | 20 | 5 | 10 | 4.526097 |
| | Algorithm 5.1 | 2 | 4 | 1 | 2 | 4.530063 |
| | Algorithm 2.1 | 54 | 108 | 27 | 54 | 4.500000 |
| | Algorithm 5.1 | 4 | 8 | 2 | 4 | 4.500000 |

**Table 4.1: Summary of Numerical Results**

The abbreviations in the table have the following meanings:

NF: Number of objective function evaluations.

NG: Number of constraint function evaluations.

NDF: Number of gradient evaluations of the objective function.

NDG: Number of gradient evaluations of the constraints.

FV: Value of the objective function at the final iterate.

Each constraint was counted separately in the tabulation of NG and NDG. Bounds on the variables, i.e., $x^i \le 0$, were not included in the tabulation.

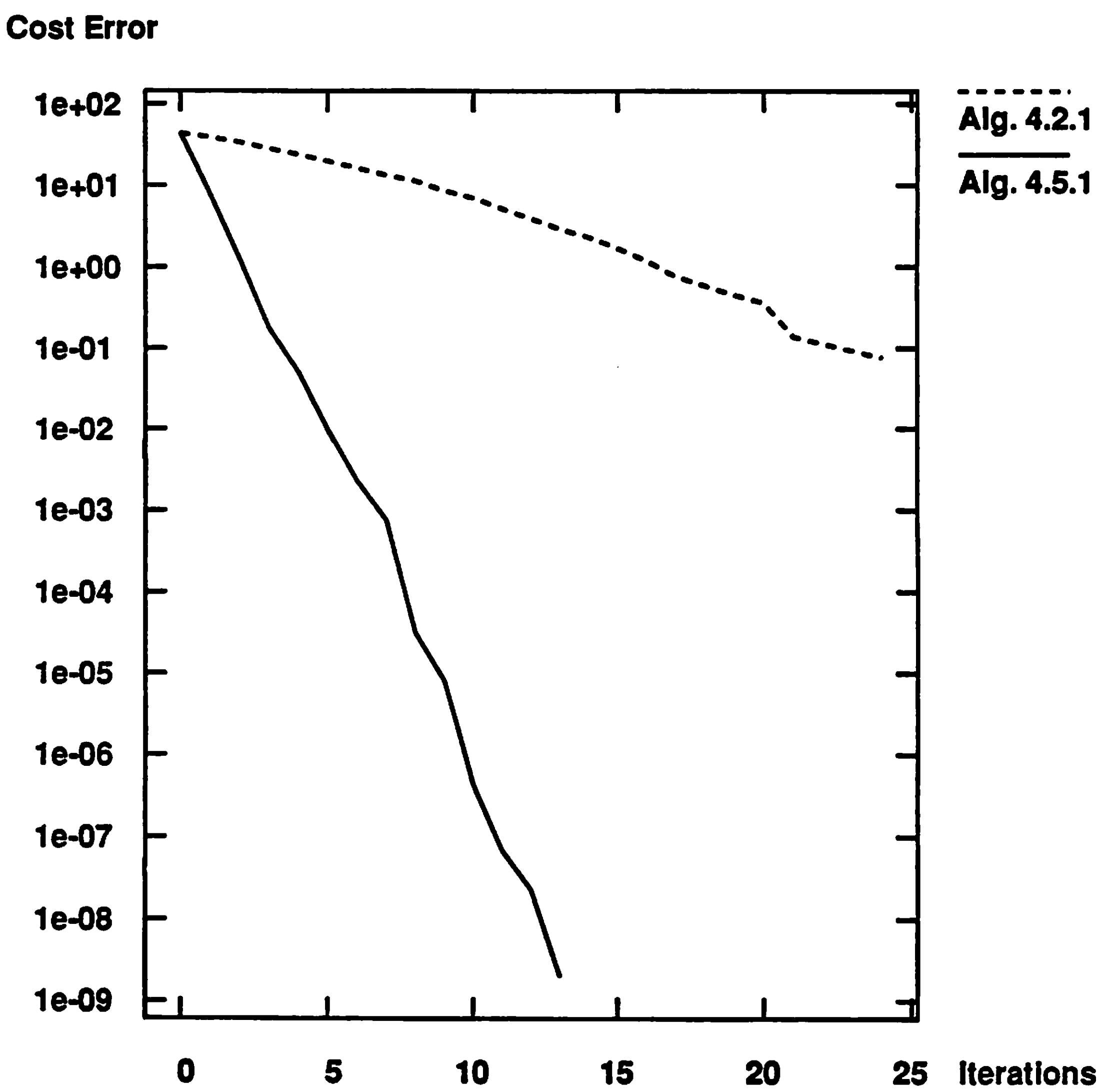


Figure 4.1: Rosen-Suzuki Problem

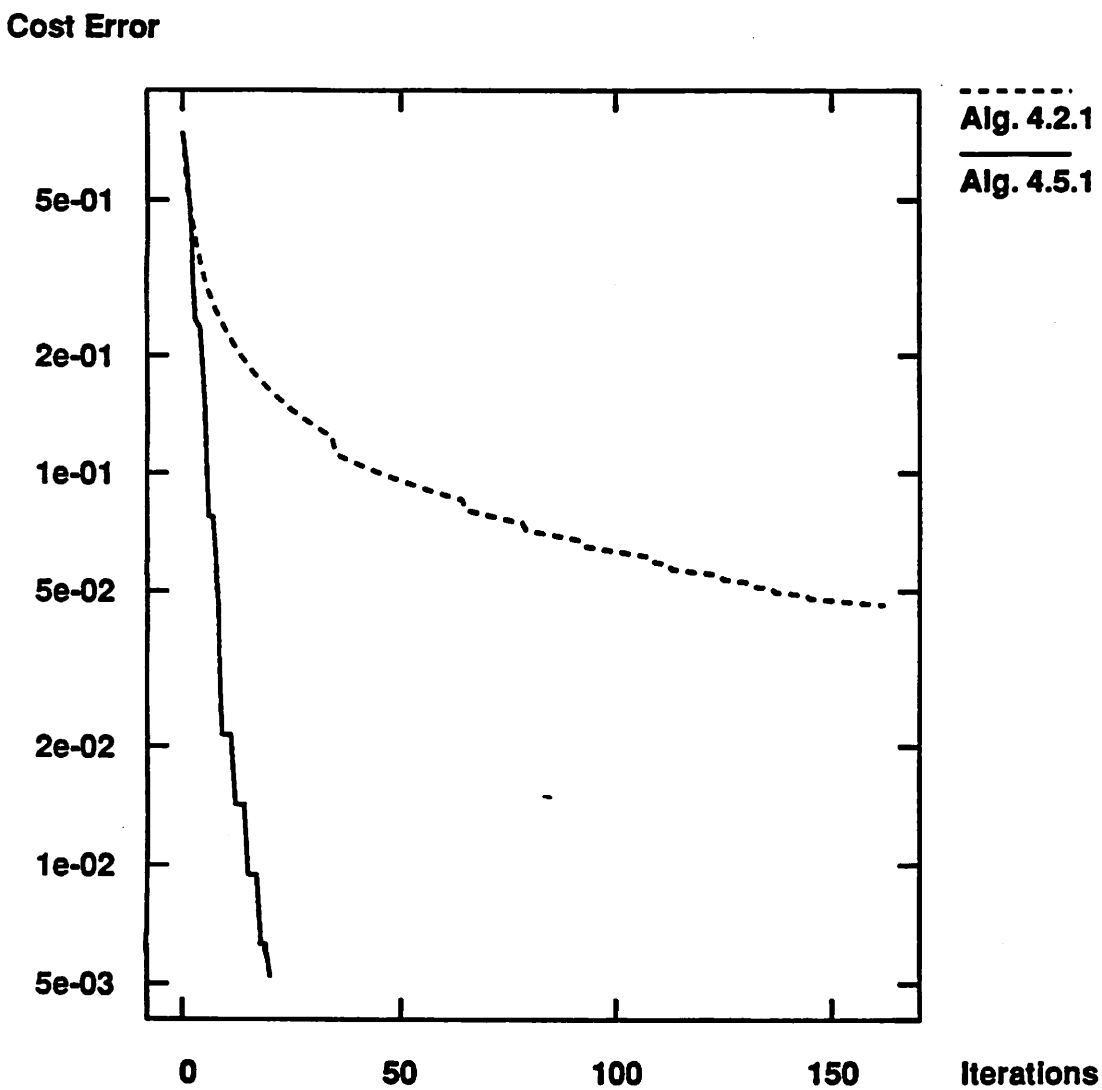


Figure 4.2: Kuhn-Tucker Problem

# CHAPTER 5

# SUPERLINEARLY CONVERGENT GENERALIZED QUADRATIC PROGRAMMING-BASED METHODS

## 5.1 INTRODUCTION

We consider the inequality-constrained nonlinear programming problem,

$$\text{ICP} \qquad \min_{x \in \mathbb{R}^n} \{ f^0(x) \mid f^j(x) \leq 0, \forall j \in \mathbf{p} \} , \tag{5.1.1}$$

where p denotes the set $\{ 1, \ldots, p \}$ and each function $f^j : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. In this chapter, we develop second-order algorithms based upon successive approximation to the problem ICP, as proposed in [Pol.10]. The search direction subproblem for such algorithms is obtained from ICP by replacing all of the functions $f^j(\cdot)$ by quadratic approximations,

$$\tilde{f}^j(x \mid x_k, H) \triangleq f^j(x_k) + \langle \nabla f^j(x_k), x - x_k \rangle + \tfrac{1}{2} \langle x - x_k, H_k^j(x - x_k) \rangle , \tag{5.1.2}$$

where $H_k = [H_k^0, \ldots, H_k^p]$ and each $H_k^j \in \mathbb{R}^{n \times n}$ is a matrix approximating $\nabla^2 f_j(x_k)$.[1] The result is a quadratically constrained quadratic program which we call a *generalized quadratic program* (GQP),

$$\text{GQP}(x_k, H): \qquad \min_{x \in \mathbb{R}^n} \{ \tilde{f}^0(x \mid x_k, H) \mid \tilde{f}^j(x \mid x_k, H) \leq 0, \ \forall j \in \mathbf{p} \} . \tag{5.1.3}$$

The GQP-based algorithms which we develop are novel in their use of the full second-order information. As a rule, the second-order information used by other algorithms is limited to an estimate of the Lagrangian Hessian (or a submatrix thereof). The Lagrangian Hessian is a linear combination of the Hessians of the individual functions. If exact second derivatives or finite difference approximations are used, the Lagrangian Hessian estimate is generally formed by combining the Hessians of the individual functions. If the Lagrangian Hessian estimate is updated as

---

[1] Note that this definition of $\tilde{f}^j(\cdot \mid \cdot)$ differs from that used in Chapter 4, not only in the addition of the argument $H$, but also in

in a variable metric method, then the update is a function of vectors which are linear combinations of vectors which could be used to update individual Hessian estimates. In both cases, the Lagrangian Hessian estimate is formed by combining information which could be used to provide estimates of the individual Hessians. Hence, these algorithms *discard* information about the differences among the curvatures of the functions $f^j(\cdot)$.[2] In contrast, the GQP subproblem naturally incorporates all of this second-order information. This information will not, of course, enable GQP-based algorithms to converge faster than quadratically, but we undertook this research in the hope that the use of this information would speed convergence when the algorithm was far from a solution, where quadratic terms are not dominated by linear terms.

Algorithms based on GQP subproblems have been proposed before. In [Pol.10], GQP-based algorithms are proposed, and a convergence theory is developed for them. Minimax algorithms have been based on subproblems obtained in a similar way from the minimax problem [Pol.9, Pol.11, Pan.3] and the constrained minimax problem [Pan.4].[3] Rates of convergence are obtained in [Pol.9, Pol.11] under assumptions of uniform convexity. It is shown in [Pan.4] that, on uniformly convex problems, the norms of the search directions constructed by a conceptual GQP-based algorithm converge superlinearly to zero as the iterates approach a solution.

In this chapter, we develop a comprehensive theory of convergence and rate of convergence for a class of algorithms based on second-order GQP subproblems. Our convergence rate theory shows that these algorithms will achieve rates of convergence ranging from Q-superlinear [Ort.1] to Q-order 3/2, depending on the accuracy of the Hessian approximations. The results hold for a class of algorithms characterized by an algorithm model and a set of generic conditions. The assumptions made about the problem in our main theorem are weaker than the assumptions usually made in superlinear convergence theorems. Our convergence rate theorem requires neither

the origin of the first argument.

[2]Consider the fact that the Lagrangian Hessian consists of $n^2$ numbers, while the individual Hessians contain $(p+1)\times n^2$ numbers.

[3]Quadratic constraints have also appeared in the subproblems of trust region algorithms [Mor.1-2]. However, in these algorithms, they function to limit the search direction, rather than to represent the constraints of the problem.

strict complementarity nor linear independence of the gradients of the active constraints at the solution. The sufficiency condition which it assumes is weaker than that in [Theorem 3.2 of Rob.1], and the Mangasarian-Fromovitz constraint qualification used is also quite weak. However, the assumption made in our main theorem about the convergence of Hessian estimates, Algorithm Property 5.4.7, is stronger than usual.

We also propose an efficient method for solving the GQP subproblem. The GQP-based algorithms proposed in [Pol.9-11, Pan.3-4] were conceptual, that is, they assumed that the GQP subproblem is solved exactly. These algorithms were not implemented (to the author's knowledge) because no finite step procedures for solving the GQP subproblem were known [Pol.4, Pan.4].[4] We resolve this difficulty in two steps. First, the convergence rate theory which we develop requires that the GQP be solved only to within a prescribed accuracy. Second, we propose an efficient method for approximating the solution of the GQP to this accuracy. Our approximation method requires the solution of one quadratic program and the inversion of a $(n+p)\times(n+p)$ matrix. Our proof that the approximation method provides the required accuracy does rely on a strict complementarity condition.

The class of algorithms we describe are *feasible descent* algorithms for ICP. They are shown to converge superlinearly. The only other such algorithms known to the author are those in [Pan.1-2].

In Section 2, we present an algorithm model; in Section 3, we prove that algorithms which conform to the model and which possess certain properties converge. In Section 4, we derive rates of convergence for algorithms conforming to the algorithm model. In Section 5, we develop a method for approximately solving the GQP and incorporate it into an example algorithm which conforms to the algorithm model. In Section 6, we present the results of numerical experiments

[4] The GQP-based minimax algorithm of [Pol.9] was implemented, but because no efficient finite method for approximating the solution to the GQP subproblem was known, a nonlinear programming algorithm was applied to solve the subproblem. Because no stopping rule was available, the subproblem was solved to high precision at each iteration. As a result, the search direction computation consumed so much time that the algorithm's overall efficiency was little better than that of a first-order minimax algorithm. The convergence rate theory developed in [Pol.11] resolved this problem by including a stopping rule for the solution of the search direction subproblem which preserves the rate of convergence of the conceptual algorithm.

with the example algorithm.

## 5.2 A PHASE II ALGORITHM MODEL

In this section, we present an algorithm model for a class of feasible descent algorithms for solving ICP. Each algorithm in the class is characterized by a *search arc function*, $A(x, H)$, and by the method used to construct Hessian approximations, $H_j$. The search arc function,

$$A : \mathbb{R}^n \times \mathbb{R}^{n \times (p+1)n} \to PC([0,1], \mathbb{R}^n), \tag{5.2.1}$$

maps the parameters $(x, H)$ (which define the subproblem GQP$(x, H)$) into a piecewise continuous arc in $\mathbb{R}^n$. We also require that $A(x, H)(s)$ be differentiable at $s = 0$. We give an example of a search arc function in a familiar setting. An algorithm which improves on an iterate $x_i$ by performing a line search in the direction $h(x_i, H_i)$ has a search arc function given by $A(x, H)(s) = x + sh(x, H)$.

We define $\mathbf{I} = [I_n, ..., I_n] \in \mathbb{R}^{n \times (p+1)n}$ where $I_n$ is the $n \times n$ identity. Also, for compactness we will write $a(\cdot)$ for $A(x, H)(\cdot)$, suppressing the $x$ and $H$ dependence, and we will let $\dot{a}(s)$ denote $da(s) / ds$. We define $\psi(x) \triangleq \max_{j \in \mathbf{p}} f^j(x)$.

**Algorithm Model 5.2.1**

**Data:** $x_0 \in \mathbb{R}^n$ such that $\psi(x_0) \le 0$, $H_0 \in \mathbb{R}^{n \times n(p+1)}$, $\alpha, \beta \in (0, 1)$.

**Step 1:** Construct a search arc $a_k = A(x_k, H_k)$.

**Step 2:** Take the step length $s_k$ to be the largest element $s$ of the set $\{ 1, \frac{1}{2}\beta^k, 1 - \frac{1}{2}\beta^k \}_{k \in \mathbb{N}}$ such that

$$\psi(a_k(s)) \le 0, \tag{5.2.2a}$$

$$f^0(a_k(s)) - f^0(x_k) \le s\,\rho(\|\dot{a}_k(0)\|) \max_{j \in \mathbf{p} \cup 0} \tilde{f}^j(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathbf{I}), \tag{5.2.2b}$$

where $\rho(t) \triangleq \alpha t / (1+t)$.

Step 3: Set $x_{k+1} = a_k(s_k)$, update $H_k$, replace $k$ by $k+1$ and go to Step 1.

□

The ½ preceding $\dot{a}_k(s)$ in (5.2.2b) and (5.3.2b) was chosen for the convenience of our proofs; it could be replaced by any positive constant.

In Algorithm Model 5.2.1, each iteration involves the computation of a search arc and the selection of a step size along the search arc using an Armijo-like step size rule [Arm.1, Pol.4]. The step size rule differs from more familiar Armijo-type rules in [Pan.1-2] in two ways. First, $s = 1$, as well as $s = 0$, is an accumulation point of the set of trial step sizes. This allows a near-unity step size to be accepted when $s = 1$ would yield a slightly infeasible point, and prevents the degradation of superlinear convergence to linear. Second, the quantity on the right-hand side of the inequality is a generalization of the usual Armijo-rule term $s\alpha\langle \nabla f^0(x_k), \dot{a}_k(0)\rangle$. The coefficient $\rho(\|\dot{a}_k(0)\|)$ tends to zero as $x_k$ tends to a solution of ICP, which ensures that near-unity step sizes satisfy (5.3.2b) for $x_k$ near such a solution.

To simplify the exposition below, we will use the expression "*for x near y*" to mean "for all $x$ in a neighborhood of $y$". We will also use the expression "*for large k*" to mean "for every $k \in \mathbb{N}$ greater than some $k_0 \in \mathbb{N}$". Finally, we will have no need to distinguish among many of the constants which appear below. To avoid accumulating long expressions for them, we adopt the following shorthand. A single symbol $K$ will denote any *large* positive real constant. Hence, $K + K = K$ and $K^2 = K$. We will avoid trouble by refraining from subtracting $K$ or dividing by $K$. Similarly, a single symbol $\delta$ will denote any *small* positive real constant.

## 5.3 GLOBAL CONVERGENCE

In this section, we prove a convergence result which applies to any algorithm which conforms to Algorithm Model 5.2.1 and which is based on a search arc function possessing certain properties. We restate the Fritz John necessary conditions for optimality for ICP. Letting $\mathbf{p}\cup 0$ denote the set $\mathbf{p}\cup\{0\}$, we denote the *standard unit simplex* by

$\Sigma_{p+1} \triangleq \{ \mu \in \mathbb{R}^{p+1} \mid \sum_{j \in \mathbf{p} \cup 0} \mu^j = 1, \mu \geq 0 \}$ . Note that we index the components of vectors in $\Sigma_{p+1}$ from 0 to $p$.

**Theorem 5.3.1 [Cla.1, Dem.3, Joh.1, Pol.4]:** *If $\bar{x}$ solves* ICP, *then there exists $\mu \in \Sigma_{p+1}$ such that*

$$\sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(\bar{x}) = 0 \tag{5.3.1a}$$

$$\sum_{j \in \mathbf{p}} \mu^j f^j(\bar{x}) = 0 . \tag{5.3.1b}$$

Any feasible point which satisfies conditions (5.3.1a-b) is a *stationary point* for the ICP.

In order to show that an algorithm conforming oto Algorithm Model 5.2.1 converges, we must assume that the search arc function $A(\cdot, \cdot)$ has the following properties. For any $x$ and $H$, the arc $A(x, H)$ must begin at the point $x$, and the beginning of the arc must be both smooth and tangent to a feasible descent direction. We will describe the variety of search arc functions possessing Algorithm Property 5.3.2 in Section 5. In that section, we will exhibit an example of a search arc function and prove that it possesses Algorithm Property 5.3.2.

**Algorithm Property 5.3.2:** *Consider any $\bar{x} \in \mathbb{R}^n$ and compact set $\mathbf{H} \subset \mathbb{R}^{(p+1) \times n \times n}$. The following hold for $x$ near $\bar{x}$ and $H \in \mathbf{H}$ (using the notation $a \triangleq A(x, H)$ and $\dot{a}(s) = da(s)/ds$).*

(i) *There exist $\delta > 0$ and $K > 0$ such that,*

$$\| a(s) - ( a(0) + \dot{a}(0)s ) \| \leq Ks^2 , \tag{5.3.2a}$$

*for $s$ near 0. Also, $a(0) = x$ and $\delta < \| \dot{a}(0) \| < K$.*

(ii) *If $\bar{x} \in \mathbb{R}^n$ is not a stationary point of* ICP, *there exists $\delta > 0$ such that*

$$\tilde{f}^j(x + \tfrac{1}{2}\dot{a}(0) \mid x, I) < -\delta \quad , \tag{5.3.2b}$$

*for all $j \in \mathbf{p} \cup 0$.*

**Theorem 5.3.3:** *Suppose that an algorithm conforming to Algorithm Model 5.2.1 constructs a sequence* $\{ x_k \}_{k \in \mathbf{N}}$ *starting from a feasible point* $x_0$, *and that*

(i) *the functions* $f^j(\cdot)$ *are Lipschitz continuously differentiable,*

(ii) *the search arc function* $A(\cdot,\cdot)$ *satisfies Algorithm Property 5.3.2,*

(iii) *the set* $\{ H_k \}_{k \in \mathbf{N}}$ *is bounded.*

*If the sequence* $x_k$ *has an accumulation point* $\bar{x}$, *then* $\bar{x}$ *satisfies the Fritz John necessary conditions for optimality* (5.3.1a-b).

Proof: Suppose that the subsequence $\{ x_k \}_{k \in L}$, where $L \subset \mathbf{N}$, converges to $\bar{x}$, but that $\bar{x}$ does not satisfy the necessary conditions (5.3.1a-b). Since the functions $f^j(\cdot)$ are Lipschitz continuously differentiable, there exists $K > 0$ such that

$$f^j(a_k(s)) \le f^j(x_k) + \langle \nabla f^j(x_k), a_k(s) - x_k \rangle + K \| a_k(s) - x_k \|^2$$

$$\le f^j(x_k) + \langle \nabla f^j(x_k), s\dot{a}_k(0) \rangle + \| \nabla f^j(x_k) \| \, \| a_k(s) - x_k - s\dot{a}_k(0) \| + K \| a_k(s) - x_k \|^2 \qquad (5.3.3a)$$

using the triangle inequality. By Algorithm Property 5.3.2(i) and the triangle inequality, there exists $K > 0$ such that

$$\| \nabla f^j(x_k) \| \, \| a_k(s) - x_k - s\dot{a}_k(0) \| + K \| a_k(s) - x_k \|^2 \le ( \| \nabla f^j(x_k) \| + K ) \, \| a_k(s) - x_k - s\dot{a}_k(0) \| + K \| s\dot{a}_k(0) \|^2 \qquad (5.3.3b)$$

$$( \| \nabla f^j(x_k) \| + K ) \, Ks^2 + Ks^2 + 2Ks^3 ,$$

for $s$ near 0, large $k \in L$ and $j \in \mathbf{p}$. Substituting (5.3.3b) into (5.3.3a) yields, for some $K > 0$,

$$f^j(a_k(s)) \le f^j(x_k) + s \langle \nabla f^j(x_k), \dot{a}_k(0) \rangle + Ks^2$$

$$\le 2\tilde{f}^j(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathrm{I}) + Ks^2 , \qquad (5.3.3c)$$

for $s$ near 0, large $k \in L$ and $j \in \mathbf{p}$. Similarly,

$$f^0(a_k(s)) - f^0(x_k) \le 2s\tilde{f}^0(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathrm{I}) + Ks^2 , \qquad (5.3.4)$$

for $s$ near 1. By the descent requirement of Algorithm Property 5.3.2(ii), there exist $\delta > 0$ such that

$$\tilde{f}^{\,j}(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathrm{I}) < -\delta\,, \tag{5.3.5}$$

for $j \in \mathbf{p}\cup 0$, $s$ near 0 and large $k \in L$. Since $f^{\,j}(x_k) \le 0$, substituting (5.3.5) into inequality (5.3.3) yields

$$f^{\,j}(a_k(s)) \le 0\,, \tag{5.3.6}$$

for $j \in \mathbf{p}$, $s$ near 0 and large $k \in L$. Similarly, since $\rho(\cdot) \le 1$,

$$f^{\,0}(a_k(s)) - f^{\,0}(x_k) \le s\,\rho(\|\dot{a}_k(0)\|)\tilde{f}^{\,0}(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathrm{I})\,, \tag{5.3.7}$$

for $s$ near 0 and large $k \in L$. Inequalities (5.3.6) and (5.3.7) and Step 2 of Algorithm Model 5.2.1 imply that there exists $\bar{s} > 0$ such that $s_k \ge \bar{s}$ for large $k \in L$. By Algorithm Property 5.3.2(i), there exists $\delta > 0$ such that $\|\dot{a}_k(0)\| > \delta$ for large $k \in L$. Applying this fact and (5.3.3) to (5.2.2b) yields

$$f^{\,0}(a_k(s)) - f^{\,0}(x_k) \le s_k\rho(\|\dot{a}_k(0)\|)\tilde{f}^{\,0}(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathrm{I}) \le -s_k\rho(\|\dot{a}_k(0)\|)\delta \le -\bar{s}\delta^2\,, \tag{5.3.8}$$

for large $k \in L$. Since $f^{\,0}(x_k)$ is nonincreasing, (5.3.8) implies that $f^{\,0}(x_k) \to -\infty$. However, by the continuity of $f^{\,0}(\cdot)$, $f^{\,0}(x_k) \to f^{\,0}(\bar{x})$ as $k \to \infty$, $k \in L$. This contradiction proves that $\bar{x}$ must satisfy the necessary conditions (5.3.1a-b).

□

## 5.4 RATE OF CONVERGENCE

We now derive rates of convergence for algorithms which conform to Algorithm Model 5.2.1 and which posses a further set of properties. The rate of convergence obtained varies from superlinear to 3/2 depending on the accuracy of the Hessian approximations. In the remainder of this chapter, we assume that the problem ICP satisfies the following hypotheses on the problem ICP. In addition, we will introduce assumptions about the behavior of the algorithm, labeled Algorithm Properties, as we need them.

**Hypothesis 5.4.1:** Suppose that

(i) *the functions $f^j(\cdot)$ are twice locally Lipschitz-continuously differentiable,*

(ii) *there exist $T_1, T_2 > 0$ such that the set $V \triangleq \{ x \in \mathbb{R}^n \mid \psi(x) \le T_1$ and $f^0(x) < T_2 \}$ is bounded and such that there exists a single point, $\hat{x}$, in the set $V$ which satisfies the necessary conditions (5.3.1a-b),*

(iii) *the point $\hat{x}$ is feasible.*

The following definitions are needed in order to frame further assumptions about the problem ICP. We assume below that $\hat{x}$ is as defined in Hypothesis 5.4.1.

**Definition:** We denote by $\hat{U}$ the subset of $\Sigma_{p+1}$ which, together with $\hat{x}$, satisfies (5.3.1a-b),

$$\hat{U} \triangleq \{ \mu \in \Sigma_{p+1} \mid \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(\hat{x}) = 0 \text{ and } \sum_{j \in \mathbf{p}} \mu^j f^j(\hat{x}) = 0 \} . \tag{5.4.1}$$

Using the definition $J(\mu) \triangleq \{ j \in \mathbf{p} \mid \mu^j > 0 \}$, we define the set of indices corresponding to strictly positive multipliers,

$$\hat{J} \triangleq \bigcup_{\mu \in \hat{U}} J(\mu) = \{ j \in \mathbf{p} \mid \exists \mu \in \hat{U} : \mu^j > 0 \} . \tag{5.4.2}$$

Furthermore, we let $B \triangleq \left[ Span \{ \nabla f^j(\hat{x}) \}_{j \in \hat{J}} \right]^{\perp}$, the null space of the matrix with columns $\{ \nabla f^j(\hat{x}) \}_{j \in \hat{J}}$. We also define $F^j(x) \triangleq \nabla^2 f^j(x)$ for each $j \in \mathbf{p} \cup 0$, and we denote by $F(x)$ the matrix $[F^0(x), \ldots, F^p(x)]$.

□

**Hypothesis 5.4.2:** *Let $\hat{x}$ be as defined in Hypothesis 5.4.1, and suppose that*

(i) *there exists $m > 0$ such that*

$$m \| h \|^2 < \langle h, \left[ \sum_{j \in \mathbf{p} \cup 0} \mu^j F^j(\hat{x}) \right] h \rangle \qquad \forall h \in B, h \ne 0, \ \forall \mu \in \hat{U} , \tag{5.4.3}$$

(ii) *there exists* $d \in \mathbb{R}^n$ *such that* $f^j(\hat{x}) + \langle \nabla f^j(\hat{x}), d \rangle < 0$ *for all* $j \in p$.

Hypotheses 5.4.1-2 constitute a strengthened version of the standard second-order sufficiency conditions for $\hat{x}$ to be a local minimizer of ICP. However, this condition is *weaker* than the strong second-order sufficiency condition used in [Cha.1, Theorem 3.2 of Rob.1]. In their second-order condition, the inequality in (5.4.3) must hold for every $h \in \{ \nabla f^j(\hat{x}) \}_{j \in J(\mu)})^{\perp}$, which is a larger set than

$$B = \bigcap_{\mu \in \hat{U}} \left[ Span \{ \nabla f^j(\hat{x}) \}_{j \in J(\mu)} \right]^{\perp}.$$

Hypothesis 5.4.2(ii) is equivalent to the Mangasarian-Fromovitz constraint qualification, which is quite weak.

In the following definition, we define an optimality (or merit) function which, as Lemmas 5.4.3 and 5.4.5 show, gauges the distance from the minimizer $\hat{x}$.

Definition: We define a measure of constraint violation by $\psi_+(x) \triangleq \max \{ \psi(x), 0 \}$ , and we will use the optimality function $\theta(\cdot)$ discussed in Chapter 4 with change of sign (and symbol to avoid confusion),

$$\sigma(x) \triangleq \psi_+(x) - \min_{x' \in \mathbb{R}^n} \max_{j \in p \cup 0} \tilde{f}^j(x' \mid x, \mathrm{I}), \tag{5.4.4}$$

where $\mathrm{I} = [I_n, ..., I_n] \in \mathbb{R}^{n \times (p+1)n}$ and $I_n$ is the $n \times n$ identity. The function $\sigma(\cdot)$ is nonnegative and is zero only at points satisfying the necessary conditions for optimality (5.3.1a-b). The function $\sigma(\cdot)$ can be rewritten using Theorem 2.7.1,

$$\sigma(x) = \min_{\mu \in \Sigma_{p+1}} \psi_+(x) - \sum_{j \in p} \mu^j f^j(x) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} \mu^j \nabla f^j(x) \|^2. \tag{5.4.5}$$

We define analogous quantities for the GQP subproblem. Let $\tilde{\psi}(x \mid \bar{x}, H) \triangleq \max_{j \in p} \tilde{f}^j(x \mid \bar{x}, H)$ and $\tilde{\psi}_+(x \mid \bar{x}, H) \triangleq \max_{j \in p} \{ \tilde{f}^j(x \mid \bar{x}, H), 0 \}$ . We will denote the quantity corresponding to $\sigma(\cdot)$ by

$$\tilde{\sigma}(x \mid \bar{x}, H) = \min_{\mu \in \Sigma_{p+1}} \tilde{\psi}_+(x \mid x, H) - \sum_{j \in \mathbf{p}} \mu^j \tilde{f}^j(x \mid \bar{x}, H) + \tfrac{1}{2} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla \tilde{f}^j(x \mid \bar{x} H) \|^2 . \quad (5.4.6)$$

□

We assume in the remainder of this section that Hypotheses 5.4.1-2 hold, with $T_1, T_2, \hat{x}$ and $m$ as defined there. We will also assume that an algorithm which conforms to Algorithm Model 5.2.1 constructs a sequence $\{ x_k \}_{k \in \mathbf{N}}$ starting from a feasible point $x_0$, and that the algorithm is based upon a search arc function $A(\cdot, \cdot)$ which possesses another set of properties, Algorithm Properties 5.4.6(i-iv) below. We will denote $A(x_k, H_k)$ by $a_k$. The convergence rate proof for such an algorithm is structured as follows.

The main convergence rate theorem, Theorem 5.4.12, is proved by combining two facts, which are proved in Lemmas 5.4.10 and 5.4.11, respectively. The first is that an unstabilized version of Algorithm Model 5.2.1, i.e., an algorithm for which $x_{k+1} \triangleq a_k(1)$, converges superlinearly in a neighborhood of the solution. The second is that the step size $s_k$ converges to one.

Several relations are combined in the proof of Lemma 5.4.10 to show that the unstabilized algorithm converges superlinearly. (1) Algorithm Property 5.4.6(iii) ensures that, at each iteration, the search arc $a_k$ ends at a point $a_k(1)$ which is nearly stationary for the subproblem $GQP(x_k, H_k)$. This is a point at which the optimality function for the subproblem, $\tilde{\sigma}(a_k(1) \mid x_k, H_k)$, is small. (2) By Lemma 5.4.9, this implies that $a_k(1)$ decreases the optimality function $\sigma(\cdot)$ for the main problem ICP. (3) Lemmas 5.4.5 and 5.4.3 translate this decrease in $\sigma(\cdot)$ into a decrease in the distance from the minimizer $\hat{x}$.

We begin by relating the optimality function $\sigma(x)$ to the distance from the minimizer $\| x - \hat{x} \|$ using a nondifferentiable exact penalty function.

**Definition:** We will denote the $l_\infty$ exact penalty function for ICP by

$$p_c(x) \triangleq f^0(x) + c\,\psi_+(x) , \quad (5.4.7)$$

with $c > 0$. The function $p_c(\cdot)$ can be represented as a max function, $p_c(x) \triangleq \max_{j \in \mathbf{p}} \{ f^0(x) + c f^j(x), f^0(x) \}$. Let $\underline{\mu}^0 \triangleq \min \{ \mu^0 \mid \mu \in \hat{U} \}$.

□

Lemma 5.4.3 shows that $\hat{x}$ is a local minimizer of the penalty function as well. The lemma is used in the proofs of Lemmas 5.4.5 and 5.4.10.

**Lemma 5.4.3:** *For any $c > (1-\underline{\mu}^0)/\underline{\mu}^0$, there exists $\delta > 0$ such that $p_c(x) - p_c(\hat{x}) \geq \delta \| x - \hat{x} \|^2$ for $x$ near $\hat{x}$.*

**Proof:** See Lemma 4.3.2. □

**Definition:** We define

$$U(x) \triangleq \operatorname*{argmin}_{\mu \in \Sigma_{p+1}} \psi_+(x) - \sum_{j \in p} \mu^j f^j(x) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} \mu^j \nabla f^j(x) \|^2 . \tag{5.4.8}$$

Note that, from (5.4.5), $U(x)$ is the set of multiplier vectors such that

$$\sigma(x) = \psi_+(x) - \sum_{j \in p} \mu^j f^j(x) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} \mu^j \nabla f^j(x) \|^2 . \tag{5.4.9a}$$

The following definition is used in the proofs of several of the lemmas.

$$\tilde{U}(x \mid \bar{x}, H) \triangleq \operatorname*{argmin}_{\mu \in \Sigma_{p+1}} \tilde{\psi}_+(x \mid \bar{x}, H) - \sum_{j \in p} \mu^j \tilde{f}^j(x \mid \bar{x}, H) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} \mu^j \nabla \tilde{f}^j(x \mid \bar{x}, H) \|^2 \tag{5.4.9b}$$

□

The remaining lemmas are established in the Appendix. Lemma 5.4.4 is a technical result which is used in the proofs of Lemmas 5.4.5, 4.11 and 5.1.

Lemma 5.4.4: *For any $c > (1-\underline{\mu}^0)/\underline{\mu}^0$,*

(a) $\min \{ \mu^0 \mid \mu \in U(x) \} > 1/(1+c)$ *for $x$ near $\hat{x}$*,

(b) $\psi_+(x) \leq (1+c)\sigma(x)$ *for $x$ near $\hat{x}$*,

(c) $\tilde{\psi}_+(x \mid \bar{x}, H) \leq (1+c)\tilde{\sigma}(x \mid \bar{x}, H)$ *for $x$ near $\hat{x}$, $\bar{x}$ near $\hat{x}$ and $H$ near $F(\hat{x})$.*

Lemmas 5.4.5 and 5.4.3 show that $\sigma(x)$, which depends only upon first-order information at $x$, can be used as a merit function to gauge progress toward $\hat{x}$. Lemma 5.4.5 is used in the proof that a step size of one yields superlinear convergence (Lemma 5.4.10) and in the proof that the step size converges to one (Lemma 5.4.11).

Lemma 5.4.5: *For any* $c > \max\{1, (1-\underline{\mu}^0)/\underline{\mu}^0\}$, *there exist* $\delta > 0$ *and* $K > 0$ *such that*

$$\delta\sigma(x) \le p_c(x) - p_c(\hat{x}) \le K\sigma(x), \tag{5.4.10}$$

*for* $x$ *near* $\hat{x}$.

In order to be assured of a superlinear rate of convergence, the search arc constructed in Algorithm Model 5.2.1 must possess further properties. For $x$ near a solution, the arc $A(x, H)$ is required to be well-behaved at $s = 1$ and end at an approximate solution to the subproblem GQP$(x, H)$. It is also required that, at $a_k(1)$, the arc be tangent to a descent direction for $\tilde{\psi}(\cdot)$. This property ensures that $\psi(a_k(s))$ is sensitive to changes in the step size. Hence, when $A(x, H)$ is slightly infeasible at $s = 1$, feasibility can be recovered with only a small change in the step size. The additional properties we need are as follows.

Algorithm Property 5.4.6: *For any* $\varepsilon > 0$ *and for* $x$ *near* $\hat{x}$ *and* $H$ *near* $F(\hat{x})$, *there exist* $K > 0$ *and* $\delta > 0$ *such that (with the notation* $a = A(x, H)$ *and* $\dot{a}(s) = da(s)/ds$*),*

(i) $\|a(s) - a(1)\| \le K\,|1-s|\,\|\dot{a}(1)\|$, (5.4.11a)

(ii) $\delta\|a(1) - x\|^2 \le \|\dot{a}(1)\| \le K\|a(1) - x\|$, (5.4.11b)

(iii) $\tilde{\sigma}(a(1) \mid x, H) \le K\|a(1) - x\|^3 \le \varepsilon$, (5.4.11c)

(iv) $\tilde{\psi}(a(s) \mid x, H) \le \tilde{\psi}(a(1) \mid x, H) + (s-1)\delta\|\dot{a}(1)\|$, (5.4.11d)

*for* $s$ *near* 1.

The right-hand inequality in Algorithm Property 5.4.6(iii) ensures that $A(x, H)(1)$ converges to $\hat{x}$ as $x$ converges to $\hat{x}$ and $H$ converges to $F(\hat{x})$.

We must also assume that the Hessian estimates converge.

Algorithm Property 5.4.7: *If an algorithm conforming to Algorithm Model 5.2.1 constructs a sequence* $\{x_i\}_{i \in \mathbb{N}}$ *which converges to* $\bar{x} \in \mathbb{R}^n$, *then the sequence* $\{H_k^j\}_{k \in \mathbb{N}}$ *converges to*

*$F^j(\bar{x})$ for each $j \in$ p∪0.*

Note that that Algorithm Property 5.4.7 is stronger than the assumption usually made about the convergence of estimates of second-order information. Whereas the analyses in [Pan.1-2, Pow.1] assume only that the reduced Lagrangian Hessian estimate converges, we assume that all of the Hessian estimates converge to the Hessians of the respective functions at the solution. This assumption will be satisfied if exact Hessians or secant updates [Bre.1, Pol.12] are used, but cannot be guaranteed if variable metric updates are used.

We assume in the remainder of this section that an algorithm which conforms to Algorithm Model 5.2.1 and which satisfies Algorithm Properties 5.3.1, 5.4.6 and 5.4.7 constructs a sequence $\{ x_k \}_{k \in \mathbf{N}}$ starting from a feasible point $x_0$. We will denote $A(x_k, H_k)$ by $a_k$.

**Lemma 5.4.8:** *As $k \to \infty$,* (a) $x_k \to \hat{x}$, (b) $\sigma(x_k) \to 0$, (c) $a_k(1) \to \hat{x}$ .

**Definition:** We define

$$\eta(x, H) \triangleq \max_{j \in \mathbf{p} \cup 0} \max \{ \| H^j - F^j(x') \| \mid x' \in S(x, 2\| A(x, H)(1) - x \|) \} , \tag{5.4.12}$$

where $S(x, \varepsilon)$ denotes the ball in $\mathbb{R}^n$ with center $x$ and radius $\varepsilon$. For sequences $\{ x_k \}_{k \in \mathbf{N}}$ and $\{ H_k \}_{k \in \mathbf{N}}$, we define $\eta_k \triangleq \eta(x_k, H_k)$.

□

The quantity $\eta(x, H)$ is a measure of the error with which $H$ approximates $F(x')$ for $x'$ near $x$. Note that $\eta(x, H)$ is equal to zero only if $F(\cdot)$ is constant near $x$.

**Lemma 5.4.9:** *There exists $K > 0$ such that, for large $k$,*

$$\sigma(a_k(1)) \le \tilde{\sigma}(a_k(1) \mid x_k, H_k) + K \eta_k \| a_k(1) - x_k \|^2 . \tag{5.4.13}$$

In Lemma 5.4.10, four relations are combined to show that an unstabilized version of Algorithm Model 5.2.1, i.e., an algorithm in which $x_{k+1} = a_k(1)$, converges superlinearly in a neighborhood of the solution. Algorithm Property 5.4.6(iii) shows that $a_k(1)$ is nearly stationary for $GQP(x_k, H_k)$. By Lemma 5.4.9, this implies that $a_k(1)$ decreases the optimality function $\sigma(\cdot)$.

Lemmas 5.4.5 and 5.4.3 translate this decrease into the amount of progress made toward the solution.

**Lemma 5.4.10:** *There exists $K > 0$ such that, for large $k$,*

$$\|a_k(1) - \hat{x}\| \le K \sqrt{\max\{\eta_k, \|a_k(1) - x_k\|\}}\, \|a_k(1) - x_k\| . \tag{5.4.14}$$

An immediate corollary is s Applying the triangle inequality to (5.4.14) yields superlinear convergence of the sequence $a_k(1)$, $\|a_k(1) - \hat{x}\| \le o(\|x_k - \hat{x}\|)$ .

The next lemma shows that the step size converges to unity.

**Lemma 5.4.11:** *There exists $K > 0$ such that, for large $k$,*

$$s_k \ge 1 - K \max\{\eta_k, \|a_k(1) - x_k\|\} \frac{\|a_k(1) - x_k\|^2}{\|\dot{a}_k(1)\|} . \tag{5.4.15}$$

It is apparent from the proof of Lemma 5.4.11 (see the Appendix) that the interval of step sizes which are guaranteed to satisfy (5.2.2a-b) shrinks rapidly as the sequence converges (with $\|a_k(1) - x_k\|$). This analysis suggests, and our numerical experiments confirm, that a careless line search may miss this interval of acceptable near-unity step sizes. If this happens, the algorithm's rate of convergence is degraded to linear. If the algorithm were modified by replacing the variable $s$ in the search arc by $(1 - \sqrt{\eta_k}\|a_k(1) - x_k\|)s$, then a step size of 1 would eventually be accepted by the line search and only a single step size would need to be tested when very near the solution. However, this would slow convergence when farther from the solution. We chose not to disguise the computational difficulty caused by the narrowness of the interval of acceptable step sizes in this way.

**Theorem 5.4.12:** (a) *If $H_k^j \to F^j(\hat{x})$ as $k \to \infty$ for each $j \in$ p$\cup$0, then*

$$\limsup_{k \to \infty} \|x_{k+1} - \hat{x}\| / \|x_k - \hat{x}\| = 0 . \tag{5.4.16}$$

(b) *If, for some* $M > 1$, *some* $q \geq 1$ *and each* $j \in \mathbf{p} \cup 0$, $\| H_k^j - F^j(x_k) \| \leq M \max_{i=1,\ldots,q} \| x_{k-i} - x_k \|$, *then there exists* $K > 0$ *such that*

$$\limsup_{k \to \infty} \| x_{k+1} - \hat{x} \| / \| x_k - \hat{x} \|^{\tau} \leq K , \tag{5.4.17}$$

*where* $\tau$ *is the unique positive solution to* $t^{q+1} - t^q - ½ = 0$.

(c) *If* $H_k^j = F^j(x_k)$ *for each* $j \in \mathbf{p} \cup 0$, *then there exists* $K > 0$ *such that*

$$\limsup_{k \to \infty} \| x_{k+1} - \hat{x} \| / \| x_k - \hat{x} \|^{1.5} \leq K . \tag{5.4.18}$$

Proof: We prove this theorem by combining the fact that the unstabilized algorithm converges superlinearly (Lemma 5.4.10) with the fact that the step size converges to one (Lemma 5.4.11). By the triangle inequality, Algorithm Property 5.4.6(i), Lemma 5.4.11 and Lemma 5.4.10, there exists $K > 0$ such that

$$\begin{aligned}
\| x_{k+1} - \hat{x} \| &\leq \| a_k(s_k) - a_k(1) \| + \| a_k(1) - \hat{x} \| \\
&\leq (1 - s_k) \| \dot{a}_k(1) \| + \| a_k(1) - \hat{x} \| \\
&\leq \left[ K \max \{ \eta_k , \| a_k(1) - x_k \| \} \frac{\| a_k(1) - x_k \|^2}{\| \dot{a}_k(1) \|} \right] \| \dot{a}_k(1) \| + \| a_k(1) - \hat{x} \| \\
&\leq \left[ K \max \{ \eta_k , \| a_k(1) - x_k \| \} \| a_k(1) - x_k \|^2 \right] + \left[ K \sqrt{\max \{ \eta_k , \| a_k(1) - x_k \| \}} \| a_k(1) - x_k \| \right] \\
&\leq K \sqrt{\max \{ \eta_k , \| a_k(1) - x_k \| \}} \| a_k(1) - x_k \| ,
\end{aligned} \tag{5.4.19}$$

for large $k$. We now estimate the size of the coefficient $\sqrt{\max \{ \eta_k , \| a_k(1) - x_k \| \}}$ in (5.4.19) for the cases (a-c).

(a) Since $x_k$ and $a_k(1)$ converge to $\hat{x}$, $\eta_k$ converges to 0. The result follows immediately from (5.4.19).

(b) By the definition of $\eta_k$, the triangle inequality and the local Lipschitz continuity of $F^j(\cdot)$,

$$\eta_k \le K\|x_k - \hat{x}\| + \max_{j \in p \cup 0} \max\{\, \|F^j(x) - F^j(x_k)\| \mid x \in S(x_k, 2\|a_k(1) - x_k\|)\,\} + \|F^j(x_k) - H_k^j\|$$

$$\le K\|a_k(1) - x_k\| + M \max_{i = 1, \dots, q} \{\, \|x_{k-i} - x_k\| \,\}\ , \tag{5.4.20}$$

for some $K > 0$ and large $k$. By Lemma 5.4.10, $\|a_k(1) - x_k\| \le 2\|x_k - \hat{x}\|$ for large $k$. Applying this and the triangle inequality to (5.4.20), there exists $K > 0$ such that

$$\eta_k \le 2K\|x_k - \hat{x}\| + M \max_{i = 1, \dots, q} \{\, \|x_{k-i} - x_k\| \,\}$$

$$\le K \max_{i = 1, \dots, q} \{\, 3\|x_k - \hat{x}\| + \|x_{k-i} - \hat{x}\| \,\}$$

$$\le K \max_{i = 0, 1, \dots, q} \|x_{k-i} - \hat{x}\|\ , \tag{5.4.21}$$

for large $k$. Substituting (5.4.21) into (5.4.19) yields

$$\|x_{k+1} - \hat{x}\| \le K \sqrt{\max_{i = 0, 1, \dots, q} \{\, \|x_{k-i} - \hat{x}\| \,\}}\ \|x_k - \hat{x}\|\ , \tag{5.4.22}$$

for large $k$. Since $x_k$ converges to $\hat{x}$, (5.4.22) implies that $\|x_{k+1} - \hat{x}\| \le \|x_k - \hat{x}\|$ for large $k$. Hence, $\max_{i = 0, 1, \dots, q} \{\, \|x_{k-i} - \hat{x}\| \,\} = \|x_{k-q} - \hat{x}\|$ for large $k$. Substituting this into (5.4.22) yields

$$\|x_{k+1} - \hat{x}\| \le K \sqrt{\|x_{k-q} - \hat{x}\|}\ \|x_k - \hat{x}\|\ , \tag{5.4.23}$$

It follows from (5.4.23) and Theorem 3.1 in [Pot.1] that there exists $K > 0$ such that

$$\limsup_{k \to \infty} \|x_{k+1} - \hat{x}\| / \|x_k - \hat{x}\|^{\tau} \le K\ , \tag{5.4.24}$$

where $\tau$ is the unique positive root of the polynomial $t^{q+1} - t^q - \frac{1}{2}$.

(c) By the local Lipschitz continuity of $F^j(\cdot)$, there exists $K > 0$ such that $\eta_k \le K \|a_k(1) - x_k\| \le K \|x_k - \hat{x}\|$, and therefore (5.4.19) implies

$$\|x_{k+1} - \hat{x}\| \le K \|x_k - \hat{x}\|^{1.5} , \tag{5.4.25}$$

for large $k$.

□

## 5.5 AN EXAMPLE OF A GQP-BASED ALGORITHM

Algorithm Model 5.2.1 and Algorithm Properties 5.3.2, 5.4.6 and 5.4.7 define a class of feasible descent algorithms for ICP. Each algorithm in the class is characterized by its search arc function, $A(x, H)$, and by the method it uses to construct Hessian approximations. The choice of a method for approximating the Hessian matrices is independent of the choice of search arc function. In this section, we describe the variety of admissible search arc functions. We will present one search arc function in detail and show that it possesses Algorithm Properties 5.3.2 and 5.4.6.

First, however, we discuss methods for approximating the Hessians. In order for the global convergence result, Theorem 5.3.3, to hold, the method used must ensure that the Hessian estimates are bounded. Furthermore, in order for superlinear convergence to be assured, the method must satisfy Algorithm Property 5.4.7. The use of exact Hessians or sufficiently accurate finite-difference approximations yields an asymptotic rate of convergence of at least 1.5. Because of the high cost of such approximations, it may be more efficient to use a secant method [Bre.1, Pol.12] to update the Hessian approximations. The secant method of [Pol.12] forms Hessian estimates on the basis of gradients from the past $n$ iterations. By Theorem 5.4.12, such a scheme has a potential rate of convergence of $\tau$ where $\tau$ is the positive root of the polynomial $\tau^{n+1} - \tau^n - \frac{1}{2}$. Variable metric methods are the most commonly used means for updating Hessian approximations. A variety of updates can be considered for use in GQP-based algorithms because the Hessian approximations need not be positive definite. However, it is not clear that any variable metric update will ensure that Algorithm Property 5.4.7 is satisfied.

We now discuss search arc functions. Algorithm Properties 5.3.2 and 5.4.6 place restrictions on four parts of the search arc function: the initial point $A(x, H)(0)$, the initial direction $dA(x, H)(0)/ds$, the final point $A(x, H)(1)$ and the final direction $dA(x, H)(1)/ds$. Hence, a search arc function can be defined by an algorithm which, given $x$ and $H$, computes suitable points and directions and assembles them into an arc. These points and directions can be computed in various ways.

Algorithm Property 5.4.6(iii) specifies that the arc $A(x, H)$ end at a good approximation to a stationary point for the subproblem GQP($x, H$). Any method for nonlinear programming could be used to obtain such a point. However, the method used should satisfy (5.4.11c) in a bounded, and preferably small, number of iterations. This limits consideration to superlinearly convergent methods with a Q-rate [Ort.1] strictly greater than one. Since high-order derivatives of the functions appearing in GQP($x_0, H$) can be computed trivially, extensions of the high-order root-finding methods like those of Chebyshev [Tra.1] and Halley [Cuy.1] to the nonlinear complementarity problem formed by the first-order necessary conditions for GQP($x, H$) would be particularly desirable.

The initial direction $dA(x, H)(0)/ds$ must be a feasible descent direction. A feasible descent direction can be computed by any of the methods described in [Zou.1, Pir.1]. Each of these requires the solution of a quadratic or linear program. The final direction $dA(x, H)(0)/ds$ must be a descent direction for the constraint violation function $\tilde{\psi}(\cdot \mid x, H)$, and can be computed using similarly techniques.

There are many ways to connect two given points with a curve tangent to a given vector at each end. Algorithm Properties 5.3.2(i) and 5.4.6(i-ii) require only that the arc $A(x, H)(s)$ be smooth at its beginning and end, uniformly with respect to $x$, $H$ and $s$. As illustrated by Algorithm 5.5.1 below, it is not necessary for the arc to be continuous. However, it is possible to construct an admissible search arc which smoothly interpolates between the initial and final points.

In general, there does not exist a quadratic curve (i.e. - a curve of the form $c(s) = z_0 + sz_1 + s^2 z_2$ with $z_0, z_1, z_2 \in \mathbb{R}^n$) which interpolates between two given points and is tangent to given directions at the endpoints. For each pair of points and directions, however, there is a two-parameter family of cubic curves which perform the interpolation. (The parameters correspond to the *norms* of the velocities of the curves at the endpoints.) Alternatively, if an interior point method is used to solve GQP($x_0, H$), a search arc can be constructed by piecewise linear interpolation between the iterates which it generates.

The most efficient way of obtaining suitable initial and final directions for the search arc is to approximate the solution to GQP($x, H$) in a way which generates such directions automatically. Interior point methods, which are now in vogue for linear programming, automatically construct a feasible search arc which satisfies the descent conditions (5.3.2a-b) and (5.4.11d) in the course of solving GQP(x, $H$). At present, however, no superlinearly convergent interior point methods are known.

We now develop a search arc function in detail, and we show that it satisfies Algorithm Properties 5.3.2 and 5.4.6. The search arc function is defined by Algorithm 5.5.1 below. Algorithm 5.5.1 requires only a moderate amount of computation to construct the search arc for a given $x_0$ and $H$. Algorithm 5.5.1 computes an approximate solution to GQP($x_0, H$) using a root-finding method which is *quartically* convergent, i.e. - of order four. A feasible descent direction at $x$ and a direction of descent for $\tilde{\psi}(\cdot)$ at the approximate solution are generated in the course of the computation of the approximate solution. A search arc $A(x_0, H)$ is then constructed from the four quantities, the current point, the feasible descent direction, the approximate solution and the $\tilde{\psi}$-descent direction.

We turn to the development of Algorithm 5.5.1. Condition 4.6(iii) requires that we compute a point which is nearly stationary for GQP($x_0, H$), i.e., a point $x$ for which $\tilde{\sigma}(x \mid x_0, H)$ is small. By Lemma 5.5.1 below, problem GQP($x_0, H$) possesses a stationary point $\tilde{x}$ for $x_0$ near a solution $\hat{x}$ and for $H$ near $F(\hat{x})$. Furthermore, $\tilde{x}$ is the unique stationary point for GQP$(x_0, H)$ in a neighborhood of $\hat{x}$. We can satisfy Condition 4.6(iii) by computing an approximation to $\tilde{x}$. (Note that $\tilde{x}$ is a function of $x_0$ and $H$; we suppress this dependence to simplify our notation.)

We take an active set approach to approximating $\tilde{x}$. We denote the set of indices of constraints active at $\tilde{x}$ by $J_{\mathrm{GQP}}(x_0, H) \triangleq \{ j \in \mathbf{p} \mid \tilde{f}^j(\tilde{x} \mid x_0, H) = 0 \}$. For any $J \subset \mathbf{p}$, we define an equality-constrained subproblem,

$$\mathrm{P}(x_0, H, J)\text{:} \quad \min_{x \in \mathbb{R}^n} \{ \tilde{f}^0(x \mid x_0, H) \mid \tilde{f}^j(x \mid x_0, H) = 0, \forall j \in J \} . \tag{5.5.1}$$

Since $\tilde{x}$ is a stationary point of GQP$(x_0, H)$, $\tilde{x}$ must be a stationary point of $P(x_0, H, J_{\mathrm{GQP}}(x_0, H))$. We can write the first-order necessary conditions for $P(x_0, H, J_{\mathrm{GQP}}(x_0, H))$ in a convenient form. For any set $J = \{ j_1, \ldots, j_r \} \subset \mathbf{p}$, $x_0 \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times (p+1)n}$, we define the function $C(\cdot \mid x_0, H, J) \colon \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^{n+r}$ by

$$C(x, u \mid x_0, H, J) \triangleq \begin{bmatrix} \nabla \tilde{f}^0(x \mid x_0, H) + \sum_{j \in J} u^j \nabla \tilde{f}^j(x \mid x_0, H) \\ \tilde{f}^{j_1}(x \mid x_0, H) \\ \vdots \\ \tilde{f}^{j_r}(x \mid x_0, H) \end{bmatrix} . \tag{5.5.2a}$$

Since $\tilde{x}$ is stationary for P$(x_0, H, J_{\mathrm{GQP}}(x_0, H))$, there exists $\tilde{u} \in \mathbb{R}^p$ such that

$$C(\tilde{x}, \tilde{u} \mid x_0, H, J_{\mathrm{GQP}}(x_0, H)) = 0 . \tag{5.5.2b}$$

This system of quadratic equations may be solved by any root-finding method. We discuss the method used in Algorithm 5.5.1 later.

Of course, the set $J_{\mathrm{GQP}}(x_0, H)$ is not known *a priori*. Nor is a good point at which to initialize a root-finding technique for the solution of (5.5.2b) known. We propose to approximate $J_{\mathrm{GQP}}(x_0, H)$ and obtain such a starting point as follows. Consider the minimax problem,

$$\min_{x \in \mathbb{R}^n} \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \mathbf{p} \cup 0} \mu^j \tilde{f}^j(x \mid x_0, \mathbf{I}) , \tag{5.5.3}$$

where $\mathbf{I} = [I_n, \ldots, I_n] \in \mathbb{R}^{n \times (p+1)n}$ and $I_n$ is the $n \times n$ identity. Problem (5.5.3) is the search direction problem of the Pironneau-Polak method of feasible directions [Pir.1]. It can be solved by conversion to its dual form (see [Pol.4])

$$\max_{\mu \in \Sigma_{p+1}} \sum_{j \in \mathbf{p}} \mu^j f^j(x_0) - \tfrac{1}{2} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x_0) \|^2 . \tag{5.5.4a}$$

Recall that (5.4.8) defines $U(x_0)$ as the set of solutions to (5.5.4a). Problem (5.5.4a) is a positive semi-definite quadratic program, and a vector $\mu \in U(x_0)$ can be computed by a variety of methods [Gil.1, Hig.1, von.1, Kiw.2-3].

Let the function $\mu_1 : \mathbb{R}^n \to \Sigma_{p+1}$ be any selection from the set-valued map $U : \mathbb{R}^n \to 2^{\Sigma_{p+1}}$. We define $J_1(x_0) \triangleq J(\mu_1(x_0))$, where, for any $\mu \in \Sigma_{p+1}$, We define $J(\mu) \triangleq \{ j \in \mathbf{p} \mid \mu^j > 0 \}$. The set $J_1(x_0)$ is used in Algorithm 5.5.1 to approximate $J_{\mathrm{GQP}}(x_0, H)$. We will show in Lemma 5.5.1 that $J_1(x_0) = J_{\mathrm{GQP}}(x_0, H)$ for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, provided that the local minimizer of ICP, $\hat{x}$, satisfies Hypothesis 5.5.1 below.

The unique solution to the primal problem (5.5.3)

$$x_1 \triangleq \operatorname*{argmin}_{x \in \mathbb{R}^n} \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \mathbf{p} \cup 0} \mu^j \tilde{f}^j(x \mid x_0, \mathbf{I}) , \tag{5.5.4b}$$

can be obtained from any solution $\mu$ to the dual problem (5.5.4a) [Pol.4] by

$$x_1 = x_0 - \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x_0) . \tag{5.5.4c}$$

We define a function which maps Fritz-John multiplier vectors into Kuhn-Tucker multiplier vectors, $u_1 : \Sigma_{p+1} \to \mathbb{R}^p_+$,

$$u_1(\mu) \triangleq \begin{cases} (\mu^1, \ldots, \mu^p) / \mu^0 & \text{if } \mu^0 > 0 \\ 0 & \text{if } \mu^0 = 0 \end{cases} . \tag{5.5.5}$$

By Lemma 5.4.4(a), $\mu_1^0(x_0) > 0$ for $\mu_1 \in U(x_0)$ and $x_0$ near a solution, $\hat{x}$, to ICP. The pair

$(x_1, u_1)$ obtained from $\mu_1(x_0)$ according to (5.5.4b) and (5.5.5) is used in Algorithm 5.5.1 as a starting point for the root-finding method.

Now we discuss the root-finding technique which is applied in Algorithm 5.5.1 to

$$C(x, u \mid x_0, H, J_1(x_0)) = 0, \tag{5.5.6}$$

in order to approximate $(\tilde{x}, \tilde{u})$. The accuracy to which we are required by (5.4.11c) to approximate $\tilde{x}$ could be achieved by the application of two iterations of Newton's method to the system of equations (5.5.6). Instead, we use a single iteration of a higher order method, Chebyshev's method [Tra.1]. This simplifies the analysis, and requires less computation than two iterations of Newton's method.[5] Chebyshev's method requires the computation of the second derivative of $C(\cdot)$. This computation is trivial since the system (5.5.6) is quadratic. Chebyshev's method has convergence of order three in general, but, as is evident from the proof of Proposition 5.2, it has convergence of order four on quadratic systems.[6]

We denote by $(x_2, u_2)$ the approximate solution to (5.5.6) which results from the application of one iteration of Chebyshev's method to (5.5.6) starting from the point $(x_1, u_1)$. (We define $\nabla C(x, u \mid x_0, H, J) \triangleq \partial C(x, u \mid x_0, H, J) / \partial(x, u)$ and we denote by $A^\dagger$ the pseudoinverse of any matrix $A$.) The Chebyshev step, $(x_2, u_2) - (x_1, u_1)$, is the sum of the Newton step

$$h_n(x_0, H) \triangleq -\nabla C(x_1, u_1 \mid x_0, H, J_1(x_0))^\dagger C(x_1, u_1 \mid x_0, H, J_1(x_0)), \tag{5.5.7}$$

(where $A^\dagger$ denotes the pseudoinverse of the matrix $A$) and a correction step which compensates for the curvature of the function $C(\cdot \mid x_0, H, \hat{J})$,

$$h_c(x_0, H) \triangleq -\tfrac{1}{2}\nabla C(x_1, u_1 \mid x_0, H, J_1(x_0))^\dagger \nabla^2 C(x_1, u_1 \mid x_0, H, J_1(x_0))[h_n, h_n]. \tag{5.5.8}$$

The term $\nabla^2 C(x, u \mid x_0, H, J)$ denotes the second-derivative of the function $C(x, u \mid x_0, H, J)$

[5] A single iteration of Chebyshev's method requires the solution of two linear systems with the same coefficient matrix, while two iterations of Newton's method requires the solution of two linear systems with different coefficient matrices.

[6] An alternative is Halley's method [Cuy.1] which is based on rational approximation of the function $C(\cdot)$. Algebraic approaches to the solution of systems of algebraic equations are discussed in [Can.1, Kob.1], among other places.

with respect to $(x, u)$. It is a bilinear operator of two vector arguments, defined by (suppressing the dependence on $x_0$, $H$ and $J$),

$$\lim_{t \to 0^+} \| \nabla C((x,u)+th)^T g - \nabla C(x,u)^T g - t\nabla^2 C(x,u)[g,h] \| / t = 0 , \tag{5.5.9}$$

for all $g, h \in \mathbb{R}^{n+p}$.

The computation of an approximation $x_2$ to $\tilde{x}$ described above is incorporated into the following algorithm which defines a search arc function. Let $\mathbf{e}$ denote the vector of ones of length $p$.

**Algorithm 5.5.1:**

**Data:** $x_0 \in \mathbb{R}^n$ such that $\psi(x_0) \le 0$, $H \in \mathbb{R}^{n \times (p+1)n}$.

**Step 1:** Compute any $\mu_1 \in U(x_0)$.

**Step 2:** Obtain from $\mu_1$, $J_1 = J(\mu_1)$, $x_1$ according to (5.5.4c) and $u_1$ according to (5.5.5).

**Step 3:** Compute $h_n$ and $h_c$ according to (5.5.7-8) and set

$$\begin{bmatrix} x_2 \\ u_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ u_1 \end{bmatrix} + h_n + h_c . \tag{5.5.10}$$

**Step 4:** Set $d_0 = x_1 - x_0$, and set $d_2$ equal to the first $n$ components of $-\nabla C(x_1, u_1 \mid x_0, H, J_1)^{\dagger} \begin{bmatrix} 0 \\ \mathbf{e} \end{bmatrix} \| x_2 - x_0 \|^2$.

**Step 5:** Set $A(x_0, H)$ equal to

$$A(x_0, H)(s) = \begin{cases} x_0 + (2s-1)d_0 & \text{if } s \le ½ \\ x_2 + (2-2s)d_2 & \text{if } s > ½ \end{cases} . \tag{5.5.11}$$

□

Steps 1 and 3 are the only computationally expensive steps of this algorithm. Step 1 requires the solution of a semi-definite quadratic program, (5.5.4). A number of methods are available for this task [Gil.1, von.1, Hig.1, Kiw.2-3, Rus.1]. Step 3 requires the solution of two linear systems which have dimension no greater than $n+p$ and which have the same coefficient matrix,

$\nabla C(x_1, u_1 \mid x_0, H, J_1)$.

Now we verify that the search arc function $A(x, H)$ defined by Algorithm 5.5.1 possesses Algorithm Properties 5.3.2 and 5.4.6. We need the following additional assumption in order to prove that Algorithm Property 5.4.6 is satisfied.

**Hypothesis 5.5.1:** *Suppose that Hypotheses 5.4.1-2 hold with $\hat{x}$ as defined there, and, in addition, that strict complementary slackness holds at $\hat{x}$*[7] *and that the vectors $\{ \nabla f^j(\hat{x}) \}_{j \in \hat{J}}$ are linearly independent. Without loss of generality, suppose that $\hat{J} = \{ 1, \ldots, r \}$.*

Hypothesis 5.5.1 implies that $\hat{U}$ is a singleton, $\{ \hat{\mu} \}$. By Hypothesis 5.4.2(ii), $\hat{\mu}^0 > 0$, and hence the unique Kuhn-Tucker multiplier vector for the problem ICP associated with $\hat{x}$ is

$$\hat{u} \triangleq u_1(\hat{\mu}) = (\hat{\mu}^1, \ldots, \hat{\mu}^p) / \hat{\mu}^0 . \tag{5.5.12}$$

By strict complementary slackness and the uniqueness of $\hat{\mu}$, $\hat{J} = J(\hat{\mu}) = \{ j \in \mathbf{p} \mid f^j(\hat{x}) = 0 \}$.

The following lemma will be useful in showing that Algorithm 5.5.1 satisfies Algorithm Property 5.4.6. It is proved in the Appendix.

**Lemma 5.5.1:** *If Hypotheses 5.4.1-2 and 5.5.1 hold with $\hat{x}$ as defined there, then*

(a) $J_1(x_0) = J_{GQP}(x_0, H) = \hat{J}$ *for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$,*

(b) *the matrix $\nabla C(x, u \mid x_0, H, \hat{J})^{\dagger}$ is continuous and*

$$\nabla C(x, u \mid x_0, H, \hat{J})^{\dagger} = \begin{bmatrix} \left[ \dfrac{\partial C(x, u \mid x_0, H, \hat{J})}{\partial(x, u_1, \ldots, u_r)} \right]^{-1} \\ 0 \end{bmatrix} , \tag{5.5.13a}$$

---

[7] Strict complementary slackness is said to hold at a stationary point $\hat{x}$ if there exists $\mu \in \hat{U}$ such that $\mu^j > 0$ for every $j \in \mathbf{p}$ such that $f^j(\hat{x}) = 0$.

*for $x$ and $x_0$ near $\hat{x}$, $u$ near $\hat{u}$ and $H$ near $F(\hat{x})$,*

(c) *for any selection $\mu_1:\mathbb{R}^n \to \Sigma_{p+1}$ from $U:\mathbb{R}^n \to 2^{\Sigma_{p+1}}$,*

$$\lim_{x_0 \to \hat{x}} \begin{bmatrix} x_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} \hat{x} \\ \hat{u} \end{bmatrix}, \tag{5.5.13b}$$

(d) *there exists a neighborhood $W$ of $\hat{x}$ such that, for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, the problem* GQP$(x_0, H)$ *has a unique stationary point, $\tilde{x}(x_0, H)$, in $W$, and there exists a unique Kuhn-Tucker multiplier vector $\tilde{u}$ associated with $\tilde{x}$. Furthermore,*

$$\lim_{\substack{x_0 \to \hat{x} \\ H \to F(\hat{x})}} \begin{bmatrix} \tilde{x} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} \hat{x} \\ \hat{u} \end{bmatrix}. \tag{5.5.13c}$$

Lemma 5.5.1(c), Hypothesis 5.4.2(ii) and equation (5.5.5) show that $u_1$ is continuous at $\hat{x}$. Lemma 5.5.1(b-c) shows that $h_n$ and $h_c$ are continuous at $x_0 = \hat{x}$ and $H = F(\hat{x})$.

**Proposition 5.2:** *The search arc function $A(\cdot,\cdot)$ defined by the Algorithm 5.5.1 possesses Algorithm Property 5.3.2. If Hypotheses 5.4.1-2 and 5.5.1 hold with $\hat{x}$ as defined there, then Algorithm 5.5.1 possesses Algorithm Property 5.4.6 with respect to $\hat{x}$.*

**Proof:** Let $a(s) \triangleq A(x_0, H)(s)$. We consider each property separately.

**Algorithm Property 5.3.2(i):** By construction, $a(0) = x$ and $\|a(s) - a(0) - \dot{a}(0)s\| = 0$ for $s \in [0, \frac{1}{2}]$. From [Pol.4], $x_1$ is a continuous function of $x_0$, independent of $H$, and hence $\|\dot{a}(0)\|$ is bounded on any compact set of $(x_0, H)$.

**Algorithm Property 5.3.2(ii):** Since $\dot{a}(0) = 2(x_1 - x_0)$,

$$\max_{j \in p \cup 0} \tilde{f}^j(x_0 + \tfrac{1}{2}\dot{a}(0) \mid x_0, I) = \max_{j \in p \cup 0} \tilde{f}^j(x_1 \mid x_0, I)$$

$$= \min_{x \in \mathbb{R}^n} \max_{j \in p \cup 0} \tilde{f}^j(x \mid x_0, I), \tag{5.5.14}$$

by the definition of $x_1(x_0)$ in (5.5.4b). From [Pol.4], the right-hand side of (5.5.14) is continuous in $x_0$ and is strictly negative at any point $x_0$ which is nonstationary for ICP. Therefore, Algorithm 5.5.1 has Algorithm Property 5.3.2(ii).

**Algorithm Property 5.4.6(i):** For $s \in [\frac{1}{2}, 1]$, $\|a(s) - a(1)\| = 2\,|1-s|\,\|\dot{a}(1)\|$.

**Algorithm Property 5.4.6(ii):** From Algorithm 5.5.1,

$$\|\dot{a}(1)\| = 2\|d_2\| = \left\|\nabla C(x_1, u_1 \mid x_0, H, J_1(x_0))^{\dagger} \begin{bmatrix} 0 \\ e \end{bmatrix}\right\| \|x_2 - x_0\|^2 . \tag{5.5.15}$$

By Lemma 5.5.1(b), the quantity $\left\|\nabla C(x_1, u_1 \mid x_0, H, J_1(x_0))^{\dagger} \begin{bmatrix} 0 \\ e \end{bmatrix}\right\|$ is continuous for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. It follows from Lemma 5.5.1(a-d) that this quantity is nonzero for $x_0 = \hat{x}$ and $H = F(\hat{x})$. Therefore, there exists $\delta > 0$ and $K > 0$ such that

$$\delta\|x_0 - x_2\|^2 \le \|\dot{a}(1)\| \le K\|x_0 - x_2\|^2 , \tag{5.5.16}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Now $\|x_0 - x_2\| \le \|h_n\| + \|h_c\|$, where $h_n$ and $h_c$ are as defined in (5.5.7-8). From (5.5.7) and (5.5.8), therefore, $\|x_2 - x_0\|$ is bounded for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Therefore, there exists $K > 0$ such that

$$\delta\|x_0 - x_2\|^2 \le \|\dot{a}(1)\| \le K\|x_0 - x_2\| , \tag{5.5.17a}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$.

**Algorithm Property 5.4.6(iii):** We derive a relationship between $\|x_2 - \tilde{x}\|$ and $\|x_k - \tilde{x}\|$. In the paragraph below, we suppress the dependence of $C(\cdot)$ on $x_0$, $H$ and the index set $J$. The analysis below is local, and we restrict consideration to $x_0$ in a neighborhood of $x$ in which the three index sets $J_1$, $J_{GQP}$ and $\hat{J}$ are equal. Such a neighborhood exists by Lemma 5.5.1(a). We treat only the special case where $\hat{J} = \mathbf{p}$. In this case, the inverse of the matrix $\nabla C(x, u \mid x_0, H, \hat{J})$ exists, and hence the inverse of the function $C(\cdot \mid x_0, H, \hat{J})$ exists. The result we derive can be obtained for the general case, but the proof is more complicated and we omit it.

First, we derive an upper bound on $\|\tilde{x} - x_2\|$. Chebyshev's method for solving systems of equations is an *inverse* method. It approximates the root $(\tilde{x}, \tilde{u})$ of $C(\cdot \mid x_0, H, \hat{f})$ using a second-order Taylor series approximation to the *inverse* of the function $C(\cdot \mid x_0, H, \hat{f})$. From the definitions in (5.5.7) and (5.5.8) and our assumption that $\nabla C(x, u \mid x_0, H, \hat{f})^{\dagger} = \nabla C(x, u \mid x_0, H, \hat{f})^{-1}$, it follows that

$$h_n = -\nabla C(x_1, u_1)^{-1} C(x_1, u_1) = \frac{d}{dt} C^{-1}((1-t)C(x_1, u_1)) \big|_{t=0}, \qquad (5.5.17b)$$

$$h_c = -\tfrac{1}{2} \nabla C(x_1, u_1)^{-1} \nabla^2 C(x_1, u_1)[h_n, h_n] = \tfrac{1}{2} \frac{d^2}{dt^2} C^{-1}((1-t)C(x_1, u_1)) \big|_{t=0}. \qquad (5.5.18)$$

Then, since $C^{-1}((1-t)C(x_1, u_1)) \big|_{t=0} = (x_1, u_1)$, we can write $(x_2, u_2)$ as

$$\begin{bmatrix} x_2 \\ u_2 \end{bmatrix} \triangleq \begin{bmatrix} x_1 \\ u_1 \end{bmatrix} + h_n + h_c$$

$$= C^{-1}((1-t)C(x_1, u_1)) \big|_{t=0} + \frac{d}{dt} C^{-1}((1-t)C(x_1, u_1)) \big|_{t=0}$$

$$+ \tfrac{1}{2} \frac{d^2}{dt^2} C^{-1}((1-t)C(x_1, u_1)) \big|_{t=0}. \qquad (5.5.19)$$

Note that the right-hand side of (5.5.19) is a second-order Taylor series expansion of $C^{-1}((1-t)C(x_1, u_1))$ about $t = 0$ evaluated at $t = 1$. By Taylor's theorem and since $C^{-1}((1-t)C(x_1, u_1)) \big|_{t=1} = C^{-1}(0) = (\tilde{x}, \tilde{u})$,

$$\left\| \begin{bmatrix} \tilde{x} \\ \tilde{u} \end{bmatrix} - \begin{bmatrix} x_2 \\ u_2 \end{bmatrix} \right\| = \left\| \tfrac{1}{2} \int_0^1 (1-s)^2 \frac{d^3}{dt^3} C^{-1}((1-t)C(x_1, u_1)) \big|_{t=s} ds \right\|. \qquad (5.5.20)$$

Now, since $\nabla^2 C(x, \mu)$ is constant,

$$\frac{d^3}{dt^3} C^{-1}((1-t)C(x_1, u_1)) = \nabla C(\xi_t)^{-1} \nabla^2 C(x_1, u_1)[\tilde{h}, \tilde{h}], \qquad (5.5.21)$$

where

$$\tilde{h} \triangleq \frac{d^2}{dt^2} C^{-1}((1-t)C(x_1, u_1))$$

$$= -\nabla C(\xi_t)^{-1} \nabla^2 C(x_1, u_1)[\nabla C(\xi_t)^{-1} C(\xi_t), \nabla C(\xi_t)^{-1} C(\xi_t)] \,, \tag{5.5.22}$$

and $\xi_t \triangleq C^{-1}((1-t)C(x_1, u_1))$. (The curve $\xi_t$ exists and is continuous by the Inverse Function Theorem and the fact that $\nabla C(\hat{x}, \hat{u})$ is nonsingular.) By Lemma 5.5.1(b), $\nabla C(\xi_t)^{-1}$ is bounded above for $x_0$ near $\hat{x}$, $H$ near $F(\hat{x})$ and $t \in [0, 1]$. Therefore, (5.5.20) and (5.5.21) imply that there exists $K > 0$ such that

$$\|\tilde{x} - x_2\| \le \left\| \begin{bmatrix} \tilde{x} \\ \tilde{u} \end{bmatrix} - \begin{bmatrix} x_2 \\ u_2 \end{bmatrix} \right\| \le K \max_{t \in [0,1]} \|C(\xi_t)\|^4 \le K \|C(x_1, u_1)\|^4 \,, \tag{5.5.23}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$.

Now we derive an upper bound on $\|C(x_1, u_1 \mid x_0, H, \mathbf{f})\|$. By its definition in (5.5.2a),

$$\|C(x_0, \tilde{u} \mid x_0, H, \mathbf{f})\| = \sum_{j \in \mathbf{p}} |\tilde{f}^j(x_0 \mid x_0, H)| + \|\nabla \tilde{f}^0(x_0 \mid x_0, H) + \sum_{j \in \mathbf{p}} \tilde{u}^j \nabla \tilde{f}^j(x_0 \mid x_0, H)\|$$

$$= \sum_{j \in \mathbf{p}} |f^j(x_0)| + \|\nabla f^0(x_0) + \sum_{j \in \mathbf{p}} \tilde{u}^j \nabla f^j(x_0)\| \,. \tag{5.5.24a}$$

Let $\tilde{\mu} \triangleq (1, \tilde{u}^1, \ldots, \tilde{u}^p) / (1 + \sum_{j \in \mathbf{p}} \tilde{u}^j)$. Then, substituting $\tilde{u}^j = \tilde{\mu}^j / \tilde{\mu}^0$ into (5.5.24a),

$$\|C(x_0, \tilde{u} \mid x_0, H, \mathbf{f})\| = \sum_{j \in \mathbf{p}} |f^j(x_0)| + \frac{1}{\tilde{\mu}^0} \|\sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0)\|$$

$$\ge \sum_{j \in \mathbf{p}} |f^j(x_0)| + \|\sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0)\|$$

$$\ge \sum_{j \in \mathbf{p}} \tilde{\mu}^j |f^j(x_0)| + \|\sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0)\| \,, \tag{5.5.24b}$$

since $\tilde{\mu}^j \le 1$ for all $j \in \mathbf{p} \cup 0$. Since $|t| \ge -t$ for any $t \in \mathbb{R}$,

$$\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \geq -\sum_{j \in \mathbf{p}} \tilde{\mu}^j f^j(x_0) + \| \sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0) \| . \tag{5.5.24c}$$

The quantity $\nabla f^0(x_0) + \sum_{j \in \mathbf{p}} \tilde{u}^j \nabla f^j(x_0)$ converges to zero as $x_0$ converges to $\hat{x}$ and $H$ converges to $F(\hat{x})$. Hence,

$$\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \geq -\sum_{j \in \mathbf{p}} \tilde{\mu}^j f^j(x_0) + ½\| \sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0)\|^2$$

$$\geq -\left[ \sum_{j \in \mathbf{p}} \tilde{\mu}^j f^j(x_0) - ½\| \sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0)\|^2 \right] , \tag{5.5.25}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Therefore,

$$\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \geq -\left[ \sum_{j \in \mathbf{p}} \tilde{\mu}^j f^j(x_0) - ½\| \sum_{j \in \mathbf{p} \cup 0} \tilde{\mu}^j \nabla f^j(x_0)\|^2 \right]$$

$$\geq -\left[ \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \mathbf{p}} \mu^j f^j(x_0) - ½\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x_0)\|^2 \right] , \tag{5.5.26}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. From [Pol.4] (or compare (5.4.4) with (5.4.5)),

$$\max_{\mu \in \Sigma_{p+1}} \sum_{j \in \mathbf{p}} \mu^j f^j(x_0) - ½\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x_0)\|^2 = \min_{x \in \mathbb{R}^n} \max_{j \in \mathbf{p} \cup 0} \tilde{f}^j(x \mid x_0, I) , \tag{5.5.27}$$

and, hence, using the definition of $x_1$ given in (5.5.4b),

$$\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \geq -\max_{j \in \mathbf{p} \cup 0} \tilde{f}^j(x_1 \mid x_0, I) , \tag{5.5.28}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Since $J_1(x_0) = \hat{J}$ for $x_0$ near $\hat{x}$, $\tilde{f}^j(x_1 \mid x_0, I)$ is equal to the maximum for each $j \in \hat{J}$. Therefore, (since this maximum is negative)

$$\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \geq \frac{1}{p} \sum_{j \in \hat{J}} |\tilde{f}^j(x_1 \mid x_0, I)| , \tag{5.5.29}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Since $x_1$ satisfies necessary conditions for (5.5.3) and

$J_1(x_0) = \hat{J}$, $\|\sum_{j \in \hat{J}} \mu_j \nabla \tilde{f}^j(x_1 \mid x_0, I)\| = 0$. Hence, there exists $K > 0$ such that

$$\|\sum_{j \in p(x_0)} \mu_j \nabla \tilde{f}^j(x_1 \mid x_0, H)\| \le K \|x_0 - x_1\|^2 . \tag{5.5.30}$$

Furthermore, there exists $K > 0$ such that $|\tilde{f}^j(x_1 \mid x_0, I) - \tilde{f}^j(x_1 \mid x_0, H)| \le K \|x_0 - x_1\|^2$ for $x_0$ near $\hat{x}$. Applying this and (5.5.30) to the definition of $C(\cdot)$ in (5.5.2a), we have that

$$\|C(x_1, u_1 \mid x_0, H, \hat{J})\| \le \sum_{j \in \hat{J}} |\tilde{f}^j(x_1 \mid x_0, I)| + K \|x_0 - x_1\|^2 , \tag{5.5.31a}$$

for $x_0$ near $\hat{x}$. Substituting (5.5.31a) into (5.5.29) yields

$$\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \ge \delta \|C(x_1, u_1 \mid x_0, H, \hat{J})\| - K \|x_0 - x_1\|^2 . \tag{5.5.31b}$$

Since $C(\cdot)$ is Lipschitz continuous and $C(\tilde{x}, \tilde{u} \mid x_0, H) = 0$, there exists $K > 0$ such that $\|C(x_0, \tilde{u} \mid x_0, H, \hat{J})\| \le K \|x_0 - \tilde{x}\|$. Substituting this into (5.5.31b) yields

$$K \|x_0 - \tilde{x}\| \ge \delta \|C(x_1, u_1 \mid x_0, H, \hat{J})\| - K \|x_0 - x_1\|^2 . \tag{5.5.32}$$

Substituting $\|x_0 - x_1\| \le \|x_0 - \tilde{x}\| + \|\tilde{x} - x_1\|$ into (5.5.32) yields

$$K\,(\|x_0 - \tilde{x}\| + \|x_0 - \tilde{x}\|^2 + \|x_0 - \tilde{x}\|\|x_1 - \tilde{x}\|) \ge \delta \|C(x_1, u_1 \mid x_0, H, \hat{J})\| - K \|x_1 - \tilde{x}\|^2 \tag{5.5.33}$$

By Lemma 5.5.1(b), there exists $\delta > 0$ such that

$$\|C(x_1, u_1 \mid x_0, H, \hat{J})\| \ge \delta \|(x_1, u_1) - (\tilde{x}, \tilde{u})\| \ge \|x_1 - \tilde{x}\| , \tag{5.5.34}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. By (5.5.34) and Lemma 5.5.1(c-d),

$$\delta \|C(x_1, u_1 \mid x_0, H, \hat{J})\| - K \|x_1 - \tilde{x}\|^2 \ge \tfrac{1}{2}\delta \|C(x_1, u_1 \mid x_0, H, \hat{J})\| . \tag{5.5.35}$$

Substituting (5.5.35) into (5.5.33) yields

$$K\,(\|x_0 - \tilde{x}\| + \|x_0 - \tilde{x}\|^2 + \|x_0 - \tilde{x}\|\|x_1 - \tilde{x}\|) \ge \tfrac{1}{2}\delta \|C(x_1, u_1 \mid x_0, H, \hat{J})\| . \tag{5.5.36}$$

Also by Lemma 5.5.1(c),

$$2K\|x_0-\tilde{x}\| \geq K\left(\|x_0-\tilde{x}\|+\|x_0-\tilde{x}\|^2+\|x_0-\tilde{x}\|\|x_1-\tilde{x}\|\right), \tag{5.5.37}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Substituting (5.5.37) into (5.5.36) yields

$$K\|x_0-\tilde{x}\| \geq \|C(x_1,u_1 \mid x_0,H,\hat{f})\|, \tag{5.5.38}$$

for some $K>0$ and for all $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$.

Combining (5.5.23) with (5.5.38), there exists $K>0$ such that

$$\|\tilde{x}-x_2\| \leq \left\|\begin{bmatrix}\tilde{x}\\ \tilde{u}\end{bmatrix}-\begin{bmatrix}x_2\\ u_2\end{bmatrix}\right\| \leq K\|x_0-\tilde{x}\|^4, \tag{5.5.39}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Inequality (5.5.39) implies that $\|x_0-\tilde{x}\| \leq 2\|x_0-x_2\|$. Hence,

$$\|\tilde{x}-x_2\| \leq K\|x_0-x_2\|^4, \tag{5.5.40}$$

for some $K>0$, $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. By (5.5.40), s Since $\tilde{\sigma}(x \mid x_0,H)$ is Lipschitz continuous in $x$, uniformly in $x_0$ and $H$, and since $\tilde{\sigma}(\tilde{x})=0$, there exists $K>0$ such that

$$\tilde{\sigma}(x_2 \mid x_0,H) \leq K\|\tilde{x}-x_2\|. \tag{5.5.41a}$$

Substituting (5.5.40) into (5.5.41a) yields that there exists $K>0$ such that

$$\tilde{\sigma}(x_2 \mid x_0,H) \leq K\|x_0-x_2\|^4, \tag{5.5.41b}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Hence the left-hand inequality in (5.5.4.11c) holds.

We turn to the right-hand inequality in (5.5.4.11c). By (5.5.39) and Lemma 5.5.1(d),

$$\lim_{x_0\to\hat{x},\,H\to F(\hat{x})}\|\tilde{x}-x_2\|=0 \quad\text{and}\quad \lim_{x_0\to\hat{x},\,H\to F(\hat{x})}\|\tilde{x}-\hat{x}\|=0.$$

Hence

$$\lim_{x_0\to\hat{x},\,H\to F(\hat{x})}\|x_2-x_0\|=0. \tag{5.5.41c}$$

**Algorithm Property 5.4.6(iv):** If $\psi(\hat{x}) < 0$, then (5.5.4.11d) holds automatically for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. We assume now that $\psi(\hat{x}) = 0$. With $a(s) \triangleq A(x_0, H)(s)$,

$$\tilde{f}^j(a(s) \mid x_0, H) = \tilde{f}^j(x_2 \mid x_0, H) + (2-2s)\langle \nabla \tilde{f}^j(x_2 \mid x_0, H), d_2 \rangle$$

$$+ \tfrac{1}{2}(2-2s)^2 \langle d_2, H_{\hat{k}} d_2 \rangle, \tag{5.5.42}$$

for $s > \frac{1}{2}$. It follows from Lemma 5.5.1(b) and the definition of $d_2$ in Algorithm 5.5.1 that there exists $K > 0$ such that

$$\|\dot{a}(1)\| = 2\|d_2\| \leq K\|x_2 - x_0\|^2, \tag{5.5.43}$$

for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Therefore, by the Lipschitz continuity of $\nabla\tilde{f}^j(\cdot \mid x_0, H)$,

$$\langle \nabla\tilde{f}^j(x_2 \mid x_0, H), d_2 \rangle \leq \langle \nabla\tilde{f}^j(x_1 \mid x_0, H), d_2 \rangle + K\|x_2 - x_1\|\|x_2 - x_0\|^2. \tag{5.5.44}$$

By the definition of $d_2$, $\langle \nabla\tilde{f}^j(x_1 \mid x_0, H), d_2 \rangle = -\|x_2 - x_0\|^2$ for $j \in \hat{J}$. Hence,

$$\langle \nabla\tilde{f}^j(x_2 \mid x_0, H), d_2 \rangle \leq -\|x_2 - x_0\|^2 + K\|x_2 - x_1\|\|x_2 - x_0\|^2. \tag{5.5.45a}$$

Lemma 5.5.1(c) and (5.5.41c) imply that $\lim_{x_0 \to \hat{x}, H \to F(\hat{x})} \|x_2 - x_1\| = 0$. Hence,

$$\langle \nabla\tilde{f}^j(x_2 \mid x_0, H), d_2 \rangle \leq -\tfrac{1}{2}\|x_2 - x_0\|^2, \tag{5.5.45b}$$

for $j \in \hat{J}$, $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Substituting (5.5.45b) into (5.5.42),

$$\tilde{f}^j(a(s) \mid x_0, H) \leq \tilde{f}^j(x_2 \mid x_0, H) - \tfrac{1}{4}(2-2s)\|x_2 - x_0\|^2, \tag{5.5.46}$$

for all $j \in \hat{J}$, $x_0$ near $\hat{x}$, $H$ near $F(\hat{x})$ and $s$ near 1. Using (5.5.43) and the fact that $f^j(\hat{x}) < 0$ for $j \notin \hat{J}$, there exists $\delta > 0$ such that

$$\tilde{f}^j(a(s) \mid x_0, H) \leq \tilde{\psi}(x_2 \mid x_0, H) - (1-s)\delta\|\dot{a}(1)\|, \tag{5.5.47}$$

for all $j \in \mathbf{p}$, $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, and $s$ near 1.

□

## 5.6 NUMERICAL EXPERIMENTS

An implementation of Algorithm Model 5.2.1 with the search arc function defined by Algorithm 5.5.1 was used to solve the inequality-constrained problems in [Hoc.1] for which feasible starting points are given. The results are compared in Table 5.1 with those for the algorithm of [Pan.1], which was the first superlinearly convergent feasible descent algorithm. The search direction computation for the algorithm of [Pan.1] involves the solution of two quadratic programs (involving the same quadratic term), an extra evaluation of the constraint functions, and the solution of a linear least-squares problem. This effort is comparable to that required to construct the search arc for Algorithm 5.5.1.

In the experiments in [Pan.1], the BFGS variable metric update was used to update the Lagrangian Hessian estimate. In our experiments with Algorithm Model 5.2.1, we also used a variable metric method to estimate the Hessians. Because the search direction computation described in Algorithm 5.5.1 does not require that the Hessian estimates $H_j^k$ be positive definite, we were free to use a wide variety of updating techniques. The rank-one update described in [Lue.1] was used when it was defined; otherwise the BFGS update was used. The Hessian estimates $H_k^j$ were each initialized to the identity.

In the experiments, the algorithm parameters were set to $\alpha = 0.35$ and $\beta = 0.9$. As we mentioned after the proof of Lemma 5.5.1, locating the narrow interval of acceptable near-unity step sizes can be difficult and requires a careful line search. Quadratic interpolation was used to reduce the number of trial step sizes tested.

Several modifications were made to Algorithm 5.5.1 to improve its performance. Although they would not invalidate the results of Section 5, their inclusion there would have complicated the analysis. A scaled norm, $\|\cdot\|_{L_k}$, was used in (5.5.4), where $L_k^{-1} \triangleq \sum_{j \in \mathbf{p} \cup 0} \mu_{k-1}^j H_k^j + \varepsilon_k I$ and where $\varepsilon_k > 0$ was selected to ensure that $L_k$ remained positive definite. Because the full Chebyshev step can be "wild" when $x_0$ is far from a solution to the GQP, we added the following step to Algorithm 5.5.1,

Step 3½ If $\|C((x_1, u_1) + h_n + h_c \mid x_0, H, J_1)\| > \|C(x_0, u_1 \mid x_0, H, J_1)\|$, then set $h_n = \bar{t}h_n$ and $h_c = \bar{t}^2 h_c$ where $\bar{t}$ solves

$$\min_{t \in \mathbb{R}} \|C((x_1, u_1) + th_n + t^2 h_c \mid x_0, H, J_1)\|^2 . \tag{6.1}$$

This minimization is relatively inexpensive, since the objective function of problem (6.1) is a fourth order polynomial in a single variable. Near a solution $\hat{z}$ satisfying Hypothesis 5.5.1, $\|C((x_1, u_1) + h_n + h_c \mid x_0, H, J_1)\| < \|C(x_0, u_1 \mid x_0, H, J_1)\|$, and $t = 1$. Hence, Proposition 5.2 holds for Algorithm 5.5.1 with this additional step. Finally, in the computation of a descent direction for $\tilde{\psi}(\cdot)$, we divide $d_2$ by $1 + \|x_2 - x_0\|$ to prevent $d_2$ from becoming excessively large when $x_0$ is far from a solution to ICP.

This implementation of Algorithm Model 5.2.1 proved to be competitive with the superlinear feasible descent method of [Pan.1] on all but a few of the problems tested. In general, the GQP-based algorithm performed better than the algorithm of [Pan.1] on convex problems, and more poorly than the algorithm of [Pan.1] on nonconvex problems. This may be partly due to the choice of the search arc function. Algorithm 5.5.1, which defines the search arc function used in the experiments, selects a feasible descent direction using the Pironneau-Polak algorithm, which uses *uniformly convex* approximations to the functions $f^j(\cdot)$.

## 5.7 CONCLUSIONS

The numerical experiments which we performed constitute a "proof of principle". They show that a GQP-based algorithm is competitve with a sophisticated method based upon successive approximation to optimality conditions. The mixed performance of the GQP-based algorithm in comparison with the algorithm of [Pan.1] suggests two things. First, the search arc function defined by Algorithm 5.5.1 can be improved upon. Algorithm Properties 5.3.2 and 5.4.6 offer considerable latitude in this task. Second, the extra evaluations of the constraints performed by the algorithm of [Pan.1], together with the Lagrangian Hessian, seems to contain as much curvature information as is useful in optimization, and the additional curvature information contained in the individual Hessian estimates does not seem to be useful even far from a solution.

## 5.8 APPENDIX

Proof of Lemma 5.4.4: (a) By assumption, $\underline{\mu}^0 > 1/(1+c)$. (Since Hypothesis 5.4.2(ii) implies that $\underline{\mu}^0 \triangleq \min\{ \mu^0 \mid \mu \in \hat{U} \} > 0$, such a $c$ exists.) The set-valued map $U(\cdot)$ is upper semicontinuous and compact-valued [Pol.4], and $U(\hat{x}) = \hat{U}$. Hence, $\min\{ \mu^0 \mid \mu \in U(x) \} > 1/(1+c)$ for $x$ near $\hat{x}$.

(b) For any $\mu \in \Sigma_{p+1}$, we have $\psi_+(x) \geq \sum_{j \in \mathbf{p}} \mu^j f^j(x) / (1-\mu^0)$, since $\psi_+(x) \geq f^j(x)$ for all $j \in \mathbf{p}$. Hence, $\psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x) \geq \mu^0 \psi_+(x)$. We apply this to the expression for $\sigma(\cdot)$ in (5.4.5). For any $x \in \mathbb{R}^n$ and any $\mu \in U(x)$,

$$\sigma(x) = \psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^0(x) \|^2$$

$$\geq \psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x)$$

$$\geq \mu^0 \psi_+(x) \,. \tag{5.8.1}$$

The conclusion follows from (5.8.1) and part (a).

(c) As in (b) above,

$$\tilde{\sigma}(x \mid \bar{x}, H) \geq \mu^0 \tilde{\psi}_+(x \mid \bar{x}, H) \,, \tag{5.8.2}$$

for any $\mu \in \tilde{U}(x \mid \bar{x}, H)$. Since $\tilde{U}(\cdot \mid \bar{x}, H)$ is upper semicontinuous in $x$, uniformly in $\bar{x}$ and $H$, and since $\tilde{U}(\hat{x} \mid \hat{x}, F(\hat{x})) = \hat{U}$, $\min\{ \mu^0 \mid \mu \in \tilde{U}(x \mid \bar{x}, H) \} > 1/(1+c)$ for $x$ near $\hat{x}$, $\bar{x}$ near $\hat{x}$ and $H$ near $F(\hat{x})$.

□

Proof of Lemma 5.4.5: First, we prove right hand inequality in (5.4.10). We define $d^j(x) \triangleq f^0(x) + cf^j(x)$ for $j \in \mathbf{p}$ and $d^0(x) \triangleq f^0(x)$. Note that $p_c(x) = \max_{j \in \mathbf{p} \cup 0} d^j(x)$. In Chapter 2, an optimality function for the minimax problem, $\min_{x \in \mathbb{R}^n} p_c(x)$, is defined by

$$\theta_c(x) = \min_{h \in \mathbb{R}^n} \max_{j \in \mathbf{p} \cup 0} d^j(x) + \langle \nabla d^j(x), h \rangle + \tfrac{1}{2} \| h \|^2 - p_c(x) . \tag{5.8.7}$$

The function $\theta_c(\cdot)$ is nonpositive and is zero only at stationary points of $p_c(\cdot)$ [Pol.4].

To relate $\theta_c(\cdot)$ to $p_c(x) - p_c(\hat{x})$, we will make use of Lemma 2.3.3. For this purpose, we will show that the assumptions of Lemma 2.3.3 hold. Hypothesis 5.4.1 ensures that assumptions (i) and (ii) of Lemma 2.3.3 hold with respect to the minimax problem, $\min_{x \in \mathbb{R}^n} p_c(x)$ at $\hat{x}$. We turn to assumption (iii) of Lemma 2.3.3.

We associate with the minimax problem, $\min_{x \in \mathbb{R}^n} p_c(x)$, the set of multiplier vectors $V_c(\hat{x})$ consisting of those $\nu \in \Sigma_{p+1}$ for which

$$\nu^0 \nabla f^0(\hat{x}) + \sum_{j \in \mathbf{p}} \nu^j \{ \nabla f^0(\hat{x}) + c \nabla f^j(\hat{x}) \} = 0 , \tag{5.8.8a}$$

$$\nu^0 f^0(\hat{x}) + \sum_{j \in \underline{p}} \nu^j \{ f^0(\hat{x}) + cf^j(\hat{x}) \} = p_c(\hat{x}) . \tag{5.8.8b}$$

The sets $U_c(\hat{x})$ and $U(\hat{x})$ are related as follows. Since $\psi_+(\hat{x}) = 0$, (5.8.8a-c) can be rewritten as

$$\nabla f^0(\hat{x}) + \sum_{j \in \mathbf{p}} \nu^j c \nabla f^j(\hat{x}) = 0 , \tag{5.8.9a}$$

$$\sum_{j \in \mathbf{p}} \nu^j c f^j(\hat{x}) = 0 . \tag{5.8.9b}$$

Then, since $1 - \nu^0 = \sum_{j \in \mathbf{p}} \nu^j$, $(1, c\nu^1, \ldots, c\nu^p) / (1 + c(1 - \nu^0)) \in U(\hat{x})$, for any $\nu \in V_c(\hat{x})$. It follows from Hypothesis 5.4.2 that, with $B$ as defined in Hypothesis 5.4.2,

$$m \| h \|^2 < \langle h , \left[ \frac{1}{1+c(1-\nu^0)} F^0(\hat{x}) + \sum_{j \in \mathbf{p}} \frac{c\nu^j}{1+c(1-\nu^0)} F^j(\hat{x}) \right] h \rangle, \quad \forall\, h \in B,\ h \neq 0, \tag{5.8.10}$$

for any $\nu \in V_c(\hat{x})$. Inequality (5.8.10) and the fact that $B = H$ imply that for any $\nu \in V_c n(\hat{x})$,

$$m_c \| h \|^2 < \langle h , \left[ \nu^0 F^0(\hat{x}) + \sum_{j \in \mathbf{p}} \nu^j \{ F^0(\hat{x}) + cF^j(\hat{x}) \} \right] h \rangle \quad \forall\, h \in H ,\ h \neq 0 , \tag{5.8.11}$$

where $\quad m_c \triangleq \min \{ m\ (1+c(1-\nu^0)) \mid \nu \in V_c(\hat{x}) \} = m\ (1+c(1-\max \{ \nu^0 \mid \mu \in V_c(\hat{x}) \} ))$.

Hence, assumption (iii) of Lemma 2.3.3 is satisfied at $\hat{x}$ for the minimax problem, and it follows from Lemma 2.3.3 that

$$\limsup_{i \to \infty} \frac{\theta_c(x_i)}{p_c(x_i) - p_c(\hat{x})} \leq - \frac{\min \{ m_c , (1+c) \}}{(1+c)} . \tag{5.8.12}$$

This implies that there exists $K > 0$ such that

$$\theta_c(x) \leq K\ (p_c(\hat{x}) - p_c(x))\ , \tag{5.8.13}$$

for $x$ near $\hat{x}$.

Now we relate $\theta_c(\cdot)$ to $\sigma(\cdot)$. It follows from Theorem 2.7.1, which is an extension to the von Neumann Minimax Theorem, that

$$\theta_c(x) = \max_{\mu \in \Sigma_{p+1}} \mu^0 ( f^0(x) - p_c(x)) + \sum_{j \in \mathbf{p}} \mu^j\ ( f^0(x) + cf^j(x) - p_c(x))$$

$$- \tfrac{1}{2} \| \mu^0 \nabla f^0(x) + \sum_{j \in \mathbf{p}} \mu^j\ (\nabla f^0(x) + c \nabla f^j(x)) \|^2 . \tag{5.8.14}$$

Rearranging (5.8.14), we have

$$\theta(x) = - \left[ \min_{\mu \in \Sigma_{p+1}} c\, \psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j\, cf^j(x) + \tfrac{1}{2} \| \nabla f^0(x) + \sum_{j \in \mathbf{p}} \mu^j\, c \nabla f^j(x) \|^2 \right] . \tag{5.8.15}$$

Using any $\mu \in U(x)$, we can define another multiplier vector as follows.

$$\nu \triangleq \left[ 1 - \sum_{j \in \mathbf{p}} \frac{\mu^j}{c\mu^0}, \frac{\mu^1}{c\mu^0}, \ldots, \frac{\mu^p}{c\mu^0} \right] \in \Sigma_{p+1}. \tag{5.8.16}$$

By Lemma 5.4.4(a), $c > (1-\mu^0)/\mu^0$ for any $\mu \in U(x)$ for $x$ near $\hat{x}$. Then, since $c > (1-\mu^0)/\mu^0$, $\nu^0 > 0$. Substituting this $\nu$ into the min in (5.8.15) and using (5.8.16) yields

$$-\theta(x) \le c\,\psi_+(x) - \sum_{j \in \mathbf{p}} \nu^j c f^j(x) + \tfrac{1}{2}\|\nabla f^0(x) + \sum_{j \in \mathbf{p}} \nu^j c\, \nabla f^j(x)\|^2$$

$$= c\,\psi_+(x) - \sum_{j \in \mathbf{p}} \frac{\mu^j}{\mu^0} f^j(x) + \tfrac{1}{2}\|\nabla f^0(x) + \sum_{j \in \mathbf{p}} \frac{\mu^j}{\mu^0} \nabla f^j(x)\|^2$$

$$= \left[ \mu^0 \{ c\mu^0\psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x) \} + \tfrac{1}{2}\|\mu^0\nabla f^0(x) + \sum_{j \in \mathbf{p}} \mu^j\, \nabla f^j(x)\|^2 \right] / (\mu^0)^2$$

$$\le \left[ c\mu^0\psi_+(x) - \textstyle\sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2}\|\mu^0\nabla f^0(x) + \displaystyle\sum_{j \in \mathbf{p}} \mu^j\, \nabla f^j(x)\|^2 \right] / (\mu^0)^2 \tag{5.8.17}$$

since $\mu^0 c > \sum_{j \in \mathbf{p}} \mu^j$ and $\psi_+(x) \ge f^j(x)$. Since $\mu \in U(x)$, $\sigma(x) = \psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2}\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x)\|^2$. Hence,

$$-\theta(x) \le \{ \sigma(x) + (c\bar{\mu}^0 - 1)\psi_+(x) \} / (\mu^0)^2, \tag{5.8.18a}$$

for $x$ near $\hat{x}$. By Lemma 5.4.4(b),

$$-\theta(x) \le \{ \sigma(x) + (c\bar{\mu}^0 - 1)(1+c)\sigma(x) \} / (\bar{\mu}^0)^2 \le (1+c)^3\sigma(x), \tag{5.8.18b}$$

for $x$ near $\hat{x}$, since $c > c\bar{\mu}^0 - 1$. Substituting (5.8.18b) into (5.8.13) yields

$$p_c(x) - p_c(\hat{x}) \le (1+c)^3 K\sigma(x), \tag{5.8.19}$$

for $x$ near $\hat{x}$. This is the right-hand inequality in (5.4.10).

Now we prove the left-hand inequality in (5.4.10). By Lemma 5.4.3, $\hat{x}$ is a strict local minimizer of $p_c(\cdot)$. Therefore, there exists a neighborhood $W$ of $\hat{x}$ such that

$$p_c(x) - p_c(\hat{x}) = \max_{\bar{x} \in \overline{W}} p_c(x) - p_c(\bar{x}) \,, \tag{5.8.20}$$

where $\overline{W}$ denotes the closure of the open set $W$. Since the functions $f^j(\cdot)$ are twice continuously differentiable, there exists $K \geq 1$ such that

$$f^j(\bar{x}) \leq f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2 \,, \tag{5.8.21}$$

for $x, \bar{x} \in \overline{W}$. Therefore,

$$p_c(x) - p_c(\hat{x}) \geq \max_{\bar{x} \in \overline{W}} p_c(x) - \Big\{ f^0(x) + \langle \nabla f^0(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2$$

$$+ c \max_{j \in \mathbf{p}} \{ 0, f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2 \} \Big\}. \tag{5.8.22}$$

By the definition of $p_c(\cdot)$,

$$p_c(x) - p_c(\hat{x}) \geq \max_{\bar{x} \in \overline{W}} c\, \psi_+(x) - \Big\{ \langle \nabla f^0(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2$$

$$+ c \max_{j \in \mathbf{p}} \{ 0, f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2 \} \Big\}. \tag{5.8.23}$$

We can replace the max over $j \in \mathbf{p}$ with a max over $\mu \in \Sigma_{p+1}$ (which becomes a min when pulled through the minus sign to the front),

$$p_c(x) - p_c(\hat{x}) \geq \max_{\bar{x} \in \overline{W}} \min_{\mu \in \Sigma_{p+1}} c\, \psi_+(x) - \Big\{ \langle \nabla f^0(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2$$

$$+ \sum_{j \in \mathbf{p}} \mu^j c \left( f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2 \right) \Big\}. \tag{5.8.24}$$

Now the function

$$\Phi(\bar{x} \mid x) \triangleq \min_{\mu \in \Sigma_{p+1}} c\, \psi_+(x) - \Big\{ \langle \nabla f^0(x), \bar{x} - x \rangle + \tfrac{1}{2} K \| \bar{x} - x \|^2$$

$$+ \sum_{j \in \mathbf{p}} \mu^j c \left( f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + ½K \| \bar{x} - x \|^2 \right) \Bigg\}, \quad (5.8.25)$$

is concave in $\bar{x}$. Therefore, the function $\operatorname{argmax}_{\bar{x} \in \overline{W}} \Phi(\bar{x} \mid x)$ is an upper semicontinuous set-valued map in $x$ (by the Maximum Theorem in [Ber.1]). Since $\operatorname{argmax}_{\bar{x} \in \overline{W}} \Phi(\bar{x} \mid \hat{x}) = \{ \hat{x} \}$, this implies that $\operatorname{argmax}_{\bar{x} \in \overline{W}} \Phi(\bar{x} \mid x) \subset W$ for $x$ near $\hat{x}$. Therefore we can substitute $\mathbb{R}^n$ for $\overline{W}$ in the max in (5.8.24),

$$p_c(x) - p_c(\hat{x}) \ge \max_{\bar{x} \in \mathbb{R}^n} \min_{\mu \in \Sigma_{p+1}} c\,\psi_+(x) - \Bigg\{ \langle \nabla f^0(x), \bar{x} - x \rangle + ½K \| \bar{x} - x \|^2$$

$$+ \sum_{j \in \mathbf{p}} \mu^j c \left( f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + ½K \| \bar{x} - x \|^2 \right) \Bigg\}. \quad (5.8.26)$$

By Theorem 2.7.1, we can interchange the min and max in (5.8.26). Hence,

$$p_c(x) - p_c(\hat{x}) \ge \min_{\mu \in \Sigma_{p+1}} \max_{\bar{x} \in \mathbb{R}^n} c\,\psi_+(x) - \Bigg\{ \langle \nabla f^0(x), \bar{x} - x \rangle + ½K \| \bar{x} - x \|^2$$

$$+ \sum_{j \in \mathbf{p}} \mu^j c \left( f^j(x) + \langle \nabla f^j(x), \bar{x} - x \rangle + ½K \| \bar{x} - x \|^2 \right) \Bigg\}. \quad (5.8.27)$$

Rearranging (5.8.27),

$$p_c(x) - p_c(\hat{x}) \ge \min_{\mu \in \Sigma_{p+1}} \max_{\bar{x} \in \mathbb{R}^n} \Big( c\,\psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j c f^j(x) \Big) - \Big\langle \nabla f^0(x) + \sum_{j \in \mathbf{p}} \mu^j c \nabla f^j(x), \bar{x} - x \Big\rangle$$

$$- ½(1 + c(1 - \mu^0))K \| \bar{x} - x \|^2 . \quad (5.8.28)$$

Solving the inner max, we have

$$p_c(x) - p_c(\hat{x}) \ge \min_{\mu \in \Sigma_{p+1}} \Big( c\,\psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j c f^j(x) \Big)$$

$$+\frac{1}{2(1+c(1-\mu^0))K}\|\nabla f^0(x)+\sum_{j\in \mathbf{p}}\mu^j c\nabla f^j(x)\|^2 . \tag{5.8.29a}$$

Therefore, for some $\mu \in \Sigma_{p+1}$,

$$p_c(x)-p_c(\hat{x}) \geq (c\,\psi_+(x)-\sum_{j\in \mathbf{p}}\mu^j c f^j(x))$$

$$+\frac{1}{2(1+c(1-\mu^0))K}\|\nabla f^0(x)+\sum_{j\in \mathbf{p}}\mu^j c\nabla f^j(x)\|^2 . \tag{5.8.29b}$$

Since $\psi_+(x) \geq f^j(x)$ for $j \in \mathbf{p}$ and $K \geq 1$,

$$p_c(x)-p_c(\hat{x}) \geq \frac{1}{K}\Big\{c\,\psi_+(x)-\sum_{j\in \mathbf{p}}\mu^j c f^j(x)$$

$$+\frac{1}{2(1+c(1-\mu^0))}\|\nabla f^0(x)+\sum_{j\in \mathbf{p}}\mu^j c\nabla f^j(x)\|^2\Big\}. \tag{5.8.30}$$

Let

$$\nu \triangleq \left[1,\mu^1 c,\ldots,\mu^p c\right]/(1+c(1-\mu^0)), \tag{5.8.31}$$

and note that $\nu \in \Sigma^{p+1}$. We substitute $\mu^j = \nu^j(1+c(1-\mu^0))/c$ for $j \in \mathbf{p}$ into (5.8.30),

$$p_c(x)-p_c(\hat{x}) \geq \frac{1}{K}\Big\{c\,\psi_+(x)-(1+c(1-\mu^0))\sum_{j\in \mathbf{p}}\nu^j f^j(x)$$

$$+\frac{1+c(1-\mu^0)}{2}\|\sum_{j\in \mathbf{p}\cup 0}\nu^j\nabla f^j(x)\|^2\Big\}. \tag{5.8.32}$$

Now, $p_c(x)-p_c(\hat{x}) \geq c\,\psi_+(x)$ for $x$ near $\hat{x}$, since $\hat{x}$ is a local minimizer of ICP and $\psi_+(\hat{x})=0$. Therefore,

$$\frac{1+c(1-\mu^0)}{K}(p_c(x)-p_c(\hat{x})) \geq \frac{1+c(1-\mu^0)}{K}\psi_+(x). \tag{5.8.33}$$

Adding (5.8.33) to (5.8.32) and rearranging yields

$$p_c(x) - p_c(\hat{x}) \geq \frac{(1 + c(1 - \mu^0))/K}{1 + (1 + c(1 - \mu^0))/K} \left\{ (\psi_+(x) - \sum_{j \in p} v^j f^j(x)) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} v^j \nabla f^j(x) \|^2 \right\}$$

$$\geq \frac{1}{1 + c + K} \left\{ (\psi_+(x) - \sum_{j \in p} v^j f^j(x)) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} v^j \nabla f^j(x) \|^2 \right\}. \tag{5.8.34a}$$

We can take the min over $v \in \Sigma_{p+1}$,

$$p_c(x) - p_c(\hat{x}) \geq \frac{1}{1 + c + K} \left\{ \min_{v \in \Sigma_{p+1}} (\psi_+(x) - \sum_{j \in p} v^j f^j(x)) + \tfrac{1}{2} \| \sum_{j \in p \cup 0} v^j \nabla f^j(x) \|^2 \right\} \tag{5.8.34b}$$

Therefore, by the characterization of $\sigma(\cdot)$ in (5.4.5), there exists $\delta > 0$ such that

$$p_c(x) - p_c(\hat{x}) \geq \delta \sigma(x) , \tag{5.8.35}$$

for $x$ near $\hat{x}$.

□

**Proof of Lemma 5.4.8:** (a) Since $\psi(x_0) \leq 0$, $\psi(x_k) \leq 0$ and $f^0(x_k) \leq f^0(x_0)$ for all $k \in \mathbb{N}$, by Step 2 of Algorithm Model 5.2.1 and Algorithm Property 5.3.2(ii). This implies that the sequence $\{ x_k \}_{k \in \mathbb{N}}$ is contained in $V$ and hence is bounded. Therefore, $\{ x_k \}_{k \in \mathbb{N}}$ converges to the set of its accumulation points. Since the conditions of Theorem 5.3.3 are implied by Hypothesis 5.4.1, Algorithm Property 5.4.7 and Algorithm Property 5.3.2, any accumulation point of the sequence $\{ x_k \}_{k \in \mathbb{N}}$ is a stationary point of ICP. By Hypothesis 5.4.1(ii), $\hat{x}$ is the only stationary point in $V$. Hence $x_k \to \hat{x}$ as $k \to \infty$.

(b) This follows from (i) of this lemma, the continuity of $\sigma(\cdot)$ [Pol.4] and the fact that $\sigma(\hat{x}) = 0$.

(c) From (ii) of this lemma and Algorithm Property 5.4.6(iii), $\tilde{\sigma}(a_k(1) \mid x_k, H_k) \to 0$ as $k \to \infty$. The set-valued map $M(\bar{x}, H, r) \triangleq \{ x \in \mathbb{R}^n \mid \tilde{\sigma}(x \mid \bar{x}, H) \leq r \}$ is upper semicontinuous in $\bar{x}$, $r$ and $H$, $M(\cdot)$ is compact-valued and $M(\hat{x}, F(\hat{x}), 0) = \{ \hat{x} \}$. Therefore, $a_k(1) \to \hat{x}$ as $k \to \infty$.
□

**Proof of Lemma 5.4.9:** By the definition of $\sigma(\cdot)$ in (5.4.5) and the triangle inequality,

$$\sigma(x) = \min_{\mu \in \Sigma_{p+1}} \psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2}\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2$$

$$\leq \psi_+(x) - \sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2}\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2$$

$$\leq \tilde{\psi}_+(x \mid x_k, H_k) - \sum_{j \in \mathbf{p}} \mu^j \tilde{f}^j(x \mid x_k, H_k) + \tfrac{1}{2}\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla \tilde{f}^j(x \mid x_k, H_k) \|^2$$

$$+ \delta_1(x, x_k, H_k) + \delta_2(x \mid x_k, H_k) + \delta_3(x \mid x_k, H_k) + \delta_4(x \mid x_k, H_k), \tag{5.8.36}$$

for any $x \in \mathbb{R}^n$ and $\mu \in \Sigma_{p+1}$, where

$$\delta_1(x, x_k, H_k) \triangleq \psi_+(x) - \tilde{\psi}_+(x \mid x_k, H_k), \tag{5.8.37}$$

$$\delta_2(x \mid x_k, H_k) \triangleq \sum_{j \in \mathbf{p}} \mu^j (\tilde{f}^j(x \mid x_k, H_k) - f^j(x)) \tag{5.8.38}$$

$$\delta_3(x \mid x_k, H_k) \triangleq \tfrac{1}{2}\| \sum_{j \in \mathbf{p} \cup 0} \mu^j (\nabla f^j(x) - \nabla \tilde{f}^j(x \mid x_k, H_k)) \|^2, \tag{5.8.39}$$

$$\delta_4(x \mid x_k, H_k) \triangleq \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla \tilde{f}^j(x \mid x_k, H_k) \| \| \sum_{j \in \mathbf{p} \cup 0} \mu^j (\nabla f^j(x) - \nabla \tilde{f}^j(x \mid x_k, H_k)) \| \tag{5.8.40}$$

For any $x \in \mathbb{R}^n$ and any $\mu \in \tilde{U}(x \mid x_k, H_k)$, therefore,

$$\sigma(x) \leq \tilde{\sigma}(x \mid x_k, H_k) + \delta_1(x \mid x_k, H_k) + \delta_2(x \mid x_k, H_k) + \delta_3(x \mid x_k, H_k) + \delta_4(x \mid x_k, H_k), \tag{5.8.41}$$

using the definition of $\tilde{U}(\cdot)$ in (5.4.9b). Now, for any $x \in S(x_k, 2\|a_k(1) - x_k\|)$ and each $j \in \mathbf{p}$,

$$f^j(x) - \tilde{f}^j(x \mid x_k, H_k) \leq \tfrac{1}{2}\eta_k \|x - x_k\|^2. \tag{5.8.42}$$

Therefore, $\delta_1(x \mid x_k, H_k) \leq \eta_k \|x - x_k\|^2$ and $\delta_2(x \mid x_k, H_k) \leq \eta_k \|x - x_k\|^2$ for any $x \in S(x_k, 2\|a_k(1) - x_k\|)$. Also,

$$\| \sum_{j \in \mathbf{p} \cup 0} \mu^j \{ \nabla f^j(x) - \nabla \tilde{f}^j(x \mid x_k, H_k) \} \| \leq \eta_k \|x - x_k\|, \tag{5.8.43}$$

and hence $\delta_3(x \mid x_k, H_k) \le (\eta_k \| x - x_k \|)^2$ for any $x \in S(x_k, 2\|a_k(1) - x_k\|)$. By the definition of $\tilde{\sigma}(\cdot)$ and Algorithm Property 5.4.6(iii), there exists $K > 0$ such that

$$\| \sum_{j \in p \cup o} \mu^j \nabla \tilde{f}^j(a_k(1) \mid x_k, H_k) \| \le \sqrt{\tilde{\sigma}(a_k(1) \mid x_k, H_k)} \le K \| a_k(1) - x_k \|^{1.5} \le K \| a_k(1) - x_k \| . \quad (5.8.44)$$

for large $k$. Inequalities (5.8.43) and (5.8.44) imply that $\delta_4(a_k(1) \mid x_k, H_k) \le \eta_k \| x - x_k \|^2$ for large $k$. Substituting these bounds on $\delta_j(a_k(1) \mid x_k, H_k)$ into (5.8.41) yields

$$\sigma(a_k(1)) \le \tilde{\sigma}(a_k(1) \mid x_k, H_k) + \eta_k \| a_k(1) - x_k \|^2 + \tfrac{1}{2}(\eta_k \| a_k(1) - x_k \|)^2 + K \eta_k \| a_k(1) - x_k \|^2 . \quad (5.8.45)$$

Since $\eta_k$ converges to 0 as $k \to \infty$ by Algorithm Property 5.4.7, $\eta_k \le 1$ for large $k$ and hence

$$\sigma(a_k(2)) \le \tilde{\sigma}(a_k(1) \mid x_k, H_k) + K \eta_k \| a_k(1) - x_k \|^2 . \quad (5.8.46)$$

□

Proof of Lemma 5.4.10: We combine the previous lemmas. By Lemma 5.4.8, the sequences $a_k(1)$ and $x_k$ converge to $\hat{x}$. Therefore, by Lemma 5.4.3, there exists $K > 0$ such that

$$\| a_k(1) - \hat{x} \|^2 \le K \, p_c(a_k(1)) - p_c(\hat{x}) , \quad (5.8.47)$$

for any $c > (1 - \underline{\mu}^0) / \underline{\mu}^0$ and large $k$. Combining Lemmas 5.4.5 and 5.4.9 with (5.8.47), there exists $K > 0$ such that

$$\| a_k(1) - \hat{x} \|^2 \le K \sigma(a_k(1)) \le K \left[ \tilde{\sigma}(a_k(1) \mid x_k, H_k) + \eta_k \| a_k(1) - x_k \|^2 \right] \quad (5.8.48)$$

for large $k$. Applying Algorithm Property 5.4.6(iii) to (5.8.48), there exists $K > 0$ such that

$$\| a_k(1) - \hat{x} \|^2 \le K \| a_k(1) - x_k \|^3 + K \eta_k \| a_k(1) - x_k \|^2$$

$$\le K \max \{ \eta_k , \| a_k(1) - x_k \| \} \| a_k(1) - x_k \|^2 . \quad (5.8.49)$$

for large $k$. Taking the square root of both sides of (5.8.49) yields (5.4.14).

□

Proof of Lemma 5.4.11: We show that inequalities (5.2.2a-b) of Step 2 of Algorithm Model 5.2.1 hold for step lengths $s$ satisfying (5.4.15). Since $x_0$ is feasible for ICP, every $x_k$ is feasible.

By Algorithm Properties 5.4.6(i-ii), $\|a_k(s)-x_k\| \le 2\|a_k(1)-x_k\|$, for large $k$ and $s$ near 1. Therefore, there exists $K > 0$ such that

$$\psi(a_k(s)) \le \tilde{\psi}(a_k(s) \mid x_k, H_k) + \max_{j \in \mathbf{p}\cup 0} \langle a_k(s)-x_k, \left[\int_0^1 (1-t) F^j(x_k + t(a_k(s)-x_k))dt - H_k^j\right](a_k(s)-x_k)\rangle$$

$$\le \tilde{\psi}(a_k(s) \mid x_k, H_k) + \eta_k \|a_k(s)-x_k\|^2, \tag{5.8.50}$$

for $s$ near 1 and large $k$. By Algorithm Property 5.4.6(iv), Lemma 5.4.4(c) and Algorithm Property 5.4.6(iii), there exist $\delta > 0$ and $K > 0$ such that

$$\psi(a_k(s)) \le \tilde{\psi}(a_k(1) \mid x_k, H_k) + (s-1)\delta\|\dot{a}_k(1)\| + \eta_k\|a_k(1)-x_k\|^2$$

$$\le (1+c)\tilde{\sigma}(a_k(1) \mid x_k, H_k) + (s-1)\delta\|\dot{a}_k(1)\| + \eta_k\|a_k(1)-x_k\|^2$$

$$\le K\|a_k(1)-x_k\|^3 + (s-1)\delta\|\dot{a}_k(1)\| + \eta_k\|a_k(1)-x_k\|^2 \tag{5.8.51}$$

for $s$ near 1 and large $k$. Therefore, there exists $K > 0$ such that $\psi(a_k(s)) \le 0$ for large $k$ and for $s$ near 1 such that

$$1-s \ge K \max\{\eta_k, \|a_k(1)-x_k\|\}\, \|a_k(1)-x_k\|^2 / \|\dot{a}_k(1)\|. \tag{5.8.52}$$

Hence, for such $s$ and large $k$, inequality (5.2.2a) of Step 2 of Algorithm Model 5.2.1 holds.

Now we determine which step lengths $s$ provide sufficient decrease in the objective function to satisfy inequality (5.2.2b). Let $c > \max\{1, (1-\underline{\mu}^0)/\underline{\mu}^0\}$. For $s$ such that $\psi(a_k(s)) \le 0$,

$$f^0(a_k(s)) - f^0(x_k) = p_c(a_k(s)) - p_c(x_k)$$

$$= p_c(a_k(s)) - p_c(a_k(1)) + p_c(a_k(1)) - p_c(\hat{x}) + p_c(\hat{x}) - p_c(x_k). \tag{5.8.53}$$

Now we bound each of the three terms on the right-hand side of (5.8.55). By Lemma 5.4.5 and the fact that $x_k$ converges to $\hat{x}$, there exists $\tau > 0$ such that

$$p_c(\hat{x}) - p_c(x_k) \le -\tau\sigma(x_k), \tag{5.8.54}$$

for large $k$.

By Lemma 5.4.5, there exists $K > 0$ such that

$$p_c(a_k(1)) - p_c(\hat{x}) \le K\,\sigma(a_k(1)) \,, \tag{5.8.55}$$

for large $k$. Applying Lemma 5.4.9 and Algorithm Property 5.4.6(iii) to (5.8.55) yields

$$p_c(a_k(1)) - p_c(\hat{x}) \le K \max\{\, \eta_k \,, \|a_k(1) - x_k\| \,\} \, \|a_k(1) - x_k\|^2 \,, \tag{5.8.56}$$

for large $k$. By Lemmas 5.4.5 and 5.4.7, there exists $K > 0$ such that

$$K\,\sigma(x_k) \ge \|x_k - \hat{x}\,\|^2 \,, \tag{5.8.57}$$

for large $k$. By Lemma 5.4.10, $\|a_k(1) - x_k\| \le 2\|x_k - \hat{x}\,\|$ for large $k$. Substituting this into (5.8.57) yields

$$K\,\sigma(x_k) \ge \|a_k(1) - x_k\|^2 \,, \tag{5.8.58}$$

for large $k$ and some $K > 0$. Substituting (5.8.58) into (5.8.56) yields

$$p_c(a_k(1)) - p_c(\hat{x}) \le K \max\{\, \eta_k \,, \|a_k(1) - x_k\| \,\} \, \sigma(x_k) \,, \tag{5.8.59}$$

for large $k$ and $s$ near 1. Since $a_k(1) - x_k$ converges to 0,

$$p_c(a_k(1)) - p_c(\hat{x}) \le \tfrac{1}{4}\tau\sigma(x_k) \,, \tag{5.8.60}$$

for large $k$.

By the Lipschitz continuity of the functions $f^j(\cdot)$ and Algorithm Property 5.4.6(i), there exists $K > 0$ such that

$$p_c(a_k(s)) - p_c(a_k(1)) \le K\,|1 - s|\,\|\dot{a}_k(1)\| \tag{5.8.61}$$

for $s$ near 1 and large $k$. Let $\rho > 0$ be arbitrary. For $s$ such that

$$|1 - s| \le \rho \frac{\|a_k(1) - x_k\|^2}{\|\dot{a}_k(1)\|} \,, \tag{5.8.62}$$

(5.8.61) becomes

$$p_c(a_k(s)) - p_c(a_k(1)) \leq \rho K \| a_k(1) - x_k \|^2 . \tag{5.8.63}$$

Substituting (5.8.58) into (5.8.63),

$$p_c(a_k(s)) - p_c(a_k(1)) \leq \rho K \sigma(x_k) , \tag{5.8.64a}$$

for large $k$ and $s$ near 1 satisfying (5.8.62). Since $\rho > 0$ was arbitrary, we can choose $\rho$ small enough that

$$p_c(a_k(s)) - p_c(a_k(1)) \leq \tfrac{1}{4}\tau\sigma(x_k) . \tag{5.8.64b}$$

Substituting (5.8.54), (5.8.60) and (5.8.64b) into (5.8.53) yields

$$f^0(a_k(s)) - f^0(x_k) \leq -\tfrac{1}{2}\tau\sigma(x_k) , \tag{5.8.65}$$

for large $k$ and $s$ near 1 satisfying (5.8.62). By (5.4.4) and the fact that $\psi_+(x_k) = 0$,

$$\sigma(x) = - \max_{\mu \in \Sigma_{p+1}} \sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2 \tag{5.8.66}$$

$$= - \sum_{j \in \mathbf{p}} \mu^j f^j(x) + \tfrac{1}{2} \| \sum_{j \in \mathbf{p} \cup 0} \mu^j \nabla f^j(x) \|^2 ,$$

for any $\mu \in U(x)$. Then, for any $\mu \in U(x)$,

$$\sigma(x) = - \min_{x' \in \mathbb{R}^n} \sum_{j \in \mathbf{p} \cup 0} \mu^j \tilde{f}^j(x' \mid x, I) . \tag{5.8.67}$$

Substituting $x_k + \tfrac{1}{2}\dot{a}_k(0)$ for $x'$ in the min in (5.8.67),

$$\sigma(x) \geq - \sum_{j \in \mathbf{p} \cup 0} \mu^j \tilde{f}^j(x + \tfrac{1}{2}\dot{a}(0) \mid x, I) \geq -\mu^0 \tilde{f}^0(x + \tfrac{1}{2}\dot{a}(0) \mid x, I) , \tag{5.8.68}$$

since Algorithm Property 5.3.2(iii) implies that $\tilde{f}^j(x + \tfrac{1}{2}\dot{a}(0) \mid x, I) < 0$ for $j \in \mathbf{p}$. By Lemma 5.4.4(a), there exists $\delta > 0$ such that $\mu^0 > \delta$ for $\mu \in U(x)$ and $x$ near $\hat{x}$. Hence, $\sigma(x) \geq -\delta \tilde{f}^0(x + \tfrac{1}{2}\dot{a}(0) \mid x, I)$ for $x$ near $\hat{x}$. Substituting this into (5.8.65),

$$f^0(a_k(s)) - f^0(x_k) \le \tfrac{1}{2}\pi\delta\tilde{f}^0(x + \tfrac{1}{2}\dot{a}(0) \mid x, \mathcal{I}), \tag{5.8.69}$$

for large $k$ and $s$ near 1 satisfying (5.8.58). Algorithm Properties 5.4.6(ii-iii) together imply that $\dot{a}_k(0)$ converges to zero as $x_k$ converges to $\hat{x}$. Therefore,

$$f^0(a_k(s)) - f^0(x_k) \le s\rho(\|\dot{a}_k(0)\|) \max_{j \in p \cup 0} \tilde{f}^j(x_k + \tfrac{1}{2}\dot{a}_k(0) \mid x_k, \mathcal{I}), \tag{5.8.70}$$

for large $k$ and $s$ such that

$$1 - s \le \rho \frac{\|a_k(1) - x_k\|^2}{\|\dot{a}_k(1)\|}. \tag{5.8.71}$$

By (5.8.70), (5.8.71) and (5.8.51-52), inequalities (5.2.2a-b) are satisfied for large $k$ by all $s$ in the interval

$$1 - \rho \frac{\|a_k(1) - x_k\|^2}{\|\dot{a}_k(1)\|} \le s \le 1 - K \max\{\eta_k, \|a_k(1) - x_k\|\} \frac{\|a_k(1) - x_k\|^2}{\|\dot{a}_k(1)\|}, \tag{5.8.72}$$

for some $\delta > 0$ and $K > 0$. Since $\eta_k \to 0$, $a_k(1) \to \hat{x}$ and $x_k \to \hat{x}$, this interval is nonempty for large $k$.

Now we must account for the quantization of the step size. For each $k$, let $\bar{s}_k$ denote the maximum element of $[0, 1]$ satisfying (5.2.2a-b). By (5.8.72), there exists $K > 0$ such that

$$1 - \bar{s}_k \le K \max\{\eta_k, \|a_k(1) - x_k\|\} \frac{\|a_k(1) - x_k\|^2}{\|\dot{a}_k(1)\|}, \tag{5.8.73}$$

for large $k$. For large $k$, either $s_k = 1$ or $s_k = 1 - \beta^{j_k}$ for some $j_k \in \mathbf{N}$. In both cases, $1 - s_k \le (1 - \bar{s}_k)/\beta$. Therefore, (5.4.15) holds for large $k$.

□

**Proof of Lemma 5.5.1:**

(a) This part follows by the same argument as Lemma 4.5.1.

(b) Since we have assumed, without loss of generality, that $\hat{\mathcal{I}} = \{1, \ldots, r\}$, the matrix $\nabla C(x, u \mid \hat{x}, F(\hat{x}), \hat{\mathcal{I}})$ has the form

$$\nabla C(x, u \mid x_0, H, \hat{\mathcal{J}}) = \left[ \frac{\partial C(x, u \mid x_0, H, \hat{\mathcal{J}})}{\partial(x, u_1, \ldots, u_r)} \;\; 0 \right], \tag{5.8.74a}$$

where

$$\frac{\partial C(x, u \mid x_0, H, \hat{\mathcal{J}})}{\partial(x, u_1, \ldots, u_r)} = \begin{bmatrix} F^0(\hat{x}) + \sum_{j \in \hat{\mathcal{J}}} u^j F^j(\hat{x}) & \nabla \tilde{f}^{j_1}(x \mid x_0, H) \ldots \nabla \tilde{f}^{j_r}(x \mid x_0, H) \\ \nabla \tilde{f}^1(x \mid x_0, H)^T & \\ \vdots & 0 \\ \nabla \tilde{f}^r(x \mid x_0, H)^T & \end{bmatrix} \tag{5.8.74b}$$

The matrix $\partial C(x, u \mid x_0, H, \hat{\mathcal{J}}) / \partial(x, u_1, \ldots, u_r)$ has the form of a "bordered Hessian". By Hypothesis 5.4.2(i), $F^0(\hat{x}) + \sum_{j \in \hat{\mathcal{J}}} u^j F^j(\hat{x})$ is positive definite on the subspace orthogonal to the vectors $\{ \nabla \tilde{f}^j(\hat{x} \mid \hat{x}, F(\hat{x})) \}_{j \in \hat{\mathcal{J}}}$. Hence, the matrix $\partial C(x, u \mid x_0, H, \hat{\mathcal{J}}) / \partial(x, u_1, \ldots, u_r)$ is nonsingular [Lue.1]. It follows from the Moore-Penrose conditions which characterize the pseudoinverse that

$$\begin{bmatrix} A & 0 \end{bmatrix}^{\dagger} = \begin{bmatrix} A^{-1} \\ 0 \end{bmatrix}, \tag{5.8.74c}$$

if $A$ is a square, nonsingular matrix. The result (b) then follows from the continuity of the matrix inverse at nonsingular matrices.

(c) It was shown in [Pol.4] that $x_1$, which was defined in (5.4b), is a continuous function of $x_0$, and that the compact-valued, set-valued map $U(\cdot)$ is an upper semicontinuous function of $x_0$. By inspection of (5.4.8) and by Theorem 5.3.1, $U(\hat{x}) = \hat{U}$. Hence, $\lim_{x_0 \to \hat{x}} U(x_0) = \{ \hat{\mu} \}$, where $\hat{\mu}$ is the unique Fritz-John multiplier vector associated with $\hat{x}$. Since $\hat{\mu}^0 > 0$ by Lemma 5.4.4(a), $\mu_1^0(x_0) > 0$ for $x_0$ near $\hat{x}$, and hence $u_1 = (\mu_1^1, \ldots, \mu_1^r) / \mu_1^0$ and $u_1$ is continuous for $x_0$ near $\hat{x}$. Since $\lim_{x_0 \to \hat{x}} \mu_1(x_0) = \hat{\mu}$, $\lim_{x_0 \to \hat{x}} u_1(\mu_1(x_0)) = \hat{u}$. Since $\mu_1(\hat{x}) = \hat{\mu}$, it follows from (5.4c) that $x_1(\hat{x}) = \hat{x}$. It follows from this and the continuity of $x_1$ that $\lim_{x_0 \to \hat{x}} x_1(x_0) = \hat{x}$.

(d) The system of equations

$$\begin{bmatrix} C(x, \mu \mid x_0, H, \hat{f}) \\ u^{r+1} \\ \cdot \\ \cdot \\ \cdot \\ u^p \end{bmatrix} = 0, \tag{5.8.75}$$

has a solution $(\hat{x}, \hat{\mu})$ for $x_0 = \hat{x}$ and $H = F(\hat{x})$. By Lemma 5.5.1(b), the gradient of the function $(C(\cdot \mid \hat{x}, F(\hat{x}), \hat{f})^T, u^{r+1}, \dots, u^p)^T$ is nonsingular at the point $(\hat{x}, \hat{u})$. Hence, there exists a neighborhood $W$ of $(\hat{x}, \hat{u})$ such that, for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, the system (5.8.75) has a unique solution in $W$, $(\tilde{x}, \tilde{u})$, which is continuous in $x_0$ and $H$.

Now $(x, u)$ is a Kuhn-Tucker pair for GQP$(x_0, H)$ if

$$\nabla \tilde{f}^0(x \mid x_0, H) + \sum_{j \in \mathbf{p}} u^j \nabla \tilde{f}^j(x \mid x_0, H) = 0, \tag{5.8.77}$$

$$\sum_{j \in \mathbf{p}} u^j \tilde{f}^j(x \mid x_0, H) = 0. \tag{5.8.78}$$

By Hypothesis 5.5.1, there exists a neighborhood $W$ of $\hat{x}$ such that $\tilde{f}^j(x \mid x_0, H) < 0$ for all $j \notin \hat{J}$, $x \in W$, $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$. Therefore, for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, $u^j = 0$ for $j \notin \hat{J}$ for every Kuhn-Tucker pair $(x, u) \in W \times \mathbb{R}_+^p$ of GQP$(x_0, H)$. Hence, for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, $(x, u)$ is a Kuhn-Tucker pair for GQP$(x_0, H)$ if and only if $(x, u)$ solves (5.8.75). This implies that, for $x_0$ near $\hat{x}$ and $H$ near $F(\hat{x})$, $(\tilde{x}, \tilde{u})$ is the unique Kuhn-Tucker pair of the problem GQP$(x_0, H)$ in the set $W \times \mathbb{R}_+^p$.

| Problem | Algorithm | NF | NDF | FV | KT |
|---|---|---|---|---|---|
| 12 | PT | 7 | 7 | -.30000000E+02 | .12E-06 |
| | GQP | 34 | 5 | -.30000000E+02 | .31E-15 |
| 29 | PT | 14 | 10 | -.22627417E+02 | .17E-06 |
| | GQP | 50 | 9 | -.22627417E+02 | .11E-11 |
| 30 | PT | 14 | 13 | .10000000E+01 | 0. |
| | GQP | 6 | 4 | .10000000E+01 | .21E-17 |
| 31 | PT | 11 | 8 | .60000000E+01 | .41E-06 |
| | GQP | 15 | 6 | .60000000E+01 | .35E-14 |
| 33 | PT | 4 | 4 | -.40000000E+01 | 0. |
| | GQP | 36 | 7 | -.40000000E+01 | .14E-14 |
| 34 | PT | 9 | 8 | -.83403245E+00 | .43E-04 |
| | GQP | 47 | 10 | -.83376811E-00 | .45E-07 |
| 43 | PT | 9 | 9 | -.44000000E+02 | .68E-04 |
| | GQP | 29 | 9 | -.44000000E+02 | .25E-10 |
| 57 | PT | 33 | 19 | .28459673E-01 | .20E-07 |
| | GQP | 52 | 20 | .30729013E-01 | .64E-04 |
| 66 | PT | 8 | 8 | .51816324E+00 | 0. |
| | GQP | 23 | 8 | .51816327E-00 | .35E-14 |
| 84 | PT | 4 | 4 | -.52803389E+07 | 0. |
| | GQP | 42 | 4 | -.23512434E+07 | .79E+06 |
| 100 | PT | 42 | 14 | .68063006E+03 | .21E-03 |
| | GQP | 110 | 14 | .68063005E+03 | .42E-06 |
| 113 | PT | 18 | 14 | .24306209E+02 | .17E-04 |
| | GQP | OVERFLOW | | | |
| 117 | PT | 28 | 16 | .3234679E+02 | .68E-04 |
| | GQP | OVERFLOW | | | |

**Table 5.1: Summary of Numerical Results**

The abbreviations in the table have the following meanings:

Problem: Number of the test problem in [Hoc.1].

Algorithm: PT denotes the algorithm in [Pan.1]. The results are quoted from that article. GQP denotes the algorithm described in Section 6 with $H_0 = I$.

NF: Number of objective function evaluations.

NDF: Number of gradient evaluations of the objective function.

FV: Value of the objective function at the final iterate.

KT: Norm of the Kuhn-Tucker vector at the final point. As defined in [Hoc.1], this is $\min_{u \geq 0} |\nabla fx^0(x_f) + \sum_{j \in \hat{J}} u^j \nabla fx^j(x_f)|$, where $x_f$ denotes the final point.

# CHAPTER 6
# CONCLUSIONS

## 6.1 CONTRIBUTIONS

We showed in Chapter 2 that PPP algorithms converge linearly even on problems which are nonconvex and on which strict complementary slackness does not hold. This weakening of assumptions is significant because few problems arising in practice are convex. Recall that, while singular matrices form a set of measure zero in $\mathbb{R}^{n \times n}$, nearly singular matrices are common and pose computational difficulties. Similarly, while problems for which strict complementarity does not hold form a "set of measure zero", many problems nearly violate strict complementarity. On such nearly degenerate problems, an optimization algorithm which is sensitive to degeneracy might converge slowly until very near to the solution. Our result shows that this will not occur with PPP algorithms.

PPP algorithms *are* sensitive to domain scaling, however. This issue was explored in Chapter 3, where we investigated the effect of a variable metric technique on the performance of a PPP algorithm on a class of composite minimax problems. Luenberger [Lue.1] advocates the evaluation of variable metric algorithms for differentiable, unconstrained optimization by considering how they affect the eigenvalue structure of the Hessian matrix. The rationale for this is that the condition number of the Hessian determines the convergence ratio of sequences constructed by the methods. We used this idea in Chapter 3 to evaluate the variable metric technique presented there. The use of the variable metric technique in conjunction with the PPP algorithm decreases the upper bound on the convergence ratio which we derived in Chapter 2. Experiments have shown that, while this bound is not tight, it is a reliable predictor of the *relative* speed of convergence on different problems. This provides a theoretical explanation for the performance improvement observed in numerical experiments.

The use of GQP subproblems in algorithms for solving problem (1.1.2b) was proposed in 1971 [Pol.10], but the algorithms were never implemented because no finite-step method for solving the subproblem was known. In Chapters 4 and 5, we presented efficient methods for approximately solving first-order and second-order GQP subproblems. Both are active set methods. The difficult inequality-constrained GQP subproblem is reduced to an easier equality-constrained problem by guessing the set of constraints which are active at the solution. The method by which we guess the active set - computing the Pironneau-Polak search direction - yields both a good point at which to initialize a root-finding method for the solution of the equality-constrained problem and a feasible descent direction which can be used in stabilizing the overall algorithm. In the first-order case, discussed in Chapter 4, the error in the approximation to the GQP solution is zero when the error in the approximation to the solution to (1.1.2b) is small. In the second-order case, discussed in Chapter 5, the error in the approximation to the GQP solution is of the order of the fourth power of the error in the approximation to the solution to (1.1.2b).

The convergence rate theory developed in Chapter 4 is another example of the usefulness of convergence ratio bounds for comparing linearly convergent algorithms. The convergence ratio bounds derived for the GQP-based algorithm are smaller than those obtained for the Pironneau-Polak algorithm. The GQP-based algorithms proved superior in numerical experiments as well.

The superlinear convergence rate theory developed in Chapter 5 is fairly general. The results apply to an entire class of algorithms. (Admittedly, proving that an algorithm is a member of the class requires some calculation.) A range of convergence rates from superlinear to 3/2 is obtained, depending upon the degree of accuracy of the Hessian approximations. In contrast, other superlinearly convergent feasible descent algorithms have been shown to converge only two-step superlinearly [Pan.1-2].[1]

---

[1]We do assume, however, that each of the Hessian estimates converges to the actual Hessian at the solution; this is stronger than the corresponding assumption made in [Pan.1-2].

## 6.2 FUTURE RESEARCH

Asymptotic convergence rate results, like the ones proved in this dissertation, provide a partial ordering of algorithms according to their overall performance. Even when far from a solution, a linearly convergent algorithm with a small convergence ratio bound can be expected to converge faster than a sublinearly convergent algorithm or a linearly convergent algorithm with a convergence ratio bound near one. Similarly, a superlinearly convergent algorithm can be expected to converge faster than a linearly convergent algorithm. However, there is a limit to the usefulness of asymptotic theories in predicting pre-asymptotic performance. Small differences in the convergence ratio bounds do not reliably indicate the relative overall speed of linearly convergent algorithms. A two-step superlinearly convergent algorithm may well be faster in the early phase of optimization than a quadratically convergent algorithm. As a result, other gauges of algorithm performance are needed.[2]

Some *non*-asymptotic efficiency results have been obtained. In the course of their complexity research, Nemirovsky and Yudin [Nem.1] have derived bounds on the total number of iterations needed by certain algorithms to reduce the cost-to-go, $f^0(x_i) - f^0(\hat{x})$, by a certain fraction. However, such results are difficult to obtain, and it is not clear how conservative they are.

We suggest the investigation of an alternative measure of pre-asymptotic performance: the rate of escape or *divergence* of an algorithm from the neighborhood of a nonminimum stationary point. Such situations are encountered in practice, and many iterations may be spent on such an escape. Furthermore, an algorithm's performance in this test situation may be indicative of its ability to pass quickly through a poorly scaled region of the domain. An algorithm's rate of divergence from nonminimum stationary points is not, in general, the same as its rate of convergence to minimizers. For example, Newton's method converges quadratically to both minimizers and maximizers. In addition, algorithms with the same asymptotic rate of convergence to

[2] Measures of algorithm performance like rate of convergence are not useful merely as means for ranking algorithms. The insight gained in proving that an algorithm attains a certain level of performance, for example, is superlinearly convergent, leads to the construction of better algorithms.

minimizers may have different rates of divergence from nonminimizers. The rate of divergence may be a better predictor of the relative behavior of algorithms during the most time-consuming portion of the computation, reaching the vicinity of the solution. The potential usefulness of this measure could be evaluated by numerically estimating the rates of divergence of well-known algorithms and correlating the results with their overall performances on, for example, the test problems in [Hoc.1]. Deriving the rate of divergence of an algorithm should be no more complex than deriving its rate of convergence; both tasks would rely heavily on the use of Taylor series expansions.

# CHAPTER 7

# REFERENCES


[Arm.1] L. Armijo, *Minimization of Functions Having Continuous Partial Derivatives*, Pacific Journal of Mathematics, Vol. 16, pp. 1-3, 1966

[Bak.1] T. Baker, *Algorithms for Optimal Control of Systems Described by Partial and Ordinary Differential Equations*, Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, University of California, Berkeley, 1988.

[Ber.1] C. Berge, *Topological Spaces*, Macmillan, New York, New York, 1963.

[Bre.1] R. P. Brent, *Some Efficient Algorithms for Solving Systems of Nonlinear Equations*, SIAM Journal on Numerical Analysis, Vol. 10, No. 2, pp. 327-344, 1974.

[Byr.1] R. Byrd, J. Nocedal and Y.-X. Yuan, *Global Convergence of a Class of Quasi-Newton Methods on Convex Problems*, SIAM Journal of Numerical Analysis, Vol. 24, No. 5, pp. 1171-1190, 1987.

[Can.1] J. Canny, *Generalized Characteristic Polynomials*, pp. 293-299, in Symbolic and Algebraic Computation ISSAC '88, P. Gianni, ed., Springer-Verlag, Berlin, 1989.

[Cha.1] R. W. Chaney, *On the Pironneau-Polak Method of Centers*, Journal of Optimization Theory and Applications, Vol. 20, No. 3, pp. 269-295, 1976.

[Cha.2] R. W. Chaney, *A Hybrid Algorithm for Nonlinear Programming*, SIAM Journal on Control and Optimization, Vol. 15, No. 3, pp. 384-399, 1977.

[Cla.1] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, New York, 1983.

[Col.1] T. F. Coleman and A. R. Conn, *Nonlinear Programming Via an Exact Penalty Function: Asymptotic Analysis*, Mathematical Programming, Vol. 24, pp. 123-136, 1982.

[Col.2] T. F. Coleman and A. R. Conn, *Nonlinear Programming Via an Exact Penalty Function: Global Analysis*, Mathematical Programming, Vol. 24, pp. 137-161, 1982.

[Con.1] A. R. Conn, *Constrained Optimization Using a Nondifferentiable Penalty Function*, SIAM Journal of Numerical Analysis, Vol. 10, No. 4, pp. 760-784, 1973.

[Con.2] A. R. Conn and T. Pietrzykowski, *A Penalty Function Method Converging Directly to a Constrained Optimum*, SIAM Journal of Numerical Analysis, Vol. 14, No. 2, pp. 348-375, 1977.

[Cuy.1] A. Cuyt, *Pade Approximants for Operators: Theory and Applications*, Berlin, Springer-Verlag, 1984.

[Dau.1] V. Daugavet, and V. Malozemov, *Quadratic Rate of Convergence of a Linearization Method for Solving Discrete Minimax Problems*, USSR Comput. Maths. Phys., Vol. 21, No. 4, pp. 19-28, 1981.

[Dem.1] V. Demyanov, *On the Solution of Certain Minimax Problems I, II*, Cybernetics, Vol. 2, No. 6, pp. 47-53, 1966, and Vol. 3, No. 3, pp. 51-54, 1967.

[Dem.2] V. Demyanov, *Algorithms for Some Minimax Problems*, Journal of Computer System Sciences, Vol. 2, No. 4, pp. 342-380, 1968.

[Dem.3] V. Demyanov and V. Malezemov, *Introduction to Minimax*, John Wiley and Sons, Inc., New York, 1974.

[Den.1] J. Dennis and J. More, *Quasi-Newton Methods: Motivation and Theory*, SIAM Review, Vol. 19, pp. 46-89, 1977.

[Des.1] C.A. Desoer, R.W. Liu, J. Murray, and R. Saeks "Feedback System Design: The Fractional Representation Approach to Analysis and Synthesis", *IEEE Transactions on Automatic Control*, Vol. AC-25, pp. 399-412, 1980.

[Dix.1] L. C. W. Dixon, *Variable Metric Algorithms: Necessary and Sufficient Conditions for Identical Behavior on Nonquadratic Functions*, Journal of Optimization Theory and Applications, Vol. 10, pp. 34-40, 1972.

[Fle.1] R. Fletcher, *Second Order Corrections for Nondifferentiable Optimization*, in Numerical Analysis, Dundee 1981, G. A. Watson, ed., Berlin: Springer-Verlag, 1982, pp.85-114.

[Fuk.1] M. Fukushima, *A Successive Quadratic Programming Algorithm with Global and Superlinear Convergence Properties*, Mathematical Programming, Vol. 35, pp. 253-264, 1986.

[Gab.1] D. Gabay, *Reduced Quasi-Newton Methods with Feasibility Improvement for Nonlinearly Constrained Optimization*, Mathematical Programming Study 16, pp. 18-44, 1982.

[Gil.1] P. Gill, J. Hammarling, W. Murray, M. Saunders and M. Wright, *User's Guide for LSSOL (Version 1.0): A Fortran Package for Constrained Linear Least-squares and Convex Quadratic Programming*, Technical report SOL 86-1, Department of Operations Research, Stanford University, Stanford, January 1986.

[Gol.1] A. A. Goldstein, *Constructive Real Analysis*, Harper and Row, New York, N. Y., 1967.

[Han.1] S.-P. Han, *Variable Metric Methods for Minimizing a Class of Nondifferentiable Functions*, Mathematical Programming, Vol. 20, pp. 1-13, 1981.

[Han.2] S.-P. Han and O. L. Mangasarian, *Exact Penalty Functions in Nonlinear Programming*, Mathematical Programming, Vol. 17, pp. 251-269, 1979.

[He.1] He, L., *Discretization Strategies for Semi-Infinite Optimization*, Dissertation, Electrical Engineering and Computer Science Department, University of California, Berkeley, 1990.

[Her.1] J. N. Herskovits, *A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization*, Mathematical Programming, Vol. 36, pp. 19-38, 1986.

[Hig.1] J. Higgins and E. Polak, *Minimizing Pseudo-Convex Functions on Convex Compact Sets*, Journal of Optimization Theory and Applications, Vol. 65, No. 1, pp. 1-27, 1990.

[Hoa.1] P. Hoang, *Computational Results on the Optimal Design of Feedback Compensators*, Masters Thesis, Department of Electrical Engineering and Computer Science, University of California, Berkeley, 1987.

[Hoc.1] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems 187, Springer-Verlag, Berlin, 1981.

[Hua.1] P. Huard, *Method of Centers*, Nonlinear Programming, J. Abadie, ed., North-Holland Publishing Company, Amsterdam, Holland, 1967.

[Hua.2] P. Huard, *Programmation Mathematique Convexe*, Revue Francaise d'Informatique et de Recherche Operationnelle, Vol. 7, pp. 43-59, 1968.

[Joh.1] F. John, *Extremum problems with inequalities as side conditions*, in Courant Anniversary Volume, K. O. Friedrichs, O. E. Neugebauer and J. J. Stoker, eds., Wiley-Interscience, New York, 1948, pp. 187-204.

[Kiw.1] K. Kiwiel, *A Phase I - Phase II Method for Inequality Constrained Minimax Problems*, Control and Cybernetics, Vol. 12, pp. 55-75, 1983.

[Kiw.2] K. Kiwiel, *A Method for Solving Certain Quadratic Programming Problems Arising in Nonsmooth Optimization*, IMA Journal of Numerical Analysis, Vol. 6, pp. 137-152, 1986.

[Kiw.3] K. Kiwiel, *A Dual Method for Solving Certain Positive Semi-Definite Quadratic Programming Problems*, SIAM Journal of Scientific and Statistical Computing, Vol. 10, pp. 175-186, 1989.

[Kob.1] H. Kobyashi, S. Moritsugu and R. Hogan, *Solving Systems of Algebraic Equations*, p. 139-149 in Symbolic and Algebraic Computation ISSAC '88, P. Gianni, ed., Springer-Verlag, Berlin, 1989.

[Lue.1] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Massachusetts, 1973.

[Mad.1] K. Madsen and H. Schjaer-Jacobsen, *Linearly Constrained Minimax Optimization*, Mathematical Programming, Vol. 14, pp. 208-223, 1978.

[Mar.1] N. Maratos, *Exact Penalty Function Algorithms for Finite Dimensional Optimization Problems*, Dissertation, Imperial College of Science and Technology, University of London, 1978.

[May.1] D. Q. Mayne and E. Polak, *A Superlinearly Convergent Algorithm for Constrained Optimization Problems*, Mathematical Programming Study 16, pp. 45-61, 1982.

[McC.1] G. McCormick, *Second Order Conditions for Constrained Minima*, SIAM Journal on Applied Mathematics, Vol. 15, pp. 641-652, 1967.

[Mey.1] G. G. L. Meyer, *An Efficient Method of Feasible Directions*, SIAM Journal on Control and Optimization, Vol. 21, No. 1, pp. 153-162, 1983.

[Mif.1] R. Mifflin, *Rates of Convergence for a Method of Centers Algorithm*, Journal of Optimization Theory and Applications, Vol. 18, No. 2, pp. 199-227, 1976.

[Mor.1] J. J. More and D. C. Sorensen, *Computing a Trust Region Step*, SIAM Journal of Scientific and Statistical Computing, Vol.4, No. 3, pp. 553-572, 1983.

[Mor.2] J. J. More, *Recent Developments in Algorithms and Software for Trust Region Methods*, Mathematical Programming: The State of the Art, Bonn, 1982, A. Bachem, M. Grotschel and B. Korte, eds., Berlin, Springer-Verlag, 1983.

[Muk.1] H. Mukai and E. Polak, *A Second Order Algorithm for Unconstrained Optimization*, Journal of Optimization Theory and Applications, Vol. 26, No. 4, 1978.

[Muk.2] H. Mukai and E. Polak, *A Second Order Algorithm for the General Nonlinear Programming problem*, Journal of Optimization Theory and Applications, Vol. 26, No. 4, 1978.

[Nem.1] A. Nemirovsky and D. Yudin, *Problem Complexity and Method Efficiency in Optimization*, John Wiley, New York, N. Y., 1979.

[Nye.1] W. T. Nye and A. L. Tits, *An Application-Oriented, Optimization-Based Methodology for Interactive Design of Engineering Systems*, International J. Control, Vol. 43, pp. 1693-1721, 1986.

[Ort.1] J. Ortega and W. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, New York: Academic Press, 1970. John Wiley, New York, N. Y., 1979.

[Pan.1] E. R. Panier and A. L. Tits, *A Superlinearly Convergent Feasible Method for the Solution of Inequality Constrained Optimization Problems*, SIAM Journal on Control and

Optimization, Vol. 25, No. 4, pp. 934-950, 1987.

[Pan.2] E. Panier, A. Tits and J. Herskovits, *A QP-Free, Globally Convergent, Locally Superlinearly Convergent Algorithm for Inequality Constrained Optimization*, SIAM Journal on Control and Optimization, Vol. 26, No. 4, pp. 788-811, 1988.

[Pan.3] V. M. Panin, *On a Second-Order Method for the Discrete Min-Max Problem*, USSR Computational Mathematics and Mathematical Physics, Vol. 19, No. 1, pp. 90-100, (Zh. vychisl. Mat. mat. Fiz., Vol. 19, No. 1, pp. 88-98, 1979).

[Pan.4] V. M. Panin, *Some Methods of Solving Convex Programming Problems*, USSR Computational Mathematics and Mathematical Physics, Vol. 21, No. 2, pp. 57-72, 1981, (Zh. vychisl. Mat. mat. Fiz., Vol. 21, No. 2, pp. 315-328, 1981).

[Pev.1] A. Pevnyi, *The Convergence Speed of Some Methods of Finding Minimaxes and Saddle Points*, Cybernetics, (A translation of Kibernetika) Vol. 8, No. 4, pp. 635-640, 1972.

[Pir.1] O. Pironneau and E. Polak, *On the Rate of Convergence of Certain Methods of Centers*, Mathematical Programming, Vol. 2, No. 2, pp. 230-258, 1972.

[Pol.1] B. Polyak, *Introduction to Optimization*, (in Russian), Nauka, Moscow (1983).

[Pol.2] E. Polak and E. Wiest, *Domain Rescaling Techniques for the Solution of Affinely Parametrized Nondifferentiable Optimal Design Problems*, Proceedings of the 27th IEEE Conference on Decision and Control, Austin, Texas, 1988.

[Pol.3] E. Polak, *Basics of Minimax Algorithms*, Proc., Summer School on Nonsmooth Optimization and Related Topics, Erice, Italy, June 19 - July 8 1988.

[Pol.4] E. Polak, *On the Mathematical Foundations of Nondifferentiable Optimization in Engineering Design*, SIAM Review, Vol. 29, No.1, pp. 21-91, March 1987.

[Pol.5] E. Polak and S. Salcudean, *On The Design of Linear Multivariable Feedback Systems via Constrained Nondifferentiable Optimization in $H^\infty$ Spaces*, IEEE Transactions on Automatic Control, Vol. 34, No.3, pp 268-276, 1989.

[Pol.6] E. Polak, J. Higgins and D. Q. Mayne, *A Barrier Function Method for Minimax Problems*, University of California, Berkeley, Electronics Research Laboratory Memo No. UCB/ERL M88/64, 20 October 1988.

[Pol.7] E. Polak, R. Trahan and D. Q. Mayne, *Combined Phase I - Phase II Methods of Feasible Directions*, Mathematical Programming, Vol. 17, No. 1, pp. 32-61, 1979.

[Pol.8] E. Polak and L. He, *A Unified Phase I Phase II Method of Feasible Directions for Semi-infinite Optimization*, University of California, Electronics Research Laboratory, Memo UCB/ERL M89/7, 3 February 1989.
To appear in *JOTA*.

[Pol.9] E. Polak, D. Q. Mayne and J. Higgins, *A Superlinearly Convergent Algorithm for Min-Max Problems*, University of California, Berkeley, Electronics Research Laboratory Memo No. M86/103, Nov. 15, 1986. Proc. 28th IEEE CDC, Tampa, Florida, Dec. 1989. Full length version to appear in JOTA.

[Pol.10] E. Polak, Computational Methods in Optimization: A Unified Approach, Academic Press, 329 pages, 1971.

[Pol.11] E. Polak, D. Q. Mayne, and J. Higgins, *On the Extension of Newton's Method to Semi-Infinite Minimax Problems*, University of California, Electronics Research Laboratory, Memo UCB/ERL M89/92, 1 August, 1989. Submitted to SICOPT.

[Pol.12] E. Polak, *A Modified Secant Method for Unconstrained Optimization*, Mathematical Programming, Vol. 6, No. 3, pp. 264-280, 1974.

[Pol.13] E. Polak and E. J. Wiest, *Variable Metric Technique for the Solution of Affinely Parametrized Nondifferentiable Optimal Design Problems*, Journal of Optimization Theory and Applications, Vol. 66, No. 3, pp. 391-411, 1990.

[Pot.1] F. A. Potra, *On the Q-Order and R-Order of Convergence*, Journal of Optimization Theory and Applications, Vol. 63, No. 3, pp. 415-432, 1989.

[Pow.1] M. J. D. Powell, *The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations*, in Nonlinear Programming 3, O. L. Mangasarian, R. R. Meyer and S. M. Robinson, eds., Academic Press, New York, 1978, pp. 27-63.

[Psh.1] Pshenichnyi, B. N. and Danilin, Yu. M., *Numerical Methods in Extremal Problems*, Nauka, Moscow, 1975.

[Psh.2] B. Pshenichnyi, *Algorithms for General Mathematical Programming Problems*, Cybernetics, Vol. 6, No. 5, pp. 677-684, 1970, (a translation of Kibernetika, No. 5, pp. 120-125, 1970.)

[Rin.1] A. H. G. Rinnooy Kan and G. T. Timmer, *Stochastic Global Optimization Methods Part I: Clustering Methods*, Mathematical Programming, Vol. 39, pp. 27-56, 1987.

[Rin.2] A. H. G. Rinnooy Kan and G. T. Timmer, *Stochastic Global Optimization Methods Part II: Multi Level Methods*, Mathematical Programming, Vol. 39, pp. 57-78, 1987.

[Rob.1] S. M. Robinson, *Generalized Equations and their Solutions, Part II: Applications to Nonlinear Programming*, Mathematical Programming Study 19, pp. 200-221, North-Holland Publishing Co., 1982.

[Rus.1] A. Ruszczynski, *A Regularized Decomposition Method for Minimizing a Sum of Polyhedral Functions*, Mathematical Programming, Vol. 35, pp. 46-61, 1986.

[Sal.1] S. E. Salcudean, "Algorithms for Optimal Design of Feedback Compensators", Ph.D. Thesis, Department of Electr. Engr. and Comp. Sci., University of California, Berkeley, California, December 1986.

[Sho.1] N. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.

[Top.1] D. M. Topkis and A. F. Veinott, *On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming*, SIAM Journal on Control and Optimization, Vol. 5, No. 2, pp. 268-279, 1967.

[Tra.1] J. F. Traub, *Iterative Methods for the Solution of Equations*, Englewood Cliffs, Prentic-Hall, 1964.

[von.1] B. von Hohenbalken, *Simplical Decomposition in Nonlinear Programming Algorithms*, Mathematical Programming, Vol. 13, pp. 49-68, 1977

[Wie.1] E. J. Wiest and E. Polak, *On the Rate of Convergence of Two Minimax Algorithms*, University of California, Electronics Research Laboratory, Memo UCB/ERL M89/111, 15 August 1989. To appear in JOTA.

[Wie.2] E. J. Wiest and E. Polak, *A Phase I - Phase II Method for Inequality-Constrained Optimization Based on Generalized Quadratic Programs*, University of California, Electronics Research Laboratory, Memo UCB/ERL M90/46, 24 May 1990. Submitted to Applied Mathematics and Optimization.

[Wil.1] R. B. Wilson, *A Simplified Algorithm for Concave Programming*, Dissertation, Harvard University, Cambridge, Massachusetts, 1963.

[Wom.1] Womersley, R. S., *Optimality Conditions for Piecewise Smooth Functions*, Mathematical Programming Study 17, pp. 13-27, 1982.

[Wuu.1] Tzyh-Lih Wuu, *DELIGHT.MIMO: An Interactive System for Optimization-Based Multivariable Control System Design*, Dissertation, Department of Electrical Engineering and Computer Science, University of California, Berkeley, 1986.

[You.1] D.C. Youla, H. Jabr and J.J. Bongiorno, Jr., "Modern Wiener-Hopf Design of Optimal Controllers - Part II", *IEEE Transactions on Automatic Control*, Vol. AC-21, pp. 319-338, 1967.

[Zou.1] G. Zoutendijk, *Mathematical Programming Methods*, North Holland, Amsterdam, 1976.