ON THE MATHEMATICAL FOUNDATIONS OF NONDIFFERENTIABLE OPTIMIZATION IN ENGINEERING DESIGN

by

E. Polak

Memorandum No. UCB/ERL M85/17

28 February 1985
(Revised - October 1985)

ON THE MATHEMATICAL FOUNDATIONS OF NONDIFFERENTIABLE OPTIMIZATION IN ENGINEERING DESIGN

by

E. Polak

Memorandum No. UCB/ERL M85/17

28 February 1985
(Revised - October 1985)

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE MATHEMATICAL FOUNDATIONS OF NONDIFFERENTIABLE OPTIMIZATION IN ENGINEERING DESIGN.

by

E. Polak

Department of Electrical Engineering
and Computer Sciences
University of California
Berkeley, Ca. 94708

ABSTRACT

It is shown by example that a large class of engineering design problems can be transcribed into the form of a canonical optimization problem with inequality constraints involving max functions. Such problems are commonly referred to as semi-infinite optimization problems. The bulk of this paper is devoted to the development of a mathematical theory for the construction of first order nondifferentiable optimization algorithms, related to phase I - phase II methods of feasible directions, which solve these semi-infinite optimization problems. The applicability of the theory is illustrated with examples that are relevant to engineering design.

0. INTRODUCTION

0.1. Evolution of Optimization-Based Engineering Design

Over the years, engineering design has been increasing in complexity. This constant growth in complexity is due to several factors, such as, (i) progressively increasing expectations in product performance, (ii) progressively more restrictive constraints imposed by environmental and resource cost considerations, and (iii) progressively more and more ambitious projects being launched.

For example, in structural engineering, the increase in design complexity is due to the need to ensure the earthquake survivability of sky scrapers and nuclear reactors at reasonable cost; in control engineering and electronics to the need for reliable, high performance, worst case designs; in the automotive world, to the need to conserve energy while eliminating pollution; and in the area of space exploration, to attempts to design complex shaped, highly flexible, large space structures and their control systems simultaneously, to unprecedented performance standards.

Fortunately, over the last decade, while material and labor costs have grown rapidly, computing costs have decreased dramatically and hence, not surprisingly, engineers have been turning more and more frequently to the computer for assistance in design. As a result, a new, interdisciplinary engineering specialty has emerged which is commonly referred to as computer-aided design (CAD). Most of the existing CAD methodology is based on computer-aided analysis, with the design parameter selection carried out by the designer on a trial and error basis. Since decision making in a multiparameter space is very difficult, the trial and error approach is not very effective. Therefore, there is growing hope that considerable benefits in engineering design might be obtained from the use of sophisticated optimization tools. However, the effective use of optimization algorithms in engineering design is predicated on the supposition that engineering design problems are transcribable into a suitable canonical optimization problem.

Fortunately, as we shall shortly illustrate by example, engineering design specifications can frequently be expressed as inequalities in terms of a finite dimensional *design vector* $x \in \mathbb{R}^n$. These inequalities are either of the form

$$g(x) \leq 0 \tag{0.1.1}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, or of the form

$$\varphi(x,y) \leq 0, \ \forall \, y \in Y, \tag{0.1.2a}$$

or, equivalently

$$\max_{y \in Y} \varphi(x,y) \leq 0, \tag{0.1.2b}$$

where $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ is locally Lipschitz continuous and $Y \subset \mathbb{R}^p$ is compact. Constraints of the form (0.1.1) often express simple bounds on the design variable or a "static" design condition. Constraints of the form (0.1.2b) can be used to express bounds on time and frequency responses of a dynamical system as well as tolerancing or uncertainty conditions in worst case design. Consequently, a rather large number of engineering design problems are transcribable into the following *canonical optimization problem:*

$$\min\{ f(x) \,|\, g^i(x) \leq 0, i \in \underline{k}; \varphi^j(x,y_j) \leq 0, y_j \in Y_j, j \in \underline{m} \} \tag{0.1.3}$$

where we use the notation $\underline{k} \triangleq \{1,2,...,k\}$, for any positive integer k. At a minimum, the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g^i : \mathbb{R}^n \to \mathbb{R}$, $i \in \underline{k}$ and $\varphi^j : \mathbb{R}^n \times \mathbb{R}^{p_j} \to \mathbb{R}$, $j \in \underline{m}$, must be assumed to be locally Lipschitz continuous, while the sets $Y_j \subset \mathbb{R}^{p_j}$ must be assumed to be compact.

Occasionally one encounters equality constraints as well, in engineering design. These can be removed by means of exact penalty function techniques which we will not discuss in this paper. For examples of exact penalty functions in conjunction with algorithms of the type that we will describe in this paper, the reader is referred to [May.2, May.4].

Problems of the form (0.1.1) are often referred to as *semi-infinite* optimization problems, or SIP for short, because the *design vector* x is finite dimensional, while the number of constraints is infinite.

A number of optimal control problems with state space constraints also have the formal form of (0.1.3), except that the design vector x is a control (in $L_\infty^m[0,1]$, say) rather than a finite dimensional vector. Although the theory that we will present will be entirely in terms of problems in which the design vector x is finite dimensional, it is very easy to extend the algorithms that we will be presenting, both formally and analytically, to the case where x is a control. In Example 5.34 we shall illustrate this fact.

0.2. Factors in the Selection of an the Axiomatic Structure

Problem (0.1.3) has a great deal of structure. The effect of this structure is particularly pronounced when, as is so often the case in engineering design, the functions $f(\cdot)$, $g^i(\cdot)$ and $\varphi^j(\cdot,\cdot)$ are differentiable. In this case, (0.1.3) is a *differentiable* optimization problem with an infinite number of constraints. Hence (0.1.3) is best solved by algorithms which exploit its structure to the limit (such as [Gon.1, Pol.10]), rather than by general purpose nondifferentiable optimization algorithms (such as [Lem.1, Lem.2, Lem.3], [Mif.1], [Kiw.1] [Sho.1]). Algorithms which exploit the structure of (0.1.3) can be first order or higher order. First order algorithms tend to be extensions of methods of feasible directions (see e.g. [Gon.1, Kiw.1, Pol.10]), while higher order methods are extensions of Newton's method or sequential quadratic programming methods (see e.g., [Het.2, May.7, Pol.11]).

In this paper we present an axiomatic approach to a class of nondifferentiable optimization algorithms for solving problem (0.1.3). These algorithms can be viewed as extensions of combined phase I - phase II, methods of feasible directions for differentiable problems, introduced in [Pol.6]. Because the literature contains only a few examples of methods of feasible directions (see, e.g., [Ben.1, Pol.1, Zou.1]), it is, generally, not realized that it is possible to define a very large number of such methods. We shall now demonstrate how one can generate whole families of methods of feasible directions.

Consider the problem

$$\min\{ f^0(x) \mid f^j(x) \le 0,\ j = 1,2,\ldots,m \}, \tag{0.2.1}$$

where, for $j = 0,1,2,..,m$, $f^j : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable, and suppose that $\hat{x}$ is a local minimizer for (0.2.1). Let us adopt the notation that vectors $\mu \in \mathbb{R}^{m+1}$ have components $\mu^0, \mu^1, \ldots, \mu^m$, and let us define $\Sigma_{m+1} \triangleq \{\mu \in \mathbb{R}^{m+1} \mid \mu^j \ge 0 \text{ for } j = 1,2,\ldots,m,\ \sum_{j=0}^{m} \mu^j = 1\}$. Finally, we define $J(\hat{x}) \triangleq \{j \in \mathbf{m} \mid f^j(\hat{x}) = 0\} \cup \{0\}$. Then one can state the F. John condition of optimality [Joh.1] in two ways:

$$0 \in \text{co}\{\nabla f^j(\hat{x})\}_{j \in J(\hat{x})}, \tag{0.2.2}$$

where "co" denotes the convex hull of the set, or, equivalently, that for some $\hat{\mu} \in \Sigma_{m+1}$, the following two equations must be satisfied:

$$\sum_{j=0}^{m} \hat{\mu}^j \nabla f^j(\hat{x}) = 0, \tag{0.2.3a}$$

$$\sum_{j=1}^{m} \hat{\mu}^j f^j(\hat{x}) = 0. \tag{0.2.3b}$$

To define a family of methods of feasible directions, we derive from either (0.2.2) or (0.2.3a,b) a family of search direction finding problems. For $i = 1,2,\ldots,m$ and $j = 0,1,2,\ldots,m$, let $s^i, t^j : \mathbb{R} \to \mathbb{R}$ be continuous functions such that (i) $s(z) = 0$ if and only if $z = 0$, $sgn\, s^i(z) = sgn\, z$, (ii) $t^j(z) > 0$ for all $z \ne 0$. Next, let B be any compact set in $\mathbb{R}^n$ containing the origin in its interior, and let $\alpha,\beta > 0$ be arbitrary. Now consider the functions

$$\Theta^1 \triangleq \min_{h \in B} \max_{j \in J(\hat{x})} t^j(\|\nabla f^j(x)\|)\langle \nabla f^j(x), h \rangle, \tag{0.2.4a}$$

$$\Theta^2 \triangleq \min_{h \in \mathbb{R}^n} \max_{j \in J(\hat{x})} \{ \tfrac{1}{2}\|h\|^2 + s^j(f^j(x)) + t^j(\|\nabla f^j(x)\|)\langle \nabla f^j(x), h \rangle \}, \tag{0.2.4b}$$

$$\Theta^3 \triangleq \min_{\mu \in \Sigma_{m+1}} \{ \alpha \left[-\sum_{j=1}^{m} \mu^j s^j(f^j(x)) \right]^k + \beta \left\| \left[\sum_{j=0}^{m} \mu^j t^j(\|\nabla f^j(x)\|)\nabla f^j(x) \right] \right\|^2 \} \tag{0.2.4c}$$

for $k = 1$ or $k = 2$. It is easy to see that if $\hat{x}$ is such that $f^j(\hat{x}) \le 0$ for all $j \in \mathbf{m}$, then (0.2.2) (and hence also (0.2.3a,b)) holds if and only if $\Theta^k(\hat{x}) = 0$, where $k = 1,2,3$. We have thus created broad families of equivalent first order optimality conditions. Now, if $x \in \mathbb{R}^n$ is such that $f^j(x) \le 0$ for all $j \in \mathbf{m}$ and $\Theta^1(x) < 0$ (so that $\Theta^k(x) \ne 0$ for $k = 2,3$ also holds), then the solution vector $h(x)$ of (0.2.4a) (or of (0.2.4b)) defines a direction along which the cost can be reduced without constraint violation. The same holds true for the vector $h(x) \triangleq -\sum_{j=0}^{m} \mu_x^j t^j(\|\nabla f^j(x)\|)\nabla f^j(x)$, where μ_x is a solution of of (0.2.4c). We are thus on the way to obtaining several infinite families of methods of feasible directions. In the literature one finds methods related to (0.2.4a) for $B = \{h \in \mathbb{R}^n \mid \|h\|_\infty \le 1\}$ (see [Pol.1, Zou.1]), methods related to (0.2.4b) (see [Kiw.1]) and methods using (0.2.4c) with $k = 1$, but not with $k = 2$ (see [Pir.2, Pol.1]).

In constructing an axiomatic theory of semi-infinite optimization algorithms, we were faced with the choice of whether to place emphasis on elegance and simplicity or whether to make sure that as many of the published algorithms and possible variants of these, such as those sketched out above, could be accounted for. We have opted to emphasize axiomatic simplicity and have concentrated on algorithms which can be viewed as progressively more complex extensions of the method of steepest descent. These are related to those defined by (0.2.4b) and (0.2.4c) with $k = 2$.

A number of the algorithms stated in this paper have not been published before, because, whenever we could not find in the literature an appropriate algorithm that satisfied our axioms exactly, we made use of the above explained freedom to generate a new algorithm, by creating a variant of an existing algorithm. Usually this involved no more than replacing $k = 1$ by $k = 2$ in a form such as (0.2.4c). On the basis of experience, we expect that the computational behavior of these new variants will be found to be quite similar to that of the original algorithms.

0.3. Comments on Contents and Bibliographical Notes

This paper is reasonably self contained. In Section 1, we present four examples of the transcription of an engineering design problem into an optimization problem of the form (0.1.3). In Section 2, we present the mathematical results in the areas of continuity of point-to-point and point-to-set functions, convexity of functions and sets, and nonsmooth analysis which are essential to the understanding of our work. In Section 3, we assemble a number of results on max functions. In Section 4, we give first order conditions of optimality for problem (0.1.3). In Section 5, we show that it is possible to evolve a family of algorithms, for the solution of both unconstrained and constrained semi-infinite optimization problems, by extension of the method of steepest descent for differentiable functions. In Section 6, we present a family of more sophisticated algorithms, derived by extension of the algorithms in Section 5. These more sophisticated algorithms are characterized by search direction subprocedures that require less computational effort than the ones used by the algorithms presented in Section 5. As might be expected, the savings in search direction computation are paid for by the loss of some continuity in the search directions, which is likely to increase the number of iterations required to solve a problem to a given precision. However, there is some consensus in the community of feasible directions code users that the savings do seem to outway the losses. Finally, in Section 7, we introduce the reader to the concept of algorithm implementation and illustrate it by an example.

Our list of references includes citations to articles and books which are of immediate relevance to the material presented, as well as to some articles and books which may help the reader to see our results from a broader perspective. We conclude this section with a brief discussion of our sources and bibliography.

For further examples of engineering applications, see [Bha.1, May.5, May.6, Pol.2, Pol.4, Pol.7, Pol.10, Pol.15, War.1].

Our sources for nonsmooth analysis, optimality conditions and max functions were [Ber.1, Cla.1, Dan.1, Dem.1, Dem.2, Leb.1].

Our axioms for the convergent direction finding maps represent the culmination of a considerable effort on the theory of nondifferentiable optimization algorithms. For earlier results see [Pol.12, Pol.16, Pol.17]. Our convergence proofs are based on general convergence theorems. For a sampling of such theorems, see [Pol.1, Pol.9, Pol.16, Tish.1].

The algorithm implementation theory presented in Section 7 is based on results in [Kle.1, Muk.1, Pol.1, Pol.8, Tra.1].

For further examples of nonlinear and semi-infinite programming algorithms which have contributed to our development of the theory in this paper, see [Bert.1, Gon.1, May.4, May.8, Oet.1, Pir.1, Pir.2, Pol.1, Pol.3, Pol.6, Pol.10, Tra.1, War.1].

For optimal control algorithms which can be understood in terms of the algorithm models presented in this paper, see [May.2, May.3, Warg.1, Wil.1].

The possible variety of "direct" alternatives for constructing nondifferentiable optimization algorithms is illustrated by [Het.1, Het.3, Kiw.1, Lem.2, Lem.3, May.9, Mif.1, Mif.3, Polj.1, Polj.2, Polj.3, Sho.1].

There is a good deal of work on curve fitting which is closely related to that done in the area of semi-infinite optimization, see, e.g., [Opf.1]. We refer the reader to the book by Hettich and Zencke [Het.3] for a bibliography of this work and some discussion.

An "indirect" alternative for solving semi-infinite optimization problems is presented by various versions of problem decomposition by means of outer approximations, see, e.g., [Gon.2, May.1, May.5, May.8].

For a few examples of superlinearly converging nondifferentiable optimization algorithms, which will not be covered in this paper see [Het.2, Lem.1, Lem.4, Lem.5, May.7, Mif.3, Pol.11].

Finally, we would like to draw the reader's attention to the following major works: the book by Clarke [Cla.1] which presents a comprehensive theory of nonsmooth analysis and optimality conditions for nondifferentiable optimization problems, the doctoral dissertation by Lemarechal [Lem.3], which presents an interesting collection of algorithms for the solution of nondifferentiable, convex problems, the book by Kiwiel [Kiw.1] which is a compendium of various nondifferentiable optimization algorithms, and the book by Shor [Sho.1] which presents some very original alternatives, including space dilation methods which have subsequently led to the so called ellipsoid methods (see, e.g.,[Sho.4, Iud.1, Blan.1]) including the famous Khachian algorithm [Kha.1]. For additional bibliographic information, the reader may consult [Nur.2].

1. DESIGN EXAMPLES

We shall now illustrate by means of a few simple examples how SIP problems of the form (0.1.3) arise in a variety of engineering design situations.

1.1. Design of Earthquake Resistant Structures

One of the simplest examples of a problem of the form (0.1.3) is found in the design of braced frame buildings which are expected to withstand small earthquakes with no damage and large ones with repairable damage. A simple three story braced frame is shown in Fig 0.1.1. The components of the design vector x are the stiffnesses of the frame members, as indicated in Fig.0.1.1. Under the hypotheses of a lumped parameter model, the horizontal floors and roof are assumed to be rigid and to concentrate the mass of the structure. The relative displacements of the three floors and roof form the components of the displacement vector y. The lumped parameter model of the braced frame obeys a second order vector differential equation of the form:

$$M\ddot{y}(t,x) + D(y,\dot{y},x)\,\dot{y}(t,x) + K(y,\dot{y},x)\,y(t,x) = F(t), \qquad (1.1.1)$$

where $F(t)$ represents the seismic forces. When F is small, i.e., when the earthquake is small D and K can be taken to be constant so that (1.1.1) is a linear differential equation, but when F is large, the bending of steel introduces gross nonlinearities due to its hysteretic behavior. It is common to consider a whole family of earthquakes $\{F_k\}_{k \in K}$, both large and small, in carrying out a design. When an earthquake is small, a building is expected to remain elastic and no structural damage is allowed. When an earthquake is large, survival of occupants becomes a major consideration and large, energy absorbing, non eleastic deformations are accepted, short of outright failure of the structure. A simple optimal design problem consists of minimizing the weight of the structure subject to bounds on the relative floor displacements over the entire duration of the family of earthquakes considered as well as simple bounds on the stiffness of the structural members. This leads to a SIP of the form

$$\min\{f(x) \mid 0 < \alpha \le x^i \le \beta, \ \forall\, i \in \underline{n}; \\ |y^{j+1}(t,x,F_k) - y^j(t,x,F_k)| \le d_k^j, \\ \forall\, t \in [0,T], \ \forall\, k \in K, \ j = 0,1,2\}. \tag{1.1.2}$$

1.2. Design of a MIMO Control System

We shall now consider a simple design of a multi-input multi-output (MIMO) control system, with specifications both in time and frequency domains. Consider the feedback configuration in Fig. 1.2.1, where $C(x,s)$ is a compensator transfer function matrix that needs to be designed. The equations governing the behavior of this system in the time domain are of the form

$$\dot{z}_p = A_p z_p + B_p u_p \tag{1.2.1a}$$

$$y_p = C_p z_p \tag{1.2.1b}$$

$$\dot{z}_c = A_c(x) z_c + B_c(x) u_c \tag{1.2.2a}$$

$$y_c = C_c(x) z_c \tag{1.2.2b}$$

$$u_p = y_c \tag{1.2.3a}$$

$$u_c = r - y \tag{1.2.3b}$$

$$y = y_p + d \tag{1.2.3c}$$

where (1.2.1a,b) represents the plant, (1.2.2a,b) represents the compensator to be designed and (1.2.3a-c) are the interconnection relations. We assume that r, u_p, u_c, y_p, y_c are all m-dimensional vectors and that the matrices A_c, B_c, C_c are continuously differentiable in the design vector x which, most likely, consists of the "free" elements of these matrices.

The most elementary requirement is that of closed loop stability. With

$$G_p(s) = C_p(sI - A_c)^{-1}B_p\,, \tag{1.2.4a}$$

$$G_c(x,s) = C_c(x)(sI - A_c(x))^{-1}B_c(x)\,, \tag{1.2.4b}$$

it can be shown that the eigenvalues of the closed loop system are the zeros of the polynomial in s

$$\chi(x,s) \triangleq \det(sI-A_p)\det(sI-A_c(x))\det(I+G_p(s)G_c(x,s)). \qquad (1.2.5)$$

To ensure that the zeros of $\chi(x,s)$ are all in the open left half plane, we make use of the modified Nyquist stability test introduced in [Pol.14]. For this purpose, let $d(s)$ be a monic polynomial of the same degree as $\chi(s)$, such that all zeros of $d(s)$ are in the open left half plane. Let $T(x,s) \triangleq \chi(x,s)/d(s)$. The closed loop system is stable if the locus of $T(x,j\omega)$, traced out in the complex plane for $\omega \in (-\infty,\infty)$, does not pass through or encircle the origin. A sufficient condition for ensuring this (see [Pol.14]) consists of keeping the locus of $T(x,j\omega)$ out of a parabolic region containing the origin (see Fig. 1.2.2) by imposing the semi-infinite inequality:

$$-d\,Re[T(x,j\omega)]^2 + Im[T(x,j\omega)] + c \le 0 \;\forall\, \omega \ge 0. \qquad (1.2.6)$$

where $c, d > 0$.

Next, for a set of specified inputs $\{r_k(\cdot)\}_{k\in K}$, the designer may require that the zero initial conditions response error be limited as follows (see Fig. 1.2.3):

$$\underline{b}_k^i(t) \le y_p^i(t;x,r_k) - r_k^i(t) \le \overline{b}_k^i(t) \qquad (1.2.7)$$

for all $k \in K$ and $i = 1,2,\dots,m$, with the $\underline{b}_k^i$, $\overline{b}_k^i$ piecewise continuous functions.

Finally, for the purpose of expressing insensitivity to the disturbance d, we set $r = 0$, which leads to the Laplace transform equation

$$\hat{y}(s) = [I + P(s)\,C(x,s)]^{-1}\,\hat{d}(s) \triangleq Q(x,s)\,\hat{d}(s) \qquad (1.2.8a)$$

$$\hat{u}_p(s) = -C(x,s)\,Q(x,s)\,\hat{d}(s) \triangleq R(x,s)\,\hat{d}(s)\;, \qquad (1.2.8b)$$

where $\hat{u}_p(s)$, $\hat{d}(s)$, $\hat{y}(s)$ denote the Laplace transforms of $u_p(t)$, $d(t)$, $y(t)$, respectively.

Let $\overline{\sigma}H$ denote the largest singular value of a complex $m\times m$ matrix H. Since the largest singular value of a matrix is its induced L_2 norm, to make the response y of the system small for a large class of disturbances d, without unduly saturating the system as a result of u becoming too large, control system designers strive to keep $\bar{\sigma}[Q(x,j\omega)]$ small and $\bar{\sigma}[R(s,j\omega)]$ bounded over the frequency range $[\omega', \omega'']$ in which the energy of the disturbances is known to be concentrated. This leads to the following formulation of the MIMO control system design problem:

minimize $f(x)$,

where

$$f(x) \triangleq \max\{\bar{\sigma}[Q(x,j\omega)] \mid \omega \in [\omega',\omega'']\} \tag{1.2.9}$$

subject to (1.2.6), (1.2.7) and

$$\bar{\sigma}[R(x,j\omega)] \le b(\omega), \ \forall \ \omega \in [\omega',\omega''], \tag{1.2.10}$$

$$\underline{x}^i \le x^i \le \bar{x}^i , \tag{1.2.11}$$

where $b(\omega)$ is a continuous, real valued function.

In addition, there could be constraints expressing *decoupling* i.e., the requirement that when only a single component of the input vector is a nonzero function, only the corresponding component of the output vector is nonzero, as well as stability robustness requirements, all of which are semi-infinite in form. We note that from an algorithmic point of view, since singular values are nondifferentiable, the optimization problem corresponding to MIMO control system design is considerably more difficult than the one corresponding to structural design.

1.3. Design of a Wide Band Amplifier

The design of a wide band amplifier usually involves three transfer functions: the input impedance $Z_{in}(x,s)$, the output impedance, $Z_{out}(x,s)$ and the gain, $A(x,s)$, which are all proper rational functions in the complex variable s.

The design vector $x \in \mathbb{R}^n$ determines certain critical component values (e.g., resistor, capacitor values) in the circuit, which affect the impedances and the gain. Thus, the coefficients of the rational functions Z_{in}, Z_{out} and A are functions of the design vector x.

The simplest formulation of a wide band amplifier design has the form

$$\begin{aligned} \max_{(x,\omega_f)} \{\omega_f \mid \underline{b}_{in} \le\ & |Z_{in}(x,j\omega)|^2 \le \bar{b}_{in},\ \forall\, \omega \in [\omega_0, \omega_f]\ ; \\ \underline{b}_{out} \le\ & |Z_{out}(x,j\omega)|^2 \le \bar{b}_{out},\ \forall\, \omega \in [\omega_0, \omega_f]\ ; \\ \underline{A} \le\ & |A(s,j\omega)|^2 \le \bar{A},\ \forall\, \omega \in [\omega_0, \omega_f]; \\ \underline{x}^i \le\ & x^i \le \bar{x}^i\ ,\ i = 1,2,\ldots,n\}. \end{aligned} \tag{1.3.1a}$$

As stated, this problem is not quite of the form (0.1.3). To bring it in line with the canonical form (0.1.3), we augment the design variable by one component, x^0, to $\bar{x} = (x^0, x) \in \mathbb{R}^{n+1}$. Problem (1.3.1a) can then seen to be equivalent to the problem:

$$\begin{aligned} \min_{x} \{-x^0 \mid \underline{b}_{in} \le\ & |Z_{in}(x,j(\omega_0+yx^0))|^2 \le \bar{b}_{in},\ \forall\, y \in [0,1]; \\ \underline{b}_{out} \le\ & |Z_{out}(x,j(\omega_0+yx^0))|^2 \le \bar{b}_{out},\ \forall\, y \in [0,1]; \\ \underline{A} \le\ & |A(x,j(\omega_0+yx^0))|^2 \le \bar{A},\ \forall\, y \in [0,1]; \\ \underline{x}^i \le\ & x^i \le \bar{x}^i,\ i = 1,2,\ldots,n\}. \end{aligned} \tag{1.3.1b}$$

1.4. Robot Arm Path Planning

In designing a sequence of moves to be carried out by a robot manipulator in a manufacturing situation, it is necessary to find a number of paths which take the robot arm from one location to another without collision with the workpiece. We shall describe a simple problem involving a two link robot manipulator and a circular workpiece obstacle in $\mathbb{R}^2$. Let $\vartheta^1(t)$, $\vartheta^2(t)$ be the angles at time t between reference rays and the robot links (see Fig. 1.4.1), and let $\vartheta(t) \triangleq (\vartheta^1(t), \vartheta^2(t))$. Then the dynamics of the robot have the form

$$M(\vartheta(t))\ddot{\vartheta}(t) = \tau(t) - C(\vartheta(t),\dot{\vartheta}(t))\dot{\vartheta}(t) + G(\vartheta(t)) \tag{1.4.1}$$

where $M(\cdot)$ and $C(\cdot,\cdot)$ are 2×2 continuously differentiable matrices, and

$G:\mathbb{R}^2 \to \mathbb{R}^2$ is continuously differentiable and $\tau(t) \in \mathbb{R}_2$ is a torque vector, with $\tau^1(t)$ the torque applied at the first joint and $\tau^2(t)$ the torque applied at the second joint. The circular workpiece is described by an inequality of the form

$$h(x) \le 0 , \tag{1.4.2}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$h(x) = 1 - (x^1 - a)^2 - (x^2 - b)^2 . \tag{1.4.3}$$

for some $a, b \in \mathbb{R}$.

Now suppose that we are given that at t = 0, the angles are $\vartheta^1(0) = \vartheta_0^1$, $\vartheta_2(0) = \vartheta_0^2$, and that we are supposed to find a torque vector $\tau(t)$, $t \in [0,1]$, which results in a collision free path that takes the robot manipulator from these initial angles to the angles $\vartheta_1(1) = \vartheta_f^1$, $\vartheta^2(1) = \vartheta_f^2$ at time t = 1, with $|\tau^j(t)| \le c$, j = 1,2, for $t \in [0,1]$. We assume that $\tau(\cdot)$ is an $L_\infty^2[0,1]$ function.

Let us denote the solution of (1.4.1), which satisfies the initial condition $\vartheta(0) = \vartheta_0$, and which corresponds to the torque $\tau(\cdot)$ by $\vartheta^\tau(\cdot)$. We can now express our problem in the form

$$\min\{f(\tau) \mid g^j(\tau) \le 0 , j = 1,2; \varphi^k(\tau,y) \le 0, k = 1,2, \ \forall \ y \in Y\} \tag{1.4.4a}$$

where $f : L_\infty^2[0,1] \to \mathbb{R}$ is defined by

$$f(\tau) \triangleq \|\vartheta^\tau(1) - \vartheta_f\|^2 ; \tag{1.4.4b}$$

the $g^j : L_\infty^2[0,1] \to \mathbb{R}$, j = 1,2 are defined by

$$g_1(\tau) \triangleq \max_{t \in [0,1]} |\tau(t)| - c ; \tag{1.4.4c}$$

$$g_2(\tau) \triangleq \max_{t \in [0,1]} |\tau^2(t)| - c ; \tag{1.4.4d}$$

$Y = [0,1] \times [0,1] \subset \mathbb{R}^2$ and, for $k = 1,2$, and $y \triangleq (s,t)$, $\varphi^k : L_\infty^2[0,1] \times \mathbb{R}^2 \to \mathbb{R}$ are defined, by

$$\varphi^1(\tau,y) \triangleq h(sl_1\cos \vartheta^{1\tau}(t), sl_1\cos \vartheta^{2\tau}(t)) \tag{1.4.4e}$$

$$\varphi^2(\tau,y) \triangleq h((l_1\cos \vartheta^{1\tau}(t) + sl_2\cos(\vartheta^{1\tau}(t)+\pi-\vartheta^{2\tau}(t)), \; l_1\sin \vartheta^{1\tau}(t) + sl_2\sin(\vartheta^{1\tau}(t) + \pi-\vartheta^{2\tau}(t)) \tag{1.4.4f}$$

where l_1 is the length of the first link and l_2 is the length of the second link. The function $\varphi^1(\cdot,\cdot)$ is used to ensure that the *entire first link* will avoid collision with the workpiece, while the functon $\varphi^2(\cdot,\cdot)$ is used to insure that the entire second link will avoid collision with the workpiece. As stated, the design vector $\tau(\cdot)$ is a function. The problem can be made finite dimensional by representing $\tau(\cdot)$ in terms of splines, say, over a fixed set of nodes.

2. PRELIMINARY RESULTS: CONTINUITY, CONVEXITY AND NONSMOOTH ANALYSIS

We shall now summarize the various results in the theory of point-to-set maps, convexity, and nonsmooth analysis that we shall make constant use of in this paper. In the process we shall also establish the mathematical notation that we shall use.

The book by Berge [Ber.1] is an excellent reference for various point-to-set map results used in optimization. Unfortunately, it is out of print and hence we reproduce in this section a number of the most essential definitions and theorems that Berge presents. In addition, we reproduce a certain number of results in convexity theory, most of which can be found in [Roc.1] or [Ber.1]. Finally, we extract from [Cla.1] a few basic results in nonsmooth analysis.

2.1. Continuity

We begin by summarizing the various concepts of continuity which play a role in optimization theory. Since in the context of optimization algorithms one generally deals with sequences rather than with neighborhoods, we shall give sequential alternatives whenever possible.

Definition 2.1.1. A function $f:\mathbb{R}^n \to \mathbb{R}$ is said to be *upper semi-continuous (u.s.c.) at* $\hat{x}$ if for every $\delta>0$ there exists a $\hat{\rho}>0$ such that

$$f(x)-f(\hat{x}) \le \delta \quad \forall\ x \in B(\hat{x},\hat{\rho}), \tag{2.1.1a}$$

where

$$B(\hat{x},\hat{\rho}) \triangleq \{x \in \mathbb{R}^n \mid \|x-\hat{x}\| \le \hat{\rho}\} \tag{2.1.1b}$$

A function $f(\cdot)$ is said to be u.s.c. if it is u.s.c. at all $x \in \mathbb{R}^n$.

■

Proposition 2.1.1. A function $f:\mathbb{R}^n \to \mathbb{R}$ is u.s.c. at $\hat{x}$ if and only if for any sequence $\{x_i\}_{i=0}^{\infty}$ in $\mathbb{R}^n$, such that $x_i \to \hat{x}$ as $i \to \infty$

$$\overline{\lim} f(x_i) \le f(\hat{x}) \tag{2.1.2}$$

■

Definition 2.1.2. A function $f:\mathbb{R}^n \to \mathbb{R}$ is said to be *lower semi-continuous* if $-f(\cdot)$ is u.s.c.

■

Proposition 2.1.2. A function $f:\mathbb{R}^n \to \mathbb{R}$ is l.s.c. at $\hat{x}$ if and only if for *any* sequence $\{x_i\}_{i=0}^{\infty}$ in $\mathbb{R}^n$, such that $x_i \to \hat{x}$ as $i \to \infty$, $\underline{\lim} f(x_i) \ge f(\hat{x})$.

■

Notation 2.1.1. We shall denote the solution set of a maximization (minimization) problem by *argmax* *(argmin)*. Thus, for example, $\operatorname*{argmax}_{y \in Y} \varphi(x,y) \triangleq \{y \in Y \mid \varphi(x,y) = \psi(x)\}$.

■

Next we turn to point-to-set functions. For example, let $\varphi:\mathbb{R}^n \mathrm{x} \mathbb{R}^m \to \mathbb{R}$ be a continuous function. We can define the point-to-set function

$$F(x) \triangleq \{y \in \mathbb{R}^m \mid \varphi(x,y) \le 0\} \tag{2.1.3}$$

which maps $\mathbb{R}^n$ into $2^{\mathbb{R}^m}$. As another example, consider the point-to-set function

$$M(x) \triangleq \underset{y \in Y}{\operatorname{argmax}}\, \varphi(x,y), \tag{2.1.4}$$

where $Y \subset \mathbb{R}^m$ is compact and $\psi(x) \triangleq \max_{y \in Y} \varphi(x,y)$, which also maps $\mathbb{R}^n$ into $2^{\mathbb{R}^m}$.

The most important concept for point-to-set maps is that of upper semi-continuity, though some use can also be made of lower semi-continuity. Note that the definitions, below, have nothing to do with the ones that we gave for functions from $\mathbb{R}^n$ into $\mathbb{R}$. Note furthermore, that the definitions, below, which are due to Berge [Ber.1], are not universally adopted.

Definition 2.1.3. A function (map) $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be *upper-semi-continuous* (u.s.c.) at $\hat{x}$ if

a) $f(\hat{x})$ is nonempty and compact, and

b) for every open set G such that $f(\hat{x}) \subset G$, there exists a $\hat{\rho}>0$ such that $f(x) \subset G$, for all $x \in B(\hat{x},\hat{\rho})$.

■

Definition 2.1.4. A function $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be *lower-semi-continuous* (l.s.c.) at $\hat{x}$ if for every open set G such that $f(\hat{x}) \cap G \neq \emptyset$,there exists a $\hat{\rho}>0$ such that $f(x) \cap G \neq \emptyset$, for all $x \in B(\hat{x},\hat{\rho})$, where $\emptyset$ denotes the empty set.

A function $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$ is u.s.c. (l.s.c.) if it is u.s.c. (l.s.c.) at every $x \in \mathbb{R}^n$.

■

Definition 2.1.5. A function $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be continuous if it is both u.s.c. and l.s.c.

■

Remark 2.1.1. Note that when $f:\mathbb{R}^n \to \mathbb{R}$ is either u.s.c. or l.s.c. in the sense of set valued maps, it is continuous in the ordinary sense.

■

Proposition 2.1.3. Suppose that $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$ is l.s.c. at $\hat{x}$ and $f(\hat{x})$ is compact. Then for any $\hat{\delta}>0$ there exists a $\hat{\rho}>0$ such that

$$f(x)\cap B(y,\hat{\delta})\neq\emptyset \quad \forall\, x \in B(\hat{x},\hat{\rho}),\ \forall\, y \in f(\hat{x}). \tag{2.1.5}$$

■

Upper and lower semi-continuity can also be given a sequential interpretation in terms of limit points and cluster points.

Definition 2.1.6. Consider a sequence of sets $\{A_i\}_{i=0}^{\infty}$ in $\mathbb{R}^n$.

a) The point $(\hat{x})$ is said to be a *limit point* of $\{A_i\}_{i=0}^{\infty}$ if $d(\hat{x},A_i) \to 0$ as $i \to \infty$, where

$$d(\hat{x},A_i) \triangleq \inf\{\|x-\hat{x}\| \mid x \in A_i\}, \tag{2.1.6}$$

i.e., if there exist $x_i \in A_i$ such that $x_i \to \hat{x}$ as $i \to \infty$.

b) The point $\hat{x}$ is a *cluster point* of $\{A_i\}_{i=0}^{\infty}$ if it a limit point of a subsequence of $\{A_i\}_{i=0}^{\infty}$.

c) We denote the set of *limit points* of $\{A_i\}$ by $\mathrm{Lim} A_i$ and the set of *cluster points* of $\{A_i\}$ by $\overline{\mathrm{Lim}}\, A_i$.

■

Proposition 2.1.4.

a) A function $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$, such that $f(x)$ is compact for all $x \in \mathbb{R}^n$ and bounded on bounded sets is u.s.c. at $\hat{x}$ if and only if for any sequence $\{x_i\}_{i=0}^{\infty}$ such that $x_i \to \hat{x}$ as $i \to \infty$, $\overline{\mathrm{Lim}}\, f(x_i) \subset f(\hat{x})$.

b) A function $f:\mathbb{R}^n \to 2^{\mathbb{R}^m}$ is l.s.c. at $\hat{x}$ if and only if for any sequence $\{x_i\}_{i=0}^{\infty}$ such that $x_i \to \hat{x}$ as $i \to \infty$, $\mathrm{Lim}\, f(x_i) \supset f(\hat{x})$.

■

2.2. Convexity

We assume that the reader has had some exposure to convexity. Thus, we assume that the reader is familiar with the definitions of convex sets and convex functions, and that the reader is aware of the following facts: (i) that convex functions are continuous, (ii) their epigraphs are convex, (iii) gradients of differentiable convex functions are normals to support hyperplanes to their epigraphs. Apart from these commonly known results, we shall make use of a few which are not in every elementary text dealing with convexity. To simplify matters, we collect in this subsection these assorted results. For proofs we refer the reader to [Roc.1] and [Ber.1].

We begin with two results involving convex sets: the Caratheodory theorem and a separation theorem.

Definition 2.2.1. Let S be a subset of $\mathbb{R}^n$. We shall denote by $\text{co}S$ the *convex hull* of S (i.e., the smallest convex set containing S).

■

Theorem 2.2.1 (Caratheodory). Let S be a subset in $\mathbb{R}^n$. If $\bar{x} \in \text{co}S$ then there exist at most $(n+1)$ distinct points $\{x_i\}_{i=1}^{n+1}$, in S such that $\bar{x} = \sum_{i=1}^{n+1} \mu^i x_i$, $\mu^i \geq 0$, $\sum_{i=1}^{n+1} \mu^i = 1$.

■

Definition 2.2.2. Let S_1, S_2 be any two sets in $\mathbb{R}^n$, and let $\nu \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ be given. We say that the hyperplane

$$H = \{x \in \mathbb{R}^n \mid \langle x, \nu \rangle = \alpha\} \tag{2.2.1a}$$

separates S_1 and S_2 if

$$\langle x, \nu \rangle \geq \alpha \ \forall \ x \in S_1, \tag{2.2.1b}$$

$$\langle y, \nu \rangle \leq \alpha \ \forall \ y \in S_2. \tag{2.2.1c}$$

The separation is said to be *strict* if one of the inequalities is satisfied strictly.

■

Definition 2.2.3 Suppose $S \subset \mathbb{R}^n$ is convex and $\nu \in \mathbb{R}^n$ is given. We say that

$$H = \{x \mid \langle x - \bar{x}, \nu \rangle = 0\} \tag{2.2.2a}$$

is a *support hyperplane* to S at $\bar{x}$ with inward (outward) normal ν if $\bar{x} \in \bar{S}$ (the closure of S) and

$$\langle x - \bar{x}, \nu \rangle \geq 0 (\leq 0) \ \forall \ x \in S. \tag{2.2.2b}$$

■

Proposition 2.2.1. Suppose that $S \subset \mathbb{R}^n$ is compact and convex and $0 \notin S$. Let

$$\hat{x} = \text{argmin} \{\|x\|^2 \mid x \in S\}. \tag{2.2.3}$$

Then the hyperplane

$$H = \{x \mid \langle \hat{x}, x \rangle = \|\hat{x}\|^2\} \tag{2.2.4}$$

is a support hyperplane to S at $\hat{x}$ which separates S from 0, i.e., $\langle \hat{x}, x \rangle \geq \|\hat{x}\|^2$ for all $x \in S$.

■

More generally, it is possible to establish the following result, (see [Roc.1] p. 97).

Theorem 2.2.2. (Separation) Suppose that $S_1, S_2 \subset \mathbb{R}^n$ are nonempty convex sets. Then there exists a hyperplane H which separates them if and only if their relative interiors have no points in common.

■

Next we turn to *support functions* which can be used to characterize convex sets and which play an important role in nondifferentiable analysis and optimization.

Definition 2.2.4. Let $S \subset \mathbb{R}^n$ be convex and compact. We define the *support function* $\sigma_S : \mathbb{R}^n \to \mathbb{R}$ of S by

$$\sigma_s(h) \triangleq \max\{\langle h, x \rangle \mid x \in S\}. \tag{2.2.5}$$

■

Proposition 2.2.2. Let $\sigma_S(\cdot)$ be defined by (2.2.5) with $S \subset \mathbb{R}^n$ convex and compact. Then

a) $\sigma_S(\cdot)$ is positive homogeneous, i.e., $\forall\ \lambda \geq 0$,

$$\sigma_S(\lambda h) = \lambda \sigma_S(h); \tag{2.2.6}$$

b) $\sigma_S(\cdot)$ is subadditive, i.e., $\forall\ h_1, h_2$,

$$\sigma_S(h_1+h_2) \leq \sigma_S(h_1) + \sigma_S(h_2); \tag{2.2.7}$$

c) $\sigma_S(\cdot)$ is convex.

■

Proposition 2.2.3. Let $S \subset \mathbb{R}^n$ be convex and compact. Suppose that for a given $h \in \mathbb{R}^n$, $x_h \in S$ is such that $\sigma_S(h) = \langle h, x_h \rangle$. Then

$$\langle x - x_h, h \rangle \leq 0 \quad \forall x \in S, \tag{2.2.8}$$

i.e., $\langle x, h \rangle = \langle x_h, h \rangle$ is a support hyperplane to S with *outward* normal h.

■

Proposition 2.2.4. Let $\sigma : \mathbb{R} \to \mathbb{R}$ be a positively homogeneous, subadditive function. Then the set

$$C = \{x \in \mathbb{R}^n \mid \langle x, h \rangle \leq \sigma(h) \ \forall h \in \mathbb{R}^n\} \tag{2.2.9}$$

is nonempty, convex, compact and $\sigma(\cdot)$ is the support function for C.

■

Minimax theorems play an important role both in game theory and in the construction of search directions subprocedures in optimization algorithms. The following result is one of the best known (see [Ber.1] for proof).

Theorem 2.2.3 (Von Neumann). Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be such that $f(x,y)$ is convex in x and concave in y and let $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$ be compact convex sets. Then

$$\min_{x \in X} \max_{y \in Y} f(x,y) = \max_{y \in Y} \min_{x \in X} f(x,y). \tag{2.2.10}$$

■

It is easy to extend the Von Neumann Theorem to the case where either X or Y is unbounded, as follows.

Corollary 2.2.1. Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be 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 that $f(x,y) \to \infty$ as $||x|| \to \infty$, uniformly in $y \in Y$. Then

$$\min_{x \in \mathbb{R}^n} \max_{y \in Y} f(x,y) = \max_{y \in Y} \min_{x \in \mathbb{R}^n} f(x,y). \tag{2.2.11}$$

■

The result for X compact and $Y = \mathbb{R}^m$ is obtained by assuming that $f(x,y) \to -\infty$ as $||y|| \to \infty$, uniformly in $x \in X$.

An extension of Von Neumann's theorem for the case where X,Y are subsets of normed spaces was given by K. Fan [Fan.1].

The minimax theorems lead to the following results which are important in optimization algorithm theory.

Proposition 2.2.5. Let S be a compact convex set in $\mathbb{R}^n$ and let $B = \{h \in \mathbb{R}^n \mid \|h\| \leq 1\}$. Then, with $\sigma_S(\cdot)$ the support function of S, we have

$$\min_{h \in B} \sigma_S(h) = -\min_{x \in S} \|x\| \tag{2.2.12}$$

and

$$\min_{h \in \mathbb{R}^n} \{½ \|h\|^2 + \sigma_S(h)\} = -\min_{x \in S} ½ \|x\|^2. \tag{2.2.13}$$

Proof. By definition of $\sigma_S(\cdot)$,

$$\min_{h \in B} \sigma_S(h) = \min_{h \in B} \max_{x \in S} \langle h, x \rangle \tag{2.2.14}$$

Since B, S are convex and compact and $\langle h, x \rangle$ is convex-concave, by the Van Neumann Theorem we get

$$\min_{h \in B} \sigma_S(h) = \max_{x \in S} \min_{h \in B} \langle x, h \rangle \tag{2.2.15}$$

Now $\min_{h \in B} \langle x, h \rangle$ is solved by $h = -x / \|x\|$. Hence, substituting in (2.2.14) we get

$$\begin{aligned} \min_{h \in B} \sigma_S(h) &= \max_{x \in S} -\|x\| \\ &= -\min_{x \in S} \|x\| \end{aligned} \tag{2.2.16}$$

Next, by Corollary 2.2.1,

$$\begin{aligned} \min_{h \in \mathbb{R}^n} \{ \tfrac{1}{2}\|h\|^2 + \sigma_S(h) \} &= \min_{h \in \mathbb{R}^n} \max_{x \in S} \{ \tfrac{1}{2}\|h\|^2 + \langle h, x \rangle \} \\ &= \max_{x \in S} \min_{h \in \mathbb{R}^n} \{ \tfrac{1}{2}\|h\|^2 + \langle h, x \rangle \} \end{aligned} \tag{2.2.17}$$

Now $\min_{h \in \mathbb{R}^n} \{ \tfrac{1}{2}\|h\|^2 + \langle h, x \rangle \}$ is solved by $h = -x$ (by taking derivatives and setting them to zero). Substituting into (2.2.16) we obtain

$$\begin{aligned} \min_{h \in \mathbb{R}^n} \{ \tfrac{1}{2}\|h\|^2 + \sigma_S(h) \} &= \max_{x \in S} -\tfrac{1}{2}\|x\|^2 \\ &= -\min_{x \in S} \tfrac{1}{2}\|x\|^2 \end{aligned}$$

■

The following obvious corollary plays an important role in the development of optimality conditions for optimization problems.

Corollary 2.2.2. Let S be a compact convex set in $\mathbb{R}^n$. Then $\sigma_S(h) \geq 0$ for all $h \in \mathbb{R}^n$ if and only if $0 \in S$.

■

The last result that we need is

Proposition 2.2.6. Let C, D be two convex, compact subsets in $\mathbb{R}^n$. Then $C \subset D$ if and only if $\sigma_C(h) \leq \sigma_D(h)$ for all $h \in \mathbb{R}^n$.

■

2.3. Nonsmooth Analysis

We now turn to real valued functions on $\mathbb{R}^n$ which are assumed to be only locally Lipschitz continuous (l.L.c). The results in this section culled from the book by F. H. Clarke [Cla.1]. Functions within this category that are particularly important in engineering design are the max functions discussed in Section 3, eigenvalues and singular values of various system matrices [May.8, Pol.10], and max min max functions discussed in [Pol.5, Pol.8], in connection with tolerancing and tuning problems.

Definition 2.3.1. We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is *locally Lipschitz continuous* (l.L.c.) at $\hat{x}$ if there exist $L \in [0,\infty)$, $\hat{\rho} > 0$ such that

$$\|f(x)-f(x')\| \leq L\|x-x'\| \quad \forall\, x,x' \in B(\hat{x},\hat{\rho}). \tag{2.3.1}$$

We begin by stating a key property of l.L.c. functions, the Rademacher Theorem [Ste.1].

Proposition 2.3.1. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous. Then $\nabla f(x)$ exists for almost all $x \in \mathbb{R}^n$.

■

Since a l.L.c. function may fail to have directional derivatives at a point $x \in \mathbb{R}^n$, it became necessary to extend the concept of directional derivative, as follows.

Definition 2.3.2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be l.L.c. We define the (Clarke) *generalized directional derivative of* $f(\cdot)$ at $x \in \mathbb{R}^n$ in the direction $h \in \mathbb{R}^n$ by

$$d_0 f(x;h) \triangleq \overline{\lim_{\substack{t \downarrow 0 \\ y \to x}}} \frac{f(y+th)-f(y)}{t}. \tag{2.3.2}$$

■

Since there exist $\varepsilon > 0$, $L > 0$ such that $|f(y+th)-f(y)| \leq tL\|h\|$, for all $y \in B(x,\varepsilon)$, $t < \varepsilon$, it is clear that $d_0 f(x;h)$ is well defined.

Proposition 2.3.2. Let $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuous function such that $\nabla_x \varphi(\cdot,\cdot)$ exists and is continuous and let Y be a compact subset of $\mathbb{R}^m$. Let

$$\psi(x) \triangleq \max\{\varphi(x,y) \mid y \in Y\} \quad \textbf{(2.3.3a)}$$

$$\xi(x) \triangleq \min\{\varphi(x,y) \mid y \in Y\}. \quad (2.3.3b)$$

Then for any $x, h \in \mathbb{R}^n$, the (ordinary) directional derivatives $d\psi(x;h)$, $d\xi(x;h)$ exist and satisfy

$$d\psi(x;h) = d_0\psi(x;h), \quad (2.3.4a)$$

$$d\xi(x;h) \leq d_0\xi(x;h). \quad (2.3.4b)$$

■

Proposition 2.3.3. The generalized directional derivative $d_0 f(x;h)$ of a l.L.c. function $f : \mathbb{R}^n \to \mathbb{R}$ has the following properties:

a) $h \to d_0 f(x;h)$ is positively homogeneous and subadditive on $\mathbb{R}^n$.

b) If L is a local Lipschitz constant for $f(\cdot)$ at x, then for any $h \in \mathbb{R}^n$

$$|d_0 f(x;h)| \leq L\|h\|. \quad (2.3.5)$$

c) The function $d_0 f(\cdot;\cdot)$ is u.s.c.

d) For any $x \in \mathbb{R}^n$, the function $d_0 f(x;\cdot)$ is Lipschitz continuous with constant L, where L is a local Lipschitz constant for $f(\cdot)$ at x.

e) For any x, $h \in \mathbb{R}^n$, $d_0 f(x;-h) = d_0(-f)(x;h)$.

■

Definition 2.3.3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be l.L.c. We define the (Clarke) *generalized gradient* of $f(\cdot)$ at x by

$$\partial f(x) \triangleq \{\xi \in \mathbb{R}^n \mid d_0 f(x;h) \geq \langle \xi, h \rangle, \ \forall\, h \in \mathbb{R}^n\}. \quad (2.3.6)$$

■

We now elucidate the reasons for calling the set $\partial f(x)$ the generalized gradient of $f(\cdot)$. First, suppose that $f(\cdot)$ is continuously differentiable at x. Then, $d_0 f(x;h) = df(x;h) = \langle \nabla f(x), h \rangle$ for any $h \in \mathbb{R}^n$. By definition (2.3.6), for any $\xi \in \partial f(x)$

$$\langle \nabla f(x) - \xi, h \rangle \geq 0 \quad \forall\, h \in \mathbb{R}^n . \tag{2.3.7}$$

Hence we must have $\nabla f(x) - \xi = 0$ for all $\xi \in \partial f(x)$, i.e., $\partial f(x) = \{\nabla f(x)\}$. Next, suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is l.L.c. and convex. Then its epigraph is convex and, at any point $(\hat{x}, f(\hat{x}))$ the epigraph has one or more support hyperplanes, with normal $(\xi, -1) \in \mathbb{R}^{n+1}$, such that

$$\langle (\xi,-1), (x-\hat{x}, f(x)-f(\hat{x})) \rangle \leq 0 \quad \forall\, x \in \mathbb{R}^n . \tag{2.3.8}$$

Hence

$$\langle \xi, x-\hat{x} \rangle \leq f(x) - f(\hat{x}) \quad \forall\, x \in \mathbb{R}^n . \tag{2.3.9}$$

Now let $x = \hat{x} + th$, for any $h \in \mathbb{R}^n$, $t > 0$. Then we get

$$\langle \xi, h \rangle \leq \overline{\lim_{t \downarrow 0}} \frac{f(\hat{x}+th) - f(\hat{x})}{t} \leq d_0 f(\hat{x}; h), \tag{2.3.10}$$

i.e., $\xi \in \partial f(\hat{x})$. Finally we have

Proposition 2.3.4. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is l.L.c. with constant L in a ball centered on $\hat{x}$. Then

a) $\partial f(\hat{x})$ is nonempty, convex and compact, and $\|\xi\| \leq L$ for all $\xi \in \partial f(\hat{x})$.

b) For every $h \in \mathbb{R}^n$,

$$d_0 f(\hat{x}; h) = \max\{\langle \xi, h \rangle \mid \xi \in \partial f(x)\} . \tag{2.3.11}$$

c)

$$\partial f(x) = G(x) \triangleq \text{co} \lim_{x_i \to x} \{\nabla f(x_i)\}, \tag{2.3.12}$$

where the convex hull is taken over all sequences $\{x_i\}$ converging to x, such that $\nabla f(x_i)$ exists for all $i \in \mathbb{N}$ and $\{\nabla f(x_i)\}_{i=1}^{\infty}$ converges (where $\mathbb{N} = \{0,1,2,3,\ldots\}$).

d) The generalized gradient $\partial f(\cdot)$ is u.s.c.

■

In general, when a l.L.c. function $f(\cdot)$ is the result of operations on other functions (e.g., sum, product), its generalized gradient will not be equal to the set suggested by similar operations on differentiable functions. Rather, it will only be contained in it. A sufficient condition for equality to hold is satisfied when the functions being operated on are regular.

Definition 2.3.4. A l.L.c. function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *regular* if its directional derivative $df(x;h)$ exists for all $x, h \in \mathbb{R}^n$ and $df(x;h) = d_0 f(x;h)$.

■

Thus, for example, we have the following result:

Proposition 2.3.5. Suppose that $f^1, f^2 : \mathbb{R}^n \to \mathbb{R}$, are l.L.c., then

$$\partial[f^1 + f^2](x) \subset \partial f^1(x) + \partial f^2(x). \tag{2.3.13}$$

Furthermore, if f^1, f^2 are regular then equality holds in (2.3.13).

■

Proposition 2.3.6. Suppose that $f^1, f^2, \ldots, f^m : \mathbb{R}^n \to \mathbb{R}$ are l.L.c. and let

$$\psi(x) \triangleq \max_{j \in \mathbf{m}} f^j(x). \tag{2.3.14}$$

Then

$$\partial\psi(x) \subset \text{co}\{\partial f^j(x) \mid j \in I(x)\}, \tag{2.3.15}$$

where $I(x) \triangleq \{j \in \mathbf{m} \mid f^j(x) = \psi(x)\}$ and $\mathbf{m} \triangleq \{1, 2, \ldots, m\}$. Further, if the functions $f^j(\cdot)$, $j \in \mathbf{m}$ are all regular, then equality holds in (2.3.15).

■

It is also possible to establish a chain rule.

Proposition 2.3.7. Let $h : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable, let $g : \mathbb{R}^m \to \mathbb{R}$ be l.L.c. and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) \triangleq g(h(x)) \tag{2.3.17}$$

Then

$$\partial f(x) \subset \overline{\text{co}}\,\{ \sum_{i=1}^{m} \gamma^i \eta_i \,|\, \eta_i \in \partial h^i(x),\ \gamma^i = \nabla g(h(x)) \}. \tag{2.3.18}$$

Again, equality holds in (2.3.18) when g is regular.

■

The last result in nondifferentiable analysis that we wish to establish is the Lebourg Mean Value Theorem [Leb.1].

Theorem 2.3.1. (Mean Value). Let $f : \mathbb{R}^n \to \mathbb{R}$ be l.L.c. Then, given any $x, y \in \mathbb{R}^n$,

$$f(y) - f(x) = \langle \xi, y - x \rangle \tag{2.3.19}$$

for some $\xi \in \partial f(x + s(y-x))$, with $s \in (0,1)$.

■

3. MAX FUNCTIONS

It is clear from the examples presented in Section 1 that max functions play a central role in optimization problems arising in engineering design. They are also a particularly tractable kind of nondifferentiable function. In this section we shall establish some of their most important properties, see also [Ber.1], [Cla.1], [Dan.1], [Dem.1].

Notation convention. Given a sequence $\{x_i\}_i^\infty$ and an infinite subset $K \subset \mathbb{N} \triangleq \{0,1,2,3,\ldots\}$, we shall denote by $x_i \overset{K}{\to} \hat{x}$ as $i \to \infty$ the fact that the subsequence $\{x_i\}_{i \in K}$ converges to $\hat{x}$.

■

Proposition 3.1. Let $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be continuous and $Y : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ u.s.c. Then

$$\psi(x) \triangleq \max_{y \in Y(x)} \varphi(x,y) \tag{3.1}$$

is u.s.c.

Proof. Let $\hat{x} \in \mathbb{R}^n$ be given. Let $\{x_i\}$ be an arbitrary sequence such that $x_i \to \hat{x}$ as $i \to \infty$, and let $y_i \in Y(x_i)$ be such that $\psi(x_i) = \varphi(x_i,y_i)$ for $i = 1,2,3,\ldots$. Since $Y(\cdot)$ is u.s.c. and $x_i \to \hat{x}$, $\{y_i\}$ is bounded and hence, since $\varphi(\cdot,\cdot)$ is continuous, $\overline{\lim}\, \varphi(x_i,y_i)$ exists. Suppose $y_i, i \in K \subset \{0,1,\ldots\}$ is such that $\overline{\lim}\, \varphi(x_i,y_i) = \lim_{i \in K} \varphi(x_i,y_i)$ and $y_i \overset{K}{\to} y^*$. Then (see Proposition 2.1.4), $y^* \in Y(\hat{x})$ by u.s.c. of $Y(\cdot)$ and hence

$$\psi(\hat{x}) \ge \varphi(\hat{x},y^*) = \lim_{i \in K} \varphi(x_i,y_i) = \overline{\lim}\, \psi(x_i), \tag{3.2}$$

which completes our proof.

■

Corollary 3.1. Consider $\varphi(\cdot,\cdot)$ and $Y(\cdot)$ as in Proposition 3.1 and suppose that $Y(\cdot)$ is continuous. Then $\psi(\cdot)$ is continuous.

Proof. We only need to show that $\psi(\cdot)$ is l.s.c. under the stronger assumption on $Y(\cdot)$. For the sake of contradiction, suppose there is a point $\hat{x} \in \mathbb{R}^n$ and a sequence $x_i \to \hat{x}$ as $i \to \infty$ such that $\lim \psi(x_i)$ exists and

$$\lim \psi(x_i) < \psi(\hat{x}). \tag{3.3}$$

Suppose that $\psi(\hat{x}) = \varphi(\hat{x},\hat{y})$ with $\hat{y} \in Y(\hat{x})$. Let $y_i \in Y(x_i)$ be such that $\psi(x_i) = \varphi(x_i,y_i)$ and let $\hat{y}_i = \text{argmin}\{\|y - \hat{y}\|^2 \mid y \in Y(x_i)\}$. Then, since $Y(\cdot)$ and $\varphi(\cdot,\cdot)$ are continuous, $\hat{y}_i \to \hat{y}$ as $i \to \infty$, so that $\lim \varphi(x_i,\hat{y}_i) = \varphi(\hat{x},\hat{y})$. Hence there exists an i_0 such that $\varphi(x_i,\hat{y}_i) > \psi(x_i)$, which contradicts the definition of $\psi(x_i)$.

■

Proposition 3.2. Consider the function

$$\psi(x) \triangleq \max_{y \in Y(x)} \varphi(x,y), \tag{3.4}$$

with $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ continuous and $Y : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ continuous. Let

$$\hat{Y}(x) \triangleq \{y \in Y(x) \mid \psi(x) = \varphi(x,y)\} \tag{3.5}$$

Then $\hat{Y}(\cdot)$ is u.s.c. Furthermore, if $\hat{Y}(x) = \{y(x)\}$, a singleton, then $Y(\cdot)$ ($y(\cdot)$ is continuous at x.

Proof. Clearly $\hat{Y}(\cdot)$ is bounded on bounded sets and $\hat{Y}(x)$ is compact because $Y(x)$ is compact and $\varphi(x,\cdot)$ is continuous. By Proposition 2.1.4 we only need to show that $\overline{\text{Lim}}\ \hat{Y}(x_i) \subset \hat{Y}(\hat{x})$ for any sequence $\{x_i\}_{i=0}^{\infty}$ converging to a point $\hat{x}$. Suppose this is false, i.e., there exists a point $\hat{x}$ and a sequence $x_i \to \hat{x}$ such that for $y_i \in \hat{Y}(x_i)$ we have $y_i \to \hat{y} \not\in \hat{Y}(\hat{x})$. But this means that $\psi(x_i) = \varphi(x_i,y_i) \to \varphi(\hat{x},\hat{y}) < \psi(\hat{x})$, which contradicts the continuity of $\psi(\cdot)$ (Corollary 3.1).

When $\hat{Y}(x)$ is a singleton, its continuity follows directly from the definition of upper semi-continuity, see Remark 2.1. This completes our proof.

■

Next we explore the differentiability properties of max functions of the form (3.4). First, suppose that $\varphi(x,y)$ is differentiable in x, with $\nabla_x \varphi(x,y)$ continuous, and that $Y = \{y_1,y_2, \ldots, y_m\}$. Letting $f^i(x) \triangleq \varphi(x,y_i)$, for $i \in \mathbf{m} \triangleq \{1,2,\ldots,m\}$, (3.4) becomes

$$\psi(x) = \max_{i \in \mathbf{m}} f^i(x) \tag{3.6}$$

Drawing the graph of $\psi(x+\lambda h)$, for a fixed $h \in \mathbb{R}^n$, which is a function of λ only, we obtain Fig. 3.1 and conclude that $\psi(\cdot)$ is not differentiable everywhere. However, its directional derivative seems to exist and seems to be equal to its generalized derivative (see Propositions 2.3.4, 2.3.6). From Fig. 3.1, we conclude that the directional derivative of ψ at x in the direction h is equal to the steepest slope of the "active" functions $f^i(\cdot)$, i.e., if we let $I(x) \triangleq \{i \in \mathbf{m} \mid \psi(x) = f^i(x)\}$, i.e., that

$$d\psi(x;h) \triangleq \lim_{t \downarrow 0} \frac{\psi(x+th)-\psi(x)}{t}$$
$$= \max_{i \in I(x)} df^i(x;h) \tag{3.7a}$$
$$= \max_{i \in I(x)} \langle \nabla f^i(x), h \rangle .$$

Furthermore, we conclude that its generalized gradient is given by

$$\partial\psi(x) = \underset{i \in I(x)}{\mathrm{co}} \{\nabla f^i(x)\}. \tag{3.7b}$$

These results are, in fact, correct. We shall now explore to what extent it can be generalized for the case where ψ is defined as in (3.1). First, assuming that $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous and that $\varphi(\cdot,y)$ is l.L.c. for any $y \in \mathbb{R}^m$, we define the ***partial*** directional and Clarke generalized directional derivatives at $x \in \mathbb{R}^n$ of $\varphi(\cdot,y)$ in the direction h, by

$$d_x\varphi(x,y;h) \triangleq \lim_{t \downarrow 0} \frac{\varphi(x+th,y)-\varphi(x,y)}{t} \tag{3.8a}$$

and

$$d_{x0}\varphi(x,y;h) \triangleq \overline{\lim_{\substack{x' \to x \\ t \downarrow 0}}} \frac{\varphi(x'+th,y)-\varphi(x',y)}{t} \tag{3.8b}$$

We shall denote by $\partial_x\varphi(x,y)$ the partial generalized gradient of $\varphi(\cdot,y)$.

Theorem 3.1. Consider the function

$$\psi(x) \triangleq \max_{y \in Y} \varphi(x,y). \tag{3.9}$$

Suppose that

(i) $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous and $Y \subset \mathbb{R}^m$ is compact;

(ii) for all $y \in Y$, $\varphi(\cdot,y)$ is locally Lipschitz continuous.

Then $\psi(\cdot)$ is locally Lipschitz continuous and its generalized gradient satisfies the relation

$$\partial\psi(x) \subset G(x) \triangleq \text{co}\,\{\underset{\substack{x_i \to x \\ y_i \to y}}{\text{Lim}}\, \partial_x \varphi(x_i, y_i) \,|\, y_i \in Y,\, y \in \hat{Y}(x)\}, \tag{3.10}$$

where $\hat{Y}(x)$ was defined in (3.5) and Lim in Definition 2.1.6. The convex hull in (3.10) is taken over all possible sequences $\{x_i\}$, $\{y_i\}$.

Proof. Clearly, $\varphi(\cdot, y)$ is l.L.c. near $x \in \mathbb{R}^n$, uniformly in $y \in Y$, with constant L, say. Hence, for x', x'' in the appropriate neighborhood of x,

$$\begin{aligned} \psi(x') - \psi(x'') &= \varphi(x', y') - \varphi(x'', y'') \\ &= [\varphi(x', y') - \varphi(x'', y')] + [\varphi(x'', y') - \varphi(x'', y'')] \\ &\le \varphi(x', y') - \varphi(x'', y') \le L\|x' - x''\|, \end{aligned} \tag{3.11}$$

where $y' \in \hat{Y}(x')$ and $y'' \in \hat{Y}(x'')$. Interchanging x' and x'' in (3.11) we conclude that $\psi(\cdot)$ is l.L.c.

Next, let $g(x;\cdot)$ be the support functional of $G(x)$, so that for any $h \in \mathbb{R}^n$,

$$g(x;h) = \max\{\langle \xi, h \rangle \,|\, \xi \in G(x)\}. \tag{3.12}$$

By Proposition 2.2.6 and the definition of $\partial\psi(x)$ in (2.3.6), to show that $\partial\psi(x) \subset G(x)$, we only need to show that

$$g(x;h) \ge d_0\psi(x;h) \quad \forall\, h \in \mathbb{R}^n. \tag{3.13}$$

Hence, let $h \in \mathbb{R}^n$ be arbitrary and let $x_i \to x$ and $t_i \downarrow 0$ be such that

$$\Delta_i = \frac{\psi(x_i + t_i h) - \psi(x_i)}{t_i} \tag{3.14}$$

converges to $d_0\psi(x;h)$, the generalized derivative of ψ. Let $y_i \in \hat{Y}(x_i + t_i h)$ be arbitrary. Then

$$\Delta_i \le \frac{\varphi(x_i + t_i h, y_i) - \varphi(x_i, y_i)}{t_i}. \tag{3.15}$$

It now follows from the Lebourg Mean Value Theorem 2.3.1, that there exist $\xi_i \in \partial_x \varphi(x_i + s_i t_i h, y_i)$, for some $s_i \in [0,1]$, such that

$$\Delta_i \leq \langle \xi_i, h \rangle \quad, \forall\, i \in \mathbb{N}. \tag{3.16}$$

Hence

$$d_0\psi(x;h) = \lim_{i \to \infty} \Delta_i \leq \overline{\lim_{i \to \infty}} \langle \xi_i, h \rangle \quad . \tag{3.17}$$

Since $(x_i + s_i t_i h) \to x$ as $i \to \infty$ and $\hat{Y}(\cdot)$ is u.s.c., it follows that all the accumulation points of $\{y_i\}$ are in $\hat{Y}(x)$ and hence that all the accumulation points of $\{\xi_i\}$ are in the set $G(x)$. Consequently, (3.13) holds and so does (3.10). This completes our proof.

■

In general, relation (3.10) has negative consequences from an algorithmic point of view. This is due to the fact that, as we shall see later, the accumulation points $\hat{x}$ constructed by an algorithm, minimizing ψ over $\mathbb{R}^n$, can only be guaranteed to be such that $0 \in G(\hat{x})$. Hence, when $\partial\psi(\hat{x}) \neq G(\hat{x})$, it is possible that the accumulation points are not stationary. Fortunately, in our experience, the functions entering engineering constraints are regular (see Definition 2.3.4): a fact that leads to the following, much more satisfactory result.

Theorem 3.2. Consider the function $\psi(\cdot)$ defined in (3.9). In addition to hypotheses (i) and (ii) of Theorem (3.1), suppose that

(iii) for all $y \in Y$, $\varphi(\cdot, y)$ is regular;

(iv) for any $x \in \mathbb{R}^n$ and $y \in Y$ and any sequences $\{x_i\} \subset \mathbb{R}^n$, $\{y_i\} \subset Y$, converging to x, y, respectively, $\text{co} \operatorname*{Lim}_{\substack{x_i \to x \\ y_i \to y}} \partial_x \varphi(x_i, y_i) = \partial_x \varphi(x, y)$.

Then

(a) $d\psi(x;h)$ exists for all $x \in \mathbb{R}^n$, $h \in \mathbb{R}$;

(b) $d\psi(x;h) = d_0\psi(x;h)$, and

(c) we have

$$\partial\psi(x) = G(x) = \text{co}\{\partial_x\varphi(x,y)\}_y \in \hat{Y}(x)\}, \tag{3.18}$$

where $G(x)$ was defined in (3.10).

Proof. To prove that $d\psi(x;h)$ exists and that $d\psi(x;h) = d_0\psi(x;h)$ we only need to show that

$$\begin{aligned} \alpha &\triangleq \lim_{t\downarrow 0} \frac{\psi(x+th)-\psi(x)}{t} \\ &\geq g(x;h) \geq d_0\psi(x;h) \geq \overline{\lim_{t\downarrow 0}} \frac{\psi(x+th)-\psi(x)}{t} \end{aligned} \tag{3.19}$$

where $g(x;h)$ was defined in (3.12).

First we note that because of hypothesis (iv),

$$G(x) = \text{co}\{\partial_x\varphi(x,y) \mid y \in \hat{Y}(x)\} \tag{3.20}$$

(which proves half of (3.18)). Next, let $y \in \hat{Y}(x)$ be arbitrary. Then for any $t > 0$,

$$\frac{\psi(x+th)-\psi(x)}{t} \geq \frac{\varphi(x+th,y)-\varphi(x,y)}{t}. \tag{3.21}$$

Since $d_x\varphi(x,y;h)$ exists by (iii), and $y \in \hat{Y}(x)$ is arbitrary, (3.21) yields that

$$\begin{aligned} \alpha &\geq \max\{d_x\varphi(x,y;h) \mid y \in \hat{Y}(x)\}, \\ &= \max\{d_{x0}\varphi(x,y;h) \mid y \in \hat{Y}(x)\} \\ &= \max\{\langle \xi,h \rangle \mid \xi \in G(x)\} = g(x;h) \end{aligned} \tag{3.22}$$

where we have made use of hypothesis (iv). Hence $d\psi(x;h)$ exists and $d\psi(x;h) = d_0\psi(x;h)$.

Finally, the fact that $\partial\psi(x) = G(x)$ follows from the fact that $d_0\psi(x;h) = g(x;h)$ for all $h \in \mathbb{R}^n$ and Proposition 2.2.6. This completes our proof.

■

Theorem 3.2 has an obvious, but important corollary.

Corollary 3.1. Consider the function $\psi(\cdot)$ defined in (3.9) and suppose that $\varphi : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ is continuous, $Y \subset \mathbb{R}^p$ is compact, $\nabla_x \varphi(\cdot,\cdot)$ exists and is continuous. Then

$$\partial\psi(x) = \mathrm{co}\{\nabla_x \varphi(x,y)\}_{y} \in \hat{Y}(x) \tag{3.23a}$$

and

$$\begin{aligned} d\psi(x;h) &= d_0\psi(x;h) \\ &= \max_{y \in \hat{Y}(x)} \langle \nabla_x \varphi(x,y), h \rangle . \end{aligned} \tag{3.23b}$$

■

Corollary 3.1 has an important special case.

Proposition 3.4. Suppose that with $(x,\omega) \in \mathbb{R}^n \times \mathbb{R}$,

$$\psi(x) \triangleq \max_{\omega \in \Omega} \{\bar{\sigma}[H(x,j\omega)]^2 - b(\omega)\} \tag{3.24}$$

where $H(x,j\omega)$ is an $m \times m$ continuous complex valued matrix which is differentiable in x, $\bar{\sigma}[H]$ is its largest singular value, $\frac{\partial}{\partial x^i} H(x,j\omega)$ exists and is continuous, $b(\omega)$ is continuous and $\Omega \subset \mathbb{R}$ is compact. Then

$$\begin{aligned} \partial\psi(x) = \mathrm{co}\{v \mid v^i &= \langle U(\omega)z, \frac{\partial Q(x,j\omega)}{\partial x^i} U(\omega)z \rangle, \\ &\|z\| = 1 , \ \omega \in \hat{\Omega}(x)\}, \end{aligned} \tag{3.25}$$

where $Q \triangleq H^T H$, U is an orthonormal matrix of eigenvectors corresponding to its maximum eigenvalue $\bar{\sigma}[H]^2$ and $\hat{\Omega}(x) = \{\omega \in \Omega \mid \psi(x) = \bar{\sigma}[H(x,j\omega)]^2 - b(\omega)\}$.

Proof. Let $y = (\omega, u) \in \mathbb{R}^{m+1}$ and let

$$\varphi(x,y) = \langle u, Q(x,j\omega)u \rangle - b(\omega). \tag{3.26}$$

Then $\nabla_x \varphi(x,y)$ is a vector whose i^{th} element is $\langle u, \frac{\partial Q(x,j\omega)}{\partial x^i} u \rangle$. Since $Y = \Omega \times \{u \in \mathbb{C}^m \mid \|u\| = 1\}$, $\hat{Y}(x) = \hat{\Omega}(x) \times \{u \mid Q(x,j\omega)u = \bar{\sigma}[H(x,j\omega)]^2 u,$ $\|u\| = 1, u \in \mathbb{C}^m\}$. The desired result now follows directly from Corollary 3.1.

■

Referring to [Clar.1], Section 2.8, we see that the continuity assumption on $b(\omega)$ in (3.24) can be relaxed to upper semi-continuity. This fact is of significance in engineering design.

4. FIRST ORDER OPTIMALITY CONDITIONS FOR PROBLEM (0.1.3)

We shall now develop first order optimality conditions for the canonical optimization problem (0.1.3). Since in problems of engineering design, the hypotheses introduced in Theorems 3.1 and 3.2 are usually satisfied, we shall adopt them in the derivation of optimality conditions as well.

Definition 4.1. Consider the problem P: $\min_{x \in X} f(x)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and $X \subset \mathbb{R}^n$. We shall say that $\hat{x}$ is a *global solution* to P if $\hat{x} \in X$ and $f(\hat{x}) \le f(x)$ $\forall x \in X$. We shall say that $\hat{x}$ is a *local solution* to P if $\hat{x} \in X$ and there exists a $\hat{\rho} > 0$ such that $f(\hat{x}) \le f(x)$ for all $x \in X$ such that $\|x - \hat{x}\| < \hat{\rho}$.

■

Proposition 4.1. Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{4.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is l.L.c. Suppose that $\hat{x}$ is a local solution for (4.1), then $0 \in \partial f(\hat{x})$.

Proof. Suppose $\hat{x}$ solves (4.1). Then we must have

$$d_0 f(\hat{x};h) \geq 0, \ \forall\, h \in \mathbb{R}^n \tag{4.2}$$

for otherwise there would be a direction $\hat{h}$ such that

$$d_0 f(\hat{x};\hat{h}) \triangleq \overline{\lim_{\substack{x \to \hat{x} \\ t \downarrow 0}}} \frac{f(x+t\hat{h})-f(x)}{t} < 0, \tag{4.3}$$

and hence for a finite $\hat{t} > 0$, $f(\hat{x}+t\hat{h}) < f(\hat{x})$ would hold for all $t \in (0,\hat{t}]$, contradicting our hypothesis that $\hat{x}$ is a local solution. Referring to Corollary 2.2.2, we see that (4.2) implies that $0 \in \partial f(\hat{x})$. This completes our proof.

■

The following characterization of a local solution $\hat{x}$ to a general form of problem P, in Definition 4.1, is suggested by Fig. 4.1 for the simple case where the set X is defined by a finite number of differentiable inequalities, i.e., $X = \{x \in \mathbb{R}^n \mid g^j(x) \leq 0,\ j = 1,2,\ldots,m\}$. Note that for the "active" gradients in Fig. 4.1, the origin is moved to the optimal point $\hat{x}$ and the result suggested by this figure is that if $\max_{j \in \mathbf{m}} g^j(\hat{x}) = 0$, then $0 \in \text{co}\{\nabla f(\hat{x}), \nabla g^j(\hat{x})\}_{j \in I(\hat{x})}$, where $I(\hat{x}) \triangleq \{j \in \mathbf{m} \mid g^j(\hat{x}) = 0\}$.

Theorem 4.1. Consider the problem

$$\min\{f(x) \mid \varphi(x,y) \leq 0 \ \forall\, y \in Y\} \tag{4.4}$$

where

(i) $f(\cdot)$ is l.L.c.;

(ii) $\varphi(\cdot,\cdot)$ is continuous and $Y \subset \mathbb{R}^m$ is compact;

(iii) for all $y \in Y$, $\varphi(\cdot,y)$ is l.L.c.;

(iv) for all $y \in Y$, $\varphi(\cdot,y)$ is regular;

(v) for any sequences $\{x_i\} \subset \mathbb{R}^n$, $\{y_i\} \subset Y$, such that $x_i \to x \in \mathbb{R}^n$ and $y_i \to y \in Y$, $\text{co}\, \overline{\text{Lim}}_{i \to \infty}\, \partial_x \varphi(x_i, y_i) = \partial_x \varphi(x,y)$.

If $\hat{x}$ is a local solution to (4.4), then

$$0 \in \text{co}\{\partial f(\hat{x});\ \partial_x \varphi(\hat{x},y),\ y \in \hat{Y}(\hat{x})\} \quad \text{if } \psi(\hat{x}) = 0 \tag{4.5a}$$

and

$$0 \in \partial f(\hat{x}) \quad \text{if } \psi(\hat{x}) < 0, \tag{4.5b}$$

where, as in Section 3,

$$\psi(x) \triangleq \max_{y \in Y} \varphi(x,y), \tag{4.6}$$

and

$$\hat{Y}(x) \triangleq \{y \in Y \mid \varphi(x,y) = \psi(x)\}. \tag{4.7}$$

Proof. Let $\rho > 0$ be such that $f(\hat{x}) \le f(x)\ \forall\ x \in \mathbb{R}^n$ such that $\psi(x) \le 0$ and $\|x - \hat{x}\| \le \rho$. Let

$$\begin{aligned} F(x) &\triangleq \max\{f(x) - f(\hat{x}), \psi(x)\} \\ &= \max\{f(x) - f(\hat{x});\ \varphi(x,y),\ y \in Y\} \end{aligned} \tag{4.8}$$

Note that $F(\hat{x}) = 0$, since $\psi(\hat{x}) \le 0$ and that $F(x) \ge 0$ for all $x \in \mathbb{R}^n$ such that $\|x - \hat{x}\| \le \rho$, because $f(x) - f(\hat{x}) \ge 0$ when $\|x - \hat{x}\| \le \rho$ and $\psi(x) \le 0$. Hence $\hat{x}$ is a local minimizer of $F(x)$ and hence, by Proposition 4.1, we must have

$$d_0 F(\hat{x};h) \ge 0\ \forall\ h \in \mathbb{R}^n \tag{4.9}$$

The desired result now follows from (3.18) and Proposition 2.3.5.

■

The following special case follows directly by means of Caratheodory's Theorem (2.2.1).

Corollary 4.2. Suppose that $\hat{x}$ is a local solution to (4.4) and that $\nabla f(\cdot)$, $\nabla_x \varphi(\cdot,\cdot)$ exist and are continuous. Then there exist at most $(n+2)$ points $\nabla f(\hat{x})$, $\nabla_x \varphi(\hat{x},y_i)$, $i = 1,2,\ldots,n+1$, with $y_i \in \hat{Y}(\hat{x})$, such that

$$\mu^0 \nabla f(\hat{x}) + \sum_{i=1}^{n+1} \mu^i \nabla_x \varphi(\hat{x}, y_i) = 0 \qquad (4.10)$$

where $\mu^i \geq 0$ for $i = 1,\ldots,n+1$ and $\sum_{i=0}^{n+1} \mu^i = 1$. Furthermore, if $\psi(\hat{x}) < 0$, then $\mu^i = 0$ for $i = 1,2,\ldots,n+1$; if $\psi(\hat{x}) = 0$ and $0 \not\in \{\nabla_x \varphi(\hat{x},y) \mid y \in \hat{Y}(\hat{x})\}$, then $\mu^0 > 0$.

■

5. SEMI-INFINITE OPTIMIZATION ALGORITHMS I: BASICS

We now return to the optimization problem (0.1.3). So as to avoid obscuring clarity by excessive notation, we shall consider in detail only the simplest form of problem (0.1.3), which captures all the essential features of problems in this class. Thus, consider the problem

$$\min\{f(x) \mid \varphi(x,y) \leq 0 \quad \forall \; y \in Y\}, \qquad (5.1)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and Y satisfies the hypotheses of Theorem 3.1 and Theorem 3.2, viz.

Assumption 5.1. We shall assume that

(i) the function $f(\cdot)$ is l.L.c.;

(ii) $\varphi(\cdot,\cdot)$ is continuous and $Y \subset \mathbb{R}^m$ is compact;

(iii) for all $y \in Y$, $\varphi(\cdot,y)$ is l.L.c.;

(iv) for all $y \in Y$, $\varphi(\cdot,y)$ is regular;

(v) for any sequences $\{x_i\} \in \mathbb{R}^n$, $\{y_i\} \subset Y$, such that $x_i \to x \in \mathbb{R}^n$ and $y_i \to y \in Y$, co $\overline{\text{Lim}}_{i \to \infty}\, \partial_x \varphi(x_i, y_i) = \partial_x \varphi(x,y)$.

■

If we define

$$\psi(x) \triangleq \max_{y \in Y} \varphi(x,y), \tag{5.2a}$$

we can express (5.1) in the equivalent form

$$\min\{f(x) \mid \psi(x) \le 0\}. \tag{5.2b}$$

Unless otherwise stated, we shall assume that Assumption 5.1 is satisfied throughout the next two sections. We recall that first order optimality conditions for the problem (5.1) were given in Theorem 4.1. In this section we turn to the development of algorithms for solving problems of the form (5.1). All the algorithms that we will present can be thought of as being evolved from the method of steepest descent for *unconstrained differentiable* optimization. We therefore begin by recalling this method of steepest descent (which is attributed to Cauchy).

Consider the problem

$$\min_{x \in \mathbb{R}^n} \psi(x), \tag{5.3}$$

where $\psi : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable.

Algorithm 5.1. (Differentiable Steepest Descent for Problem (5.3))

Data: $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: Compute the *search direction*

$$h_i = h(x_i) \triangleq \operatorname*{argmin}_{h \in \mathbb{R}^n} \{ \tfrac{1}{2}\|h\|^2 + d\psi(x_i; h) \} = -\nabla\psi(x_i). \tag{5.4}$$

Step 2: Compute the *step size*

$$\lambda_i \in \lambda(x_i) \triangleq \operatorname*{argmin}_{\lambda \ge 0} \psi(x_i + \lambda h_i). \tag{5.5}$$

Step 3: Update:

$$x_{i+1} = x_i + \lambda_i h_i \ . \tag{5.6}$$

Replace i by $i+1$ and go to Step 1.

■

All of our convergence theorems will be stated in terms of subsequences constructed by an algorithm, which are conveniently handled by the notation introduced in Section 3, ie., given a sequence $\{x_i\}_i^{\infty}$ and an infinite subset $K \subset \mathbb{N} \triangleq \{0,1,2,3,...\}$, we shall denote by $x_i \xrightarrow{K} \hat{x}$ as $i \to \infty$ the fact that the subsequence $\{x_i\}_{i \in K}$ converges to $\hat{x}$.

Theorem 5.1. Consider a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 5.1. If $x_i \xrightarrow{K} \hat{x}$ as $i \to \infty$, then $\nabla\psi(\hat{x}) = 0$.

Proof. Suppose that $\nabla\psi(\hat{x}) \neq 0$. Then

$$d\psi(\hat{x};h(\hat{x})) = -\|\nabla\psi(\hat{x})\|^2 < 0 \ . \tag{5.7}$$

Hence, any $\hat{\lambda} \in \lambda(\hat{x})$ satisfies $\hat{\lambda} > 0$ and there exists a $\hat{\delta}>0$ such that

$$\psi(\hat{x} + \hat{\lambda} h(\hat{x})) - \psi(\hat{x}) = -\hat{\delta} < 0 \ . \tag{5.8}$$

Since $h(\cdot) = -\nabla\psi(\cdot)$ is continuous by assumption, the function $\psi(x + \hat{\lambda} h(x)) - \psi(x)$ is continuous in x and hence there exists an i_0 such that for all $i \in K$, $i \geq i_0$,

$$\psi(x_{i+1}) - \psi(x_i) \leq \psi(x_i + \hat{\lambda} h(x_i)) - \psi(x_i) \leq -\frac{\hat{\delta}}{2} . \tag{5.9}$$

Now, by construction, $\{\psi(x_i)\}_{i=0}^{\infty}$ is monotone decreasing and $\psi(x_i) \xrightarrow{K} \psi(\hat{x})$ as $i \to \infty$ by continuity of $\psi(\cdot)$; we must therefore have that $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$. But this contradicts (5.9). Hence we must have that $\nabla\psi(\hat{x}) = 0$.

■

Remark 5.1. We must point out at this time that practical algorithms do not use the stepsize rule (5.5), but the much more efficient *Armijo* step legth rule [Arm. 1], which uses two parameters α, $\beta \in (0,1)$ and which is defined by

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, f(x_i + \lambda h_i) - f(x_i) \leq -\lambda\alpha\|h_i\|^2\}. \quad (5.10)$$

The geometry of this stepsize rule is given in Fig. 5.1.

■

The convergence analysis of Algorithm 5.1, modified to accept the Armijo Step length rule, is only somewhat more complex than the analysis presented in the proof of Theorem 5.1. The reader may look it up in [Pol. 1].

Now suppose that $\psi(\cdot)$ in (5.3) is only l.L.c. Since in this case the gradient $\nabla\varphi(x)$ need not exist for all x, a first attempt at generalizing Algorithm 5.1 to the nondifferentiable case would consist of replacing, in (5.4), the term $d\psi(x_i;h)$ $(=<\nabla\varphi(x)),h>)$ by $d_0\psi(x,h)$ $(=\max<\xi,h> \mid \xi \in \partial\psi(x)\})$. This amounts to computing the search direction according to the formula

$$\begin{aligned} h_i = h(x_i) &\triangleq \underset{h \in \mathbb{R}^n}{\operatorname{argmin}}\{½\|h\|^2 + d_0\psi(x_i;h)\} \\ &= \underset{h \in \mathbb{R}^n}{\operatorname{argmin}} \max_{\xi \in \partial\psi(\xi)} \{½\|h\|^2 + <\xi,h>\} \\ &= \underset{\xi \in \partial\psi(\xi)}{\operatorname{argmax}} \min_{h \in \mathbb{R}^n} \{½\|h\|^2 + <\xi,h>\} \\ &= -\operatorname{argmin}\{½\|h\|^2 \mid h \in \partial\psi(x)\}, \end{aligned} \quad (5.11)$$

where we have interchanged the min and max operations on the basis of Corollary 2.2.1 and have eliminated the min by making use of the fact that if h_ξ solves $\min\{½\|h\|^2 + \langle\xi,h\rangle \mid h \in \mathbb{R}^n\}$, then $h_\xi = -\xi$, so that $½\|h_\xi\|^2 + <\xi,h_\xi> = -½\|h_\xi\|^2$.

Because $\partial\psi(\cdot)$ is not continuous, $h(\cdot)$, as defined by (5.11) is not continuous. Hence it is not possible to simply mimic the proof of Theorem 5.1 in trying to show that the extended Algorithm 5.1 is convergent in the sense that $x_i \overset{K}{\to} \hat{x}$ implies that $0 \in \partial\psi(\hat{x})$. In fact, there are known counter examples in the literature on methods of feasible directions, which show that the accumulation points

$\hat{x}$ constructed by the extension of Algorithm 5.1 using (5.11) fail to satisfy $0 \in \partial\psi(\hat{x})$. Clearly, a much more sophisticated approach than using (5.11) is needed for extending Algorithm 5.1 to the nondifferentiable case of problem (5.3).

To try to obtain some intuitive insight into techniques for generating continuous search directions, let us examine the simple case where $\psi(x) = \max_{j \in \mathbf{m}} f^j(x)$, with the $f^j : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable and, as before, $\mathbf{m} \triangleq \{1,2,\dots,m\}$. In this case, (see Proposition 2.3.6), $\partial\psi(x) = \text{co}\{\nabla f^j(x)\}_{j \in I(x)}$, where $I(x) \triangleq \{j \in \mathbf{m} \mid \psi(x) - f^j(x) = 0\}$. Since the index set $I(x)$ can change abruptly, it is clear that $\partial\psi(\cdot)$ is not continuous. Now, if $\hat{x}$ is a minimizer of $\psi(\cdot)$ over $\mathbb{R}^n$, then by Proposition 4.1, we have $0 \in \partial\psi(\hat{x})$, i.e., for some $\mu^j \geq 0$, $j \in I(x)$ such that $\sum_{j \in I(x)} \mu^j = 1$, we have $\sum_{j \in I(x)} \mu^j \nabla f^j(\hat{x}) = 0$. A commonly used device for avoiding the introduction of the index set $I(x)$ into this optimality condition, is to express the optimality condition in the equivalent form of two equations

$$\sum_{j=1}^{m} \mu^j \nabla f^j(\hat{x}) = 0 \tag{5.12a}$$

$$\sum_{j=1}^{m} \mu^j (\psi(\hat{x}) - f^j(\hat{x})) = 0 \tag{5.12b}$$

with the $\mu^j \geq 0$ such that $\sum_{j=1}^{m} \mu^j = 1$. Since $\mu^j \geq 0$ and $\psi(\hat{x}) - f^j(\hat{x}) \geq 0$, (5.12b) implies that $\mu^j = 0$ for all $j \notin I(\hat{x})$. Now, (5.12a) and (5.12b) state that 0 is an element of the set $\overline{G}\psi(\hat{x}) \subset \mathbb{R}^{n+1}$ defined by

$$\overline{G}\psi(\hat{x}) \triangleq \text{co}\{\bar{\xi}_j \in \mathbb{R}^{n+1} \mid \bar{\xi}_j = (\psi(\hat{x}) - f^j(\hat{x}), \nabla f^j(\hat{x})), j \in \mathbf{m}\}. \tag{5.12c}$$

we shall denote vectors in $\mathbb{R}^{n+1}$ as $\bar{\xi} = (\xi^0, \xi)$ with $\xi \in \mathbb{R}^n$. The set valued map $\overline{G}\psi(\cdot)$ is continuous (see Example 5.1 further on) and hence,

$\bar{h}(x) = (h^0(x), h(x))$, with $h(x) \in \mathbb{R}^n$, defined uniquely by $\bar{h}(x) \triangleq -\arg_{\bar{h} \in \bar{G}\psi(x)}$ $\min \frac{1}{2}\|\bar{h}\|^2$, is also continuous by Proposition 3.2. Hence, the ***principle of wishful thinking*** leads us to the correct guess that $h(x)$ must be a "good" search direction for solving $\min_{x \in \mathbb{R}^n} \psi(x)$. We shall now present an axiomatic structure which emanates from this guess and which enables us to construct algorithms for the solution of the general case of problem (5.3). In the next section we shall present a more complex axiomatic structure which leads to computationally more efficient algorithms.

Definition 5.1. Let $\psi: \mathbb{R}^n \to \mathbb{R}$ be l.L.c.. We shall say that $\bar{G}\psi: \mathbb{R}^n \to 2^{\mathbb{R}^{n+1}}$ is an *augmented convergent direction finding* (a.c.d.f.) map for $\psi(\cdot)$ if:

(a) $\bar{G}\psi(\cdot)$ is continuous (i.e., both u.s.c. and l.s.c.) and $\bar{G}\psi(x)$ is convex for all $x \in \mathbb{R}^n$.

(b) For any $x \in \mathbb{R}^n$, if $\bar{\xi} = (\xi^0, \xi) \in \mathbb{R}^{n+1}$ is an element of $\bar{G}\psi(x)$, then $\xi^0 \geq 0$.

(c) For any $x \in \mathbb{R}^n$, a point $\bar{\xi} = (0, \xi)$ is an element of $\bar{G}\psi(x)$ if and only if $\xi \in \partial\psi(x)$.

■

Proposition 5.1. Suppose that $\psi: \mathbb{R}^n \to \mathbb{R}$ is l.L.c. and $\bar{G}\psi(\cdot)$ is an a.c.d.f. map for $\psi(\cdot)$. Then for any $x \in \mathbb{R}^n$,

(a) $0 \in \partial\psi(x)$ if and only if $0 \in \bar{G}\psi(x)$.

(b) The functions $\Theta: \mathbb{R}^n \to \mathbb{R}$ and $\bar{h}: \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by

$$\Theta(x) \triangleq \min\{\tfrac{1}{2}\|\bar{\xi}\|^2 \mid \bar{\xi} \in \bar{G}\psi(x)\} \qquad (5.13a)$$

$$\bar{h}(x) \triangleq -\text{argmin}\{\tfrac{1}{2}\|\bar{\xi}\|^2 \mid \bar{\xi} \in \bar{G}\psi(x)\} \qquad (5.13b)$$

are both continuous; furthermore, $\Theta(x) = 0 \iff 0 \in \partial\psi(x)$.

(c) Writing $\bar{h}(x) = (h^0(x), h(x))$, with $h(x) \in \mathbb{R}^n$,

$$d_0\psi(x; h(x)) \leq -\Theta(x), \quad \forall \ x \in \mathbb{R}^n . \tag{5.13c}$$

Proof.

a) Let $\xi = 0$. Then the desired result follows directly from Definition 5.1 (c).

b) Since $\bar{G}\psi(\cdot)$ is continuous, it follows from Corollary 3.1 and Proposition 3.2 that $\Theta(x)$ is continuous and $\bar{h}(x)$ is u.s.c. Since the solution of (5.13a) is unique, it follows that $\bar{h}(x)$ is a point-to-point map and hence continuous.

c) By definition (5.13b) we have

$$<-\bar{h}(x), \bar{\xi}> \ \geq \ \tfrac{1}{2}\|\bar{h}(x)\|^2 = \Theta(x) \quad \forall \ \bar{\xi} \in \bar{G}\psi(x) . \tag{5.14}$$

Now suppose that $\bar{\xi} = (0, \xi) \in \bar{G}\psi(x)$, so that $\xi \in \partial\psi(x)$. Then

$$\langle -\bar{h}(x), \bar{\xi} \rangle = \langle -h(x), \xi \rangle \geq \ \Theta(x) . \tag{5.15}$$

Consequently we have

$$d_0\psi(x; h(x)) = \max_{\xi \in \partial\psi(x)} <h(x), \xi> \ \leq \ -\Theta(x) , \tag{5.16}$$

which completes our proof.

■

We shall now see that if we modify the search direction computation in (5.4) as shown below, we obtain an algorithm for solving (5.3) under the assumption that $\psi(\cdot)$ is only l.L.c. The convergence proof of this algorithm mimics the proof of Theorem 5.1.

Algorithm 5.2. (Nondifferentiable Steepest Descent for Problem (5.3). Requires an a.c.d.f. map $\bar{G}\psi(\cdot)$).

Data: $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: Compute the *augmented search direction* $\bar{h}(x_i) = (h^0(x_i), h(x_i))$ according to (5.13b), i.e.,

$$\bar{h}(x_i) = -\text{argmin}\{\tfrac{1}{2}\|\bar{\xi}\|^2 \mid \bar{\xi} \in \bar{G}\psi(x_i)\} \tag{5.17a}$$

and set the *actual search direction* $h_i = h(x_i)$.

Step 2: Compute the *step length*

$$\lambda_i \in \lambda(x_i) \triangleq \underset{\lambda \geq 0}{\text{argmin}}\, \psi(x_i + \lambda h_i). \tag{5.17b}$$

Step 3: Update:

$$x_{i+1} = x_i + \lambda_i h_i\,; \tag{5.17c}$$

replace i by $i+1$ and go to step 1.

■

Remark 5.2. The Armijo step length rule (5.10) can be modified for use in Algorithm 5.2 as well. For Algorithm 5.2 it assumes the form

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, f(x_i + \lambda h_i) - f(x_i) \leq -\lambda\alpha\Theta(x_i)\}, \tag{5.17d}$$

where $\alpha, \beta \in (0,1)$.

■

Theorem 5.2. Suppose that $\psi: \mathbb{R}^n \to \mathbb{R}$ is l.L.c. Consider a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 5.2. If $x_i \xrightarrow{K} \hat{x}$ as $i \to \infty$, then $0 \in \partial\psi(\hat{x})$.

Proof. Suppose that $0 \notin \partial\psi(\hat{x})$. Then by Proposition 5.1 $0 \notin \partial\bar{G}\psi(\hat{x})$ and therefore $\Theta(\hat{x}) > 0$. Consequently,

$$d_0\psi(\hat{x}, h(\hat{x})) \leq -\Theta(\hat{x}) < 0. \tag{5.18a}$$

Hence, the stepsize $\hat{\lambda} \in \lambda(\hat{x})$ computed at $\hat{x}$, satisfies $\hat{\lambda} > 0$ and

$$\psi(\hat{x} + \hat{\lambda} h(\hat{x})) - \psi(\hat{x}) = -\hat{\delta} < 0. \tag{5.18b}$$

Since $\psi(\cdot)$ is continuous by assumption and $h(\cdot)$ is continuous by Proposition 5.1, it follows that there exists an i_0 such that for all $i \in K$, $i \geq i_0$.

$$\psi(x_{i+1}) - \psi(x_i) \leq \psi(x_i + \hat{\lambda} h(x_i)) - \psi(x_i) \leq -\hat{\delta}/2\,. \qquad (5.18c)$$

Now $\{\psi(x_i)\}_{i=0}^{\infty}$ is a monotonically descreasing sequence and $\psi(x_i) \overset{K}{\to} \psi(\hat{x})$ as $i \to \infty$. Hence $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$. But this contradicts (5.18c), and hence we must have $0 \in \partial\psi(\hat{x})$.

■

It is only slightly more difficult to establish the convergence properties of Algorithm 5.2, with (5.17b) replaced by (5.17d), as we shall now show.

Theorem 5.2b. Suppose that $\psi : \mathbb{R}^n \to \mathbb{R}$ is l.L.c.. Consider a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by Algorithm 5.2, with (5.17b) replaced by (5.17d), i.e., using the Armijo type step length rule. If $x_i \overset{K}{\to} \hat{x}$ as $i \to \infty$, then $\nabla\psi(\hat{x}) = 0$.

Proof. Suppose that $0 \notin \partial\psi(\hat{x})$. Then (5.18) must hold. It now follows from the Lebourg Mean Value Theorem 2.3.1 that for any $\lambda > 0$,

$$\begin{aligned}
&\psi(\hat{x}+\lambda h(\hat{x})) - \psi(\hat{x}) + \lambda\alpha\Theta(\hat{x}) \qquad (5.19a)\\
&\quad = \lambda\langle \xi_{\lambda s}, h(\hat{x})\rangle + \lambda\alpha\Theta(\hat{x})\\
&\quad \leq \lambda(d_0\psi(\hat{x} + s\lambda h(\hat{x}); h(\hat{x})) - d_0\psi(\hat{x}; h(\hat{x})) - (1-\alpha)\Theta(\hat{x})),
\end{aligned}$$

where $\xi_{\lambda s} \in \partial\psi(\hat{x}+s\lambda h(\hat{x}))$ and $s \in (0,1)$. Hence, since by Proposition 2.3.3 $d_0(\cdot\,;\cdot)$ is u.s.c., there exists a $\hat{k} \in \mathbb{N}$ such that

$$\psi(\hat{x} + \beta^{\hat{k}} h(\hat{x})) - \psi(\hat{x}) + \beta^{\hat{k}}\Theta(\hat{x}) \leq -\beta^{\hat{k}}(1-\alpha)\Theta(\hat{x})/2. \qquad (5.19b)$$

Hence, since $\psi(\cdot)$, $h(\cdot)$ and $\Theta(\cdot)$ are continuous, there is an i_0 such that for all $i \in K$, $i \geq i_0$,

$$\psi(x_i + \beta^{\hat{k}} h(x_i)) - \psi(x_i) + \beta^{\hat{k}}\Theta(x_i) \leq 0\,, \qquad (5.19c)$$

so that $\lambda_i \geq \beta^{\hat{k}}$ for all $i \in K$, $i \geq i_0$. Next, Since $\Theta(\cdot)$ is continuous, there exists an $i_1 \geq i_0$ such that for all $i \in K$, $i \geq i_1$, $\Theta(x_i) \geq \Theta(\hat{x})/2$. It therefore follows that for all $i \in K$, $i \geq i_1$,

$$\psi(x_{i+1}) - \psi(x_i) \leq -\beta^{\hat{k}} \alpha \Theta(\hat{x})/2 . \tag{5.19d}$$

Since (5.19d) implies that $\psi(x_i) \to -\infty$ as $i \to \infty$, we have a contradiction of the fact that $\psi(x_i) \to \psi(\hat{x})$, which follows from the fact that $\{\psi(x_i)\}_{i=0}^{\infty}$ is a monotone decreasing sequence with an accumulation point. This completes our proof.

■

The applicability of Algorithm 5.2 to a specific nondifferentiable optimization problem depends on the availability of an appropriate a.c.d.f. map. We shall now present three examples which show that for the max functions which occur in problems of engineering design, it is quite easy to construct a.c.d.f. maps. Our first example deals with the simple max function that triggered the introduction of Definition 5.1.

Example 5.1. Suppose that $\psi:\mathbb{R}^n \to \mathbb{R}$ is defined by

$$\psi(x) \triangleq \max_{j \in \mathbf{m}} f^j(x) \tag{5.20}$$

where the $f^j:\mathbb{R}^n \to \mathbb{R}$ are continuously differentiable. We shall show that

$$\overline{G}\psi(x) \triangleq \underset{j \in \mathbf{m}}{\text{co}} \left\{ \begin{pmatrix} \psi(x) - f^j(x) \\ \nabla f^j(x) \end{pmatrix} \right\} \tag{5.21}$$

is an a.c.d.f. map for $\psi(\cdot)$, i.e., that it satisfies the requirements (a), (b) and (c) of Definition 5.1.

(a) By construction, $\overline{G}\psi(x)$ is convex. Next, for all $j \in \mathbf{m}$, let $\bar{\xi}_j:\mathbb{R}^n \to \mathbb{R}^{n+1}$ be defined by $\bar{\xi}_j(x) = (\xi_j^0(x), \xi_j(x)) = (\psi(x) - f^j(x), \nabla f^j(x))$. Then $\bar{\xi}_j(\cdot)$ is continuous. Let

$$\Sigma_m \triangleq \{\mu \in \mathbb{R}^m \mid \mu \geq 0, \sum_{j \in \mathbf{m}} \mu^j = 1\} \tag{5.22a}$$

and let $x \in \mathbb{R}^n$ be arbitrary. Then $\overline{G}\psi(x)$ can be expressed in the form

$$\overline{G}\psi(x) = \{\overline{\xi} \in \mathbb{R}^{n+1} \mid \overline{\xi} = \sum_{j \in \mathbf{m}} \mu^j \overline{\xi}_j(x),\ \mu \in \Sigma_m\}. \tag{5.22b}$$

Since Σ_m is compact and the $\overline{\xi}_j(\cdot)$ are continuous, $\overline{G}\psi(x)$ is bounded on bounded sets. Now suppose that $x_i \to x$ as $i \to \infty$, and that $\overline{\xi}_i \in \overline{G}\psi(x_i)$ are such that $\overline{\xi}_i \to \overline{\xi}$ as $i \to \infty$. Then for some $\mu_i \in \Sigma_m$, $\overline{\xi}_i = \sum_{j \in \mathbf{m}} \mu_i^j \overline{\xi}_j(x_i)$ and, since Σ_m is compact, there exists an infinite $K \subset \mathbb{N}$ such that $\mu_i \overset{K}{\to} \mu \in \Sigma_m$ as $i \to \infty$. Clearly, $\overline{\xi} = \sum_{j \in \mathbf{m}} \mu^j \overline{\xi}_j(x)$ and hence $\overline{\xi} \in \overline{G}\psi(x)$ Hence it follows that $\overline{\mathrm{Lim}}\ \overline{G}\psi(x_i) \subset \overline{G}\psi(x)$ which proves that $\overline{G}\psi(\cdot)$ is upper semi-continuous.

Next, let $\overline{\xi} \in \overline{G}\psi(x)$ be arbitrary. Then $\overline{\xi} = \sum_{j \in \mathbf{m}} \mu^j \overline{\xi}_j(x)$ for some $\mu \in \Sigma_m$. Since $\overline{\xi}_i \triangleq \sum_{j \in \mathbf{m}} \mu^j \overline{\xi}_j \in \overline{G}\psi(x_i)$ and $\overline{\xi}_i \to \overline{\xi}$ as $i \to \infty$, we conclude that $\overline{G}\psi(x) \subset \underline{\mathrm{Lim}}\overline{G}\psi(x_i)$, i.e., that $\overline{G}\psi(\cdot)$ is l.s.c. Since it is both u.s.c. and l.s.c., it is continuous.

Properties (b) and (c) of Definition 5.1 follow from the fact that $\psi(x) - f^j(x) \geq 0$ for all $j \in \mathbf{m}$ and (3.18). This concludes the proof that $\overline{G}\psi(\cdot)$ is an a.c.d.f. map.

To compute the search direction $h(x)$, defined in (5.17a), for the function $\psi(\cdot)$ defined in (5.21a), we can proceed in two steps. First we solve the *finite* quadratic program

$$\min_{\mu \in \Sigma_m} \tfrac{1}{2}\{(\sum_{j \in \mathbf{m}} \mu^j [\psi(x) - f^j(x)])^2 + \| \sum_{j \in \mathbf{m}} \mu^j \nabla f^j(x)\|^2\}, \tag{5.23}$$

for a solution $\mu_x \in \Sigma_m$. Since the quadratic form in (5.23) may be only positive semi-definite, standard quadratic programming codes, such as [Gil.1], may fail occasionally. In that case, the Wolfe proximity algorithm [Wol.1] may be used. In either case, only a finite number of iterations are needed to solve (5.23). Once μ_x has been computed, the search direction $h(x)$ is obtained according to the formula

$$h(x) = -\sum_{j \in \mathbf{m}} \mu_x^j \nabla f^j(x) \tag{5.24}$$

Thus, we see that the computation of the search direction is quite simple for the function in (5.21a).

Example 5.2. Suppose that $\psi:\mathbb{R}^n \to \mathbb{R}$ is defined by

$$\psi(x) \triangleq \max\{\varphi(x,y) \mid y \in Y\} \tag{5.25a}$$

where $\varphi:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\nabla_x \varphi:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ are continuous and $Y \subset \mathbb{R}^m$ is compact. We shall show that

$$\overline{G}\psi(x) = \text{co}\left\{\begin{bmatrix} \psi(x) - \varphi(x,y) \\ \nabla_x \varphi(x,y) \end{bmatrix}\right\}_{y \in Y} \tag{5.25b}$$

is an a.c.d.f. map. Let $\overline{\xi}:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+1}$ be defined by $\overline{\xi}(x,y) = (\psi(x) - \varphi(x,y), \nabla_x \varphi(x,y))$. Then we see that $\overline{\xi}(\cdot,\cdot)$ is continuous. By construction, $\overline{G}\psi(x)$ is convex and bounded on bounded sets because Y is compact and $\overline{\xi}(\cdot,\cdot)$ is continuous. Suppose that $x_i \to x$ as $i \to \infty$ and that $\overline{\xi}_i \in \overline{G}\psi(x_i)$ for all $i \in \mathbb{N}$ are such that $\overline{\xi}_i \to \overline{\xi}$ as $i \to \infty$. Since by Caratheodory's Theorem 2.2.1, there exists a $\mu_i \in \Sigma_{n+2}$ (defined in (5.22)) and vectors $y_{ji} \in Y$, $j = 1,2, \cdots n+2$, such that $\overline{\xi}_i = \sum_{j \in \underline{n+2}} \mu_i^j \overline{\xi}(x_i, y_{ji})$, and since both Y and Σ_{n+2} are compact and $\overline{\xi}(\cdot,\cdot)$ is continuous, it follows that $\overline{\xi} \in \overline{G}\psi(x)$. Hence $\overline{G}\psi(\cdot)$ is u.s.c.

Now, let $\overline{\xi} \in \overline{G}\psi(x)$ be arbitrary. Then, by Caratheodory's Theorem, $\overline{\xi} = \sum_{j \in \underline{n+2}} \mu^j \overline{\xi}(x,y_j)$, with $\mu \in \Sigma_{n+2}$ and $y_j \in Y$, for all $j \in \underline{n+2}$. Clearly, $\overline{\xi}_i \triangleq \sum_{j \in \underline{n+2}} \mu_i^j \overline{\xi}(x_i, y_j)$ is an element of $\overline{G}\psi(x_i)$ and $\overline{\xi}_i \to \overline{\xi}$ as $i \to \infty$. Hence

$\underset{i \to \infty}{\text{Lim}} \overline{G}\psi(x_i) \supset \overline{G}\psi(x)$ for any sequence $x_i \to x$, which proves that $\overline{G}\psi(\cdot)$ is l.s.c. Hence, it is continuous.

b) and c) follow from the definition of $\overline{G}\psi(\cdot)$ and (3.18). This concludes the proof that $\overline{G}\psi(\cdot)$ is an a.c.d.f. map.

■

Next we must examine the problem of computing a search direction according to (5.13b). Clearly, (5.13b) no longer defines a finite dimensional quadratic program and hence $\overline{h}(x)$ must be computed by means of a proximity algorithm, such as the one stated below (see Fig. 5.2):

Proximity Algorithm 5.3.

Step 0: Compute a $\overline{\xi}_0 \in \overline{G}\psi(x)$; set $\overline{s}_0 = \overline{\xi}_0$, $i = 0$.

Step 1: Compute

$$\begin{aligned} \overline{\xi}_{i+1} = &(\psi(x) - \varphi(x, y_{i+1}), \nabla_x \varphi(x, y_{i+1})) \\ &\in \text{argmin}\{\langle \xi, \overline{s}_i \rangle \mid \xi \in \overline{G}\psi(x)\} \end{aligned} \tag{5.26a}$$

where

$$y_{i+1} \in Y_{i+1} \triangleq \text{argmin}\{s_i^0[\psi(x) - \varphi(x, y)] + \langle \nabla_x \varphi(x, y), s_i \rangle \mid y \in Y\}. \tag{5.26b}$$

Step 2: Compute

$$\overline{s}_{i+1} = \arg\min\{\|s\|^2 \mid s \in co\{\overline{\xi}_i, \overline{\xi}_{i+1}\}\}. \tag{5.26c}$$

Step 3: Replace i by $i+1$ and go to Step 1.

■

Proposition 5.2. The sequence $\{\overline{s}_i\}_{i=0}^{\infty}$ constructed by Algorithm 5.3 converges to $-\overline{h}(x)$, defined by (5.13b).

■

Remark 5.3. Formula (5.26a) is based on the observation that for any compact set $S \subset \mathbb{R}^{n+1}$, $\min_{\xi \in S} \langle \xi, \overline{s}_i \rangle = \min_{\xi \in coS} \langle \xi, \overline{s}_i \rangle$.

When y_{i+1} satisfying (5.26b) is not unique, a more efficient selection would be to set

$$\bar{\xi}_{i+1} = \text{argmin}\{\|\bar{x}\|^2 \mid \bar{\xi} \in \text{co}\left\{\begin{pmatrix}\psi(x_i) - \varphi(x_i,y) \\ \nabla_x \varphi(x_i,y)\end{pmatrix}\right\}_{y \in Y_{i+1}} \tag{5.26d}$$

We note that the computation of $\bar{h}(x)$ by means of the Proximity Algorithm 5.3 is no longer a finite process (unlike the case in Example 5.1) and hence *implementation* procedures must eventually be introduced (see, e.g., [Kle.1, Muk.1, Pol.1, Pol.8, Tra.1]). Also, the computation of y_{i+1} according to (5.26b) may or may not be practical. For example, when $Y \subset \mathbb{R}$, y_{i+1} can be computed by scanning the linear segment Y, with $\langle \nabla_x \varphi(x,y), s_i \rangle = d_x \varphi(x,y;s_i)$ approximated by a finite difference. However, when $y = (\omega, u)$, with $\omega \in \mathbb{R}$, $u \in \mathbb{R}^m$, and $\varphi(x,y) = \langle u, Q(x,j\omega)u \rangle$ with Q symmetric and positive definite (as in Proposition 3.4), the computation in (5.26b) appears to be prohibitive.

■

Finally, we shall show that it is possible to extend the concept of an a.c.d.f. map to max functions defined on an infinite dimensional space.

Example 5.3. Suppose we are given a dynamical system

$$\dot{z}(t) = f(z(t), u(t)), \quad t \in [0,1], \; z(0) = z_0, \tag{5.27}$$

where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is continuously differentiable, and suppose that we are required to find a control $u \in L_\infty[0,1]$ such that $g(z(t)) \le 0$ for $t \in [0,1]$, with $g : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable. First, denoting the solution of (5.27) by $z^u(t)$, we define $\varphi : L_\infty[0,1] \times \mathbb{R} \to \mathbb{R}$ by

$$\varphi(u,t) \triangleq g(z^u(t)) \tag{5.28a}$$

and $\psi : L_\infty[0,1] \to \mathbb{R}$ by

$$\psi(u) \triangleq \max_{t \in [0,1]} \varphi(u,t). \tag{5.28b}$$

Next, we define $\overline{G}\psi(u)$ as in (5.25b), with u replacing x. To obtain an expression for $\nabla_u \varphi(u,t)(\cdot)$, we note, formally, that to first order (in $L_\infty[0,1]$)

$$\begin{aligned} \varphi(u+s,t) - \varphi(u,t) &= g(z^{u+s}(t)) - g(z^u(t)) \\ &\approx \frac{\partial g}{\partial z}(z^u(t))\delta z^s(t) \\ &\triangleq \langle \nabla_u \varphi(u,t), s \rangle_2, \end{aligned} \tag{5.29a}$$

where $\langle \cdot,\cdot \rangle_2$ denotes the L_2 scalar product and

$$\begin{aligned} \delta \dot{z}^s(t) &= \frac{\partial f}{\partial z}(z^u(t),u(t))\delta z^s(t) + \frac{\partial f}{\partial u}(z^u(t),u(t))s(t), \quad t \in [0,1], \\ \delta z(0) &= 0. \end{aligned} \tag{5.29b}$$

Hence,

$$\begin{aligned} \nabla_u \varphi(u,t)(\tau) &= \frac{\partial f}{\partial u}(z^u(t),u(t))^T p^{u,t}(\tau) \quad \text{for } 0 \le \tau \le t \\ &= 0 \quad \text{for } t < \tau \le 1, \end{aligned} \tag{5.30}$$

where, for $\tau \in [0,t]$, $p^{u,t}(\tau)$ is determined by the adjoint equation

$$\frac{d}{dt} p^{u,t}(\tau) = -\frac{\partial f}{\partial z}(z^u(t),u(t))^T p^{u,t}(\tau), \tag{5.31a}$$

$$p^{u,t}(t) = \nabla g(z^u(t)). \tag{5.31b}$$

Next, referring to the operations in the proximity Algorithm 5.3, in step 1, (with $y = t$) we determine

$$t_{i+1} \in \operatorname*{argmin}_{t \in [0,1]} \{ s_i^0(\psi(u) - g(z^u(t)) + dg(z^u(t); \delta z^{s_i}(t)) \}, \tag{5.32a}$$

where the directional derivative $dg(z^u(t);\delta z^{s_i}(t))$ can be approximated by the finite difference $\frac{1}{\lambda}[g(z^{u+\lambda s_i}(t)) - g(z^u(t))]$ for $\lambda > 0$ small. The computation of $\overline{s}_{i+1}$ in (5.26b) reduces to solving for $\lambda \in [0,1]$ the problem

$$\min_{\lambda \in [0,1]} \{ [\psi(u) + \lambda \varphi(u,t_i) + (1-\lambda)\varphi(u,t_{i+1}))]^2 + \|\lambda \nabla_u \varphi(u,t_i) + (1-\lambda)\nabla_u \varphi(u,t_{i+1})\|_2^2 \} \tag{5.32b}$$

which is a simple problem.

Note that as defined by (5.26b), because $\nabla_u \varphi(u,t)(\tau)$ is possibly discontinuous at $\tau = t$, it follows that the search directions $s_i(t)$ are only piecewise continuous and may not have an accumulation point in $L_\infty[0,1]$. Consequently, one must eventually introduce the concept of relaxed controls [Warg.1, Wil.1] in analyzing the convergence properties of the extension of Algorithm 5.2 to optimal control. Once this is done, one finds that a corresponding restatement of Theorem 5.2 remains valid.

■

We are now ready to tackle problem (5.2). First we shall treat the solution of problem (5.2) as a two phase process. In phase I, an algorithm of the form of Algorithm 5.2 is used to find a point x_0 such that $\psi(x_0) \le 0$. Note that if $\min_{x \in \mathbb{R}^n} \psi(x) < 0$, then such an x_0 is obtained in a finite number of iterations. In phase II, we construct a minimizing sequence $\{x_i\}_0^\infty$ such that $\psi(x_i) \le 0 \; \forall \, i$. Then we shall show that the phase I and phase II processes can be combined into a more efficient single process.

For the purpose of defining search directions for the phase II process, we need to postulate a continuous set valued map $\overline{G}f_{II}^{,\psi}(x)$ such that $0 \in \overline{G}f_{II}^{,\psi}(x)$ holds if and only if the optimality condition (4.5) holds. We proceed by extension from the unconstrained case.

Definition 5.2. Let $f:\mathbb{R}^n \to \mathbb{R}$, $\psi:\mathbb{R}^n \to \mathbb{R}$ be l.L.c. and let $F \triangleq \{x \in \mathbb{R}^n \mid \psi(x) \le 0\}$. We shall say that $\overline{G}f_{II}^{,\psi}:\mathbb{R}^n \to 2^{\mathbb{R}^{n+1}}$ is a *phase II augmented convergent direction finding* map for (5.2b) if

(a) $\overline{G}_{II}^{f,\psi}(\cdot)$ is continuous and $\overline{G}_{II}^{f,\psi}(x)$ is convex for all $x \in F$.

(b) For any $x \in F$, if $\overline{\xi} = (\xi^0, \xi) \in \mathbb{R}^{n+1}$ is an element of $\overline{G}_{II}^{f,\psi}(x)$, then $\xi^0 \geq 0$.

(c) For any $x \in F$, a point $\overline{\xi} = (0, \xi)$ is an element of $\overline{G}_{II}^{f,\psi}(x)$ if and only if either $\xi \in \partial f(x)$ or $\xi \in \text{co}\{\partial f(x), \partial\psi(x)\}$ and $\psi(x) = 0$.

(d) For any $x \in F$, such that $\psi(x) < 0$, a point $\overline{\xi} = (-\psi(x), \xi)$ is an element of $\overline{G}_{II}^{f,\psi}(x)$ for all $\xi \in \partial\psi(x)$.

■

Proposition 5.3. Suppose that $f, \psi: \mathbb{R}^n \to \mathbb{R}$ are l.L.c. functions and that $\overline{G}^{f,\psi}(\cdot)$ is a phase II a.c.d.f. map for (5.2b). Then for any $x \in \mathbb{R}^n$ such that $\psi(x) \leq 0$,

(a) (i) if $\psi(x) < 0$, $0 \in \partial f(x) \iff 0 \in \overline{G}_{II}^{f,\psi}(x)$.

(ii) if $\psi(x) = 0$, $0 \in \text{co}\{\partial f(x), \partial\psi(x)\} \iff 0 \in \overline{G}_{II}^{f,\psi}(x)$.

(b) The functions $\Theta: \mathbb{R}^n \to \mathbb{R}$ and $\overline{h}: \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by

$$\Theta(x) \triangleq \min\{½\|\overline{\xi}\|^2 \mid \overline{\xi} \in \overline{G}_{II}^{f,\psi}(x)\} \tag{5.33a}$$

$$\overline{h}(x) \triangleq -\text{argmin}\{½\|\overline{\xi}\|^2 \mid \overline{\xi} \in \overline{G}_{II}^{f,\psi}(x)\} \tag{5.33b}$$

are both continuous.

(c) Writing $\overline{h}(x) = (h^0(x), h(x))$, with $h(x) \in \mathbb{R}^n$, we have

$$-h^0(x)\psi(x) + d_0\psi(x; h(x)) \leq -\Theta(x), \tag{5.33c}$$

$$d_0 f(x; h(x)) \leq -\Theta(x). \tag{5.33d}$$

Proof. a) (i) If $0 \in \partial f(x)$, then $0 \in \overline{G}_{II}^{f,\psi}$ because of (c) in Definition 5.2. Now suppose that $0 \in \overline{G}_{II}^{f,\psi}(x)$. Then, because of (c) in Definition 5.2, we must have $0 \in \partial f(x)$. (ii) This part follows directly from c) in Definition 5.2.

b) The continuity of $\Theta(\cdot)$ and $\overline{h}(\cdot)$ follows from Corollary 3.1 and the fact that the argmin in (5.33b) is a singleton.

c) By definition (5.33b), $\bar{h}(x)$ satisfies

$$\langle -\bar{h}(x), \bar{\xi} \rangle \geq \tfrac{1}{2}\|\bar{h}(x)\|^2 = \Theta(x), \quad \forall\ \bar{\xi} \in \bar{G}f_I{}^{,\psi}(x). \tag{5.34}$$

Now, let $\xi \in \partial f(x)$. Then, $\bar{\xi} = (0,\xi) \in \bar{G}_{f,\psi}(x)$ and, from (5.34) we get that

$$h^0(x)0 + \langle h(x), \xi \rangle \leq -\Theta(x). \tag{5.35}$$

Maximizing the left hand side of (5.35) over $\xi \in \partial f(x)$ we obtain (5.33d). Next, suppose that $\xi \in \partial\psi(x)$. Then $(-\psi(x), \xi) \in \bar{G}f_I{}^{,\psi}(x)$ and hence (5.34) yields

$$-h^0(x)\psi(x) + \langle h(x), \xi \rangle \leq -\Theta(x). \tag{5.36}$$

Maximizing the left hand side of (5.36) over $\xi \in \partial\psi(x)$, we obtain (5.33c).

■

Example 5.4. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are differentiable functions and that $\psi(x) \triangleq \max_{y \in Y} \varphi(x,y)$ with $Y \subset \mathbb{R}^m$ compact. The reader can easily verify that the map

$$\bar{G}f_I{}^{,\psi}(x) \triangleq \text{co}\left\{ \begin{pmatrix} 0 \\ \nabla f(x) \end{pmatrix}, \begin{pmatrix} -\varphi(x,y) \\ \nabla_x \varphi(x,y) \end{pmatrix} \right\}_{y \in Y} \tag{5.37}$$

satisfies the assumptions of Definition 5.2.

■

Next we state a phase II algorithm model for solving (5.2b) and give a proof of its convergence.

Algorithm 5.4. (Phase II for problem (5.2b). Requires a phase II a.c.d.f. map $\bar{G}f_I{}^{,\psi}(\cdot)$).

Data: $x_0 \in \mathbb{R}^n$ such that $\psi(x_0) \leq 0$.

Step 0: Set $i = 0$.

Step 1: Compute the *augmenteed search direction* $\bar{h}(x_i) = (h^0(x_i), h(x_i))$, according to

$$\bar{h}(x_i) \triangleq -\text{argmin}\{½\|\bar{\xi}\|^2 \mid \bar{\xi} \in \bar{G}f_I{}^{,\psi}(x_i)\} \tag{5.38a}$$

and set the *actual search direction* $h_i = h(x_i)$.

Step 2: Compute the *step length*

$$\lambda_i \in \lambda(x_i) \triangleq \underset{\lambda \geq 0}{\text{argmin}} \{f(x_i+\lambda h_i) \mid \psi(x_i+\lambda h_i) \leq 0\} \tag{5.38b}$$

Step 3: Update:

$$x_{i+1} = x_i + \lambda_i h_i, \tag{5.38c}$$

replace i by $i+1$ and go to step 1.

■

Remark 5.4. Again one can replace the exact minimization step size rule (5.38b) with a phase II Armijo step size rule, without altering the conclusions of the theorem below. The phase II Armijo step size rule is defined as follows:

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, f(x_i + \lambda h_i) - f(x_i) \leq -\lambda\alpha\Theta(x_i), \; \psi(x_i+\lambda h_i) \leq 0\}. \tag{5.38d}$$

where $\alpha, \beta \in (0,1)$.

■

Theorem 5.3. Suppose that $f, \psi : \mathbb{R}^n \to \mathbb{R}$ are l.L.c., that $\bar{G}f_I{}^{,\psi}(\cdot)$ is a phase II a.c.d.f. map for (5.2b). If $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 5.4 and $x_i \overset{K}{\to} \hat{x}$ as $i \to \infty$, then $\psi(\hat{x}) \leq 0$ and $0 \in \bar{G}f_I{}^{,\psi}(\hat{x})$, (i.e., $\hat{x}$ satisfies the first order condition of optimality (4.5)).

Proof. To obtain a contradiction, suppose that $0 \notin \bar{G}f_I{}^{,\psi}(\hat{x})$. Clearly, since $\psi(x_i) \leq 0$ for all i, we must have $\psi(\hat{x}) \leq 0$. We consider two cases:

a) $\psi(\hat{x}) < 0$. Then, since $\Theta(\hat{x}) > 0$, (see 5.33a), we have from (5.33d) that

$$d_0 f(\hat{x};h(\hat{x})) = -\Theta(\hat{x}) < 0 \tag{5.39}$$

Consequently, since $f(\cdot)$, $\psi(\cdot)$ and $h(\cdot)$ are continuous, there exist a $\hat{\rho} > 0$, a $\hat{\lambda} > 0$, and a $\hat{\delta} > 0$ such that

$$f(x+\hat{\lambda}h(x)) - f(x) \le -\hat{\delta}, \tag{5.40a}$$

$$\psi(x+\hat{\lambda}h(x)) \le 0, \tag{5.40b}$$

for all $x \in B(\hat{x},\hat{\rho})$. Hence, since $x_i \xrightarrow{K} \hat{x}$ as $i \to \infty$, and $f(x_{i+1}) \le f(x_i+\hat{\lambda}h(x_i))$ forall $i \in K$, there exists an i_0 such that

$$f(x_{i+1}) - f(x_i) \le -\hat{\delta} \quad \forall\, i \ge i_0, \quad i \in K. \tag{5.41}$$

Now $\{f(x_i)\}_{i=0}^{\infty}$ is monotone decreasing and $f(x_i) \xrightarrow{K} f(\hat{x})$ because $f(\cdot)$ is continuous, hence $f(x_i) \to f(\hat{x})$. But this contradicts (5.41) and therefore we must have $0 \in \overline{G}_I^{f,\psi}(\hat{x})$.

b) $\psi(\hat{x}) = 0$. In this case, since $0 \notin \overline{G}_I^{f,\psi}(\hat{x})$, it follows from (5.33c) that

$$d_0 \psi(\hat{x};h(\hat{x})) \le -\Theta(\hat{x}) < 0 \tag{5.42}$$

holds in addition to (5.39). It now follows from the continuity of $f(\cdot)$, $\psi(\cdot)$ and $h(\cdot)$ that for some $\hat{\rho} > 0$, $\hat{\lambda} > 0$, $\hat{\delta} > 0$ (5.40a), (5.40b) hold for all $x \in B(\hat{x},\hat{\rho})$. Hence, we again obtain a contradiction as for case a). This completes our proof.

■

The main disadvantage to using a two phase approach, is that the search for an initial feasible solution (phase I) does not, in any way take into account the values of the cost function, which must be minimized in phase II. A second disadvantage is that the programming of a two phase approach is somewhat cumbersome. A well executed, combined phase I - phase II approach tends to considerably alleviate both of these disadvantages.

In order to construct a phase I phase II algorithm for solving problem (5.2), we must endow a single set valued a.c.d.f. map $\overline{G}^{f,\psi}(\cdot)$ with the combined properties of the maps $\overline{G}\psi(\cdot)$ and $\overline{G}f_{II}^{,\psi}(\cdot)$, together with a "cross over" mechanism. The cross over mechanism ensures that when $\psi(x)>0$, the algorithm behaves as a phase I method which takes the cost into account, so that when $\psi(x)\gg 0$, $\overline{G}^{f,\psi}(\cdot)$ closely approximates $\overline{G}\psi(x)$, while when $\psi(x) \leq 0$, $\overline{G}^{f,\psi}(x) = \overline{G}f_{II}^{,\psi}$ and the algorithm becomes a phase II method. The construction of phase I - phase II a.c.d.f. maps can be done in a number of ways, mostly differing in the cross over mechanism that is used. For example, we can adopt the following axiomatic structure.

Definition 5.3. Let $f:\mathbb{R}^n \to \mathbb{R}$, $\psi:\mathbb{R}^n \to \mathbb{R}$ be l.L.c. and let $F \triangleq \{x \in \mathbb{R}^n \mid \psi(x) \leq 0\}$, and let $\gamma > 0$. We shall say that $\overline{G}^{f,\psi}:\mathbb{R}^n \to 2^{\mathbb{R}^{n+1}}$ is a *phase I - phase II augmented convergent direction finding* map for (5.2b) if

(a) $\overline{G}^{f,\psi}(\cdot)$ is continuous and $\overline{G}^{f,\psi}(x)$ is convex for all $x \in F$.

(b) For any $x \in F$, if $\overline{\xi} = (\xi^0,\xi) \in \mathbb{R}^{n+1}$ is an element of $\overline{G}f_{II}^{,\psi}(x)$, then $\xi^0 \geq 0$.

(c) For any $x \in \mathbb{R}^n$, a point $\overline{\xi} = (0,\xi)$ is an element of $\overline{G}^{f,\psi}(x)$ if and only if

(i) $x \in F^c$ and $\xi \in \partial\psi(x)$ or

(ii) $x \in F$ and either $\xi \in \partial f(x)$ or $\xi \in co\{\partial f(x),\partial\psi(x)\}$ and $\psi(x) = 0$.

(d) For any $x \in F$, such that $\psi(x) < 0$, a point $\overline{\xi} = (-\psi(x),\xi)$ is an element of $\overline{G}^{f,\psi}(x)$ for all $\xi \in \partial\psi(x)$.

(e) For any $x \in F^c$, a point $\overline{\xi} = (\gamma\psi(x),\xi)$ is an element of $\overline{G}^{f,\psi}(x)$ if and only if $\xi \in \partial f(x)$.

■

We note that Definition 5.3 differs from Definition 5.2 only in the additional condition e) which provides the "cross over" mechanism, and the addition of (i) to condition c). Fig. 5.3 gives an elementary illustration of the effect of the term $\gamma\psi(x)$ on the search direction when $\psi(x)>0$ and both $f(\cdot)$ and $\psi(\cdot)$ are differentiable. The reader will find it easy to verify the following result.

Proposition 5.4. Suppose that $f, \psi : \mathbb{R}^n \to \mathbb{R}$ are l.L.c. functions and that $\overline{G}^{f,\psi}(\cdot)$ is a phase - phase II a.c.d.f. map for (5.2b). Then for any $x \in \mathbb{R}^n$ such that $\psi(x) \le 0$,

(a) (i) if $\psi(x) < 0$, $0 \in \partial f(x) <=> 0 \in \overline{G}^{f,\psi}(x)$.

(ii) if $\psi(x) = 0$, $0 \in \mathrm{co}\{\partial f(x), \partial\psi(x)\} <=> 0 \in \overline{G}^{f,\psi}(x)$.

(b) The functions $\Theta : \mathbb{R}^n \to \mathbb{R}$ and $\overline{h} : \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by

$$\Theta(x) \triangleq \min\{½\|\overline{\xi}\|^2 \mid \overline{\xi} \in \overline{G}^{f,\psi}(x)\} \tag{5.43a}$$

$$\overline{h}(x) \triangleq -\mathrm{argmin}\{½\|\overline{\xi}\|^2 \mid \overline{\xi} \in \overline{G}^{f,\psi}(x)\} \tag{5.43b}$$

are both continuous.

(c) Writing $\overline{h}(x) = (h^0(x), h(x))$, with $h(x) \in \mathbb{R}^n$, we have

(i) if $\psi(x) \le 0$, then

$$-h^0(x)\psi(x) + d_0\psi(x;h(x)) \le -\Theta(x), \tag{5.43c}$$

and

$$d_0 f(x;h(x)) \le -\Theta(x), \tag{5.33d}$$

(ii) if $\psi(x) > 0$, then

$$d_0\psi(x;h(x)) \le -\Theta(x). \tag{5.33d}$$

■

Example 5.5. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are differentiable functions, that $\psi(x) \triangleq \max_{y \in Y} \varphi(x,y)$, with $Y \subset \mathbb{R}^m$ compact, and that $\gamma > 0$. Let $\psi(x)_+ \triangleq \max\{0, psi(x)\}$. The reader can easily verify that the map

$$\bar{G}^{f,\psi}(x) \triangleq \text{co}\left\{ \begin{bmatrix} \gamma\psi(x)_+ \\ \nabla f(x) \end{bmatrix}, \begin{bmatrix} \psi(x)-\varphi(x,y) \\ \nabla_x \varphi(x,y) \end{bmatrix} \right\}_{y \in Y} \tag{5.44}$$

satisfies the assumptions of Definition 5.3.

■

To conclude this section, we state a combined phase I - Phase II algorithm for solving problem (5.2b).

Algorithm 5.4. (Phase I - Phase II for problem (5.2b). Requires a phase I - phase II a.c.d.f. map $\bar{G}^{f,\psi}(\cdot)$).

Data: $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: Compute the *augmenteed search direction* $\bar{h}(x_i) = (h^0(x_i), h(x_i))$, according to

$$\bar{h}(x_i) \triangleq -\text{argmin}\{ \tfrac{1}{2}\|\bar{\xi}\|^2 \mid \bar{\xi} \in \bar{G}^{f,\psi}(x_i) \} \tag{5.45a}$$

and set the *actual search direction* $h_i = h(x_i)$.

Step 2: Compute the *step length* as follows:

if $\psi(x_i) > 0$, then

$$\lambda_i \in \lambda(x_i) \triangleq \underset{\lambda \geq 0}{\text{argmin}}\ \psi(x_i + \lambda h_i), \tag{5.45b}$$

if $\psi(x_i) \leq 0$, then

$$\lambda_i \in \lambda(x_i) \triangleq \underset{\lambda \geq 0}{\text{argmin}}\ \{ f(x_i + \lambda h_i) \mid \psi(x_i + \lambda h_i) \leq 0 \}. \tag{5.45c}$$

Step 3: Update:

$$x_{i+1} = x_i + \lambda_i h_i, \tag{5.45d}$$

replace i by $i+1$ and go to step 1.

■

Theorem 5.4. Suppose that $f, \psi: \mathbb{R}^n \to \mathbb{R}$ are l.L.c., that $\overline{G}^{f,\psi}(\cdot)$ is a phase I - phase II a.c.d.f. map for (5.2b), and that for all $x \in \mathbb{R}^n$ such that $\psi(x)>0$, $0 \notin \partial\psi(x)$. If $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 5.5 and $x_i \xrightarrow{K} \hat{x}$ as $i \to \infty$, then $\psi(\hat{x}) \leq 0$ and $0 \in \overline{G}^{f,\psi}(\hat{x})$, (i.e., $\hat{x}$ satisfies the first order condition of optimality (4.5)).

Proof. First suppose that there is an i_0 such that $\psi(x_{i_0}) \leq 0$. Then, by construction, $\psi(x_i) \leq 0$ for all $i \geq i_0$. Since $\overline{G}^{f,\psi}(\cdot)$ satisfies all the conditions defining a phase II a.c.d.f. map, it follows from Theorem 5.3 that, for this case, $\psi(\hat{x}) \leq 0$ and $0 \in \overline{G}^{f,\psi}(\hat{x})$.

Next, suppose that $\psi(x_i)>0$ for all $i \in \mathbb{N}$. Then. by repeating the arguments in the proof of Theorem 5.2, we conclude that $0 \in \overline{G}^{f,\psi}(\hat{x})$. Since $\psi(\cdot)$ is continuous, we must have that $\psi(\hat{x}) \geq 0$. Suppose that $\psi(\hat{x}) > 0$. Since $0 \in \overline{G}^{f,\psi}(\hat{x})$, we see from Definition 5.3, that we must have that $0 \in \partial\psi(\hat{x})$. But this contradicts our hypothesis on the nature of problem (5.2), and hence we must have that $\psi(\hat{x}) = 0$. This completes our proof.

■

This concludes our exposition of a first approach to the construction of semi-infinite optimization algorithms. While the approach is simple, it often results in unacceptably difficult search direction finding problems. Our second approach will therefore be to reduce this computational difficulty at the expense of an increase in algorithmic complexity.

6. SEMI-INFINITE OPTIMIZATION ALGORITHMS II: REDUCTION OF COMPUTATIONAL COMPLEXITY

We devote this section to the development of semi-infinite optimization algorithms, for solving problem (5.1), and, more generally, problem (5.2b), which are computationally more efficient than the ones presented in Section 5. The increase in computational effiency will be obtained by reducing the computational complexity of the search direction finding programs at the expense of loss of continuity in the search direction. In particular, we will show that in the search direction finding programs (5.13b) and (5.33b), the a.c.d.f. maps $\overline{G}\psi(\cdot)$ and $\overline{G}_{II}^{f,\psi}(\cdot)$, introduced in Section 5, can be replaced by subsets having lower cardinality descriptions. The effect of these replacements is to make the associated, dual search direction finding, quadratic programming problems (c.f. (5.23)) easier to solve. We shall proceed in two steps: each time gaining in efficiency and each time having to face more complex convergence analysis. Those familiar with the theory of methods of feasible directions (see [Ben.1, Pir.2, Pol.1, Pol.6, Zou.1]) will recognize that we are generalizing techniques used in methods of feasible direction. Indeed, we are relying on the fact that what is true for the simple function $\psi(x) = \max_{j \in \mathbf{m}} f^j(x)$ is also true for a much broader class of max functions.

It has been known for some time (see [Pol.6, Kiw.1]),that for the function $\psi(x) = \max_{j \in \mathbf{m}} f^j(x)$, the convergence properties of Algorithm 5.2 remain unaltered when $\overline{G}\psi(x)$, defined by (5.21) is replaced in (5.13b) by $\overline{G}_\varepsilon\psi(x) \subset \overline{G}\psi(x)$, defined below, with $\varepsilon > 0$: in (5.13b)

$$\overline{G}_\varepsilon\psi(x) \triangleq \text{co}\left\{\begin{pmatrix}\psi(x)-f^j(x)\\ \nabla f^j(x)\end{pmatrix}\right\}_{j \in I_\varepsilon(x)} \tag{6.1a}$$

where

$$I_\varepsilon(x) \triangleq \{j \in \mathbf{m} \mid \psi(x)-f^j(x) \le \varepsilon\} \tag{6.1b}$$

Clearly, when $\varepsilon > 0$ is small, the set $\overline{G}\psi_\varepsilon(\cdot)$, defined in (6.1a) is a polyhedron with a smaller number of vertices than the set $\overline{G}\psi(\cdot)$, defined in (5.21). Assuming that the search direction will be computed as in (5.23), (5.24), with $I_\varepsilon(x)$

replacing

m, we see that the use of $\overline{G}_\varepsilon\psi(x)$ leads to a lower dimensional, quadratic programming search direction computation. The effect of $\varepsilon>0$ on the search direction is shown in Fig. 6.1. Note that when $\varepsilon = \infty$, (6.1a) reduces to (5.21).

Since the index set $I_\varepsilon(x)$ can change abruptly, it is clear that the map $\overline{G}\psi_\varepsilon(\cdot)$ is not continuous. However, an examination of proofs of convergence of methods of feasible directions shows that they contain proofs that $\overline{G}_\varepsilon\psi(\cdot)$ is u.s.c.and that it is "almost" l.s.c. Axiomatizing these observations, we obtain the following modified definition of an augmented convergent direction finding map (c.f. Definition 5.1).

Definition 6.1. Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be l.L.c. We shall say that $\overline{G}\psi : \mathbb{R}^n \to 2^{\mathbb{R}^{n+1}}$ is an *efficient augmented convergent direction finding* (e.a.c.d.f.) map for $\psi(\cdot)$ if:

(a) $\overline{G}\psi(x)$ is convex for all $x \in \mathbb{R}^n$.

(b) For any $x \in \mathbb{R}^n$, if $(\xi^0,\xi) \in \overline{G}\psi(x))$, then $\xi^0 \geq 0$.

(c) For any $x \in \mathbb{R}^n$, a point $\overline{\xi} = (0,\xi)$ is an element of $\overline{G}\psi(x)$ if and only if $\xi \in \partial\psi(x)$.

(d) $G\overline{\psi}(\cdot)$ is u.s.c.

(e) For any $\hat{x} \in \mathbb{R}^n$, $\hat{\delta} > 0$, there exists a $\hat{\rho} > 0$ such that for any $\hat{\overline{\xi}} = (0,\hat{\xi}) \in \overline{G}\psi(\hat{x})$ (so that $\hat{\xi} \in \partial\psi(\hat{x})$) and any $x \in B(\hat{x},\hat{\rho})$, there exists a $\overline{\xi} = (\xi^0,\xi) \in \overline{G}\psi(x)$ such that $\|\xi - \hat{\xi}\| \leq \hat{\delta}$.

■

We see that in Definition 6.1(e), the l.s.c. relation is imposed only at those $\hat{\overline{\xi}} \in \overline{G}\psi(\hat{x})$, that have zero as their first element. Furthermore, the definition involves only the last n elements of the vector $\overline{\xi} \in \overline{G}\psi(x)$. Thus $\overline{G}\psi(\cdot)$ is "almost" l.s.c. in a rather loose sense.

Before proceeding further, we show that the map $\overline{G}_\varepsilon\psi(\cdot)$, defined by (6.1), satisfies the hypotheses stated in Definition 6.1.

Example 6.1. Consider the function $\psi(x) = \max_{j \in \mathbf{m}} f^j(x)$ with $f^j : \mathbb{R}^n \to \mathbb{R}$, $j \in \mathbf{m}$, continuously differentiable. Let $\overline{G}_\varepsilon\psi(x)$ be defined by (6.1a) with $\varepsilon > 0$. Then, referring to the requirements in Definition 6.1, we find that

(a) $\overline{G}_\varepsilon\psi(x)$ is convex for all $x \in \mathbb{R}^n$ by construction, i.e., property (a) holds.

(b) Since $\psi(x) - f^j(x) \geq 0$ for all $j \in \mathbf{m}$, property (b) holds.

(c) By (3.18), for any $j \in \mathbf{m}$, $\nabla f^j(x) \in \partial\psi(x)$ if and only if $f^j(x) = \psi(x)$. Hence property (c) holds.

(d) Suppose that x is given and $j \in I_\varepsilon(x)^c$. Then $\psi(x) - f^j(x) > \varepsilon$ and there exists a $\rho_1 > 0$ such that for all $x' \in B(x, \rho_1)$, $\psi(x') - f^j(x') > \varepsilon$ for all $j \in I_\varepsilon(x)^c$, i.e., $I_\varepsilon(x)^c \subset I_\varepsilon(x')^c$ for all $x' \in B(x, \rho_1)$. Consequently, $I_\varepsilon(x') \subset I_\varepsilon(x)$ for all $x' \in B(x, \rho_1)$. Since the functions $(\psi(\cdot) - f^j(\cdot), \nabla f^j(\cdot))$, $j \in \mathbf{m}$ are continuous, it follows that given any $\delta > 0$ there exists a $\rho \in (0, \rho_1]$ such that if $x' \in B(x, \rho)$ and $\overline{\xi}' = \sum_{j \in I_\varepsilon(x')} \mu^j((\psi(x') - f^j(x')), \nabla f^j(x')) \in \overline{G}\psi(x')$, (with $\mu^i \geq 0$, $\sum_{j \in I_\varepsilon(x')} \mu^j = 1$), then $\overline{\xi} = \sum_{j \in I_\varepsilon(x')} \mu^j(\psi(x) - f^j(x), \nabla f^j(x)) \in \overline{G}\psi(x)$ and

$$\|\overline{\xi}' - \overline{\xi}\| \leq \max_{j \in I_\varepsilon(x')} \left\| \begin{bmatrix} -f^j(x') + f^j(x) \\ \nabla f^j(x') - \nabla f^j(x) \end{bmatrix} \right\| < \delta, \tag{6.2}$$

which shows that $\overline{G}_\varepsilon\psi(\cdot)$ is u.s.c. at x, and hence u.s.c. since x is arbitrary. Thus property (d) holds.

(e) Finally we show that property (e) holds. Let $\hat{x} \in \mathbb{R}^n$ and $\hat{\delta} > 0$ be given. Then there exists a $\rho_1 > 0$ such that for all $x \in B(\hat{x}, \rho_1)$, and $j \in I_0(\hat{x})$ (i.e., $\psi(\hat{x}) - f^j(\hat{x}) = 0$), $\psi(x) - f^j(x) \leq \varepsilon$, i.e., $j \in I_\varepsilon(x)$. Thus, $I_0(\hat{x}) \subset I_\varepsilon(x)$ for all $x \in B(\hat{x}, \rho_1)$. Hence there exists a $\hat{\rho} \in (0, \rho_1]$ such that for any $(0, \hat{\xi}) \in \overline{G}\psi(\hat{x})$, and any $x \in B(\hat{x},\hat{\rho})$, we have that

(i) $\hat{\xi} \in \partial\psi(\hat{x})$, and hence that $\hat{\xi} = \sum_{j \in I_0(\hat{x})} \hat{\mu}^j \nabla f^j(\hat{x})$, with all $\hat{\mu}^j \geq 0$, and $\sum_{j \in I_0(\hat{x})} \hat{\mu}^j = 1$;

(ii) The vector $\bar{\xi}$ defined by $\bar{\xi} = (\xi^0,\xi) = (\sum_{j \in I_0(\hat{x})} \hat{\mu}^j (\psi(x)-f^j(x)), \sum_{j \in I_0(\hat{x})} \hat{\mu}^j \nabla f^j(x)) \in \bar{G}_\varepsilon\psi(x))$, because $I_0(\hat{x}) \subset I_\varepsilon(x)$, and

(iii) For $\bar{\xi}$ defined as in (ii) , above, we have $\|\xi-\hat{\xi}\| \leq \max_{j \in I_0(\hat{x})} \|\nabla f^j(x)-\nabla f^j(\hat{x})\| \leq \hat{\delta}$, so that property (e) does indeed hold.

We conclude that $\bar{G}_\varepsilon\psi(x)$ is an e.a.c.d.f. map in the sense of Definition 6.1.

■

Example 6.2. Consider the function $\psi(x) = \max_{y \in Y} \varphi(x,y)$, with $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ continuous, $\nabla_x \varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ continuous and $Y \subset \mathbb{R}^m$ compact. The reader can easily deduce from our analysis to be given in Example 6.5 that the following, obvious generalization of (6.1) defines, for $\varepsilon > 0$, an e.a.c.d.f. map for $\psi(\cdot)$:

$$\bar{G}'_\varepsilon\psi(x) \triangleq \text{co}\left\{\begin{pmatrix}\psi(x)-\varphi(x,y) \\ \nabla_x \varphi(x,y)\end{pmatrix}\right\}_{y \in Y_\varepsilon(x)}, \tag{6.3a}$$

where

$$Y_\varepsilon(x) \triangleq \{y \in Y \mid \psi(x)-\varphi(x,y) \leq \varepsilon\}. \tag{6.3b}$$

Next, again with $\varepsilon > 0$, we define a much smaller set valued map by

$$\bar{G}''_\varepsilon\psi(x) \triangleq \text{co}\left\{\begin{pmatrix}\psi(x)-\varphi(x,y) \\ \nabla_x \varphi(x,y)\end{pmatrix}\right\}_{y \in \hat{Y}_\varepsilon(x)}, \tag{6.4a}$$

where

$$\hat{Y}_\varepsilon(x) \triangleq \{y \in Y_\varepsilon(x) \mid y \text{ is a local maximizer of } \varphi(x,\cdot) \text{ in } Y\}. \tag{6.4b}$$

It can be shown, by emulating the analysis to be given in Example 6.6, that if

$\hat{Y}_\varepsilon(x)$ has finite cardinality for all $x \in \mathbb{R}^n$, then $\overline{G}''_\varepsilon\psi(x)$ is an e.a.c.d.f. map.

The e.a.c.d.f. map $\overline{G}''_\varepsilon\psi(x)$ is of importance in engineering design, because by their very nature, dynamic responses, usually, are not flat and hence the corresponding set $\hat{Y}_\varepsilon(x)$ contains only a finite number of points.

■

The following result facilitates the proof that Algorithm 5.2 remains convergent in the sense of Theorem 5.2 when Definition 6.1 is substituted for Definition 5.1.

Lemma 6.1. Suppose that $\overline{G}\psi(\cdot)$ is an e.a.c.d.f. map for a l.L.c. function $\psi:\mathbb{R}^n \to \mathbb{R}$. Then given any $\hat{x} \in \mathbb{R}^n$ and $\hat{\delta} > 0$, there exists a $\hat{\rho} > 0$ such that for all $x', x'' \in B(\hat{x}, \hat{\rho})$, if $(0, \xi') \in \overline{G}\psi(x')$, then there exists a $\overline{\xi}'' = (\xi''^0, \xi'') \in \overline{G}\psi(x'')$ such that

$$\|\xi' - \xi''\| \le \hat{\delta}. \tag{6.5}$$

Proof: Since $\partial\psi(\cdot)$ is u.s.c., there exists a $\rho_1 > 0$ such that for all $x' \in B(\hat{x}, \rho_1)$,

$$\max_{\xi' \in \partial\psi(x')} \min_{\hat{\xi} \in \partial\psi(\hat{x})} \|\xi' - \hat{\xi}\| \le \hat{\delta}/2, \tag{6.6a}$$

and by (e) of Definition 6.1, there exists a $\hat{\rho} \in (0, \rho_1]$ such that for all $x'' \in B(\hat{x}, \hat{\rho})$, given a $\hat{\xi} \in \partial\psi(\hat{x})$, there exists a $\overline{\xi}'' \in \overline{G}\psi(x'')$, such that $\overline{\xi}'' = (\xi''^0, \xi'')$ and

$$\|\xi'' - \hat{\xi}\| \le \hat{\delta}/2. \tag{6.6b}$$

Hence, if $x', x'' \in B(\hat{x}, \hat{\rho})$, and $\xi' \in \partial\psi(x')$, then by (6.6a), there exists a $\hat{\xi} \in \partial\psi(\hat{x})$ such that $\|\xi' - \hat{\xi}\| \le \hat{\delta}/2$ and by (6.6b) there exists a $\overline{\xi}'' \in \overline{G}\psi(x'')$ such that $\|\hat{\xi} - \xi''\| \le \hat{\delta}/2$. Hence we see that (6.5) must hold. This completes our proof.

■

Theorem 6.1. Consider Algorithm 5.2, in which the search direction is computed as in (5.13b), with $\overline{G}\psi(\cdot)$ satisfying the hypotheses of Definition 6.1. Then any accumulation point $\hat{x}$ of a sequence $\{x_i\}_{i=0}^{\infty}$, constructed by this algorithm, satisfies $0 \in \partial\psi(\hat{x})$.

Proof. Suppose that $x_i \overset{K}{\to} \hat{x}$ as $i \to \infty$ and that $0 \notin \partial\psi(\hat{x})$. Then $0 \notin \overline{G}\psi(\hat{x})$ and hence $\Theta(\hat{x}) > 0$ defined in see 5.13a). Since $\overline{G}\psi(\cdot)$ and $\partial\psi(\cdot)$ are both u.s.c., $\Theta(\cdot)$ is l.s.c. Hence there exist an i_0 and a $b \in (0,\infty)$ such that for all $i \geq i_0$, $i \in K$, $\Theta(x_i) \geq \Theta(\hat{x})/2$ and $\|h(x_i)\| \leq b$. Next, for any $i \geq i_0$, $i \in K$, any $\lambda>0$ and any $\overline{\xi} = (\xi_i^0, \xi_i) \in \overline{G}\psi(x_i)$, we have that

$$\begin{aligned}\psi(x_{i+1}) - \psi(x_i) &\leq \psi(x_i + \lambda h(x_i)) - \psi(x_i)\\ &= \lambda\langle \xi_{i\lambda}, h(x_i)\rangle\\ &= \lambda[\langle \xi_{i\lambda} - \xi_i, h(x_i)\rangle - \xi_i^0 h^0(x_i) + \langle \xi_i, h(x_i)\rangle + \xi_i^0 h^0(x_i)]\\ &\leq \lambda[\|\xi_{i\lambda} - \xi_i\| \, \|h(x_i)\| - 2\Theta(x_i)],\end{aligned} \tag{6.7a}$$

where $\xi_{i\lambda} \in \partial\psi(x_i + s\lambda h(x_i))$, with $s \in [0,1]$, by the Lebourg Mean Value Theorem 2.3.1, and where we have made use of the fact that $\xi_i^0 h^0(x_i) \geq 0$ by Definition 6.1(b). Since for all $i \in K$, $i \geq i_0$, $\|h(x_i)\| \leq b$, it follows from Lemma 6.1 that there exists a $\hat{\lambda} > 0$ and an $i_1 \geq i_0$ such that for all $i \in K$, $i \geq i_1$, and any $\xi_{i\hat{\lambda}} \in \partial\psi(x_i + s\hat{\lambda}h(x_i))$, with $s \in (0,1)$, there exists a $\overline{\xi}_i = (\xi_i^0, \xi_i) \in \overline{G}\psi(x_i)$, satisfying $b\|\xi_{i\hat{\lambda}} - \xi_i\| \leq \Theta(\hat{x})/2$. Since $-\Theta(\cdot)$ is l.s.c., there exists an $i_2 \geq i_1$, such that for all $i \in K$, $i \geq i_2$, $-\Theta(x_i) \geq -\Theta(\hat{x})/2$. Hence, for all $i \in K$, $i \geq i_2$, we obtain from (6.7a) and the above that

$$\begin{aligned}\psi(x_{i+1}) - \psi(x_i) &\leq \psi(x_i + \hat{\lambda}h(x_i)) - \psi(x_i)\\ &\leq -\hat{\lambda}\Theta(\hat{x})/2.\end{aligned} \tag{6.7b}$$

However, $\psi(x_i) \overset{K}{\to} \psi(\hat{x})$ as $i \to \infty$, because $\psi(\cdot)$ is continuous. Therefore, since $\{\psi(x_i)\}_{i=0}^{\infty}$ is monotone decreasing, $\psi(x_i) \to \psi(\hat{x})$ as $i \to \infty$, which contradicts (6.7b). We conclude that $0 \in \partial\psi(\hat{x})$. This completes our proof.

■

Remark 6.1. In implementing Algorithm 5.2, with e.a.c.d.f. maps, the Armijo step size rule (5.10) should be used instead of the exact line search in Algorithm 5.2, since the Armijo step size rule is much more efficient and does not affect the conclusions of Theorem 6.1.

■

We leave it as an exercise for the reader to relax the continuity assumption in Definitions 5.2 and 5.3, along the lines of Definition 6.1, to obtain more efficient search direction finding rules for the constrained problem (5.2). Instead we proceed to describe an even more efficient scheme which we will develop in full. To guide our intuition we re-examine the sets defined in (6.1a), and (6.4a). Since $\varepsilon > 0$ is fixed in these expressions the cardinality of the sets $I_\varepsilon(x)$, $\hat{Y}_\varepsilon(x)$ can be much larger than that of $I_0(x)$ or $\hat{Y}_0(x)$. Consequently, the question arises whether it is not possible to drive ε to zero as a solution point is approached and, progressively, obtain simpler and simpler search direction finding problems. We note that as $\varepsilon \to 0$, $\psi(x) - \varphi(x,y) \to 0$ for all $y \in \hat{Y}_\varepsilon(x)$, in (6.4a), and hence it seems that the term $\psi(x) - \varphi(x,y)$ can be replaced by 0 in $\bar{G}''_\varepsilon(x)$ without any ill effect. When this is done $\bar{G}''_\varepsilon\psi(x)$ effectively becomes a subset of $\mathbb{R}^n$ rather than a subset of $\mathbb{R}^{n+1}$. These considerations (supported by the fact that they are known to work (see [Gon.1], [Pol.10])) lead to the following definition of convergent direction finding maps which will eventually have to be used in conjunction with an ε -reduction rule.

Definition 6.2. We shall say that $\{G_\varepsilon\psi(\cdot)\}_{\varepsilon\geq 0}$, where $G_\varepsilon\psi:\mathbb{R}^n \to 2^{\mathbb{R}^n}$, is a family of *convergent direction finding* (c.d.f.) *maps* for the l.L.c. function $\psi:\mathbb{R}^n \to \mathbb{R}$ if

(a) For all $x \in \mathbb{R}^n$, $\partial\psi(x) = G_0\psi(x)$.

(b) For all $x \in \mathbb{R}^n$, if $0 \leq \varepsilon < \varepsilon'$, then $G_\varepsilon\psi(x) \subset G_{\varepsilon'}\psi(x)$.

(c) For any $\varepsilon \geq 0$ and $x \in \mathbb{R}^n$, $G_\varepsilon\psi(x)$ is convex.

(d) For any $\varepsilon \geq 0$, $G_\varepsilon\psi(x)$ is bounded on bounded sets.

(e) $G_\varepsilon\psi(x)$ is u.s.c. in (ε, x) at $(0, \hat{x})$ for all $\hat{x} \in \mathbb{R}^n$.

(f) For any $\hat{x} \in \mathbb{R}^n$, $\hat{\varepsilon} > 0$ and $\hat{\delta} > 0$ there exists a $\hat{\rho} > 0$ such that for any $\hat{\xi} \in \partial\psi(\hat{x})$ and any $x \in B(\hat{x}, \hat{\rho})$, there exists a $\xi \in G_{\hat{\varepsilon}}\psi(x)$ such that $\|\xi - \hat{\xi}\| \leq \hat{\delta}$.

■

We note that the property (f) above is analogous to property (e) in Definition 6.1. It is also possible to define ε -families of efficient augmented convergent direction finding maps by modifying Definition 6.2 only very slightly, as follows.

Definition 6.2a. We shall say that $\{\overline{G}_\varepsilon\psi(\cdot)\}_{\varepsilon\geq 0}$, where $\overline{G}_\varepsilon\psi: \mathbb{R}^n \to 2^{\mathbb{R}^n}$, is a family of *efficient augmented convergent direction finding* (e.a.c.d.f.) *maps* for the l.L.c. function $\psi: \mathbb{R}^n \to \mathbb{R}$ if

(a) For all $x \in \mathbb{R}^n$, a point $\overline{\xi} = (\xi^0, \xi)$ is an element of $\overline{G}_0\psi(x)$ if and only if $\xi \in \partial\psi(x)$.

(b) For any $x \in \mathbb{R}^n$, if $(\xi^0, \xi) \in \overline{G}(x)$, then $\xi^0 \geq 0$.

(c) For all $x \in \mathbb{R}^n$, if $0 \leq \varepsilon < \varepsilon'$, then $\overline{G}_\varepsilon\psi(x) \subset \overline{G}_{\varepsilon'}\psi(x)$.

(d) For any $\varepsilon \geq 0$ and $x \in \mathbb{R}^n$, $\overline{G}_\varepsilon\psi(x)$ is convex.

(e) For any $\varepsilon \geq 0$, $\overline{G}_\varepsilon\psi(x)$ is bounded on bounded sets.

(f) $\overline{G}_\varepsilon\psi(x)$ is u.s.c. in (ε, x) at $(0, \hat{x})$ for all $\hat{x} \in \mathbb{R}^n$.

(g) For any $\hat{x} \in \mathbb{R}^n$, $\hat{\varepsilon} > 0$ and $\hat{\delta} > 0$ there exists a $\hat{\rho} > 0$ such that for any $\hat{\xi} \in \partial\psi(\hat{x})$ and any $x \in B(\hat{x}, \hat{\rho})$, there exists a $\overline{\xi} = (\xi^0, \xi) \in \overline{G}_{\hat{\varepsilon}}\psi(x)$ such that $\|\xi - \hat{\xi}\| \leq \hat{\delta}$.

■

Before proceeding with the construction of an algorithm based on Definition 6.2, we derive c.d.f. maps for a few functions, which frequently occur in engineering design.

Example 6.3. Suppose that $f^j:\mathbb{R}^n \to \mathbb{R}$, $j \in \mathbf{m}$ are continuously differentiable functions, and that $\psi(x) \triangleq \max_{j \in \mathbf{m}} f^j(x)$. We shall show that the family of maps $\{G_\varepsilon\psi(\cdot)\}_{\varepsilon\ge 0}$ defined by

$$G_\varepsilon\psi(x) \triangleq \text{co}\{\nabla f^j(x)\}_{j \in I_\varepsilon(x)}, \quad \varepsilon \ge 0, \tag{6.8}$$

with I_ε defined in (6.1b), is a family of c.d.f. maps for $\psi(x)$.

(a) Clearly, $G_0\psi(x) = \partial\psi(x)$ for all $x \in \mathbb{R}^n$.

(b) Since $0 \le \varepsilon < \varepsilon'$ implies that $I_\varepsilon(x) \subset I_{\varepsilon'}(x)$, we must have $G_\varepsilon\psi(x) \subset G_\varepsilon{}'\psi(x)$.

(c) $G_\varepsilon\psi(x)$ is convex by definition.

(d) For any $\varepsilon \ge 0$, $G_\varepsilon\psi(x) \subset \text{co}\{\nabla f^j(x)\}_{j \in \mathbf{m}}$. Since the $\nabla f^j(\cdot)$ are all continuous, it follows that $G_\varepsilon\psi(x)$ is bounded on bounded sets for any $\varepsilon \ge 0$.

(e) Consider the point $(0,\hat{x})$. If $j \in \mathbf{m}$ is such that $j \notin I_0(\hat{x})$, then $\psi(\hat{x}) - f^j(\hat{x}) > 0$. Hence there exists a $\hat{\rho} > 0$ and an $\hat{\varepsilon} > 0$ such that $j \notin I_\varepsilon(x)$ for all $x \in B(\hat{x},\hat{\rho})$, $\varepsilon \in [0,\hat{\varepsilon}]$; i.e., for all $x \in B(\hat{x},\hat{\rho})$ and $\varepsilon \in [0,\hat{\varepsilon}]$, $I_0(\hat{x}) \supset I_\varepsilon(x)$. Therefore, since the $\nabla f^j(\cdot)$ are continuous and finite in number, if $\varepsilon_i \to 0$ and $x_i \to \hat{x}$ as $i \to \infty$, are arbitrary sequences, then we must have $\overline{\text{Lim}}\, G_{\varepsilon_i}\psi(x_i) \subset G_0\psi(\hat{x})$, i.e., $G_\varepsilon\psi(x)$ is u.s.c. at $(0,\hat{x})$.

(f) Let $\hat{x}$, $\hat{\varepsilon} > 0$ and $\hat{\delta} > 0$ be given. First, since for all $j \in I_0(\hat{x})$, we have $\psi(\hat{x}) - f^j(\hat{x}) = 0$, there exists a $\rho_1 > 0$ such that $\psi(x) - f^j(x) \le \varepsilon$ for all $x \in B(\hat{x},\rho_1)$ and $j \in I_0(\hat{x})$, i.e., $I_0(\hat{x}) \subset I_{\hat{\varepsilon}}(x)$ for all $x \in B(\hat{x},\rho_1)$. Next, there exists a $\hat{\rho} \in (0,\rho_1]$ such that $\|\nabla f^j(x) - \nabla f^j(\hat{x})\| \le \hat{\delta}$ for all $x \in B(\hat{x},\hat{\rho})$ and all $j \in \mathbf{m}$. Hence, if $\hat{\xi} \triangleq \sum_{j \in I_0(\hat{x})} \hat{\mu}^j \nabla f^j(\hat{x})$, with $\hat{\mu}^j \ge 0$ and $\sum_{j \in I_0(\hat{x})} \hat{\mu}^j = 1$, is any point in $\partial\psi(\hat{x})$, then for any $x \in B(\hat{x},\hat{\rho})$ there exists a $\xi \triangleq \sum_{j \in I_0(\hat{x})} \hat{\mu}^j \nabla f^j(x)$ in $G_\varepsilon\psi(x)$ such that $\|\xi - \hat{\xi}\| = \|\sum_{j \in I_0} \hat{\mu}^j (\nabla f^j(x) - \nabla f^j(\hat{x}))\| \le \hat{\delta}$.

■

Example 6.4. Suppose that $\psi(x) = \max_{y \in Y} \varphi(x,y)$ where $\varphi:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous, $\nabla_x \varphi(\cdot,\cdot)$ exists and is continuous and $Y \subset \mathbb{R}^m$ is compact. We shall show that the family of maps $\{G_\varepsilon \psi(\cdot)\}_{\varepsilon \geq 0}$ defined by

$$G_\varepsilon \psi(x) \triangleq \text{co}\{\nabla_x \varphi(x,y)\}_{y \in Y_\varepsilon(x)}, \tag{6.9}$$

where $Y_\varepsilon(x)$ was defined in (6.3b), is a family of c.d.f. maps.

(a) Clearly, $G_0 \psi(x) = \partial \psi(x)$ for all $x \in \mathbb{R}^n$.

(b) Since $0 \leq \varepsilon < \varepsilon'$ implies that $Y_\varepsilon(x) \subset Y_{\varepsilon'}(x)$, we must have that $G_\varepsilon \psi(x) \subset G_{\varepsilon'} \psi(x)$.

(c) $G_\varepsilon \psi(x)$ is convex by definition.

(d) For any $\varepsilon \geq 0$, $G_\varepsilon \psi(x) \subset \text{co}\{\nabla_x \varphi(x,y)\}_{y \in Y}$, and hence, since $\nabla_x \varphi(x,y)$ is continuous and Y is compact, $G_\varepsilon \psi(x)$ is bounded on bounded sets uniformly in $\varepsilon \geq 0$.

(e) Suppose that as $i \to \infty$, $x_i \to \hat{x} \in \mathbb{R}^n$, $\varepsilon_i \to 0$, $y_i \in Y_{\varepsilon_i}(x_i)$, and $y_i \to \hat{y} \in Y$. Then $\psi(x_i) - \varphi(x_i,y_i) \leq \varepsilon_i$ and since $\psi(\cdot)$ and $\varphi(\cdot,\cdot)$ are continuous, it follows that $\hat{y} \in Y_0(\hat{x})$ and hence that $Y_\varepsilon(x)$ is u.s.c. in (ε,x) at any $(0,\hat{x})$. Since $\nabla_x \varphi(\cdot,\cdot)$ is continuous, it now follows that $G_\varepsilon \psi(x)$ is u.s.c. in (ε,x) at any $(0,\hat{x})$.

(f) Let $\hat{x} \in \mathbb{R}^n$, $\hat{\varepsilon} > 0$, $\hat{\delta} > 0$ be given. First, there exists a $\rho_1 > 0$ such that for all $x \in B(\hat{x},\rho_1)$ and $y \in Y_0(\hat{x})$, $\psi(x) - \varphi(x,y) \leq \hat{\varepsilon}$, i.e., $Y_0(\hat{x}) \subset Y_{\hat{\varepsilon}}(x)$, for all $x \in B(\hat{x},\hat{\rho})$. Next, there exists a $\hat{\rho} \in (0,\rho_1]$ such that $\|\nabla_x \varphi(x,y) - \nabla_x \varphi(\hat{x},y)\| \leq \hat{\delta}$ for all $x \in B(\hat{x},\hat{\rho})$ and all $y \in Y$. Now, if $\xi \in \partial \psi(\hat{x})$, then, by Carathedory's Theorem 2.2.1, there exist $n+1$ vectors $y_j \in Y$ and scalars $\hat{\mu}^j \geq 0$ with $\sum_{j=1}^{n+1} \hat{\mu}^j = 1$ such that $\xi = \sum_{j=1}^{n+1} \hat{\mu}^j \nabla_x \varphi(\hat{x},y_j)$. Since for $x \in B(\hat{x},\hat{\rho})$, $Y_0(\hat{x}) \subset Y_{\hat{\varepsilon}}(x)$, the vector $\xi = \sum_{j=1}^{n+1} \hat{\mu}^j \nabla_x \varphi(\hat{x},y_j) \in G_{\hat{\varepsilon}} \psi(x)$ and $\|\hat{\xi} - \xi\| \leq \hat{\delta}$.

This completes the proof that $\{G_\varepsilon \psi(\cdot)\}_{\varepsilon \geq 0}$, defined by (6.9), is a family of c.d.f. maps.

■

Example 6.5. It is worth considering a special case of the function $\psi(\cdot)$ that was dealt with in Example 6.4. This particular form of $\psi(\cdot)$ occurs in multivariable feedback control system design when the frequency variable is discretized (see [Pol.10]), as well as in mechanical vibration control problems in which beams, plates, etc. are approximated by ordinary differential equations. An extension of the result, below, has also been used in a vibration control problem in which the beam was described by a partial differential equation [War.1].

Suppose that $Q:\mathbb{R}^n \to \mathbb{C}^{m\times m}$, is continuously differentiable, and that for all $x \in \mathbb{R}^n$, $Q(x)$ is an $m\times m$, symmetric, positive definite, complex valued matrix. Let $\varphi(x,y) \triangleq \langle y, Q(x)y\rangle$ and $Y \triangleq \{y \in \mathbb{R}^m \mid \|y\| = 1\}$. Then we define

$$\psi(x) \triangleq \lambda_{\max}[Q(x)] = \max_{\|y\|=1}\, \langle y, Q(x)y\rangle \tag{6.10a}$$

where $\lambda_{\max}[\cdot]$ is the largest eigenvalue of $Q(x)$. In this case, (referring to (3.25)) we see that (6.9) assumes the form

$$G_\varepsilon\psi(x) = \mathrm{co}\{v \in \mathbb{R}^n \mid v^i = \langle U_\varepsilon(x)z, \frac{\partial Q(x)}{\partial x^i}U_\varepsilon(x)z\rangle, \quad i = 1,2,\ldots,n, \|z\| = 1\}, \tag{6.10b}$$

where $U_\varepsilon(x)$ is a matrix of orthonormal eigenvectors corresponding to the eigenvalues $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \lambda_{k_\varepsilon}$ of $Q(x)$ such that $\lambda_{\max} - \lambda_i \leq \varepsilon$ for $i = 1,2,\ldots,k_\varepsilon$ and $\lambda_{\max} - \lambda_i > \varepsilon$ for $i > k_\varepsilon$ (for an alternative definition of $G_\varepsilon\psi(x)$ see [Pol.10].

■

Example 6.6. Next, we consider another important special case of the function $\psi(x) = \max_{y\in Y}\varphi(x,y)$, where $\varphi:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ is continuously differentiable, and suppose that $Y \subset \mathbb{R}$ is a compact interval. This is one of very few cases where the construction of the set $\hat{Y}_\varepsilon(x)$, defined in (6.4b) is computationally simple. As we have seen in the examples given in the Introduction, this type of ψ function arises when $\varphi(x,\cdot)$ is the time or frequency response of a dynamical system to a given input. A characteristic of dynamic systems is that their responses do not remain flat over intervals of time/frequency and hence they have only a finite number of local maximizers within a finite time/frequency interval, so that the cardinality of the set $\hat{Y}_\varepsilon(x)$ is usually finite. We now *assume* that $\hat{Y}_\varepsilon(x)$ is finite for all $x \in \mathbb{R}^n$ and any $\varepsilon \geq 0$. We define a family of c.d.f. maps $\{G_\varepsilon\psi(\cdot)\}_{\varepsilon\geq 0}$ for the function $\psi(\cdot)$ being considered by

$$G_\varepsilon\psi(x) \triangleq \text{co}\{\nabla_x\varphi(x,y)\}_{y \in \hat{Y}_\varepsilon(x)}. \tag{6.11}$$

It remains to be shown that as defined by (6.11), the maps $G_\varepsilon\psi(\cdot)$ satisfy the postulates of Definition 6.2. We proceed one part at a time.

(a) For all $x \in \mathbb{R}^n$, $G_0\psi(x) = \partial\psi(x)$ by (3.23a).

(b) Since $0 \leq \varepsilon < \varepsilon'$ implies that $\hat{Y}_\varepsilon(x) \subset \hat{Y}_{\varepsilon'}(x)$, we must have $G_\varepsilon\psi(x) \subset G_{\varepsilon'}\psi(x)$ for all $x \in \mathbb{R}^n$.

(c) $G_\varepsilon\psi(x)$ is convex by definition for all $x \in \mathbb{R}^n$.

(d) Since $\nabla_x\varphi(\cdot,\cdot)$ is continuous and Y is compact, the set valued map $\text{co}\{\nabla_x\varphi(x,y)\}_{y \in Y}$ is bounded on bounded sets and hence $G_\varepsilon\psi(x)$, which is contained in it for all $\varepsilon \geq 0$, is bounded on bounded sets.

(e) It was shown in Example 6.4 that $Y_\varepsilon(x)$ is u.s.c. in (ε,x) at $(0,\hat{x})$ for any $\hat{x} \in \mathbb{R}^n$. Since $\hat{Y}_\varepsilon(x)$ is closed and $\hat{Y}_0(x) = Y_0(x)$ and $\hat{Y}_\varepsilon(x) \subset Y_\varepsilon(x)$, it follows that $\hat{Y}_\varepsilon(x)$ is u.s.c. in (ε,x) at $(0,\hat{x})$ for any $\hat{x} \in \mathbb{R}^n$.

(f) First, for any $\varepsilon \geq 0$, let $\mu_\varepsilon(x) \triangleq$ Lebesgue measure of $Y_\varepsilon(x)$. Then, it can be seen that $\mu_\varepsilon(\cdot)$ is continuous, that $\mu_\varepsilon(x)$ is continuous in ε,x at $(0,x)$, and that $0 \leq \varepsilon' < \varepsilon''$ implies that $\mu_{\varepsilon'}(x) \leq \mu_{\varepsilon''}(x)$. Now, let $\hat{x} \in \mathbb{R}^n$, $\hat{\varepsilon} > 0$ and $\hat{\delta} > 0$ be given. Since $\nabla_x\varphi(\cdot,\cdot)$ is continuous and Y is compact, there exists a $\rho_1 > 0$ such that for all $x \in B(\hat{x},\rho_1)$ and $\hat{y},y \in Y$ such that $|\hat{y} - y| \leq \rho_1$,

$$\|\nabla_x \varphi(x,y) - \nabla_x \varphi(\hat{x},\hat{y})\| \le \hat{\delta}. \tag{6.12}$$

Since the set $Y_0(\hat{x})$ is finite by assumption, there exists an $\varepsilon_1 \in (0,\hat{\varepsilon}]$ such that $\mu_{\varepsilon_1}(\hat{x}) \le \rho_1/2$. Hence, by the continuity of $\mu_{\varepsilon_1}(\cdot)$, there exists a $\rho_2 \in (0,\rho_1]$ such that $\mu_{\varepsilon_1}(x) \le \rho_1$ for all $x \in B(\hat{x},\rho_2)$.

Finally, referring to Example 6.4, we see that there exists a $\hat{\rho} \in (0,\rho_2)$ such that $\hat{Y}_0(\hat{x}) \subset Y_{\varepsilon_1}(x) \subset Y_{\hat{\varepsilon}}(x)$ for all $x \in B(\hat{x},\hat{\rho})$.

Let $\hat{y}_i \in \hat{Y}_0(\hat{x})$ and $x \in B(\hat{x},\hat{\rho})$ be arbitrary. Then $\hat{y}_i \in Y_{\varepsilon_1}(x) \subset Y_{\hat{\varepsilon}}(x)$. Since $\mu_{\varepsilon_1}(x) \le \rho_1$ and since each disjoint interval in $Y_{\varepsilon_1}(x)$ must contain at least one point of $\hat{Y}_{\varepsilon_1}(x) \subset \hat{Y}_{\hat{\varepsilon}}(x)$, there exists a $y_i \in \hat{Y}_{\varepsilon_1}(x) \subset \hat{Y}_{\hat{\varepsilon}}(x)$ such that $|y_i - \hat{y}_i| \le \rho_1$. Hence $\|\nabla_x \varphi(\hat{x},\hat{y}_i) - \nabla_x \varphi(x,y_i)\| \le \hat{\delta}$.

Therefore, for any $\hat{y} \in \partial\psi(\hat{x})$, we have (a) $\hat{y} = \sum_{i=1}^{k} \hat{\mu}^i \nabla\varphi_x(\hat{x},\hat{y}_i)$, where k is the cardinality of $\hat{Y}_0(\hat{x})$, $\hat{y}_i \in \hat{Y}_0(\hat{x})$, with $\hat{\mu}^i \ge 0$ for all $i \in \underline{k}$ and $\sum_{i=1}^{k} \hat{\mu}^i = 1$; and (b) there exist $y_i \in \hat{Y}_{\hat{\varepsilon}}(x)$ such that $y \triangleq \sum_{i=1}^{k} \hat{\mu}^i \nabla\varphi_x(x,y_i) \in G_{\hat{\varepsilon}}\psi(x)$ and

$$\|y - \hat{y}\| = \|\sum_{i=1}^{k} \hat{\mu}^i [\nabla_x \varphi(x,y_i) - \nabla_x \varphi(\hat{x},\hat{y}_i)]\| \le \hat{\delta}, \tag{6.13}$$

which completes our proof.

■

We note that the assumption that $\hat{Y}_\varepsilon(x)$ has finite cardinality has two effects. The first is that the search direction finding problem $h_\varepsilon(x) = -\text{argmin}\{\|h\|^2 \, h \in G_\varepsilon\psi(x)\}$, yields a simple dual quadratic programming problem which takes very little time to resolve (by a quadratic programming code, such as [Gil.1], if it does not fail due to the semi-definiteness of the quadratic form in (6.14a), below, or by the Wolfe proximity algorithm [Wol.1] otherwise):

$$\mu_\varepsilon(x) \in \arg\min_{\mu}\left\{ \tfrac{1}{2}\| \sum_{j=1}^{k} \mu^j \nabla\varphi_x(x,y_j)\|^2 \mid \mu^j \geq 0, \sum_{j=1}^{k} \mu^j = 1 \right\} \qquad \mathbf{(6.14a)}$$

which, in turn leads to

$$h_\varepsilon(x) = -\sum_{j=1}^{k} \mu_\varepsilon^j(x) \nabla_x \varphi(x,y_j). \qquad (6.14b)$$

Since the quadratic form in (6.14a) may be only positive semi-definite and hence the solution $\mu_\varepsilon(x)$ need not be unique. However, the search direction $h_\varepsilon(x)$ which it defines is unique.

The second effect of the finite cardinality assumption is to be found in the proof, given above, that property (f) of Definition 6.2 is satisfied by the sets $G_\varepsilon\psi(x)$ defined in (6.11). It it easy to construct an example which shows that when $Y_0(x)$ contains intervals of finite length, property (f) will fail to hold for $G_\varepsilon\psi(x)$ defined in (6.11).

■

Example 6.7. Consider again the case of the function $\psi : \mathbb{R}^n \to \mathbb{R}$, introduced in Example 6.5, but in somewhat greater generality, as would be the case in multivariable linear feedback control system design, when one wishes to suppress the maximum singular value of a frequency response matrix over a specified frequency range:

$$\psi(x) \triangleq \max_{\omega \in \Omega}\{\lambda_{\max}[Q(x,j\omega)] - b(\omega)\}, \qquad (6.15)$$

where $Q(x,j\omega)$, defined on $\mathbb{R}^n \times \mathbb{R}$ is an $m \times m$, symmetric, positive semi-definite, continuously differentiable, complex valued matrix, $\Omega \subset \mathbb{R}$ is compact, and $b : \mathbb{R} \to \mathbb{R}$ is continuous. Let $y \triangleq (\omega,u)$ with $u \in \mathbb{R}^m$,and let $Y \triangleq \Omega \times U$, where $U = \{u \in \mathbb{R}^m \mid \|u\| = 1\}$. If we define $\varphi(x,y) \triangleq \langle u, Q(x,j\omega)u \rangle - b(\omega)$, we find that (6.15) is equivalent to $\psi(x) = \max_{y \in Y} \varphi(x,y)$. For any $\varepsilon \geq 0$, $x \in \mathbb{R}^n$, we define

$$\hat{\Omega}_\varepsilon(x) \triangleq \{\omega \in \Omega \mid \psi(x) - (\lambda_{\max}[Q(x,j\omega)] - b(\omega)) \leq \varepsilon, \text{ and } \omega \text{ is a local maximizer of } (\lambda_{\max}[Q(x,j\omega)] - b(\omega)) \text{ in } \Omega\}. \tag{6.16a}$$

Suppose now that the cardinality of $\hat{\Omega}_\varepsilon(x)$ is finite for all $x \in \mathbb{R}^n$ and that $\varepsilon \geq 0$. Consider the maps $G_\varepsilon\psi(x)$ defined, with $\varepsilon \geq 0$, by

$$G_\varepsilon\psi(x) \triangleq \text{co}\{v \in \mathbb{R}^n \mid v^i = \langle U_\varepsilon(x,j\omega), \frac{\partial Q(x,j\omega)}{\partial x^i} U_\varepsilon(x,j\omega)z\rangle, \omega \in \hat{\Omega}_\varepsilon(x), \|z\| = 1, i \in \underline{n}\}, \tag{6.16b}$$

where $U_\varepsilon(x,j\omega)$ is a matrix of orthonormal eigenvectors of $Q(x,j\omega)$ corresponding to the eigenvalues $\lambda_k(x,j\omega)$ which satisfy $\psi(x) - \lambda_k(x,j\omega) - b(\omega) \leq \varepsilon$. The reader may wish to verify our claim that (6.16b) defines a family of c.d.f. maps.

■

We hope that we have convinced the reader that it is not overwhelmingly difficult to construct c.d.f. maps for commonly occuring problems. We shall therefore proceed with the construction of an algorithm for solving the problem $\min_{x \in \mathbb{R}^n} \psi(x)$.

Definition 6.3. Let $\psi:\mathbb{R}^n \to \mathbb{R}$ be l.L.c., let $\{G_\varepsilon\psi(x)\}_{\varepsilon\geq 0}$ be a family of c.d.f. maps for $\psi(\cdot)$, and let $\nu \in (0,1)$. We define the $\varepsilon-$ *search direction* at $x \in \mathbb{R}^n$ by

$$h_\varepsilon(x) \triangleq -\text{argmin}\{\tfrac{1}{2}\|h\|^2 \mid h \in G_\varepsilon\psi(x)\}, \tag{6.17a}$$

and the ε *-adjustment law* by

$$\varepsilon(x) \triangleq \max\{\varepsilon \in \mathrm{E} \mid \|h_\varepsilon(x)\|^2 \geq \varepsilon\}, \tag{6.17b}$$

where

$$\mathrm{E} \triangleq \{0, 1, \nu, \nu^2, \nu^3, \cdots\}. \tag{6.17c}$$

■

Before we continue, it is worth while to pause and examine the ε-search directions $h_\varepsilon(x)$ defined by (6.17a). Suppose we define the ε- *generalized directional derivative* of ψ by

$$d_\varepsilon\psi(x;h) \triangleq \max_{\xi \in G_\varepsilon\psi(x)} \langle \xi, h \rangle, \tag{6.18a}$$

so that $d_\varepsilon\psi(x;h)$ is the support function of $G_\varepsilon\psi(x)$. Then we find that, because $\partial\psi(x) \subset G_\varepsilon\psi(x)$,

$$d_0\psi(x,h) \leq d_\varepsilon\psi(x;h), \tag{6.18b}$$

and hence that any h which makes $d_\varepsilon(x;h)$ negative is a descent direction for $\psi(\cdot)$. Also, it is easy to see that $h_\varepsilon(x)$, defined as in (6.17a), satisfies

$$h_\varepsilon(x) = \underset{h \in \mathbb{R}^n}{\operatorname{argmin}}\{ \tfrac{1}{2}\|h\|^2 + d_\varepsilon\psi(x;h) \}. \tag{6.18c}$$

Consequently, when $0 \notin G_\varepsilon\psi(x)$, $h_\varepsilon(x)$ is a descent direction for $\psi(\cdot)$ at x. Finally, a comparison with (5.11), shows that (6.18c) is still fairly close to the most naive extension of the method of steepest descent to the nondifferentiable case, except that, now, as we shall see, we have generated adequate near continuity properties.

Remark 6.2. In practice, it is common to add a second parameter $\delta > 0$ to the definition of $\varepsilon(x)$, replacing the test $\|h_\varepsilon(x)\|^2 \geq \varepsilon$ by the test $\|h_\varepsilon(x)\|^2 \geq \delta\varepsilon$ in (6.17b). The parameter δ enables us to exercise greater control over the value of $\varepsilon(x)$ so as to achieve better computational behavior. In addition, as should be obvious from the discussion in Section 0.2, when the structure of the problem permits it, it is often possible to use scaling parameters to change the shape of an initial c.d.f. map $G_\varepsilon\psi(x)$ in such a way as to get a new c.d.f. map which yields a better descent direction.

For example, let $\alpha^j > 0$, $j \in \mathbf{m}$ be arbitrary scaling parameters, then

$$G_\varepsilon\psi(x) \triangleq \text{co}\{\alpha^j \nabla f^j(x)\}_{j \in I_\varepsilon(x)}, \quad \varepsilon \geq 0, \tag{6.18d}$$

is also a c.d.f. map for the function $\psi(\cdot)$ considered in Example 6.3.

■

We are ready to state an algorithm model for solving

$$\min\{\psi(x) \mid x \in \mathbb{R}^n\}, \tag{6.19}$$

with $\psi(\cdot)$ l.L.c., and establish its convergence.

Algorithm 6.1. (Requires a family of c.d.f. maps $\{G_\varepsilon\psi(\cdot)\}_{\varepsilon\geq 0}$ for $\psi:\mathbb{R}^n \to \mathbb{R}$, and $\nu \in (0,1)$ for (6.17c)).

Data: $x_0 \in \mathbb{R}^n$.

Step 1: Compute $\varepsilon(x_i)$ and the *search direction*

$$h_i = h_{\varepsilon(x_i)}(x_i), \tag{6.20a}$$

making use of (6.17a,b,c).

Step 2: Compute the *step length*

$$\lambda_i \in \lambda(x_i) \triangleq \underset{\lambda\geq 0}{\text{argmin}}\, \psi(x_i + \lambda h_i). \tag{6.20b}$$

Step 3:

Update:

$$x_{i+1} = x_i + \lambda_i h_i, \tag{6.20c}$$

replace i by $i+1$ and go to Step 1.

■

Lemma 6.2. Let $\psi:\mathbb{R}^n \to \mathbb{R}$ be l.L.c.. Then for every $\hat{x} \in \mathbb{R}^n$ such that $0 \notin \partial\psi(\hat{x})$, there exists a $\hat{\rho} > 0$ and $\hat{\varepsilon} \in E$, $\hat{\varepsilon} > 0$ such that $\varepsilon(x) \geq \hat{\varepsilon}$ for all $x \in B(\hat{x}, \hat{\rho})$.

Proof. Suppose that $0 \notin \partial\psi(\hat{x})$. Since $G_\varepsilon\psi(x)$ is u.s.c. in (ε, x) at $(0, \hat{x})$, it follows that $\|h_\varepsilon(x)\|^2$ is l.s.c. in (ε,x) at $(0,\hat{x})$, and hence that $\|h_\varepsilon(x)\|^2 - \varepsilon$ is l.s.c. in (ε,x) at $(0,\hat{x})$. Consequently, since $\|h_0(\hat{x})\| > 0$, there exist a $\hat{\rho} > 0$ and an $\hat{\varepsilon} \in$ E, with $\hat{\varepsilon} > 0$, such that $\|h_\varepsilon(x)\|^2 - \varepsilon \geq 0$ for all $x \in B(\hat{x},\hat{\rho})$ and $\varepsilon \in [0,\hat{\varepsilon}]$. But this implies that $\varepsilon(x) \geq \hat{\varepsilon}$ for all $x \in B(\hat{x},\hat{\rho})$. This completes our proof.

■

Theorem 6.2. Suppose that $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 6.1 in minimizing a l.L.c. function $\psi:\mathbb{R}^n \to \mathbb{R}$. If $x_i \xrightarrow{K} \hat{x}$ as $i \to \infty$, then $0 \in \partial\psi(\hat{x})$.

Proof. Suppose that $0 \notin \partial\psi(\hat{x})$ for the sake of obtaining a contradiction. Then $\varepsilon(\hat{x}) > 0$ and, by Lemma 6.2, there exist i_0 and $\hat{\varepsilon} > 0$ such that $\varepsilon(x_i) \geq \hat{\varepsilon} > 0$ for all $i \geq i_0$, $i \in K$. Since for any $\varepsilon > 0$, $G_\varepsilon\psi(\cdot)$ is bounded on bounded sets and since $G_\varepsilon\psi(x) \subset G_1\psi(x)$, for all $\varepsilon \in$ E, there exists a $b < \infty$ such that for all $x \in K$, $i \geq i_0$, we have $0 < \hat{\varepsilon} \leq \varepsilon(x_i) \leq \|h(x_i)\|^2 \leq b^2$. Referring to Lemma 6.1, let $\hat{\rho} > 0$ be such that for any x', $x'' \in B(\hat{x},\hat{\rho})$, and any $\xi' \in \partial\psi(x')$ there exists a $\xi'' \in G_{\hat{\varepsilon}}(x'')$ such that $b\,\|\xi' - \xi''\| \leq \hat{\varepsilon}/2$. Hence there exists an $i_1 \geq i_0$ and a $\hat{\lambda} > 0$ such that for all $i \geq i_1$, $i \in K$, $x_i \in B(\hat{x},\hat{\rho})$, $(x_i + s\hat{\lambda}h_i) \in B(\hat{x},\hat{\rho})$ for all $s \in (0,1)$ and for any $\xi_{i\hat{\lambda}} \in \partial\psi(x_i + s\hat{\lambda}h_i)$, there exists a $\xi'_{i\hat{\lambda}} \in G_{\hat{\varepsilon}}\psi(x_i) \subset G_{\varepsilon(x_i)}(x_i)$ such that $\|\xi_{i\hat{\lambda}} - \xi'_{i\hat{\lambda}}\| b \leq \hat{\varepsilon}/2$. Making use of the Lebourg Mean Value Theorem 2.3.1, we now obtain, for all $i \geq i_1$, $i \in K$, that

$$\begin{aligned}
\psi(x_i + \lambda_i h_i) - \psi(x_i) &\leq \psi(x_i + \hat{\lambda}h_i) - \psi(x_i) \\
&= \hat{\lambda}\langle h_i, \xi_{i\hat{\lambda}}\rangle \\
&= \hat{\lambda}[\langle h_i, \xi'_{i\hat{\lambda}}\rangle + \langle h_i, \xi_{i\hat{\lambda}} - \xi'_{i\hat{\lambda}}\rangle] \\
&\leq \hat{\lambda}[-\|h_i\|^2 + \|h_i\|\|\xi_{i\hat{\lambda}} - \xi'_{i\hat{\lambda}}\|] \\
&\leq \hat{\lambda}[-\varepsilon(x_i) + \|h_i\|\,\|\xi_{i\hat{\lambda}} - \xi'_{i\hat{\lambda}}\|] \\
&\leq \hat{\lambda}[-\hat{\varepsilon} + b\,\|\xi_{i\hat{\lambda}} - \xi'_{i\hat{\lambda}}\|] \\
&\leq -\hat{\lambda}\hat{\varepsilon}/2 < 0,
\end{aligned} \tag{6.22}$$

where $\xi_{i\hat{\lambda}} \in \partial\psi(x_i + s_i\hat{\lambda}h_i)$ for some $s_i \in (0,1)$ and $\xi'_{i\hat{\lambda}} \in G_{\varepsilon(x_i)}\psi(x_i)$. Now,

$\{\psi(x_i)\}_{i=0}^{\infty}$ is monotone decreasing and $\psi(x_i) \overset{K}{\to} \psi(\hat{x})$, since $\psi(\cdot)$ is continuous. But this implies that $\psi(x_i) \to \psi(x)$ as $i \to \infty$, contradicting (6.22). Hence we must have had $0 \in \partial\psi(\hat{x})$. This completes our proof.

■

Remark 6.3. In implementing Algorithm 6.1, an Armijo step size rule should be used instead of the exact line search (6.20b), since the Armijo step size rule is much more efficient. In this case, the Armijo step size rule is commonly used in the form:

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, \psi(x_i + \lambda h_i) - \psi(x_i) \le -\lambda\alpha\varepsilon(x_i)\}, \quad (6.23)$$

where $\alpha, \beta \in (0, 1)$.

We leave it is an exercise for the reader to verify our assertion that the conclusions of Theorem 6.2 are unaffected by the substitution of (6.23) for (6.20b).

■

. Next, we develop a phase II algorithm model for solving the problem

$$\min\{f(x) \mid \psi(x) \le 0\}, \quad (6.24)$$

where $f, \psi : \mathbb{R}^n \to \mathbb{R}$ are l.L.c.. Since this is going to be a phase II algorithm, it will require as data an $x_0 \in \mathbb{R}^n$ such that $\psi(x_0) \le 0$. As we have indicated earlier, whenever $\min\{\psi(x) \mid x \in \mathbb{R}^n\} < 0$, such an x_0 can be computed by means of a finite number of iterations of Algorithm 6.1.

Definition 6.4. Let $\{G_\varepsilon f(\cdot)\}_{\varepsilon \ge 0}$ and $\{G_\varepsilon \psi(\cdot)\}_{\varepsilon \ge 0}$ be given families of c.d.f. maps for the functions $f(\cdot)$ and $\psi(\cdot)$ in (6.24), and let $F \triangleq \{x \in \mathbb{R}^n \mid \psi(x) \le 0\}$. We define the family of *phase II convergent direction finding (c.d.f) maps* $\{Gf_{II}{}^{\psi}_{\varepsilon}(\cdot)\}_{\varepsilon \ge 0}$ for (6.24), where $Gf_{II}{}^{\psi}_{\varepsilon} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, by setting

$$Gf_{II}{}^{\psi}_{\varepsilon}(x) \triangleq G_\varepsilon f(x), \text{ if } \psi(x) < -\varepsilon, \quad (6.25a)$$

and

$$G f_{I}{}^{\psi}_{\varepsilon}(x) \triangleq co\{G_\varepsilon f(x), G_\varepsilon \psi(x)\}, \text{ if } \psi(x) \geq -\varepsilon . \quad \textbf{(6.25b)}$$

Next, we define the ε *-search directions* by

$$h_\varepsilon(x) \triangleq -\text{argmin}\{½\|h\|^2 \mid h \in G f_{I}{}^{\psi}_{\varepsilon}(x)\} \quad (6.25c)$$

and the ε *-adjustment law* by

$$\varepsilon(x) \triangleq \max\{\varepsilon \in \text{E} \mid \|h_\varepsilon(x)\|^2 \geq \varepsilon\} , \quad (6.25d)$$

where E was defined in (6.17c).

■

Algorithm 6.2. (Phase II for problem (6.24). Requires $\{G_\varepsilon f(\cdot)\}_{\varepsilon\geq 0}$, $\{G_\varepsilon \psi(\cdot)\}_{\varepsilon \geq 0}$ families of c.d.f. maps for $f(\cdot)$ and $\psi(\cdot)$ and $\nu \in (0,1)$ for the set E in (6.17c)).

Data: $x_0 \in \mathbb{R}^n$ such that $\psi(x_0) \leq 0$.

Step 0: Set $i = 0$.

Step 1: Compute $\varepsilon(x_i)$ according to (6.25d) and the *search direction*

$$h_i = h_{\varepsilon(x_i)}(x_i), \quad (6.26a)$$

according to (6.25c).

Step 2:

Compute the *step length*

$$\lambda_i \in \underset{\lambda\geq 0}{\text{argmin}}\{f(x_i + \lambda h_i) \mid \psi(x_i + \lambda_i h_i) \leq 0\} \quad (6.26b)$$

Step 3: *Update:*

$$x_{i+1} = x_i + \lambda_i h_i , \quad (6.26c)$$

replace i by $i+1$ and go to Step 1.

■

Remark 6.4. The following Phase II Armijo step size rule should be substituted for (6.26b) when implementing the above algorithm model:

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, \\ f(x_i + \lambda h_i) - f(x_i) \le -\lambda\alpha\varepsilon(x_i), \psi(x_i + \lambda h_i) \le 0\}, \tag{6.27}$$

where $\alpha, \beta \in (0,1)$, since it improves computational efficiency.

■

It is straightforward to extend Lemma 6.2 to the following result.

Lemma 6.3:

(a) For every $\hat{x} \in \mathbb{R}^n$ such that $0 \notin G_{\hat{f}0}^{\psi}(\hat{x})$, there exist a $\hat{\rho} > 0$ and an $\hat{\varepsilon} \in \mathrm{E}$, $\hat{\varepsilon} > 0$ such that $\varepsilon(x) \ge \hat{\varepsilon}$ for all $x \in B(\hat{x}, \hat{\rho})$.

(b) Suppose that $\hat{x}$ solves (6.24), then $\varepsilon(\hat{x}) = 0$.

■

Theorem 6.3. Suppose $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 6.2 in solving (6.24) with $f(\cdot)$, $\psi(\cdot)$ l.L.c.. If $x_i \overset{K}{\to} \hat{x}$ as $i \to \infty$, then $\psi(\hat{x}) \le 0$ and $0 \in G_0^{f,\psi}(\hat{x})$ (and, consequently, $\varepsilon(\hat{x}) = 0$).

Proof. First, since $\psi(x_i) \le 0$ for all i by construction and $\psi(\cdot)$ is continuous, we must have $\psi(\hat{x}) \le 0$ for any accumulation point $\hat{x}$ of $\{x_i\}_{i=0}^{\infty}$. For the sake of contradiction, suppose that $x_i \overset{K}{\to} \hat{x}$ and $0 \notin G_0^{f,\psi}(\hat{x})$. Then $\varepsilon(\hat{x}) > 0$ and, by Lemma 6.3, there exists an i_0 and an $\hat{\varepsilon} \in \mathrm{E}$ such that $\varepsilon(x_i) \ge \hat{\varepsilon} > 0$ for all $i \ge i_0$ and $i \in K$.

(a) Suppose that $\psi(\hat{x}) < 0$. Since the h_i are bounded for all $i \in K$, there exist an $i_1 \ge i_0$ and a $\lambda_1 > 0$ such that $\psi(x_i + \lambda h_i) \le 0$ for all $i \in K$, $i \ge i_1$ and all $\lambda \in (0, \lambda_1]$. Consequently, by essentially repeating the arguments of the proof of Theorem 6.2 we can show that there is an $i_2 \ge i_1$ and a $\hat{\lambda} \in (\lambda_1, 0]$ such that for all $i \in K$, $i \ge i_2$, $\psi(x_i + \hat{\lambda} h_i) \le 0$, while

$$f(x_i + \hat{\lambda} h_i) - f(x_i) \le -\hat{\lambda}\hat{\varepsilon}/2. \tag{6.28a}$$

Since $f(x_{i+1}) - f(x_i) \le f(x_i + \hat{\lambda} h_i) - f(x_i)$, we are led to a contradiction, exactly as in the proof of Theorem 6.2.

(b) Suppose that $\psi(\hat{x}) = 0$. Then there exists an $i_3 \ge i_0$ such that for all $i \in K$, $i \ge i_2$, $\psi(x_i) \ge -\hat{\varepsilon} \ge -\varepsilon(x_i)$, so that $G_{\varepsilon(x_i)}^{f,\psi}(x_i)$ is given by (6.25b). Consequently, for all $i \in K$, $i \ge i_2$,

$$h_i = \text{argmin}\{\tfrac{1}{2}\|h\|^2 + \max\{d_{\varepsilon(x_i)} f(x_i;h),\ d_{\varepsilon(x_i)}\psi(x_i;h)\}\} \tag{6.28b}$$

and is a descent direction for both $f(\cdot)$ and $\psi(\cdot)$. We conclude that there is an $i_4 \ge i_3$ and a $\hat{\lambda} \ge 0$ such that for all $i \in K$, $i \ge i_4$,

$$f(x_{i+i}) - f(x_i) \le f(x_i + \hat{\lambda} h_i) - f(x_i) \le -\hat{\lambda}\hat{\varepsilon}/2, \tag{6.28c}$$

$$\psi(x_{i+i}) - \psi(x_i) \le \psi(x_i + \hat{\lambda} h_i) - \psi(x_i) \le -\hat{\lambda}\hat{\varepsilon}/2. \tag{6.28d}$$

Now $\{f(x_i)\}_{i=0}^{\infty}$ is monotone decreasing and $f(x_i) \overset{K}{\to} f(\hat{x})$ as $i \to \infty$, $i \in K$. Hence $f(x_i) \to f(\hat{x})$ as $i \to \infty$, which contradicts (6.28c). This completes our proof.

■

To conclude this section, we present a combined phase I - phase II algorithm which eliminates the disadavantages of a separate, two phase approach, discussed in Section 5.

Definition 6.5. Let $\{G_\varepsilon f(\cdot)\}_{\varepsilon \ge 0}$ and $\{G_\varepsilon \psi(\cdot)\}_{\varepsilon \ge 0}$ be given families of c.d.f. maps for the functions $f(\cdot)$ and $\psi(\cdot)$ in (6.24), let $\psi(x)_+ \triangleq \max\{0, \psi(x)\}$, and let $\gamma : \mathbb{R} \to \mathbb{R}_+$ be a continuous, strictly monotone increasing function such that $\gamma(0) = 0$. We define the family of *phase I-phase II efficient augmented convergent direction finding maps* (e.a.c.d.f.) maps $\{\overline{G}_\varepsilon^{f,\psi}(\cdot)\}_{\varepsilon \ge 0}$ for (6.24), where $\overline{G}_\varepsilon^{f,\psi} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, by setting

$$\overline{G}_\varepsilon^{f,\psi}(x) \triangleq \{\bar{\xi} \in \mathbb{R}^{n+1} \mid \bar{\xi} = (0,\xi),\ \xi \in G_\varepsilon f(x)\}, \ \text{if } \psi(x) < -\varepsilon, \tag{6.29a}$$

$$\bar{G}_\varepsilon^{f,\psi}(x) \triangleq \text{co}\left\{ \{\bar{\xi} \in \mathbb{R}^{n+1} \mid \bar{\xi} = (0,\xi), \xi \in G_\varepsilon\psi(x)\} \cup \{\bar{\xi} \in \mathbb{R}^{n+1} \mid \bar{\xi} = (\gamma(\psi(x)_+),\xi)), \xi \in G_\varepsilon f(x)\} \right\} \text{ if } \psi(x) \geq -\varepsilon. \tag{6.29b}$$

Next, we define the *augmented ε-search direction* by

$$\bar{h}_\varepsilon(x) \triangleq (h_\varepsilon^0(x), h_\varepsilon(x)) \triangleq -\text{argmin}\{\tfrac{1}{2}\|\bar{h}\|^2 \mid \bar{h} \in \bar{G}_\varepsilon^{f,\psi}(x)\}, \tag{6.29c}$$

yielding the *actual search direction* $h_\varepsilon(x)$, and the ε *adjustment law* by

$$\varepsilon(x) \triangleq \max\{\varepsilon \in \mathrm{E} \mid \|\bar{h}_\varepsilon(x)\|^2 \geq \varepsilon\}. \tag{6.29d}$$

■

Note the effect of $\psi(x)$ on $\bar{h}_\varepsilon(x)$. When $\psi(x)_+$ is large, then $h_\varepsilon(x) \cong -\text{argmin}\{\tfrac{1}{2}\|h\|^2 \mid h \in G_\varepsilon\psi(x)\}$. When $\psi(x) \leq 0$, $h_\varepsilon(x)$ is the same as computed in Algorithm 6.2 (phase II). When $\psi(x) > 0$ and decreases to zero, the effect of the cost on $\bar{h}_\varepsilon(x)$ becomes progressively more pronounced. For the case where $f(\cdot)$ and $\psi(\cdot)$ are differentiable, this effect is illustrated in Figure 6.1.

We can now state a phase I-phase II algorithm for solving 6.23.

Algorithm 6.3. (Requires $\{G_\varepsilon f(\cdot)\}_{\varepsilon \geq 0}$, $\{G_\varepsilon\psi(\cdot)\}_{\varepsilon \geq 0}$ families of c.d.f. maps for $f(\cdot)$ and $\psi(\cdot)$; $\nu \in (0,1)$ for the set E in (6.17c)).

Data: $x_0 \in \mathbb{R}^n$.

Step 0: Set $i = 0$.

Step 1: Compute $\varepsilon(x_i)$ according to (6.29d) and the *search direction*

$$h_i = h_{\varepsilon(x_i)}(x_i) \tag{6.30a}$$

according to (6.29c).

Step 2: Compute the *step length* as follows:

If $\psi(x_i) > 0$, then

$$\lambda_i \in \underset{\lambda \geq 0}{\operatorname{argmin}} \psi(x_i + \lambda h_i). \tag{6.30b}$$

If $\psi(x_i) \leq 0$, then

$$\lambda_i \in \underset{\lambda \geq 0}{\operatorname{argmin}} \{f(x_i + \lambda h_i) \mid \psi(x_i + \lambda h_i) \leq 0\}. \tag{6.30c}$$

Step 3: Update:

$$x_{i+1} = x_i + \lambda_i h_i, \tag{6.30c}$$

replace i by $i+1$ and go to step 1.

■

Remark 6.5. Since it improves computational efficiency, the following Phase I-Phase II Armijo step size rule should be substituted for (6.30b,c) when implementing the above algorithm model:

If $\psi(x_i) > 0$, then

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, \psi(x_i + \lambda h_i) - \psi(x_i) \leq -\lambda\alpha\varepsilon(x_i)\}. \tag{6.31a}$$

If $\psi(x_i) \leq 0$, then

$$\lambda_i \triangleq \max\{\lambda \mid \lambda = \beta^k, k \in \mathbb{N}, \\ f(x_i + \lambda h_i) - f(x_i) \leq -\lambda\alpha\varepsilon(x_i), \psi(x_i + \lambda h_i) \leq 0\}. \tag{6.31b}$$

where $\alpha, \beta \in (0,1)$.

■

The following theorem is easy to establish and hence we leave its proof as an exercise for the reader.

Theorem 6.4. Consider Problem (6.24) and suppose that $f(cdot), \psi(\cdot)$ is l.L.c., and that $0 \notin G_0\psi(x)$ for all $x \in \mathbb{R}^n$ such that $\psi(x) \geq 0$. If $\{x_i\}_{i=0}^{\infty}$ is a sequence constructed by Algorithm 6.3 in solving (6.24) is such that $x_i \overset{K}{\to} \hat{x}$ as $i \to \infty$, then $\psi(\hat{x}) \leq 0$ and $0 \in G_0^{f,\psi}(\hat{x})$.

■

This concludes our exposition of semi-infinite optimization algorithms in conceptual form. Our next task is to sketch out an implementation technique.

7. IMPLEMENTATION OF CONCEPTUAL ALGORITHMS

In the preceding two sections, we have presented algorithms for solving semi-infinite optimization problems, the simplest of which is

$$P: \min \{ f(x) \mid \max_{y \in Y} \varphi(x,y) \leq 0 \}, \tag{7.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\varphi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ are continuous and $\nabla f(\cdot)$, $\nabla_x \varphi(\cdot,\cdot)$ exist and are locally Lipschitz continuous, and $Y \triangleq [y_0, y_f] \subset \mathbb{R}$ is compact.

In this section we shall be concerned with Algorithm 6.3. For problem P, Algorithm 6.3 requires the *exact* evaluation of $\psi(x) \triangleq \max_{y \in Y} \varphi(x,y)$ and of the ε-active local maximizer set $\hat{Y}_\varepsilon(x)$ (defined in (6.4b). Neither of these evaluations can be carried out with infinite precision in finite time and hence the Algorithm 6.3 must be viewed as being only *conceptual*. We now turn to techniques for making this algorithm (as well as others like it) *implementable*, or more precisely, to techniqes for constructing *implementations*. A general theory related to conceptual algorithm implementation can be found in [Pol.1], see also [Kle.1, Muk.1, Pol.8, Tra.1]. In this section we shall only illustrate the general approach by constructing an implementation of Algorithm 6.3, as it applies to the problem P in (7.1).

As we have already pointed out in Example 6.6, within the realm of engineering design, we are justified in making the following hypothesis:

Assumption 7.1. For every $x \in \mathbb{R}^n$, $\varepsilon \geq 0$, $\hat{Y}_\varepsilon(x)$, defined in (6.4b), is a finite set.

■

We shall construct an implementation of Algorithm 6.2, by imbedding Algorithm 6.2 in a Master Algorithm which replaces problem P with an infinite sequence of approximating problems $\{P_q\}$ for which Algorithm 6.2 is implementable as stated. For this purpose, we shall need the following definitions.

For any $q \in \mathbb{N}$ such that $q \geq 1$, let

$$Y_q \triangleq \{y_k \in Y \mid y_k = y_o + \frac{k}{q}(y_f - y_o),\ k = 0,1,\cdots,q\}, \tag{7.2a}$$

and let $\varphi_q : \mathbb{R}^n \times Y \to \mathbb{R}$ be defined, by means of Y_q, as follows:

$$\varphi_q(x,y) \triangleq \lambda\varphi(x,y_k) + (1-\lambda)\varphi(x,y_{k+1}), \tag{7.2b}$$

for $y = \lambda y_k + (1-\lambda)y_{k+1}$, with $\lambda \in [0,1]$, and $y_k \in Y_q$ for $k = 0,1,\cdots,q-1$. Next, let

$$\psi_q(x) \triangleq \max_{y \in Y} \varphi_q(x,y). \tag{7.2c}$$

Finally, for $q \in \mathbb{N}$, we define the family of *approximating problems*

$$\mathrm{P}_q : \quad \min\{ f(x) \mid \varphi_q(x,y) \leq 0,\ \forall\, y \in Y\}, \tag{7.2d}$$

which can also be written as

$$\mathrm{P}_q : \quad \min\{ f(x) \mid \psi_q(x) \leq 0\}. \tag{7.2e}$$

To distinguish the quantities associated with one value of q from another, we shall add a subscript q to all the quantities in (6.26b) - (6.29d) and in Algorithm 6.2. Thus, for any $q \in \mathbb{N}$, $q \geq 1$, we define, for any $x \in \mathbb{R}^n$ and $\varepsilon \geq 0$, the set of ε-active $y \in Y$ by

$$Y_{q,\varepsilon}(x) \triangleq \{ y \in Y \mid \varphi_q(x,y) \geq \psi_q(x) - \varepsilon\}, \tag{7.2f}$$

and the set of ε-active local maximizers by

$$\hat{Y}_{q,\varepsilon}(x) \triangleq \{y \in Y_{q,\varepsilon}(x) \mid y \text{ is a local maximizer of } \varphi_q(x,y) \text{ in } Y\}. \tag{7.2g}$$

The simplest situation for constructing an implementation for Algorithm 6.3, for problem P, arises when the following hypothesis holds.

Assumption 7.2. For every $x \in \mathbb{R}^n$, $\varepsilon \geq 0$, $q \in \mathbb{N}$ such that $q \geq 1$, the set $\hat{Y}_{q,\varepsilon}(x)$ is finite.

■

Note that it is harder to justify Assumption 7.2 than Assumption 7.1.

Since $\psi_q(x) = \max_{y \in Y} \varphi_q(x,y)$ holds, it is clear that the evaluation of $\psi_q(x)$ is a finite process. Similarly, when Assumption 7.2 holds, the construction of the set, for $\varepsilon \geq 0$,

$$G_\varepsilon \psi_q(x) \triangleq \text{co}\{ \nabla_x \varphi(x,y) \}_{y \in \hat{Y}_{q,\varepsilon}(x)} \tag{7.2h}$$

is also a finite process. Since it follows from Example 6.6 that (7.2h) defines a family of c.d.f. maps for $\psi_q(\cdot)$, and since we may take $G_\varepsilon f(x) = \{ \nabla f(x) \}$ for all $\varepsilon \geq 0$ and all $x \in \mathbb{R}^n$, we see that for problem P_q Algorithm 6.3 is, in fact, implementable.

Now consider the following master algorithm in which $\varepsilon_q(x)$ is defined by (6.29a) - (6.29d) for the problem P_q. The reason for calling it a master algorithm is that it calls Algorithm 6.3 as a subprocedure. In fact, any other algorithm can be substituted for Algorithm 6.3 in the master algorithm, as long as it computes the required quantities.

Master Algorithm 7.1. (Solves Problem (7.1)).

Data: $q_0 \in \mathbb{N}_+$, $z_{q_0} \in \mathbb{R}^n$, and a sequence $\{ \gamma_q \}_{q=0}^{\infty}$ such that $\gamma_q > 0$ and $\gamma_q \downarrow 0$ as $q \to \infty$.

Step 0: Set $q = q_0$.

Step 1: Apply Algorithm 6.3 to problem P_q, from the initial point $x_0 = z_q$, for a sufficient number of iterations, to compute a vector $z_{q+1} = x_{i_q}$ such that $\varepsilon_q(x_{i_q}) \le \gamma_q$, and $\psi_q(x_{i_q}) \le \gamma_q$.

Step 2: Replace q by $2q$ and go to step 1.

■

Note that any monotone law for increasing q can be used in Step 2 of Master Algorithm 7.1, not just the doubling one which was used for the sake of simplicity.

Theorem 7.1. Suppose that Assumptions 7.1 and 7.2 hold and consider any sequence $\{ z_q \}_{q=0}^{\infty}$ constructed by the Master Algorithm 7.1. If $\hat{z}$ is any accumulation point of $\{ z_q \}_{q=0}^{\infty}$, then $\psi(\hat{z}) \le 0$ and $\varepsilon(\hat{z}) = 0$, where $\varepsilon(\cdot)$ is defined in (6.29d), i.e. $\hat{z}$ satisfies the optimality condition stated in Theorem 4.1, as it applies to the problem P in (7.1).

Proof. Suppose that $z_q \overset{K}{\to} \hat{z}$. Then, since $\varphi(\cdot,\cdot)$ is locally Lipschitz continuous, there exists a constant $L \in (0,\infty)$, such that for all z_q, $q \in K$,

$$| \varphi(z_q,y) - \varphi_q(z_q,y) | \le L(y_f - y_o)/q, \ \forall\, y \in Y. \tag{7.3}$$

Hence, since $\psi_q(z_q) \le \gamma_q$ for all $q \in \mathbb{N}$ and since $\gamma_q \downarrow 0$ as $q \to \infty$, by construction, we must have that $\psi(\hat{z}) \le 0$. Next, for all $q \in \mathbb{N}$, there exists a vector $\bar{h}_q \in \overline{G}^{f,\psi_q}_{\varepsilon_q(z_q)/\nu}(z_q)$, defined by (6.29c) (with ν as in (6.17c)), such that $\|\bar{h}_q\|^2 < \varepsilon_q(z_q)/\nu \le \gamma_q/\nu$. Let $\bar{\varepsilon} > 0$ be arbitrary. Then, since $\varepsilon_q(z_q) \to 0$ as $q \to \infty$, there exists a $\bar{q} \in \mathbb{N}$ such that for all $q \ge \bar{q}$, $\varepsilon_q(z_q)/\nu \le \bar{\varepsilon}/2$ and, from (7.3), $|\psi(z_q) - \psi_q(z_q)| \le \bar{\varepsilon}/2$. Hence

$$\hat{Y}_{q,\varepsilon_q(z_q)/\nu}(z_q) \subset Y_{\bar{\varepsilon}}(z_q) \triangleq \{ y \in Y \mid \varphi(z_q,y) \ge \psi(z_q) - \bar{\varepsilon} \}. \tag{7.4a}$$

It now follows from Definiton 6.4 that when $\psi_q(z_q) \ge -\varepsilon_q(z_q)$,

$$\overline{G}^{f,\psi_q}_{\varepsilon_q)/\nu}(z_q) = \text{co}\left\{ \begin{bmatrix} \gamma(\psi(z_q))_+ \\ \nabla f(z_q) \end{bmatrix}, \begin{bmatrix} 0 \\ \nabla_x \varphi(z_q, y) \end{bmatrix} \right\}_{y \in \hat{Y}_{q,\varepsilon_q(z_q)/\nu}},$$
$$\subset \text{co}\left\{ \begin{bmatrix} \gamma(\psi(z_q))_+ \\ \nabla f(z_q) \end{bmatrix}, \begin{bmatrix} 0 \\ \nabla_x \varphi(z_q, y) \end{bmatrix} \right\}_{y \in Y_{\bar{\varepsilon}}}, \tag{7.4b}$$

and that when $\psi_q(z_q) < -\varepsilon_q(z_q)$,

$$\overline{G}^{f,\psi_q}_{\varepsilon_q)/\nu}(z_q) = \left\{ \begin{bmatrix} 0 \\ \nabla f(z_q) \end{bmatrix} \right\}. \tag{7.4c}$$

Since $\bar{h}_q \overset{K}{\to} 0$ as $q \to \infty$, and $Y_{\bar{\varepsilon}}(\cdot)$ is u.s.c., we conclude that either

$$\nabla f(\hat{z}) = 0 \text{ and } \psi(\hat{z}) < 0, \tag{7.4d}$$

or

$$0 \in \text{co}\{ \nabla f(\hat{z}), \nabla\varphi(\hat{z}, y) \}_{y \in Y_{\bar{\varepsilon}}(\hat{z})} \text{ and } \psi(\hat{z}) = 0. \tag{7.4e}$$

Since $\bar{\varepsilon} > 0$ was arbitrary and $Y_\varepsilon(z)$ is u.s.c. in (ε, z) at $(0, \hat{z})$, the desired result follows.

■

Next, we consider again the problem P in (7.1) and we suppose that Assumption 7.1 holds, but we *do not* suppose that Assumption 7.2 holds. In this case, $G_\varepsilon\psi_q(x)$, defined by (7.2g) is no longer a c.d.f. map for $\psi_q(x)$. We propose the following alternative set, defined for any $q \in \mathbb{N}$, $q \geq 1$, any $\varepsilon \geq 0$ and any $x \in \mathbb{R}^n$, to be used when Assumption 7.2 cannot be trusted to hold:

$$G_\varepsilon\psi_q(x) \triangleq \text{co}\{ \nabla_x \varphi(x, y) \}_{y \in \tilde{Y}_{q,\varepsilon}(x)}, \tag{7.5a}$$

where for any $p > 0$,

$$\tilde{Y}_{q,\varepsilon}(x) \triangleq \hat{Y}_{q,\varepsilon} \cup \{ y \in Y_q \mid \varphi(x, y) \geq \psi_q(x) - \varepsilon/pq^2 \}. \tag{7.5b}$$

We see that (7.5a), (7.5b) are a "two tier" definition aimed at keeping the number of vectors in $G_\varepsilon \psi_q(x)$ relatively small, but not so small as to adversely affect the behavior of the resulting algorithm. The reason for using q^2 in (7.5b) is that if $\varphi(x,y)$ can be approximated by a quadratic $\varphi(x,\hat{y}) + M(y-\hat{y})^2$, with $M > 0$, in a neighborhood of a local maximizer $\hat{y} \in Y$ of $\varphi(x,y)$, then the test $\varphi(x,y) \geq \psi_q(x) - \varepsilon/pq^2$, will tend to keep the cardinality of the set $\{ y \in Y_q \mid \varphi(x,y) \geq \psi_q(x) - \varepsilon/pq^2 \}$ in (7.5b) approximately constant for all q, for a given $\varepsilon \geq 0$. If we had replaced q^2 by q, the cardinality of that set would have grown with q, which would have been computationally undesirable.

We invite the reader to verify the following result.

Proposition 7.1. (a) The relations (7.5a), (7.5b), define a c.d.f. map for $\psi_q(\cdot)$. (b) Suppose that the c.d.f. map defined by (7.5a), (7.5b) is used in Master Algorithm 7.1, then Theorem 7.1 remains valid.

■

This brings us to the end of our elementary exposition of semi-infinite optimization algorithms, both in conceptual and in implementable form.

8. CONCLUSION

We have shown that a broad class of semi-infinite optimization algorithms can be evolved by progressively more complex extensions of the humble method of steepest descent for differentiable optimization. Our approach was based on the introduction of convergent direction finding maps which have the property that the point of smallest norm, in the sets which they define, defines a good search direction. We hope that our approach has enabled the reader to acquire a quick understanding of an important family of semi-infinite optimization algorithms.

Our approach does not enable us to account for all the "first order" algorithms in the literature, nor can it be extended to explain or generate algorithms for nondifferentiable problems defined on normed spaces that are not Hilbert spaces. We believe that many of the algorithms which solve finite dimensional nondifferentiable optimization problems that we did not account for, as well as algorithms for solving problems in function spaces such as H^∞ or L_∞, which are not Hilbert spaces, can be organized into a related, coherent structure by introducing axioms which define ***super directional derivatives*** in terms of certain supersets of the generalized gradient. The super directional derivative, in turn, can be used to define convergent search direction finding problems. To make this suggestion more concrete, we point out that the function $d\psi_\varepsilon(x;h)$, introduced in Section 6, is and example of such a super directional derivative. However, the exploration of this possibility is beyond the scope of this paper and is left as a suggestion for future research.

ACKNOWLEDGEMENT

Sincere thanks are due to Prof. John Burns and Dr. Charles Holland without whose strong encouragement this paper may not have been written. The author wishes to thank his students, S. Salcudean, T. E. Baker and T. L. Wuu, Dr. R. Vinter, and the anonimous referees for their comments on the manuscript and for their assistance in proofreading and preparing the manuscript.

The research reported herein was sponsored in part by the National Science Foundation under Grant ECS-8121149, the Air Force Office of Scientific Research under Grant AFOSR-83-0361, the Office of Naval Research under Grant N00014-83-K-0602, the Air Force Office of Scientific Research (AFSC) under Contract F49620-79-C0178, and the Semiconductor Research Consortium under Grant SRC-82-11-008.

9. REFERENCES.

[Arm.1] Armijo, L., "Minimization of functions having Lipschitz continuous first partial derivatives", *Pacific Journal of Mathematics* Vol 16, pp 1-3, 1966.

[Ben.1] Ben-Israel, A., A. Ben-Tal and S. Zlobec, *OPtimality in Nonlinear Programming: A Feasible Directions Approach*, Wiley-Interscience, New York, 1981.

[Ber.1] Berge, C., *Topological Spaces* Macmillan Co., N.Y., 1963.

[Bert.1] Bertsekas, D. P., and Mitter, S. K., "A descent numerical method for nondifferentiable cost functionals", *SIAM J. Control*, Vol. 11, pp 637-652, 1973.

[Bha.1] Bhatti, M. A., E. Polak, K. S. Pister, "Optimization of control devices in base isolation systems for seismic design", *Proc. International IUTAM Symposium on Structural Control*, University of Waterloo, Ontario, Canada, North Holland Pub. Co., Amsterdam, pp 127-138, 1980.

[Blan.1] Bland, R. G., D. Goldfarb and M. J. Todd, "The ellipsoid method: a survey", *Operations Research* Vol. 29, pp 1039-1091, 1981.

[Clar.1] Clarke, F. H., *Optimizaton and Nonsmooth Analysis*, Wiley-Interscience, New York, N.Y., 1983.

[Dan.1] Danskin, J. M., "The theory of minmax with applications", *SIAM J. of Appl. Math.*, Vol. 14, No. 4, pp 641-655, 1966.

[Dem.1] Demjanov, V. F., and V. N. Malozemov, *Introduction to Minimax* John Wiley & Sons, New York, 1974.

[Dem.2] Demjanov, V. F., and L. V. Vasilyev, *Nondifferentiable Optimization* OPtimization Software/Springer, New York, 1985

[Fan.1] Fan, K., "Minimax theorems", Proc. Nat. Acad. Sci. Vol. 39, pp 42-47, 1953

[Fia.1] Fiacco, A. V., and K. O. Kortanek, editors, *Semi-Infinite Programming and Applications*, An International Symposium, Austin, Texas 1981. Springer-Verlag Berlin-Heidelberg-New York-Tokyo, 1983.

[Gil.1] Gill, Ph. E., W. Murray, M. A. Saunders and M. H. Wright, "User's Guide for QPSOL (Version 3.2): A Fortran Package for Quadratic Programming", *Stanford University Technical Reports*, SOL 84-6 and SOL 84-7, Sept. 1984.

[Gon.1] Gonzaga, C., E. Polak, and R. Trahan, "An improved algorithm for optimization problems with functional inequality constraints", *IEEE Trans. on Automatic Control*, Vol. AC-25, No. 1, 1980.

[Gon.2] Gonzaga, C., and E. Polak, "On constraint dropping schemes and optimality functions for a class of outer approximations algorithms", *J. SIAM Control and Optimization* Vol. 17, 1979.

[Het.1] Hettich, R., *Semi-Infinite Programming*, Springer-Verlag Lectlure Notes in Control and Information Sciences, Vol. 15, 1979.

[Het.2] Hettich, R., and W. van Honstede, "On quadratically convergent methods for semi-infinite programming", in *Semi-Infinite Programming*, Springer-Verlag Lectlure Notes in Control and Information Sciences, Vol. 15, 1979, pp 97-111.

[Het.3] Hettich, R., and p. Zencke, *Numerische Methoden der Approximation und Semi-infiniten Optimierung*, Teubner, Stuttgart, 1982.

[Het.4] Hettich, R., editor, *Semi-Infinite Programming*, Proceedings of a Workshop, Bad Honne August 30 - September 1, 1978, Springer-Verlag Berlin-Heidelberg-New York, 1979.

[Iud.1] Iudin, D. B., and A. S. Nemirovskii, "Informational complexity and effective methods for solving convex extremal problems", *Matekon* Vol. 13, pp 25-45, 1977. John, F., " Extremum problems with inequalities as side conditions",

[Joh.1] John, F., " Extremum problems with inequalities as side conditions", in *Courant Anniversary Volume*, (K. O. Friedrichs, O. E. Neugebauer and J. J. Stoker, eds.), pp 187-204, Interscience, New York, 1948.

[Kar.1] Karmarkar, N., "A New Polynomial-Time Algorithm for Linear Programming," preprint, August, 1984, *Combinatorica* to appear

[Kha.1] Khachian, L. G., "A polynomial algorithm for linear programming", *Dok. Acad. Nauk SSSR*, Vol. 244, pp 1093-1096, 1979.

[Kiw.1] Kiwiel, K. C., *Methods of Descent for Nondifferentiable Optimization* Springer-Verlag, Berlin Heidelberg New York Tokyo, 1985. in press.

[Kle.1] Klessig ,R., and E. Polak, "An adaptive algorithm for unconstrained optimization with applications to optimal control", *SIAM J. Control*, Vol. 11, No. 1, pp 80-94, 1973.

[Leb.1] Lebourg, G., "Valeur Moyenne Pour Gradient Generalise", *C. R. Acad. Sci*, Paris, Vol. 281, 1975. pp 795-797

[Lem.1] Lemarechal, C., "An extension of Davidon methods to nondifferentiable optimization problems", in *Nondifferentiable Optimization*, Mathematical Programming Study 3, M. L. Balinski and Ph. Wolfe eds., North Holland, Amsterdam, pp 95-109, 1975.

[Lem.2] Lemarechal, C., "Nondifferentiable optimization, subgradient and ε subgradient methods", Lecture Notes, *Optimization and Operations Research* Springer Verlag, Berlin, 1976.

[Lem.3] Lemarechal, C., *Extensions diverses des methods de gradient et applications*, These Docteur d'Etat, University of Paris IX, 1980. Proc.

[Lem.4] Lemarechal, C., and J. Zowe, "Some remarks on the construction of higher order algorithms in convex optimization", *J. Appl. Math. and Opt., Vol. 10, pp 51-68. 1983.*

[Lem.5] Lemarechal, C., and J.J. Strodiot, "Bundle methods, cutting plane algorithms and σ-Newton directions", *Proc. IIASA Workshop on Nondifferentiable Optimization*, Sopron, Hungary, Sept 17-22, 1984.

[May.1] Mayne, D. Q., Polak, E., and Trahan, R., "An Outer approximations algorithm for computer-aided design problems", *J. of Appl. Math. and Optimization*, Vol. 28, No. 3, 1979.

[May.2] Mayne, D. Q., and Polak, E., "An exact penalty function algorithm for control problems with control and terminal inequality constraints", Part 1, *J. of Appl. Math. and Optimization*, Vol. 32 No. 2, 1980, pp. 211-246.

[May.3] Mayne, D. Q., and Polak, E., "An exact penalty function algorithm for control problems with control and terminal inequality constraints", Part 2, *J. of Appl. Math. and Optimization*, Vol. 32 No. 3, 1980, pp. 345-363.

[May.4] Mayne, D. Q., and E. Polak, "Feasible directions algorithms for optimization problems with equality and inequality constraints", *Mathematical Programming*, Vol. 11, pp 67-80, 1976.

[May.5] Mayne, D. Q., E. Polak and A. Voreadis, "A cut map algorithm for design problems with tolerances", *IEEE Trans. on Circuits and Systems*, Vol. CAS-29 No. 1, pp 35-46, 1982.

[May.6] Mayne, D. Q., E. Polak and A. Sangiovanni Vincentelli, "Computer aided design via optimization: a review", *Automatica*, Jan. 1982.

[May.7] Mayne, D. Q., and E. Polak, "A quadratically convergent algorithm for solving infinite dimensional inequalities", *J. of Appl. Math. and Optimization*, Vol. 9., pp 25-40, 1982.

[May.8] Mayne, D. Q., and E. Polak, "Algorithms for the design of control systems subject to singular value inequalities", Mathematical Programming Study 18, Algorithms and Theory in Filtering and Control, D. C. Sorensen and R. J.-B. Wets, ed., North Holland, New York, pp 112-135, 1982.

[May.9] Mayne, D. Q., and E. Polak, "Nondifferentiable optimization via adaptive smoothing", *J. of Appl. Math. and Optimization*, Vol. 43, No. 4, pp 601-614, August 1984.

[Mif.1] Mifflin, R., "An Algorithm for Constrained Optimization with Semi-Smooth Functions", *Math. of Oper. Res.*, Vol. 2, No. 2, pp 191-207, 1977.

[Mif.2] Mifflin, R., "Semismooth and semiconvex functions in constrained optimization", *SIAM J. Control and Optimization*, Vol. 15, No. 6, pp 959-973, 1977.

[Mif.3] Mifflin, R., "Better than linear convergence and safeguarding in nonsmooth minimization", in *Systems Modelling and Optimization*, P. Thoft-Christensen ed., Lecture Notes, Control and Info. Scie. 59, Springer, Berlin, 1984.

[Muk.1] Mukai, H., and E. Polak, "On the use of approximations in algorithms for optimization problems with equality and inequality constraints", *SIAM J. Num. Anal.* Vol. 1, No. 4, pp 674-693, 1978.

[Nur.1] Nurminski, E. A., *Numerical Methods for Solving Deterministic and Minimax Problems*, Naukova Dumka, Kiev (in Russian), 1979.

[Nur.2] Nurminski, E. A., "Bibliography on nondifferentiable optimization", *Progress in Nondifferentiable Optimization*, E. A. Nurminski, ed., CP-82-S8, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1982.

[Oet.1] Oettli, W., "The method of feasible directions for continuous minimax problems", Proc. *9th Intl. Math. Prog. Symp.* Budapest, Aug 23-27, pp 505, 1976.

[Opf.1] Opfer, G., "An algorithm for the construction of best approximations based on Kolmogorov's criterion", *Jour. of Approx. Theory* Vol. 23, pp 299-317, 1978.

[Pir.1] Pironneau, O., and E. Polak, " A dual method for optimal control problems with initial and final boundary constraints", *SIAM J. Control*, Vol. 11, No. 3, pp 534-549, 1973.

[Pir.2] Pironneau, O., and E. Polak, " Rate of convergence of a class of methods of feasible directions", *SIAM J. Num. Anal.*, Vol. 10, No. 1, pp 161-174, 1973.

[Pol.1] Polak, E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, 320 pages, 1971.

[Pol.2] Polak, E., K. S. Pister and D. Ray, "Optimal design of framed structures subjected to earthquakes", *Engineering Optimization*, Vol 2, pp 65-71, 1976.

[Pol.3] Polak, E., and D. Q. Mayne, "An algorithm for optimization problems with functional inequality constraints", *IEEE Transactions on Automatic Control*, Vol. AC-21, No. 2, 1976.

[Pol.4] Polak, E., and R. Trahan, "An algorithm for computer aided design of control systems", Proc. *IEEE Conf. on Dec. and Control*, 1976.

[Pol.6] Polak, E. ,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.7] E. Polak and A. Sangiovanni Vincentelli, "Theoretical and computational aspects of the optimal design centering, tolerancing and tuning problem", *IEEE Trans. on Circuits and Systems*, Vol CAS-26, No. 9, pp 795-813, 1979.

[Pol.8] Polak, E., "An implementable algorithm for the design centering, tolerancing and tuning problem", *Computing Methods in Applied Science and Engineering*, R. Glowinski, J. L. Lions, ed., North Holland, Amsterdam, 1980, pp.499-517.

[Pol.9] Polak, E., and S. Tishyadhigama, "New convergence theorems for a class of feasible directions algorithms", *J. of Appl. Math. and Optimization*, Vol. 35, No. 3, Nov. 1981.

[Pol.10] Polak, E., and Y. Wardi, "A nondifferentiable optimization algorithm for the design of control systems subject to singular value inequalities over a frequency range", *Automatica*, Vol. 18, NO. 3, pp. 267-283, 1982.

[Pol.11] Polak, E., and A. Tits, " A Recursive Quadratic Programming Algorithm for Semi-Infinite Optimization Problems", *J. Appl. Math. and Optimization*, Vol. 8, pp 325-349, 1982.

[Pol.12] Polak, E., D.Q. Mayne and Y. Wardi, "On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems", *SIAM J. Cont. & Opt.* Vol. 21, No.2, pp 179,204, 1983.

[Pol.14] Polak, E., "A Modified Nyquist Stability Criterion for Use in Computer-Aided Design", *IEEE Trans. on Automatic Control*, Vol. AC-29, No. 1, March, 1984.

[Pol.15] Polak, E., and D.M. Stimler, "On the design of linear control systems with plant uncertainty via nondifferentiable optimization", *Proc.* IX. Triennial IFAC World Congress, Budapest, July 2-6, 1984.

[Pol.16] Polak, E., and Y. Y. Wardi, "A study of minimizing sequences", *SIAM Jour. on Contr. and Opt.*, Vol.22, No. 4, pp 599-609, 1984.

[Pol.17] Polak, E., and D. Q. Mayne, "Algorithm models for nondifferentiable optimization", *SIAM Jour. on Contr. and Opt.*, Vol. 23, No. 3, 1985.

[Pol.18] Polak, E., D. Q. Mayne and D. M. Stimler, "Control System Design via Semi-Infinite Optimization", *IEEE Proceedings*, pp 1777, Dec. 1984.

[Polj.1] Poljak, B. T., "A general method for solving extremum problems", *Soviet Mathematics*, No. 8, pp. 593-597, 1966.

[Polj.2] Poljak, B. T., "Minimization of nonsmooth functionals", *U.S.S.R. Comp. Math. and Math. Phys.*, No. 9, pp 14-29, 1969.

[Polj.3] Poljak, B. T., *Introduction to Optimization*, Nauka, Moscow, (in Russian) 1983.

[Roc.1] Rockafellar, R. T., *Convex Analysis*, Princeton Univ. Press, Princeton, N.J. 1970.

[Sho.1] Shor, N. Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.

[Sho.2] Shor, N. Z., "Utilization of the operation of space dilatation in the minimization of convex functions", *Cybernetics*, Vol. 1, pp 7-15, 1970.

[Sho.3] Shor, N. Z., and Shabashova, L. P., "Solution of minmax problems by the method of generalized gradient descent with dilatation of the space", *Cybernetics*, Vol. 1, pp 88-94, 1972.

[Sho.4] Shor, N. Z., "Cut-off method with space extensions in convex programming problems", *Cybernetics* Vol. 15, pp 502-508, 1980.

[Ste.1] Stein, E. M., *Singular Integrals and Differentiability Properties of Functions*, Princeton Math. Series, No.30, Princeton Univ. Press, Princeton, N.J. 1970

[Tis.1] Tishiyadhigama, S., E. Polak and R. Klessig, "A comparative study of several general convergence conditions for algorithms modeled by point ot set maps", *Mathematical Programming Study* No. 10, pp. 172-190, 1979.

[Tra.1] Trahan, R., and Polak, E., "A Derivative Free Algorithm for a Class of Infinitely Constrained Optimization Problems", *IEEE Trans. on Automatic Control*, Vol. AC-25, No. 1, 1979.

[War.1] Wardi, Y.Y., and E. Polak, "A nondifferentiable optimization algorithm for structural problems with eigenvalue constraints", *Journal of Structural Mechanics*, Vol 11, No. 4, 1983.

[Warg.1] Warga, J., "Iterative procedures for constrained and unilateral optimization problems", *SIAM J. Cont. and Opt.*, Vol 20, No.3, pp 360-376, 1982.

[Wil.1] Williamson, L. J., and E. Polak, "Relaxed controls and the convergence of optimal control algorithms", *SIAM J. Control*, Vol. 14, No. 4, pp 737-757, 1976.

[Wol.1] Wolfe, Ph., "Finding the nearest point in a polytope", *Math. Programming*, Vol. 11, pp 128-149, 1976.

[Zou.1] Zoutendijk, G., *Mathematical Programming Methods*, North Holland, Amsterdam, 1976.

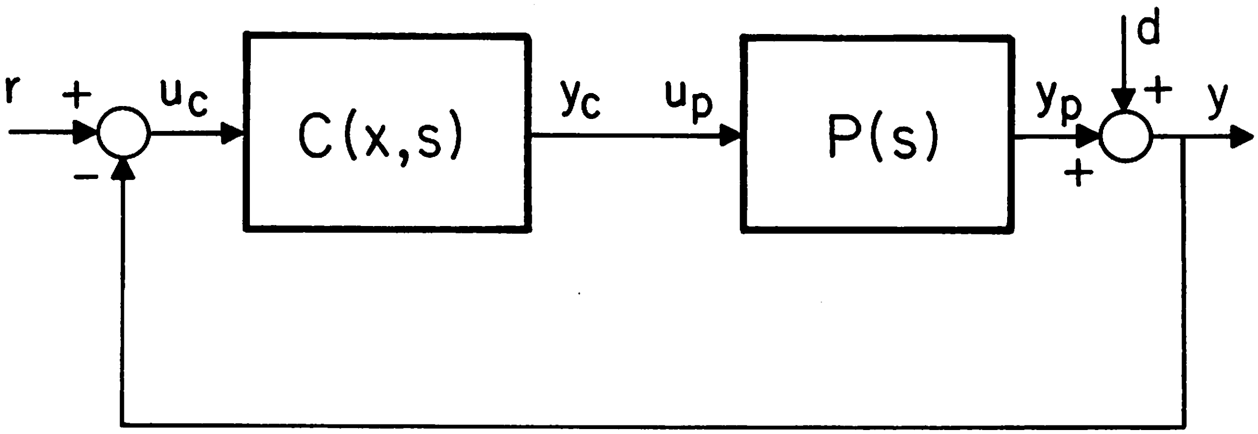
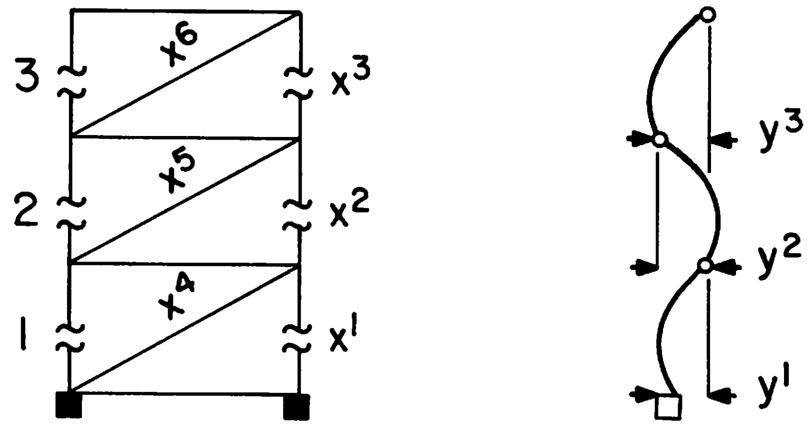


Fig. 1.2.1 Control System Block Diagram



(a) Braced Frame (b) Lumped Model

Fig. 1.1.1 Design of Earthquake Resistant Buildings

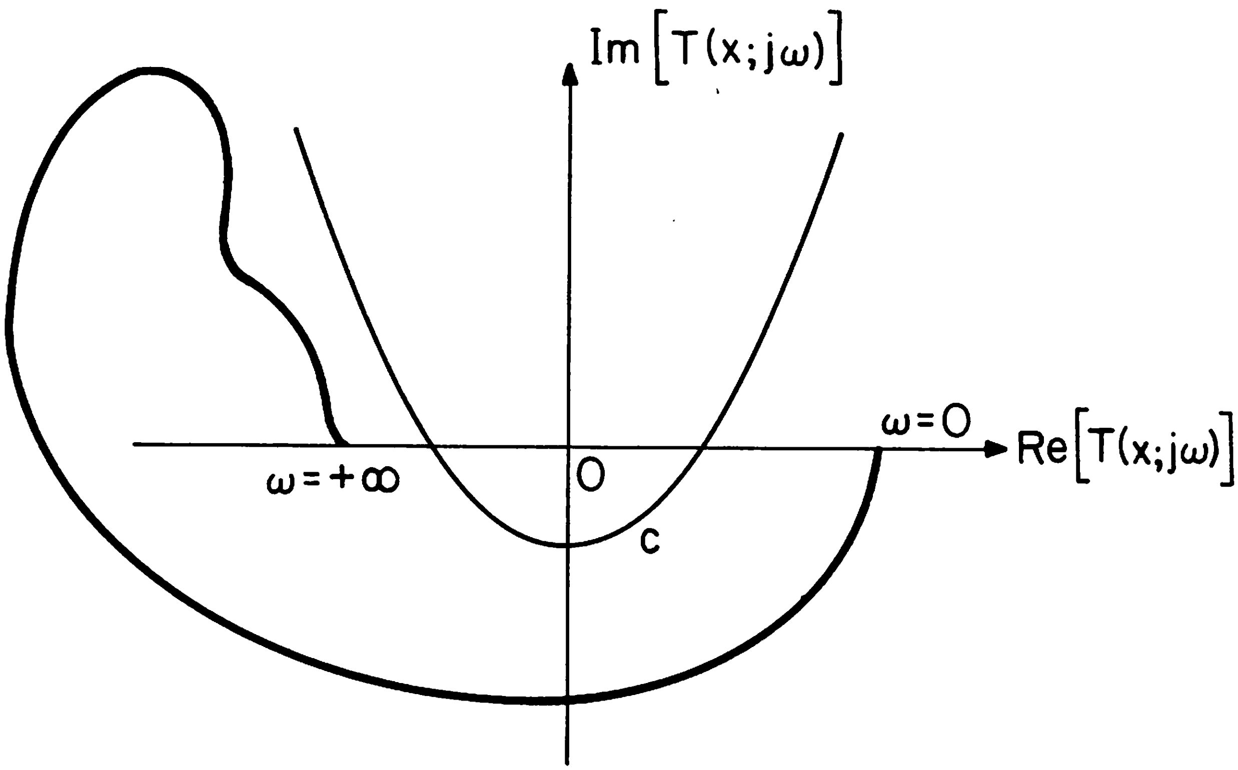


Fig. 1.2.2 Modified Nyquist Diagram

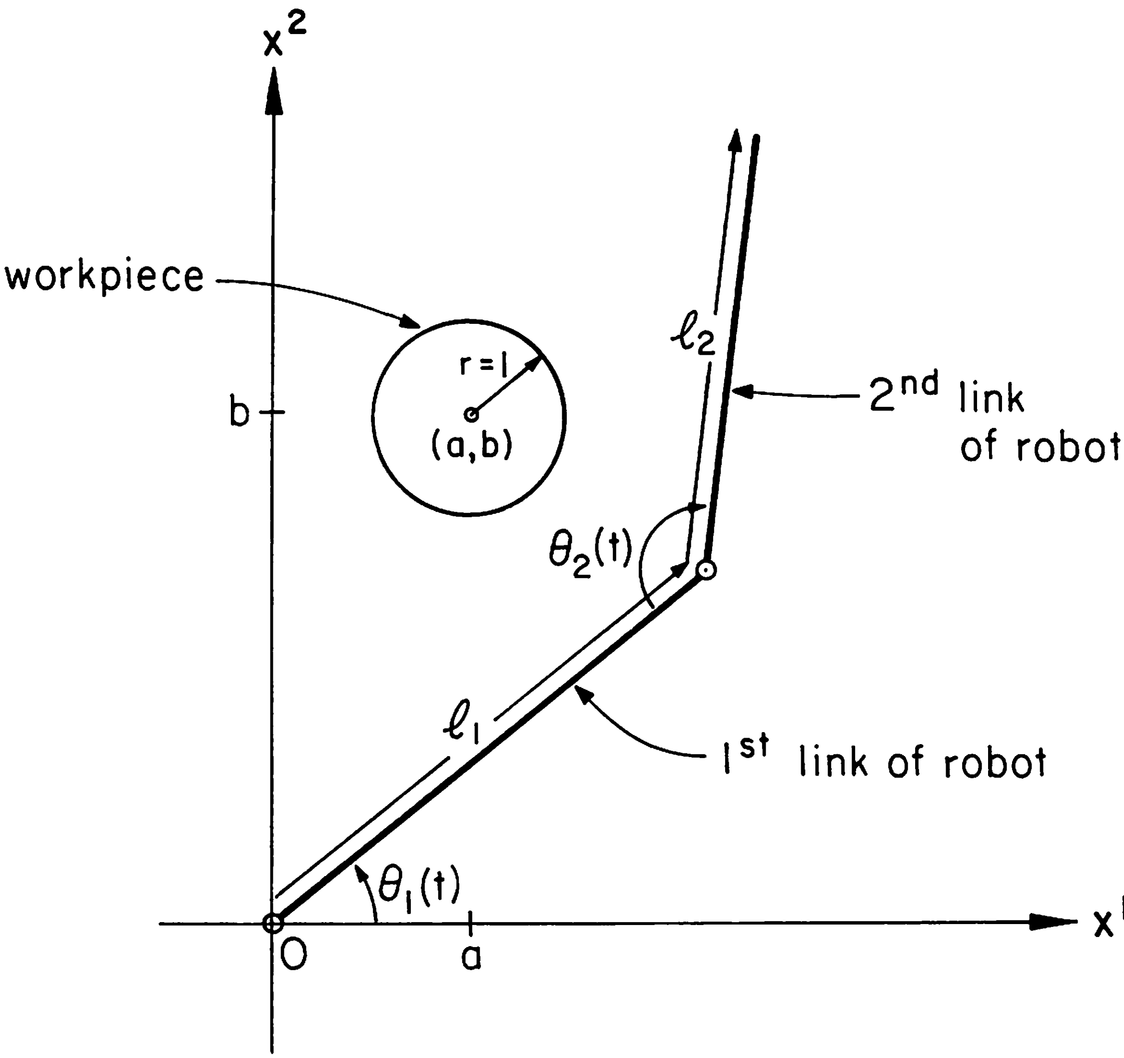
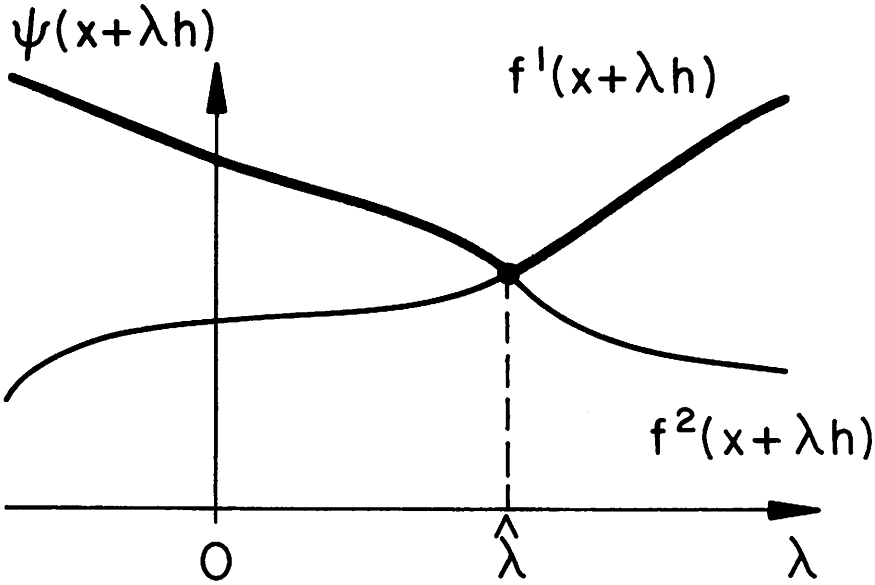


Fig. 1.4.1 Robot Arm Configuration



$\psi(\cdot)$ is not differentiable at $\hat{\lambda}$.

Fig. 3.1 Geometry of max functions.

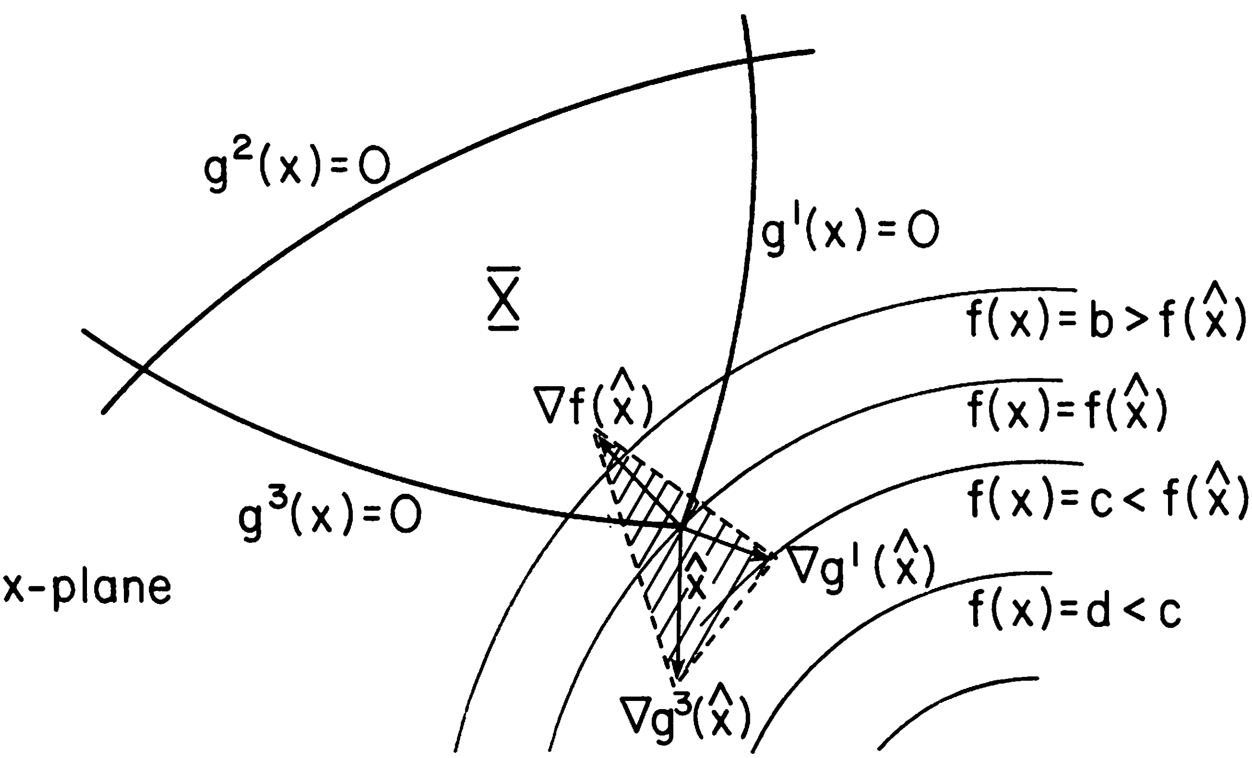


Fig. 4.1 Geometry of an optimal solution.

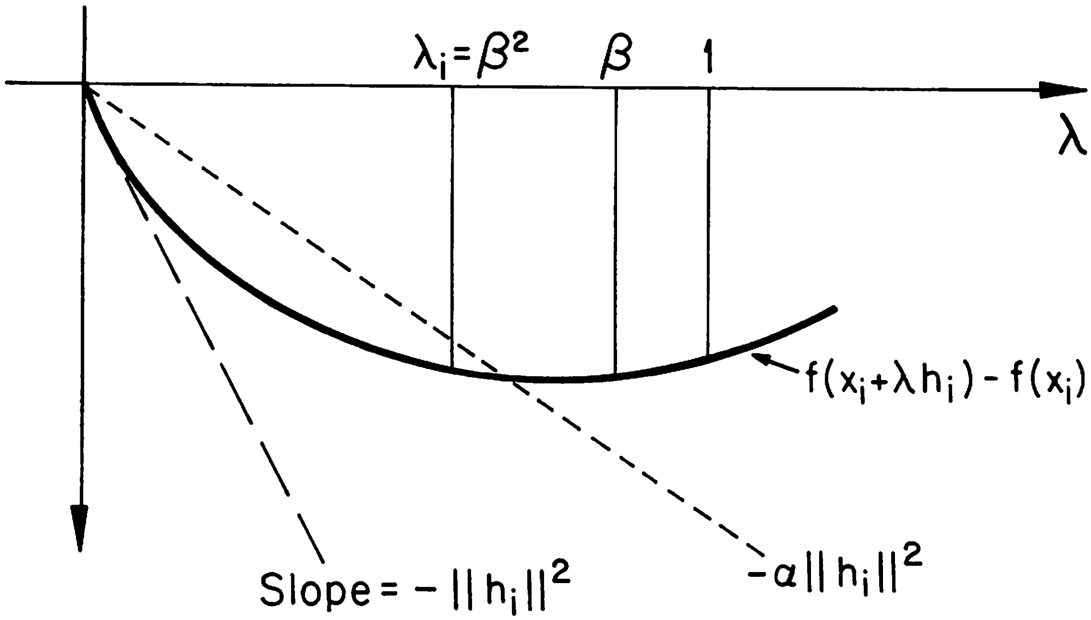


Fig. 5.1. Armijo step size calculation

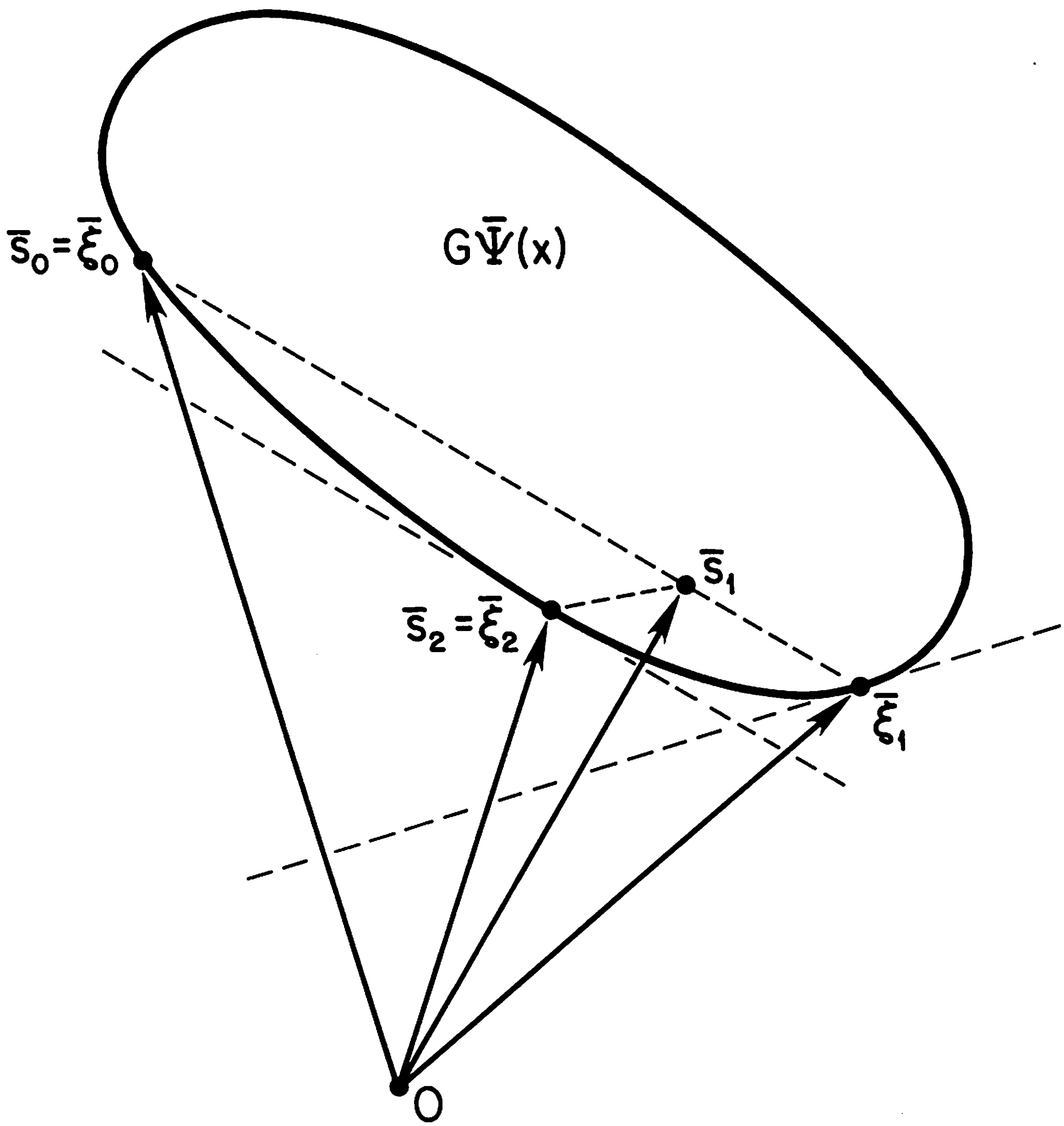
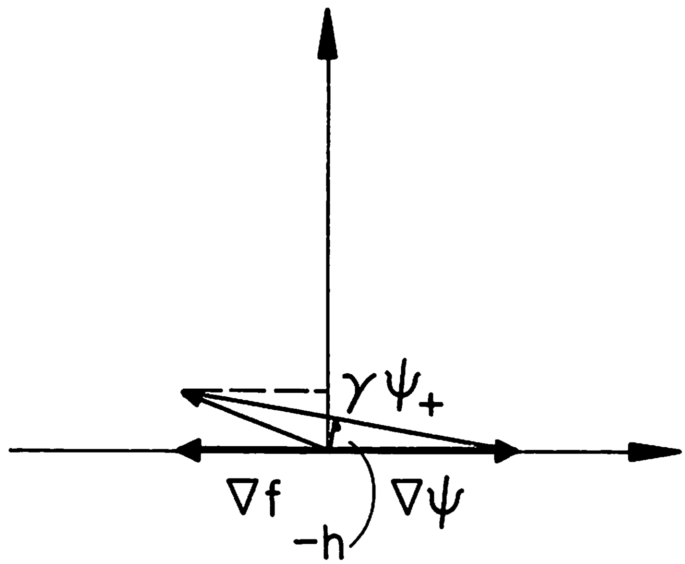
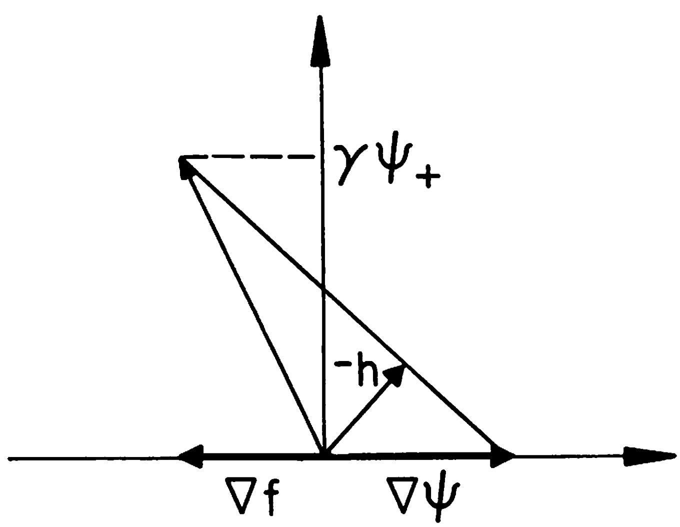


Fig. 5.2. Geometry of Proximity Algorithm 5.3.



(a) ψ_+ small



(b) ψ_+ large

Fig. 5.3 Effect of $\gamma\psi_+$ on search direction.

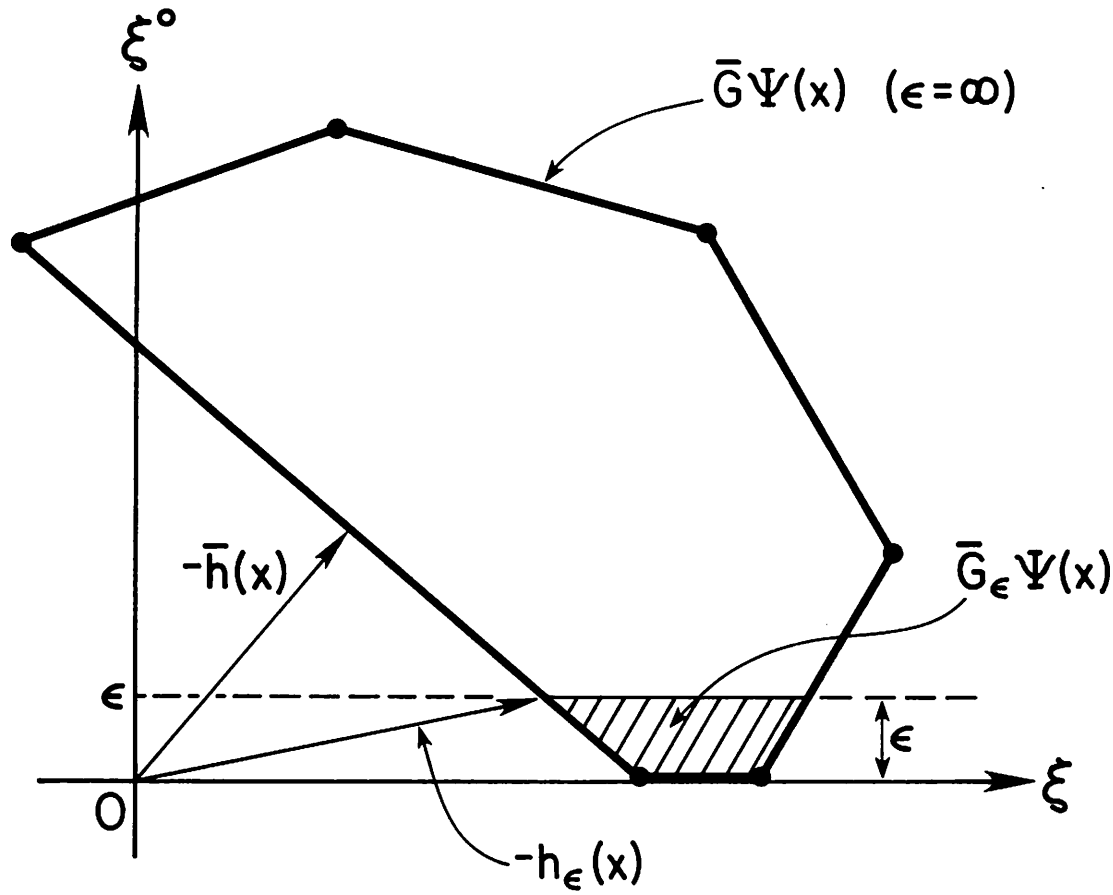


Fig. 6.1 Effect of $\epsilon > 0$ on the augmented search direction $\bar{h}_\epsilon(x)$ resulting from the use of 6.1a.