A STUDY OF MINIMIZING SEQUENCES

by

E. Polak and Y. Y. Wardi

Memorandum No. UCB/ERL M82/22

22 March 1982

A STUDY OF MINIMIZING SEQUENCES

by

E. Polak and Y. Y. Wardi

Memorandum No. UCB/ERL M82/22

22 March 1982

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A Study of Minimizing Sequences

E. Polak† and Y.Y. Wardi††

Abstract

Differentiable and nondifferentiable optimization problems in normed spaces may fail to have solutions. Even when they have solutions, optimization algorithms may produce minimizing sequences that have no accumulation points. To deal with this difficulty, this paper examines optimization problems as problems on sequences, in an extended normed space, and derives first and second order optimality conditions for them.

---

† Professor, Department of Electrical Engineering and Computer Sciences, and Electronics Research Laboratory, University of California, Berkeley, California, USA.

†† Member of Technical Staff, Bell Laboratories, Holmdel, New Jersey, USA.

1. Introduction

Engineering design periodically produces new classes of optimization problems in normed spaces. Among the earliest of such problems were those of optimal control. More recently design centering and tuning [1,2] has produced problems with maxminmax functions in the constraints, and the design of wings, turbine blades and bridges has produced problems with eigenvalue inequalities [3,4,5]. Thus, in designing the profile of a seismically resistant bridge, one may wish to minimize the weight of the structure, while considering both its low amplitude linear behaviour and large amplitude nonlinear behaviour. As a linear structure, the lowest natural frequency of the bridge must lie above a specified value; while as a nonlinear structure, its excursions, with respect to time, produced by a set of earthquakes, must be maintained within prescribed limits so as to avoid destruction. Of course, there are also constraints on the profile of the bridge, itself.

Abstractly, such problems can be viewed as being of the form

$$P : \min\{f(x) \mid x \in X\} \tag{1.1}$$

where X is a subset of $\mathcal{X}$, a "convenient" topological space, in which P may or may not have a solution, and $f : \mathcal{X} \to \mathbb{R}$ is continuous and bounded on X. $\mathcal{X}$ is convenient in the sense that it is reasonably easy to construct and analyse an optimization algorithm in its topology. The algorithm is usually accompanied by a convergence theorem which states that if a sequence $\{x_i\}_{i=0}^{\infty}$ constructed by the algorithm has accumulation points, then all these accumulation points are in X and satisfy some condition of optimality. Now, even when X is closed and bounded, a sequence $\{x_i\}_{i=0}^{\infty}$, constructed by an optimization algorithm, may fail to have accumulation points either because (1.1) has no solution in the topology of X or because of the particular process used by the algorithm in constructing the x_i. As an example of the latter, consider the case where $X = L_\infty[0,1]$, the x_i are all continuous functions, but all the local minima x^* are only piecewise continuous. Obviously, such phenomena are disturbing since they lead to the conclusion that the convergence theorems in question are vacuous.

A standard approach in dealing with the above described difficulties is to replace the space X with a suitable extension, e.g., as in the case when ordinary controls are replaced by relaxed controls in optimal control problems [17]. Unfortunately, by and large, the construction of minimal extensions can be quite difficult, for example, as was the case with relaxed controls. Consequently, the question arises whether there may not be an easier approach, one based on the concept of minimizing sequences in the original space X and utilizing elements of optimization algorithm theory. This paper explores this question.

In constructing a theory of optimality conditions for minimizing sequences we felt it was important to take into account the following facts and goals. Firstly, optimization algorithms construct sequences $\{x_i\}$ which may not be Cauchy and along which the cost $f(x_i)$ may not be monotonically decreasing. Secondly, by and large, any subsequence of a locally minimizing sequence, constructed by an optimization algorithm, is a locally minimizing sequence. Thirdly, in computing a search direction, many algorithms solve a relatively simple program whose value can be viewed as an *optimality function* $\theta(x)$. These optimality functions vanish only at stationary points. When an optimality function is continuous, it can be used loosely as an ε-solution detector.

We show that a natural extension of the space X in (1.1) is to an extended norm space X^s of sequences in X and that many of the well known optimality functions can be used to provide both first and second order optimality conditions for characterizing minimizing sequences of P.

2. The Space of Minimizing Sequences

Consider the problem

$$P: \inf\{f(x) \mid x \in X\} \tag{2.1}$$

where X is a subset of a normed space $\mathcal{X}$ and $f : \mathcal{X} \to \mathbb{R}$ is continuous and bounded from below on X.

Our definitions of global and local minimizing sequences reflect the fact that when an optimization algorithm is applied to P it may produce a sequence $\{x_i\}_{i=0}^{\infty}$ which is *not* Cauchy and for which $\{f(x_i)\}_{i=0}^{\infty}$ is *not* monotonically decreasing. However, any accumulation point x^*, that such a sequence $\{x_i\}_{i=0}^{\infty}$ may have will be a local minimum, or at least a stationary point.

Definition 2.1: $\{x_i\}_{i=0}^{\infty}$, $x_i \in X$, $i = 0, 1, 2, \ldots$, is an *eventually feasible* sequence (for P) if

$$\lim_{i\to\infty} \inf\{\|x_i - x\| \mid x \in X\} = 0 \tag{2.2}$$

□

Definition 2.2: A bounded eventually feasible sequence $\{\hat{x}_i\}_{i=0}^{\infty}$ is a *globally minimizing* sequence (for P) if for all bounded eventually feasible sequences $\{x_i\}_{i=0}^{\infty}$

$$\overline{\lim_{i\to\infty}}\, f(\hat{x}_i) \le \overline{\lim_{i\to\infty}}\, f(x_i) \tag{2.3}$$

□

The following result is obvious.

Proposition 2.1: Suppose that $\{\hat{x}_i\}_{i=0}^{\infty}$ is a globally minimizing sequence then

a) $\lim_{i\to\infty} f(\hat{x}_i)$ exists;

b) every infinite subsequence of $\{x_i\}_{i=0}^{\infty}$ is a globally minimizing sequence. □

Our definition of a locally minimizing sequence, below, ensures the property that every subsequence of a locally minimizing sequence, constructed by an algorithm, is also a locally minimizing sequence.

<u>Definition 2.3</u>: A bounded eventually feasible sequence $\{\hat{x}_i\}_{i=0}^{\infty}$ is a *locally minimizing* sequence (for P) if there exists a $\rho > 0$ such that for all eventually feasible sequences $\{x_i\}_{i=0}^{\infty}$ satisfying

$$\overline{\lim_{i\to\infty}} \; \| \hat{x}_i - x_i \| \leq \rho \tag{2.4}$$

We have

$$\overline{\lim_{\substack{K \\ i\to\infty}}} f(x_i) \leq \overline{\lim_{\substack{K \\ i\to\infty}}} \; f(x_i) \tag{2.5}$$

for all infinite subsets $K \subset \mathbb{N}_+ \triangleq \{0,1,2,\ldots\}$, with $i \overset{K}{\to} \infty$ denoting : $i \in K$, $i \to \infty$. □

<u>Proposition 2.2</u>: Suppose that $\{\hat{x}_i\}_{i=0}^{\infty}$ is a locally minimizing sequence. Then for every infinite subset $K \subset \mathbb{N}_+$, $\{\hat{x}_i\}_{i\in K}$ is a locally minimizing sequence.

<u>Proof</u>: Let $\rho > 0$ be as specified in Definition (2.3) and suppose that there is a subsequence $\{\hat{x}_i\}_{i\in K'}$, for some infinite $K' \subset \mathbb{N}_+$ which is not a minimizing sequence. Then there must exist an infinite $K'' \subset K'$ and a sequence $\{\bar{x}_i\}_{i\in K''}$ such that

$$\overline{\lim_{\substack{K'' \\ i\to\infty}}} \; \| \bar{x}_i - \hat{x}_i \| \leq \rho \tag{2.6}$$

and

$$\overline{\lim_{\substack{K'' \\ i\to\infty}}} f(x_i) < \overline{\lim_{\substack{K'' \\ i\to\infty}}} f(\hat{x}_i) \quad . \tag{2.7}$$

Let $x_i = \bar{x}_i$ for all $i \in K''$ and let $x_i = \hat{x}_i$ otherwise, then $\{x_i\}_{i=0}^{\infty}$ satisfies (2.4) but fails to satisfy (2.5) for $K = K''$, contradicting the assumption that $\{\hat{x}_i\}_{i=0}^{\infty}$ is a locally minimizing sequence. This completes our proof. □

Remark: Note that for a *locally* minimizing sequence $\{x_i\}_{i=0}^{\infty}$, the bounded sequence $\{f(x_i)\}_{i=0}^{\infty}$ may have more than one accumulation point.

□

We are now ready to put the problem P into one-to-one correspondence with a problem P^s defined in a space of sequences and thus remove the need for determining whether minimizing sequences do or do not have accumulation points in X.

Definition 2.4: a) We define $\tilde{X}$ to be the *class* of infinite sequences $\{x_i\}_{i=0}^{\infty}$, with $x_i \in X$, $i \in \mathbb{N}_+$.

b) We define $\{x_i\}_{i=0}^{\infty}$, $\{y_i\}_{i=0}^{\infty}$ in $\tilde{X}$ to be *equivalent* if

$$\overline{\lim_{i\to\infty}} \; \| x_i - y_i \| = 0$$

We shall denote this equivalence relation by the symbol $\simeq$.

c) We define the vector space X^s to be $\tilde{X}/\sim$, with addition and scalar (real) multiplication defined as follows:

$$\{x_i\}_{i=0}^{\infty} + \{y_i\}_{i=0}^{\infty} = \{x_i+y_i\}_{i=0}^{\infty} \tag{2.8}$$

$$\alpha\{x_i\}_{i=0}^{\infty} = \{\alpha x_i\}_{i=0}^{\infty} \tag{2.9}$$

□

Proposition 2.3: The operations of addition and scalar multiplication in X^s, as given by (2.8) and (2.9), are well defined, i.e. if $s_x, s'_x, s_y, s'_y \in \tilde{X}$ are such that $s_x \sim s'_x$ and $s_y \sim s'_y$, then $(s_x+s_y) \sim (s'_x+s'_y)$ and for any $\alpha \in \mathbb{R}$, $\alpha\, s_x \sim \alpha\, s'_x$. □

Next, we define the concepts of an extended norm and of an extended normed vector space.

Definition 2.5: a) Let Z be a real vector space and let $|||\cdot|||$ be a functional on Z which can take on the value ∞. We say that $|||\cdot|||$ is an *extended norm* if it has all the properties of a norm on any convex subset $B \subset Z$ on which $|||\cdot|||$ is finite, viz :

i) $|||z||| \geq 0 \quad \forall z \in Z$

ii) $|||z||| = 0 \iff z = 0,$

iii) $|||\alpha z||| = |\alpha| \; |||z|||, \qquad \forall z \in B,$

iv) $|||z_1+z_2||| \leq |||z_1||| + |||z_2|||, \quad \forall z \in B.$

b) With $|||\cdot|||$ an extended norm, we say that $(Z, |||\cdot|||)$ is an *extended normed space* if the vector addition operation is bicontinuous and the scalar multiplication is continuous on $B \times \mathbb{R}$ (with respect to $|||\cdot|||$), where B is any *convex bounded* set in Z.

□

We note that the scalar multiplication operator cannot be continuous at any $z \in Z$, Z an extended normed space, at which $|||z||| = \infty$, since for any $\alpha_i \to 0$ as $i \to \infty$, $\alpha_i > 0$, $|||\alpha_i z||| = \infty$ for all $i \in \mathbb{N}_+$, while $||| 0.z||| = 0$.

To make X^s an extended normed space we define $|||\cdot|||$ on X^s by

$$|||z||| \triangleq \overline{\lim_{i\to\infty}} |||x_i||| \tag{2.10}$$

where $\{x_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class z.

The following result is obvious in view of the definition of the equivalence relation $\sim$ and the properties of the norm $\|\cdot\|$ on X.

<u>Proposition 2.4</u>: The functional $|||\cdot|||$ defined by (2.10) is a well defined extended norm on X^s, and $(X^s, |||\cdot|||)$ is an extended normed space.

□

For the problem P to make sense in the context of X^s, it is necessary that the asymptotic behaviour of $f(\cdot)$ on $z \in X^s$ be independent of the particular choice of a sequence $\{x_i\}_{i=0}^{\infty}$ in the equivalence class defined by z. Consequently we postulate as follows.

<u>Assumption 2.1</u>: The function $f(\cdot)$ is *uniformly continuous* on bounded sets.

□

We obtain immediately

<u>Proposition 2.5</u>: If $\{x_i\}_{i=0}^{\infty} \sim \{x_i'\}_{i=0}^{\infty}$ and $|||\{x_i\}_{i=0}^{\infty}||| < \infty$, then $\overline{\lim}\, f(x_i) = \overline{\lim}\, f(x_i')$. Furthermore, if $\{x_i\}_{i=0}^{\infty}$ is eventually feasible then so is $\{x_i'\}_{i=0}^{\infty}$.

□

Let $X^s \subset \mathcal{X}^s$ be defined by

$$X^s \triangleq \{z \in \mathcal{X}^s \,|\, |||z||| < \infty \text{ and } z \text{ is eventually feasible}\} \tag{2.11}$$

and let the extended function $f^s : \mathcal{X}^s \to \overline{\mathbb{R}}$ be defined by

$$f^s(z) = \overline{\lim_{i\to\infty}}\, f(x_i) \tag{2.12}$$

where $z = \{x_i\}_{i=0}^{\infty}$. We now define the problem P^s as follows :

$$P^s : \min\{f^s(z) \,|\, z \in X^s\} \tag{2.13}$$

<u>Proposition 2.6</u>: Problem P^s has a solution if and only if P admits a bounded minimizing sequence. Furthermore, the values of P and P^s are the same.

□

We note that if $\hat{z}$ solves P^s and $\{\hat{x}_i\}_{i=0}^{\infty}$, a sequence in the equivalence class z, has an accumulation point $\hat{x}$, then $\hat{x}$ is a feasible minimizer for P. Similarly, if $\hat{x}$ is a global minimizer for P, then $\hat{z} = \{\hat{x}_i\}_{i=0}^{\infty}$, $\hat{x}_i \equiv \hat{x}$, is a solution to P^s.

The introduction of the function f^s makes the formalism(2.13) very appealing with respect to global solutions. However, it fails with respect to local solutions. Thus, if we define, as is customary, $\hat{z} \in X^s$, to be a local minimizer of P^s if for some $\rho > 0$,

$$f^s(\hat{z}) \leq f^s(z) \;\forall z \in B(\hat{z},\rho) \tag{2.14}$$

with $B(\hat{z},\rho) \triangleq \{z \,|\, |||z-\hat{z}||| \leq \rho\}$, we find that $\hat{z}$ is not a locally minimizing sequence, as defined in Definition 2.4. Consequently, we use the following.

<u>Definition 2.6</u>: We say that $\hat{z} \in X^s$ is a *local minimizer* for P^s if $\{\hat{x}_i\}_{i=0}^{\infty}$ is a locally minimizing sequence for P, where $\{\hat{x}_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class $\hat{z}$. □

Quite clearly, because of Assumption 2.1, local minimizers for P^s are well defined. Furthermore, if $\hat{z}$ is a solution to P^s and $\{\hat{x}_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class defined by z, then $\{\hat{x}_i\}_{i=0}^{\infty}$ is a minimizing sequence for P.

We are now ready to proceed to the next task : the development of necessary and sufficient conditions for characterizing local minimizers of P^s.

3. Unconstrained Minimization

We begin by assuming that $X = \mathcal{X}$ in P, so that $X^s = \mathcal{X}^s$ in P^s. We shall consider both differentiable and nondifferentiable cost functions $f(\cdot)$.

We shall characterize optimality of minimizing sequences by means of first and second order optimality functions which tend to zero along minimizing sequences. Since we will be dealing with bounded sequences which are not necessarily Cauchy, we will have to require that various properties hold uniformly on bounded sets.

We shall denote the first Frechet derivative of $f(\cdot)$ by $f_x(\cdot)$ and the second Frechet derivative by $f_{xx}(\cdot)$. We note that f_x maps $\mathcal{X}$ into $\mathcal{X}'$, the dual of $\mathcal{X}$ and that f_{xx} maps $\mathcal{X}$ into $\mathcal{X}''$ the dual of $\mathcal{X}'$.

Proposition 3.1: Suppose that $f(\cdot)$ is uniformly, continuously Frechet differentiable on bounded sets in $\mathcal{X}$. Let $\Theta^{1s}_{uc} : \mathcal{X}^s \to \mathbb{R}$ be defined on bounded $z \in \mathcal{X}^s$ by

$$\Theta^{1s}_{uc}(z) \triangleq \varliminf_{i\to\infty} \inf\{(f_x(x_i),h) \,\big|\, \|h\| \leq 1\} \tag{3.1}$$

where $\{x_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class z. If $\hat{z}$ is a local minimizer for P^s, then $\Theta^{1s}_{uc}(\hat{z}) = 0$.

Proof: First we note that, because of the assumption on $f_x(\cdot)$, the function $\Theta^{1}_{uc} : \mathcal{X} \to \mathbb{R}$ defined by

$$\Theta^{1}_{uc}(\bar{x}) \triangleq \inf\{(f_x(\bar{x}),h) \,\big|\, \|h\| \leq 1\} \tag{3.2}$$

is uniformly continuous on bounded sets and hence $\Theta^{1s}_{uc}(\cdot)$ is well defined. Also, $\Theta^{1s}_{uc}(z) \leq 0$ for any $z \in \mathcal{X}^s$. Hence suppose that $\hat{z}$ $(\doteq\{\hat{x}_i\}_{i=0}^{\infty})$ is a local minimizer for P^s, with associated radius $\hat{\rho} > 0$, and that $\Theta^{1s}_{uc}(\hat{z}) < 0$. Then there exists a $\delta > 0$ and an infinite subsequence

$\{\hat{x}_i\}_{i\in K}$ such that $\Theta^1_{uc}(\hat{x}_i) \le -\delta$ for all $i \in K$. For $i \in K$, let $h_i \in X$, be such that $\|h_i\| \le 1$ and $(f_x(\hat{x}_i), h_i) \le -\delta/2$. Then, because $f_x(\cdot)$ is uniformly continuous on bounded sets, there exists a $\bar{\lambda} \in (0,\hat{\rho}]$ such that, with $s_i \in (0,1)$, by the mean value theorem,

$$f(\hat{x}_i + \bar{\lambda}h_i) - f(x_i) = \bar{\lambda}(f_x(x_i + s_i\bar{\lambda}h_i), h_i) \le -\bar{\lambda}\,\delta/4 \qquad \text{for all } i \in K \quad (3.3)$$

Consider now the sequence $\{x_i^*\}_{i=0}^{\infty}$ defined by $x_i^* = \hat{x}_i$ for all $i \notin K$ and $x_i^* = \hat{x}_i + \bar{\lambda}h_i$ for all $i \in K$. Clearly, (2.4) fails for $x_i = x_i^*$, $i \in \mathbb{N}_+$ and hence we get a contradiction. □

Proposition 3.2: Suppose that $f(\cdot)$ is twice uniformly, continuously Frechet differentiable on bounded sets in X. Let $\Theta^{2s}_{uc} : X^s \to \mathbb{R}$ be defined on bounded $z \in X^s$ by

$$\Theta^{2s}_{uc}(z) \triangleq \varliminf_{i\to\infty} \inf\{(f_{xx}(x_i)(h),h) \mid \|h\| \le 1\} \qquad (3.3)$$

where $\{x_i\}_{i=0}^{\infty}$ is any sequence in the class z. If $\hat{z} \in X^s$ is a local minimizer for P^s then $\Theta^{2s}_{uc}(\hat{z}) \ge 0$.

Proof: Let $\Theta^2_{uc} : X \to \mathbb{R}$ be defined by

$$\Theta^2_{uc}(\bar{x}) \triangleq \inf\{(f_{xx}(\bar{x})(h),h) \mid \|h\| \le 1\} \qquad (3.4)$$

Since $\Theta^2_{uc}(\cdot)$ is uniformly continuous on bounded sets, $\Theta^{2s}_{uc}(\cdot)$ is well defined. Now suppose that $\hat{z}(=\{\hat{x}_i\}_{i=0}$ is a local minimizer for P^s, with associated radius $\hat{\rho}$, and that $\Theta^{2s}_{uc}(\hat{z}) = -\delta < 0$. Then there exists an infinite subsequence $\{\hat{x}_i\}_{i\in K}$ and a corresponding sequence $\{h_i\}_{i\in K}$, such that $\|h_i\| \le 1$ and $(f_x(x_i),h_i) \le 0$, satisfying $(f_{xx}(\hat{x}_i)(h_i),h_i) \le -\delta/2$ for all $i \in K$. Hence, making use of the uniform continuity of $f_{xx}(\cdot)$ and the Taylor formula with remainder, we find that there exist a $\bar{\lambda} \in (0,\hat{\rho}\,]$ such that

$$\begin{aligned} f(x_i+\bar{\lambda}h_i) - f(x_i) &= \bar{\lambda}(f_x(x_i),h_i) + \bar{\lambda}^2\int_0^1 (1-s)(f_{xx}(x_i+s\lambda\, h_i)(h_i),h_i)\,ds \\ &\le -\bar{\lambda}^2\,\delta/8 \;, \; \forall i \in K \end{aligned} \qquad (3.5)$$

Setting $x_i^* = \hat{x}_i$ for $i \notin K$ and $x_i^* = x_i + \bar{\lambda} h_i$ for $i \in K$, we get a contradiction of the fact that $\hat{z}$ is a local minimizer. This completes the proof. □

It is also quite easy to prove the following result.

<u>Proposition 3.3</u>: Suppose that $f(\cdot)$ is twice uniformly, continuously Frechet differentiable on bounded sets in X. Suppose that $\hat{z} \in X^s$ is such that (i) $|||\hat{z}||| < \infty$, (ii) $\underline{\lim}\, f(\hat{x}_i) > -\infty$, for $\{x_i\}_{i=0}^{\infty}$ in the equivalence class $\hat{z}$, (iii) $\Theta_{uc}^{1s}(\hat{z}) = 0$ and (iv) $\Theta_{uc}^{2s}(\hat{z}) > 0$. Then $\hat{z}$ is a local minimizer for P^s. □

Thus, for the case where $f(\cdot)$ is differentiable, we see that the standard optimality conditions for P lead directly to corresponding optimality conditions for P^s. As we shall shortly see, this is not true for the nondifferentiable case : when $f(\cdot)$ is assumed to be only uniformly Lipschitz continuous on bounded sets in X. We recall [6] that, the *generalized gradient* $\partial f(\cdot)$ of $f(\cdot)$, is defined as the subset of X' satisfying for every x, h in X the relation

$$f^o(x;h) \triangleq \varlimsup_{\substack{\lambda \downarrow 0 \\ y \to 0}} \frac{f(x+y+\lambda h)-f(x+y)}{\lambda} = \sup\{(\xi,h) \mid \xi \in \partial f(x)\} \tag{3.6}$$

where $f^o(x;h)$ is called the *generalized directional derivative* of $f(\cdot)$ at x, in the direction h. The sets $\partial f(x)$ are bounded on bounded sets in X and upper semi-continuous. The standard first order optimality condition for P, see [6,7] , is that if $\hat{x}$ is a local minimizer for P then $0 \in \partial f(\hat{x})$. A first attempt to convert this statement into an optimality function produces the following candidate $\Theta : X \to \mathbb{R}$ defined by

$$\Theta(x) \triangleq \sup_{\|h\| \leq 1} \inf_{\xi}\{(\xi,h) \mid \xi \in \partial f(x)\} \tag{3.7}$$

which is recognized as being a generalization of the function $\min\{\|h\| \mid h \in \partial f(x)\}$ in $\mathbb{R}^n$. Since $\partial f(\cdot)$ is only u.s.c., $\Theta(\cdot)$ is not continuous and it is very easy to construct functions $f(\cdot)$ and sequences $\{x_i\}$ such that $x_i \to \hat{x}$, a local minimizer, so that $0 \in \partial f(\hat{x})$ and hence $\Theta(\hat{x}) = 0$, but $\Theta(x_i) = -1$ for all i. Clearly, $\Theta(\cdot)$ cannot be used to characterize minimizing sequences for P. However, referring to [2 , 8], we find that there are other functions that can. For example, following [2 , 8], for any $\varepsilon \geq 0$, and $x \in X$, let

$$\partial_\varepsilon f(x) \triangleq \bigcup_{x' \in B(x,\varepsilon)} \partial f(x') \tag{3.8}$$

where $B(x,\varepsilon) \triangleq \{x' \in X \mid \|x'-x\| \leq \varepsilon\}$ and

$$\Theta_\varepsilon(x) \triangleq \sup_{\|h\| \leq 1} \inf_{\xi}\{(\xi,h) \mid \xi \in \partial_\varepsilon f(x)\} \tag{3.9}$$

It can be shown [2,8] that $\partial_\varepsilon f(\cdot)$ is u.s.c. and closed and bounded on bounded sets. Next, with $\beta \in (0,1)$, let

$$E \triangleq \{0,1,\beta,\beta^2,\ldots\} \tag{3.10}$$

and let $\varepsilon : X \to \mathbb{R}$ be defined by

$$\varepsilon(x) = \max\{\varepsilon \in E \mid \Theta_\varepsilon(x) \geq \varepsilon\} \tag{3.11}$$

The following result can be found in [2 , 8].

Proposition 3.4: $0 \in \partial f(x) \iff \theta(x) = 0 \iff \varepsilon(x) = 0$. □

However, while $\{\theta(x_i)\}_{i=0}^{\infty}$ generally will not converge to zero for $\{x_i\}_{i=0}^{\infty}$ a locally minimizing sequence for P, $\{\varepsilon(x_i)\}_{i=0}^{\infty}$ must converge to zero along such a sequence, as we shall shortly see. Consequently, if we define $\Theta^s_{ucnd} : X^s \to \mathbb{R}$ by

$$\Theta^s_{ucnd}(z) \triangleq \overline{\lim}\ \varepsilon(x_i) \tag{3.12}$$

with $\{x_i\}_{i=0}^{\infty}$ any sequence in the equivalence class z.

Proposition 3.5: If $\hat{z} \in X$ is a bounded local minimizer for P^s, then $\Theta^s_{ucnd}(\hat{z}) = 0$.

¤

Proposition 3.5 is a special case of a result to follow and hence its proof will not be given.

We now proceed formally.

Definition 3.1: Let $G(\cdot)$ be a map from X into $2^{X'}$ (i.e. the class of subsets of X'). We shall say that $G(\cdot)$ is *uniformly u.s.c. on bounded sets, with respect to $\partial f(\cdot)$*, if for any bounded set $B \subset X$ and any $\delta > 0$, there exists a $\rho > 0$ such that for all $x, y \in B$ satisfying $\|x-y\| < \rho$ and any $\eta \in \partial f(y)$, there exists a $\xi \in G(x)$ such that $\|\xi-\eta\| < \delta$. □

It is easily seen that if $G(\cdot)$ is *any* map which is uniformly u.s.c. on bounded sets w.r.t. $\partial f(\cdot)$, then (i) $\partial f(x) \subset G(x)$, and (ii) for any $\delta > 0$, there exists an $\varepsilon > 0$ such that $\partial_\varepsilon f(x) \subset N_\varepsilon(G(x)) \triangleq \{x' \mid \inf_{y\in G(x)} \|x'-y\| \leq \varepsilon\}$. In fact, we have the following result.

Proposition 3.6: For any $\varepsilon > 0$, the map $\partial_\varepsilon f(\cdot)$ is uniformly u.s.c. on bounded sets, w.r.t. $\partial f(\cdot)$.

□

Proof: Let $x \in X$ be arbitrary and let $\delta > 0$. Then, setting $\rho = \varepsilon > 0$ we get for any $y \in B(x,\rho)$ that $\partial f(y) \subset \partial_\varepsilon f(x)$ by definition. Hence the proposition holds.

□

Proposition 3.7: Let $G : X \to 2^{X'}$ be uniformly u.s.c. on bounded sets with respect to $\partial f(\cdot)$. If $\hat{z}$ is a local minimizer for P^s, then $\Theta^s_G(\hat{z}) = 0$, where

$$\Theta^s_G(\hat{z}) \triangleq \overline{\lim}\ \Theta_G(\hat{x}_i)$$

$$\triangleq \overline{\lim} \sup_h \inf_\xi \{(\xi,h) \mid \xi \in G(x_i),\ h \in X,\ \|h\| \leq 1\} \tag{3.13}$$

<u>Proof</u>: Clearly, $\Theta_G^s(\cdot)$ is well defined. Now suppose that $\hat{z}$ is a local minimizer for P^s, with associated radius $\hat{\rho} > 0$, and that $\Theta_G^s(\hat{z}) = \hat{\delta} > 0$. (Clearly $\Theta_G^s(z) \geq 0$ for all bounded $z \in X^s$). Then there exists an infinite subset $K \subset \mathbb{N}_+$ such that $\Theta_G(\hat{x}_i) \geq \hat{\delta}/2$ for all $i \in K$, with $\{\hat{x}_i\}_{i=0}^{\infty}$ any sequence in the class $\hat{z}$. Let $\delta = \hat{\delta}/4 > 0$ and a corresponding $\rho > 0$ satisfy the requirements of Definition 3.1. For $i \in K$, let h_i be such that $(\xi,h_i) \geq \hat{\delta}/2$ for all $\xi \in G(\hat{x}_i)$. Then, by the mean value theorem [9] ,

$$f(\hat{x}_i-\lambda h_i) - f(\hat{x}_i) = \lambda<\xi_{i\lambda},h_i> \tag{3.14}$$

with $\xi_{i\lambda} \in \partial f(\hat{x}_i-s\lambda h_i)$ for some $s \in (0,1)$. Now let $\bar{\lambda} = \min(\hat{\rho},\rho)$, then (i) $\|(\hat{x}_i-\bar{\lambda}h_i) - \hat{x}_i\| \leq \hat{\rho}$ for all $i \in K$, (ii) with $\eta_i \in G(\hat{x}_i)$ such that $\|\eta_i - \xi_{i\bar{\lambda}}\| \leq \hat{\delta}/4$, we get from (3.14)

$$\begin{aligned} f(\hat{x}_i-\bar{\lambda}h_i) - f(\hat{x}_i) &= -\bar{\lambda}(\xi_{i\bar{\lambda}},h_i) = -\lambda\,(\eta_i+(\xi_{i\bar{\lambda}}-\eta_i),h_i) \\ &\leq -\lambda(-(\eta_i,h_i) + \|\xi_{i\lambda} - -\eta_i\|) \\ &\leq \bar{\lambda}(-(\eta_{i1}h_i) + \hat{\delta}/4) \\ &\leq -\bar{\lambda}\,\hat{\delta}/4 \ . \end{aligned} \tag{3.15}$$

Clearly, the sequence

$x_i^* = \hat{x}_i$ for $i \notin K$, $x_i^* = \hat{x}_i - \bar{\lambda}h_i$, violates (2.5) for the local minimizer $\hat{z}$, and hence we have a contradiction.

□

Referring to Proposition 3.7 we note that since $\partial f(x) \subset G(x)$ is always true, $0 \in \partial f(x) \Rightarrow 0 \in G(x)$, hence $\hat{x}$ optimal for P $\Rightarrow 0 \in G(\hat{x})$. However, $0 \in G(\hat{x})$ is obviously a weaker optimality condition than $0 \in \partial f(\hat{x})$. The condition $0 \in G(\hat{x})$ can be strengthened when $G(\cdot)$, can be parametrized, as follows.

<u>Proposition 3.8</u>: For $\varepsilon \geq 0$, let $G_\varepsilon : X \to 2^{X'}$ be a family of maps that are uniformly u.s.c. on bounded sets w.r.t $\partial f(\cdot)$, such that for all $x \in X$,

(i) $G_0(x) = \partial f(x)$;

(ii) $0 \le \varepsilon < \varepsilon' \Rightarrow G_\varepsilon(x) \subset G_{\varepsilon'}(x)$,

(iii) $G_\varepsilon(x)$ is convex and weak * compact.

Next, let $\Theta : X \to \mathbb{R}$ be defined by

$$\Theta(x) \triangleq \max\{\varepsilon \in E \mid \Theta_{G_\varepsilon}(x) \ge \varepsilon\} \tag{3.16}$$

(with E as in (3.9)) and $\Theta^s : X^s \to \mathbb{R}$ by

$$\Theta^s(z) \triangleq \overline{\lim}\ \Theta(x_i) \tag{3.17}$$

If $\hat{z}$ is a local minimizer for P^s, then $\Theta^s(\hat{z}) = 0$.

<u>Proof</u>: Suppose that $\hat{z}$ is a local minimizer for P^s, but $\Theta^s(\hat{z}) = \hat{\varepsilon} > 0$, with $\hat{\varepsilon} \in E$. Then there must exist an infinite $K \subset \mathbb{N}_+$ such that $\Theta(x_i) = \hat{\varepsilon}$ for all $i \in K$, i.e. $\Theta_{G_{\hat{\varepsilon}}}(x_i) \ge \hat{\varepsilon}$ for all $i \in K$.

But by Proposition 3.7, this contradicts the local minimality of $\{\hat{x}_i\}_{i \in K}$ and hence the proof is complete. □

The maps $\partial_\varepsilon f(\cdot)$ are not the only known examples of maps that are uniformly u.s.c. with respect to $\partial f(\cdot)$ and satisfy the assumptions of Proposition 3.8. In [10, 11] we find very different maps in this class that are used for optimization problems with eigenvalue constraints.

4. Constrained Minimization

In this section we shall propose a set of optimality functions for a variety of constrained problems of the form P^s. Since the proofs associated with these optimality functions are either quite straightforward, or follow directly from existing results, they will be omitted.

We begin with equality constrained problems (c.f. [12], Chap. 8).

Proposition 4.1: Suppose that $f(\cdot)$ is uniformly continuously differentiable on bounded sets in X and that

$$X = \{x | h^i(x) = 0,\ j \in \underline{\ell}\} \tag{4.1}$$

where $\underline{\ell} \triangleq \{1,2,\ldots,\ell\}$ and $h^j : X \to \mathbb{R}$ are uniformly continuously differentiable on bounded sets. Furthermore, suppose that the functionals $h^j_x(\bar{x})$, $j \in \underline{\ell}$, $\bar{x} \in X$ are linearly independent for all $\bar{x} \in X$. If $\hat{z}$ is a local minimizer for P^s, then

$$\hat{z} \in X^s \triangleq \{z \in X^s | \lim_{i\to\infty} h^j(x_i) = 0,\ j \in \underline{\ell}\} \tag{4.2}$$

$$\begin{aligned}\Theta^{ls}_{ce}(\hat{z}) &\triangleq \lim_{i\to\infty} \inf_v \{(f_x(\hat{x}_i),v) + \tfrac{1}{2}\|v\|^2 \,|\, (h^j_x(\hat{x}_i),v) = 0\} \\ &= 0 \quad .\end{aligned} \tag{4.3}$$

where $\{\hat{x}_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class $\hat{z}$. □

When $X = \mathbb{R}^n$, the inf in (4.3) can be replaced by min and in that case the Lagrange conditions give for the "inner" problem in (4.3)

$$\nabla f(x_i) + v_i + \frac{\partial h(x_i)}{\partial x}^T \psi_i = 0 \tag{4.4}$$

where v_i solves the inner problem.

When x_i solves P, $v_i = 0$ and (4.4) reduces to the standard Lagrange condition. Thus, (4.4) is in one-to-one correspondence with the usual first-order conditions for P. We can also obtain a second order condition as follows:

Proposition 4.2: Suppose that the assumptions of Proposition 4.1 hold and that in addition $f(\cdot)$ and $h^j(\cdot)$, $j \in \underline{\ell}$, are all twice uniformly continuously differentiable on bounded sets. For any $\bar{x} \in X$, let $\mu : X \to \mathbb{R}^{\ell}$ be defined by

$$\mu(\bar{x}) \triangleq \arg\min_{\mu} \{\| f_x(\bar{x}) + \sum_{j\in\underline{\ell}} \mu^j h^j(\bar{x}) \| \} \tag{4.5}$$

and let $L : X \times \mathbb{R}^{\ell} \to \mathbb{R}$ be defined by

$$L(\bar{x},\bar{\mu}) \triangleq f(\bar{x}) + \langle\bar{\mu},h(\bar{x})\rangle \tag{4.6}$$

If $\hat{z}$ solves P^s then (i) $\hat{z} \in X^s$, (ii) $\Theta^{1s}_{ce}(\hat{z}) = 0$ and

$$\Theta^{2s}_{ce}(\hat{z}) \triangleq \underline{\lim}\ \inf\{(L_{xx}(\hat{x}_i)(v),v) \mid (h^j_x(x_i),v) = 0,\ j \in \underline{\ell},\ \|v\| = 1\} \geq 0 \tag{4.7}$$

where $\{x_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class $\hat{z}$. □

It is also quite easy to establish a second order sufficient condition for P^s under the assumptions of Proposition 4.2, viz., if $\hat{z} \in X^s$, and $\Theta^{1s}_{ce}(\hat{z}) = 0$ and $\Theta^{2s}_{ce}(\hat{z}) > 0$, then $\hat{z}$ is a local minimizer for P^s.

When inequalities are present, it is possible to deduce a broad family of optimality functions for P^s from the literature on algorithms (see e.g. [13,14]) We shall state the simplest (see [13, 14, 15]).

Proposition 4.3: Suppose that $f(\cdot)$ is uniformly continuously differentiable on bounded sets in X and that

$$X \triangleq \{x \mid h^j(x) = 0,\ j \in \underline{\ell};\ g^k(x) \leq 0,\ i \in \underline{m}\} \tag{4.8}$$

where the $h^j : X \to \mathbb{R}$ and $g^k : X \to \mathbb{R}$ are all uniformly continuously differentiable on bounded sets. Furthermore suppose that the functionals $h^j_x(\bar{x})$, $j \in \underline{\ell}$, $\bar{x} \in X$ are linearly independent for all $\bar{x} \in X$.

Let $\psi : X \to \mathbb{R}$ be defined by

$$\psi(x)_+ \triangleq \max_{k\in\underline{m}}\{0,g^k(x)\} \tag{4.9}$$

If $\hat{z}$ is a local minimizer for P^s then

$$\hat{z} \in X^s \triangleq \{z \mid \lim_{i\to\infty} h^j(x_i) = 0,\ j \in \underline{\ell},$$

$$\overline{\lim_{i\to\infty}}\ g^k(x_i) \leq 0,\ k \in m\} \tag{4.10}$$

and for $\gamma \geq 1$,

$$\Theta^{1s}_{cei}(\hat{z}) \triangleq \overline{\lim_{i\to\infty}}\ \inf\{(\tfrac{1}{2}\|v\|^2 + \max\{(f_x(\hat{x}_i),v) - \gamma\psi(\hat{x}_i)_+;\ g^k(\hat{x}_i) +$$

$$(g^k_x(\hat{x}_i),v) - \psi(\hat{x}_i),\ k \in \underline{m}\}) \Big| (h^j(x_i),v) = 0,\ j \in \underline{\ell}\} = 0 \tag{4.11}$$

□

Next we turn to a class of non-differentiable problems (see[8]).

Proposition 4.4: Suppose that $f(\cdot)$ is uniformly locally Lipschitz continuous on bounded sets and that

$$X = \{x \mid g^k(x) \leq 0,\ k \in \underline{m}\} \tag{4.12}$$

where the $g^k : X \to \mathbb{R}$ are all uniformly locally Lipschitz continuous on bounded sets. Furthermore, suppose that $\partial f(\cdot)$ and all the $\partial g^k(\cdot)$ are weak * compact and uniformly u.s.c. on bounded sets. For any $\varepsilon \geq 0$, and $\bar{x} \in X$ let

$$I_\varepsilon(\bar{x}) \triangleq \{k \in \underline{m} \mid g^k(\bar{x}) \geq \psi(\bar{x})_+ - \varepsilon\} \tag{4.13}$$

$$v^g_\varepsilon(\bar{x}) \triangleq \arg\min\{\|v\| \ \Big|\ v \in \operatorname*{co}_{k\in I_\varepsilon(\bar{x})} \partial_\varepsilon g^k(\bar{x})\} \tag{4.14}$$

$$v^f_\varepsilon(\bar{x}) \triangleq \arg\min\{\|v\| \ \Big|\ v \in \operatorname{co}\{\partial_\varepsilon f(\bar{x}) \bigcup_{k\in I_\varepsilon(\bar{x})} \partial g^k(\bar{x})\}\} \tag{4.15}$$

$$\gamma_\varepsilon(\bar{x}) \triangleq \max\{e^{-\psi(\bar{x})_+}\|v^f_\varepsilon(\bar{x})\|, (1-e^{\psi(\bar{x})_+})\|v^g_\varepsilon(\bar{x})\|\} \tag{4.16}$$

and

$$\Theta^1_{cind}(\bar{x}) = \max\{\varepsilon \in E \cdot \mid \gamma_\varepsilon(\bar{x}) \geq \varepsilon\}\ . \tag{4.17}$$

If $\hat{z}$ is a local minimizer for P^s, then

$$\hat{z} \in X^s \triangleq \{z \in X^s \Big| \overline{\lim_{i\to\infty}}\ g^k(x_i) \leq 0,\ k \in \underline{m}\} \tag{4.18}$$

and

$$\Theta^{1s}_{cind}(\hat{z}) \triangleq \underline{\lim}_{i\to\infty} \Theta^{1}_{cind}(\hat{z}_i) = 0 \tag{4.19}$$

where $\{\hat{z}_i\}_{i=0}^{\infty}$ is any sequence in the equivalence class $\hat{z}$. □

To conclude this section we state an optimality function for a simple optimal control problem, based on the Maximum-Principle and first used in [6].

Consider the dynamical system

$$\dot{x}(t) = h(x(t),u(t)), \; t \in [0,1] \tag{4.20}$$

with $x(0) = x_0$ and $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ twice uniformly continuously differentiable on bounded sets. We shall denote by $x^u(\cdot)$ the solution of (4.20) corresponding to the control $u(\cdot)$. Let Ω be a compact subset of $\mathbb{R}^m$ and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be twice uniformly continuously differentiable on bounded sets. We now define

$$X \triangleq L^m_\infty\,[0,1] \tag{4.21a}$$

$$f(u) \triangleq \phi(x^u(1)) \tag{4.21b}$$

$$X \triangleq \{u \in L^m_\infty\,[0,1] \mid u(t) \in \Omega \;\forall t \in [0,1]\} \tag{4.21c}$$

Finally, let $\lambda^u(\cdot)$ denote the solution of the adjoint equation

$$\dot{\lambda}(t) = -\frac{\partial h}{\partial x}(x^u(t),u(t))^T\lambda(t) \tag{4.22a}$$

$$\lambda(1) = \nabla\phi(x^u(1)) \tag{4.22b}$$

Proposition 4.5: Suppose $\hat{z}$ is a local minimizer for P^s, with f, X and X defined as in (4.21a-b). Then

$$\Theta^s_{oc}(\hat{z}) \triangleq \underline{\lim}\, \Theta_{oc}(\hat{u}_i) = 0 \tag{4.23}$$

where

$$\Theta_{oc}(u) \triangleq \int_0^1 \min_{v\in\Omega} \langle (h(x^u(t),v) - h(x^u(t),u(t)),\ \lambda^u(t)\rangle dt$$

□

5. Conclusions

We have shown that it is quite straightforward to construct optimality conditions for minimizing sequences by reinterpreting or modifying existing results. We hope that this work will lead to a better understanding of the behaviour of optimization algorithms.

Acknowledgement: Research sponsored by the National Science Foundation grants ECS-79-13148, and CEE-8105790, the Joint Services Electronics Program Contract F49620-79-C-0178, and the UK Science Research Council.

References

[1] J.W. Bandler, P.C. Liu and H. Tromp, "Nonlinear programming approach to optimal design centering, tolerancing and tuning", IEEE Trans. Circuits Syst., vol CAS-23, 1976 .

[2] E. Polak and A. Sangiovanni Vincentelli, "Theoretical and computational aspects of the optimal design centering, tolerancing and tuning problem", IEEE Trans. Circuit Syst., vol CAS-26, No. 9, 1979 .

[3] C. Jouron, "Sur un probleme d'optimization ou la contrainte porte sur la frequence fondamentale", R.A.I.R.O. Analyse Numerique, vol 12, No. 14, 1978, pp. 349-374.

[4] E. Haug and J.P. Rousselet, "Design sensitivity analysis in structural mechanics II, eigenvalue variations", Proc. NATO Advanced Study Institute, University of Iowa, Iowa City Iowa, May 1980.

[5] Y.Y. Wardi, "Optimization algorithms for problems with eigenvalue inequality constraints",Ph.D. Thesis, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, 1981.

[6] F. Clarke, "Generalized gradients and applications", Trans. Amer. Math. Soc., vol. 205, 1975, pp. 242-262.

[7] F. Clarke, *Nonsmooth analysis and Optimization*, Wiley, Interscience, N.Y., in press.

[8] E. Polak, D.Q. Mayne and Y.Y. Wardi, "On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems", SIAM, J. Control and Opt., in press .

[9] G. Lebourg, "Valeur moyènne pour gradient géneralisé" C.R. Acad. Sci., Paris, vol 281, 1975, pp. 795-797.

[10] J. Cullum, W.E. Donath and P. Wolfe, "The minimum of certain non-differentiable sums of eigenvalues of symmetric matrices", Math. Programming Studies, No.3, 1975, pp. 35-55.

[11] E. Polak and Y.Y. Wardi, "A nondifferentiable optimization algorithm for design of control systems subject to singular value inequalities over a frequency range", Automatica, in press.

[12] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, N.Y., 1969.

[13] E. Polak, *Computational Methods in Optimization - A Unified Approach*, Academic Press, N.Y. 1971.

[14] E. Polak, R. Trahan and D.Q. Mayne, "Combined phase I - phase II methods of feasible directions", Math. Programming. vol 17, No. 1, 1979, pp. 32-61.

[15] O. Pironneau and E. Polak, "On the rate of convergence of certain methods of centers", Math. Programming, vol 1, No. 2, 1972, pp. 230-258.

[16] D.Q. Mayne and E. Polak, "First order, strong variations algorithms for optimal control", JOTA vol 16, No. 3/4, 1975, pp. 277-301.

[17] L.J. Williamson and E. Polak, "Relaxed controls and the convergence of optimal control algorithms", SIAM J. Control, vol 14, No. 4, 1976, pp. 737-757.