A NONDIFFERENTIABLE OPTIMIZATION ALGORITHM FOR THE DESIGN OF CONTROL SYSTEMS SUBJECT TO SINGULAR VALUE INEQUALITIES OVER A FREQUENCY RANGE

by

E. Polak and Y. Wardi

Memorandum No. UCB/ERL M80/41

4 September 1980

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A NONDIFFERENTIABLE OPTIMIZATION ALGORITHM FOR THE DESIGN OF CONTROL SYSTEMS SUBJECT TO SINGULAR VALUE INEQUALITIES OVER A FREQUENCY RANGE

E. Polak and Y. Wardi

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

Abstract

It has been known for some time that many control system design requirements can be expressed as differentiable inequalities. More recently, it has been shown that important structural properties such as robustness and low noise sensitivity can be expressed as nondifferentiable inequalities involving the singular values of a system or return difference transfer function matrix. This paper presents an optimization algorithm which permits all these constraints to be considered.

† Research supported by the National Science Foundation Grant ECS-79-13148.

1. Introduction

The fact that multivariable control system design can be carried out by using constrained optimization algorithms has been known for some time, see e.g. [P1a, P21, P5, Z1, G2, G3, M1, M2]. Until recently, the constraints used were in the form of differentiable inequalities. It is now becoming clear that a number of fundamental design requirements can be expressed as inequalities involving the singular values of appropriate transfer function matrices, such as the complex valued compensator-plant transfer function matrix $G(x,\omega)$, with x denoting the design parameter [L1, L2, D1, D2, D2a, S1]. These inequalities are usually frequency (ω) dependent and have to be satisfied over a range of frequencies. Specifically, the requirement that the closed loop system remains stable in the face of additive or multiplicative perturbations of $G(x,\omega)$ is expressed in terms of singular value inequalities in [L2, D2, S1]. In [S1] we find that to ensure that a high order system is stable when modeled by a low order system via singular perturbations, it is sufficient to satisfy certain singular value inequalities. Low sensitivity to additive noise and parameter perturbations is expressed as singular value inequalities in [D2a].

From an otpimization point of view these singular value inequalities pose two serious problems. The first is due to the fact that neither singular values nor their squares are differentiable, while the second one is due to the fact that when some singular values of a matrix become close to being equal, it becomes extremely difficult to compute the corresponding singular vectors with any precision. At the present time there are no published optimization algorithms which are directly applicable to control system design in the presence of frequency singular value inequalities. In this paper we present an optimization algorithm capable of solving such problems. The algorithm makes use of outer approximations for problem decomposition [G2] and of some concepts of nondifferentiable optimization described in [C1, C2, M3, L3, P6, P3]. Computational results for simple problems such as those reported in [D1], appear to be quite encouraging.

2. The problem and its decomposition

In the context of optimization, the problem of describing a linear multivariable control system presents itself as follows. We assume that the structure of the compensators has been determined on the basis of theory such as in [D1, D2, M4], and what remains to do is to compute a compensator parameter vector $x \in \mathbb{R}^n$ which minimizes a differentiable cost function $f(x)$ subject to constraints of three kinds:

$$\text{(a)} \quad g^j(x) \leq 0, \quad j = 1,2,\ldots,m_g \tag{2.1}$$

with $g^j : \mathbb{R}^n \to \mathbb{R}^1$ continuously differentiable,

$$\text{(b)} \quad \max_{\nu \in N} \phi^k(x,\nu) \leq 0, \quad k = 1,2,\ldots,m_\phi \tag{2.2}$$

with $\phi^k : \mathbb{R}^n \times \mathbb{R}^1 \to \mathbb{R}$ continuously differentiable and $N \subset \mathbb{R}^1$ compact, and

$$\text{(c)} \quad 0 \leq \ell_\ell(\omega) \leq \sigma_\ell^i(x,\omega) \leq u_\ell(\omega) \leq u_\ell < \infty \qquad \forall i = 1,2,\ldots,m, \quad \forall \ell = 1,2,\ldots,L, \quad \forall \omega \in \Omega \tag{2.3}$$

where for $i = 1,2,\ldots,m$, $\sigma_\ell^i(x,\omega)$ is a singular value of an mxm complex valued transfer function matrix $G_\ell(x,\omega)$, $\ell_\ell(\omega)$, $u_\ell(\omega)$ are continuous functions from $\mathbb{R} \to \mathbb{R}^+$ and Ω is a frequency interval . In (b) ν may be either time or frequency (see [P5, G3, P4a, P4b]. As stated, the design problem is quite complex and hence the presentation of the algorithm for the full problem is quite cumbersome. Fortunately, there is no great loss of generality in presenting our algorithm first in terms of the highly simplified problem

$$P: \min\{f(x) \mid \ell(\omega) \leq \sigma^i(x,\omega) \leq u(\omega), \quad i = 1,2,\ldots,m, \ \omega \in \Omega\} \tag{2.4}$$

where the $\sigma^i(x,\omega)$ are the singular values of a single mxm transfer function matrix and $G(x,\omega)$ and $\Omega \subset \mathbb{R}^+$ is compact. At the end of Section 3 we shall indicate how the algorithm is to be extended for the full problem.

Our algorithm consists of two parts:

(i) a master outer approximations algorithm which decomposes P into an infinite sequence of problems

$$P_k: \min\{f(x) \mid \ell(\omega) \leq \sigma^i(x,\omega) \leq u(\omega), \ i = 1,2,\ldots,m,$$

$$\omega \in \Omega_k\} \quad k = 1,2,\ldots \tag{2.5}$$

with $\Omega_k \subset \Omega$ a finite set, and

(ii) a special nondifferentiable optimization algorithm for solving the problems P_k. We shall present the algorithm for solving P_k in the next section, while the outer approximations algorithm will be presented in section 4.

We shall need the following assumption and result which follows from it.

<u>Assumption 2.1</u>: There exists an open set $X \subset \mathbb{R}^n$ such that the transfer function matrices $G(\cdot,\omega)$: $X \to C^{mxm}$ are componentwise analytic, $\ell = 1,2,\ldots,L$. ¤

We shall assume that $\ell(\omega) < \infty$ for all $\omega \in \Omega$. Since at points $x \in X^c$ at which $G_\ell(x,\omega)$ is not analytic $G_\ell(x,\omega)$ must have at least one infinite element, it will have at least one infinite singular value. Because of this, once our algorithm is started with an $x_0 \in X$, the entire sequence it will construct will remain in X. The following result can be deduced from analytic function theory:

Proposition 2.1: Suppose that assumption 2.1 holds and that for some $\ell \in \{1,2,\ldots,L\}$, $x \in X$ and $\omega \in \Omega$, $G_\ell(x,\omega)$ has multiple singular values. Let the singular values of $G_\ell(x,\omega)$ be $\sigma^1(x,\omega) \geq \sigma^2(x,\omega) \geq \ldots \geq \sigma^m(x,\omega)$ and let $M \subset \{1,2,\ldots,m\}$ be the largest set such that $\sigma^i(x,\omega) = \sigma^i(x,\omega)$ for all $i,j \in M$. Then, given any vector $w \in \mathbb{R}^n$, there exists a $\bar{\lambda} > 0$ such that for any $i,j \in M$, either

$$\sigma^i(x+\lambda h,\omega) = \sigma^i(x+\lambda h,\omega) \quad \forall\lambda \in [0,\bar{\lambda}] \tag{2.6}$$

or

$$\sigma^i(x+\lambda h,\omega) \neq \sigma^i(x+\lambda h,\omega) \quad \forall\lambda \in (0,\bar{\lambda}] \tag{2.7}$$

We now proceed to develop an algorithm for solving the P_k.

2. An Algorithm for Solving P_k

To obtain a further simplification in exposition, we shall assume temporarily that Ω_k in (2.5) contains only one point. In this case ω can be dropped as an argument in the functions in P_k and P_k becomes

$$P_k : \min\{f(x) \mid \ell \leq \alpha^i(x) \leq u, \quad i = 1,2,\ldots,m\} \tag{3.1}$$

where the $\sigma^i(x)$ are the singular values of a complex valued mxm transfer function matrix $G(x)$. We recall that $y^i(x) \triangleq [\sigma^i(x)]^2$, $i = 1,2,3,\ldots,m$ are the eigenvalues of the matrix

$$Q(x) \triangleq G(x)^* G(x) \tag{3.2}$$

For the sake of convenience, we shall adopt the <u>convention</u> that

$$y^1(x) \geq y^2(x) \geq \dots y^m(x) \tag{3.2a}$$

The quantities $y^i(x)$ are more convenient to work with than the $\sigma^i(x)$ and hence we transform P_k into the equivalent form

$$\tilde{P}_k : \min\{f(x) \mid \ell^2 \leq y^i(x) \leq u^2,\ i \in \underline{m}\ \} \tag{3.3}$$

where $\underline{m} \triangleq \{1,2,\dots,m\}$.

It was shown in [P2] that the $y^i(\cdot)$ are locally Lipschitz continuous functions which are differentiable at all x such that $y^i(x) \neq y^j(x)$ for some $i \neq j \in \underline{m}$. When $y^i(x) = y^j(x)$ $i \neq j$ the $y^i(\cdot)$ fail to be differentiable, but, as was shown in [P.3], they are semi-smooth [M.3], i.e., they belong to the most benign class of nondifferentiable functions. Because of these facts, P_k must be treated in the context of nondifferentiable optimization. First, let

$$\psi_u(x) \triangleq \max\{y^i(x)-u^2,\ i \in \underline{m}\} \tag{3.4a}$$

let

$$\psi_\ell(x) \triangleq \max\{\ell^2-y^i(x),\ i \in \underline{m}\} \tag{3.4b}$$

let

$$\psi(x) \triangleq \max\ \{\psi_u(x),\phi_\ell(x)\} \tag{3.5a}$$

and let

$$\psi(x)_+ \triangleq \max\{0,\psi(x)\} \tag{3.5b}$$

It is well known [P.3] that if x is optimal for $\tilde{P}_k$ then, if $\psi(\hat{x}) < 0$ $\nabla f(\hat{x}) = 0$ and if $\psi(\hat{x}) = 0$, then

$$0 \in \mathrm{co}\{\nabla f(\hat{x}) \cup \partial\psi(\hat{x})\} \tag{3.6}$$

where co denotes the convex hull and $\partial\psi(\hat{x})$ is the generalized gradient [C.1] of $\psi(\cdot)$ at $\hat{x}$.

Given any $x \in \mathbb{R}^n$, $\varepsilon \geq 0$, we define, with $\underline{m}_0 \triangleq \{0,1,\ldots,m\}$,

$$k_u(x,\varepsilon) \triangleq \min\{k \in \underline{m}_0 | y^k(x)-y^{k+1}(x) > \varepsilon\} \tag{3.7a}$$

and

$$k_\ell(x,\varepsilon) \triangleq \max\{k \in \underline{m}_0 | y^{k-1}(x)-y^k(x) > \varepsilon\} \tag{3.7b}$$

Let U_u be any complex $m\mathrm{x}b_u(x,\varepsilon)$ matrix such that $U_u^* U_u = I$ and $U_u^* Q(x) U_u = \mathrm{diag}(y^1(x), y^2(x),\ldots,y^{k(x,\varepsilon)}(x))$ and let U_ℓ be any complex $m\mathrm{x}(m-k_\ell(x,\varepsilon)+1)$ matrix such that $U_\ell^* U_\ell = I$ and $U_\ell^* Q(x) U_\ell = \mathrm{diag}(y^{k_\ell(x,\varepsilon)},\ldots,y^m(x))$. For any $x \in X$ and $\varepsilon \geq 0$, we now define

$$\nabla_u^\varepsilon(x) \triangleq \mathrm{co}\{v \in \mathbb{C}^n | v^i = \langle z, U_u^* \frac{\partial Q(x)}{\partial x^i} U_u z\rangle,\quad i = 1,2,\ldots,n;\ z \in \mathbb{C}^{k_u(x,\varepsilon)},\ \|z\| = 1\} \tag{3.8a}$$

$$\nabla_\ell^\varepsilon(x) \triangleq \mathrm{co}\{v \in \mathbb{C}^n | v^i = \langle z, U_\ell^* \frac{\partial Q(x)}{\partial x^i} U_\ell z\rangle,\quad i = 1,2,\ldots,n;\ z \in \mathbb{C}^{n-k_\ell(x_1\varepsilon)+1},\ \|z\| = 1\} \tag{3.8b}$$

Note that the <u>same</u> z must be used in computing every component of v in (3.8a) or (3.8b). Also note that when U_u of U_ℓ are not unique, the definitions (3.8a,b) do not depend on the specific choice for U_u, U_ℓ.

We can now establish a characterization for $\partial\psi(x)$.

<u>Proposition 3.1</u>: Suppose that Assumption 2.1 holds. Then

$$\partial\psi(x) = \begin{cases} \nabla_u^0(x) & \text{if } \psi_u(x) > \psi_\ell(x) \\ -\nabla^0(x) & \text{if } \psi_\ell(x) > \psi_u(x) \\ \mathrm{co}\{\nabla_u^0(x) U - \nabla_\ell^0(x)\} & \text{if } \psi_\ell(x) = \Psi_u(x) \end{cases} \tag{3.9}$$

¤

The proof of this result is given in Appendix A. We now proceed to extract from the optimality condition (3.6) a method for computing a descent direction for $\tilde{P}_k$. This direction has to be a descent direction for $\psi(\cdot)$ when x is not feasible and a feasible descent direction for $f(\cdot)$ when x is feasible. The resulting algorithm will be in the family of phase I-phase II methods with ε-smearing (see [P.4]). We begin by developing an ε-approximation for $\partial\psi(x)$, with $\varepsilon > 0$. For any $x \in \mathbb{R}^n$ and $\varepsilon \geq 0$, we define

$$\nabla^{\varepsilon}(x) \triangleq \begin{cases} \nabla_u^{\varepsilon}(x) & \text{if } \psi_u(x) > \psi_\ell(x) + \varepsilon \\ -\nabla_\ell^{\varepsilon}(x) & \text{if } \psi_\ell(x) > \psi_u(x) + \varepsilon \\ \mathrm{co}\{\nabla_u(x) \cup -\nabla_\ell^{\varepsilon}(x)\} & \text{otherwise} \end{cases} \tag{3.10}$$

It will be shown in the Appendix that $\nabla^{\varepsilon}(\cdot)$ is compact and upper semicontinuous in the sense of Berge [B2]. Next, for any compact set $S \subset \mathbb{R}^n$, we define the real valued function Nr(S) by

$$Nr(S) \triangleq \min\{\|h\| \mid h \in S\} \tag{3.11}$$

and then

$$h_{\psi,\varepsilon}(x) \triangleq Nr(\nabla^{\varepsilon}(x)) \tag{3.12a}$$

$$h_{f,\varepsilon}(x) \triangleq \begin{cases} Nr(\mathrm{co}(\nabla f(x), \nabla^{\varepsilon}(x))) & \text{if } \psi(x) \geq -\varepsilon \\ \nabla f(x) & \text{otherwise} \end{cases} \tag{3.12b}$$

We will see shortly that, for ε sufficiently small $-h_{\psi,\varepsilon}(x)$ is a descent direction for $\psi(\cdot)$ and that $-h_{f,\varepsilon}(x)$ is the analog of the feasible descent direction found in methods of feasible directions [P.4]. To obtain a phase I-phase II method we need a mechanism for crossing over from $h_{\psi,\varepsilon}$ to $h_{f,\varepsilon}$ as we go from the infeasible to the feasible region for P_k. The reader familiar with the results in [P4] and in section 4.3 of [P.7] will find that we are following closely the general ideas used in phase I-phase II methods for differentiable optimization. Let $\gamma \geq 1$ we define

$$\Gamma(x) \triangleq e^{-\gamma\psi(x)_+} \tag{3.13}$$

and

$$h_\varepsilon(x) = \begin{cases} \Gamma(x)h_{f,\varepsilon}(x) + (1-\Gamma(x))h_{\psi,\varepsilon}(x) & \text{if } \psi(x) \geq -\varepsilon \\ h_{f,\varepsilon}(x) = \nabla f(x) & \text{otherwise} \end{cases} \tag{3.14}$$

note that when $\psi(x)_+ = 0$, i.e. x is feasible, $\Gamma(x) = 1$ and $h_\varepsilon(x) = h_{f,\varepsilon}(x)$, while when $\psi(x)_+$ is very large, $\Gamma(x) \approx 0$ and $h_\varepsilon(x) \approx h_{\psi,\varepsilon}(x)$. Next, for any $x \in \mathbb{R}^n$, $\varepsilon > 0$, we define

$$\theta_\varepsilon(x) \triangleq -\max\{\|\Gamma(x)h_{f,\varepsilon}(x)\|^2, \|(1-\Gamma(x))h_{\psi,\varepsilon}(x)\|^2\} \tag{3.15}$$

Finally, with $\beta \in (0,1)$ and

$$E \triangleq \{0,1,\beta,\beta^2,\dots\} \tag{3.16}$$

For any $x \in \mathbb{R}^n$, we define

$$\varepsilon(x) = \max\{\varepsilon \in E \mid \theta_\varepsilon(x) \leq -\varepsilon\}^+ \tag{3.17}$$

With this, we can now define

$$h(x) \triangleq h_{\varepsilon(x)}(x) \tag{3.18a}$$

$$\overline{\theta}(x) \triangleq \theta_{\varepsilon(x)}(x) \tag{3.18b}$$

For the algorithm to be of any use we will need the following commonly occuring

---

$^+$For computational efficiency it is often desirable to use a test of the form $\theta_\varepsilon(x) \leq -\delta\cdot\varepsilon$ with $\delta > 0$. The value of δ has no effect on the analysis.

Assumption 3.1: For every $x \in X$ such that $\psi(x) \geq 0$, $h_{\psi,0}(x) \neq 0$. ¤

We now turn to the properties of the functions $\theta_\varepsilon(\cdot)$ $\bar{\theta}(\cdot)$ and $\varepsilon(\cdot)$. These will be proved in Appendix B.

Proposition 3.2: a) For any $x \in X$, $0 \leq \varepsilon < \varepsilon' \Rightarrow \theta_\varepsilon(x) \leq \theta_{\varepsilon'}(x)$.

b) For any $\varepsilon > 0$, $\theta_\varepsilon(\cdot)$ is upper semicontinuous on X.

c) If $\hat{x}$ solves P_k, then $\varepsilon(\hat{x}) = 0$ and $\bar{\theta}(\hat{x}) = 0$.

d) If $\hat{x}$ is such that $\bar{\theta}(\hat{x}) < 0$, then there exist $\hat{\rho} > 0$ and $\hat{\varepsilon} > 0$ such that

$$\theta_\varepsilon(x) \leq -\varepsilon \text{ for all } (x,\varepsilon) \in B(\hat{x},\hat{\rho}) \times [0,\hat{\varepsilon}]. \tag{3.19}$$

¤

Corollary 3.1: If $\hat{x}$ is such that $\bar{\theta}(\hat{x}) < 0$ then there exists a $\hat{\rho} > 0$ such that

$$\varepsilon(x) \geq \beta\varepsilon(\hat{x}) \text{ for all } x \in B(\hat{x},\hat{\rho}) \tag{3.20}$$

¤

For the sake of clarity of exposition, we shall state our algorithm in three forms: first in conceptual, and hence simplest form, for the special case of problem $\tilde{P}_k$ (3.3), then in an implementable form for the special case of problem $\tilde{P}_k$ (3.3), and finally in implementable for the most general form for the problem $\tilde{P}_k$. Our first form is conceptual because it assumes that we can compute $h_{\psi,\varepsilon}(x)$ and $h_{f,\varepsilon}(x)$ exactly in finite time.

Conceptual Algorithm 3.1

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

Parameters: $\alpha,\beta,\eta \in (0,1)$, $b \gg 1$.

Step 0: Set $i = 0$.

Step 1: Compute $h(x_i)$ using the given value of β in (3.16). Stop if $h(x_i) = 0$.

Step 2: Compute the largest step size $s_i = \eta^k \in [0,M]$, with k an integer (negative values are allowed), such that if $\psi(x_i) > 0$

$$\psi(x_i - s_i h(x_i)) - \psi(x_i) \leq -s_i \alpha \| h(x_i) \|^2 \tag{3.21a}$$

If $\psi(x_i) \leq 0$

$$f(x_i - s_i h(x_i)) - f(x_i) \leq -s_i \alpha \| h(x_i) \|^2 \tag{3.21b}$$

and

$$\psi(x_i - s_i h(x_i)) \leq 0 \tag{3.21c}$$

Step 3: Set $x_{i+1} = x_i - s_i h(x_i)$, set $i = i+1$ and go to step 1. ¤

Lemma 3.1: Suppose that $x_i \in$ is such that $h(x_i) \neq 0$, then s_i, as constructed in Step 2 of Algorithm 3.1, $s_i > 0$, i.e., the algorithm is well defined.

Proof: Suppose the Lemma is false. Then there exists an $x_i \in X$ such that $h(x_i) \neq 0$ for which the appropriate test cannot be satisfied with a finite $s_i = \eta^k$ in Step 2 of Algorithm 3.1. Suppose at first that $\psi(x_i) < -\varepsilon(x_i)$, then we clearly get a contradiction because $f(\cdot)$ is differentiable on the basis of the usual arguments for methods of feasible directions. Hence, suppose that $\psi(x_i) \geq -\varepsilon(x_i)$. Then, if $\psi(x_i) \leq 0$, we must have either

$$\frac{f(x_i - \eta^k h(x_i)) - f(x_i)}{\eta^k} > -\alpha \| h(x_i) \|^2 \quad \text{for } k = 0,1,2,\ldots \tag{3.22a}$$

or

$$\frac{\psi(x_i - \eta^k h(x_i)) - \psi(x_i)}{\eta^k} > -\alpha \| h(x_i) \|^2 \quad \text{for } k = 0,1,2,\ldots \tag{3.22b}$$

or both. Since $\psi(x_i)_+ = 0$, $h(x_i) = h_{f,\varepsilon(x_i)}(x_i)$ and hence, from the strict separation property of $h_{f,\varepsilon(x_i)}(x_i)$,

$$\langle \nabla f(x_i), h(x_i) \rangle > 0 \tag{3.23}$$

It therefore follows that (3.22a) cannot take place (e.f. analysis of Armijo mehtods sec. 2.1 in [P.7]). Hence we only need to consider (3.22b). By the mean value theorem of Lebourg [L.3],

$$\psi(x_i-\eta^k h(x_i)) - \psi(x_i) = -\eta^k \langle h(x_i),\xi_k\rangle \tag{3.24}$$

where $\xi_k \in \partial\psi(x_i-\lambda\eta^k h(x_i))$ for some $\lambda \in [0,1]$. Hence from (3.22b) and (3.24)

$$\langle h(x_i),\xi_k\rangle < \alpha\| h(x_i)\|^2 \tag{3.25}$$

without loss of generality, we may assume that $\xi_k \to \hat{\xi}$ as $k \to \infty$. Since $\partial\psi(\cdot)$ is u.s.c., we must have $\hat{\xi} \subset \partial\psi(x_i) \subset \nabla^{\varepsilon(x_i)}(x_i) \subset \mathrm{co}\{\nabla f(x_i),\nabla^{\varepsilon(x_i)}(x_i)\}$. Now, taking limits in (2.25), we get that

$$\langle h(x_i), \hat{\xi}\rangle < \alpha\| h(x_i)\|^2 \tag{3.26}$$

which contradicts the fact that

$$h(x_i) = h_{f,\varepsilon(x_i)}(x_i) = Nr(\mathrm{co}\{\nabla f(x_i),\nabla^{\varepsilon(x_i)}(x_i)\}) .$$

Hence (3.22b) cannot hold either.

Now suppose that $\psi(x_i) > 0$. We will again show that (3.22b) cannot occur. By definition,

$$\langle h_{\psi,\varepsilon(x_i)}(x_i),\xi\rangle \geq \| h_{\psi,\varepsilon(x_i)}(x_i)\|^2 \qquad \forall\xi \in \nabla^{\varepsilon(x_i)}(x_i) \tag{3.27a}$$

and

$$\langle h_{f,\varepsilon(x_i)}(x_i),\xi\rangle \geq \| h_{f,\varepsilon(x_i)}(x_i)\|^2 \qquad \forall\xi \in \nabla^{\varepsilon(x_i)}(x_i) \tag{3.27b}$$

Hence, by convexity of $\|\cdot\|$, we must have

$$\langle h(x_i),\xi\rangle \geq \| h(x_i)\|^2 \qquad \forall\xi \in \nabla^{\varepsilon(x_i)}(x_i) \tag{3.27c}$$

Hence (3.26) cannot hold and we get again a contradiction.

Lemma 3.2: Let $\hat{x} \in \mathbb{R}^n$ and let $\hat{k}_n, \hat{k}_\ell \in \underline{m}$ be such that $y^{\hat{k}_u}(\hat{x}) \neq y^{\hat{k}_u+1}(x)$ and $y^{\hat{k}_\ell}(\hat{x}) \neq y^{\hat{k}_\ell-1}(\hat{x})$. Then there exists a $\hat{\rho} > 0$ such that for all $x \in \mathring{B}(\hat{x},\hat{\rho})$ the set valued

$$\hat{\nabla}^{\hat{k}_u}(x) \triangleq \mathrm{co}\{v \in C^n | v^i = \langle z, U_u^* \frac{\partial Q(x)}{\partial x^i} U_u z \rangle,$$

$$i = 1,2,\ldots,n;\ z \in C^{\hat{k}_u},\ \|z\| = 1\} \tag{3.28a}$$

$$\hat{\nabla}^{\hat{k}_\ell}(x) \triangleq \mathrm{co}\{v \in C^n | v^i = \langle z, U^* \frac{\partial Q(x_i)}{\partial x^i} U_n z \rangle,$$

$$i = 1,2,\ldots,n;\ z \in C^{\hat{k}_\ell+1},\ \|z\| = 1\} \tag{3.28b}$$

defined, with U_u, U_ℓ any matrices such that $U_u^* U_u = I, U_\ell^* U_\ell = I$ $U_u^* Q(x) U_u = \mathrm{diag}(y^1(x),\ldots,y^{\hat{k}_u}(x))$, $U_\ell^* Q(x) U_\ell = \mathrm{diag}(y^{\hat{k}_\ell}(x),\ldots,y^m(x))$. Furthermore, the maps $\hat{\nabla}^{\hat{k}_u}(\cdot)$ and $\hat{\nabla}^{\hat{k}_\ell}(\cdot)$ are continuous on $B(\hat{x},\hat{\rho})$, in the sense of Berge [B.2], i.e., they are both u.s.c. and l.s.c. ¤

The proof of this result will be given in Appendix.

We now state our main result.

Theorem 3.1: If $\hat{x}$ is an accumulation point of a sequence $\{x_i\}$ generated by the Conceptual Algorithm, then $\theta(\hat{x}) = 0$ and $\psi(\hat{x}) \leq 0$.

Proof: For the sake of contradiction, suppose that $\hat{x}$ is an accumulation point and $\theta(\hat{x}) < 0$. Then, by Proposition 3.2d) there exist $\hat{\rho} > 0$ and $\hat{\varepsilon} > 0$ such that $\varepsilon(x_i) \geq \hat{\varepsilon}$ for all $x_i \subset B(\hat{x},\hat{\rho})$. Now, suppose that $K \subset \{0,1,2,\ldots\}$ is much that $x_i \xrightarrow{k} \hat{x}$. We consider three cases:

a) Suppose that $\psi(x_i) > 0$ for all $i \in K$. By construction this implies that $\psi(x_i) > 0$ for all $i \geq 0$ and that $\psi(x_i)$ is monotonically decreasing. Since $x_i \xrightarrow{K} x$ and $\psi(\cdot)$ is continuous, it now follows that $\psi(x_i) \searrow \psi(\hat{x})$.

Next, since k_u and $k_\ell \in \underline{m}$, a finite set, there exists an infinite subset $K' \subset K$ such that $k_u(x_i, (x_i)) = \hat{k}_u$ and $k_\ell(x_i, (x_i)) = k_\ell$ for all $i \in K'$. Since $\theta(\hat{x}) < 0$, there exists an $\hat{\varepsilon} > 0$ and an i_0 such that $\varepsilon(x_i) \geq \hat{\varepsilon}$ for all $i \geq i_0$, $i \in K'$. Consequently, since $x_i \xrightarrow{K'} \hat{x}$, the sets $\hat{\nabla}^{\hat{k}_u}(\hat{x})$ and $\nabla^{\hat{k}_\ell}(\hat{x})$ are well defined and, by Lemma 3.2 $\hat{\nabla}^{\hat{k}_u}(x_i) \xrightarrow{K'} \hat{\nabla}^{\hat{k}_u}(\hat{x})$, $\hat{\nabla}^{\hat{k}_\ell}(x_i) \xrightarrow{K'} \hat{\nabla}^{\hat{k}_\ell}(x)$. Hence, since $\nabla_u^{\varepsilon(x_i)}(x_i) = \hat{\nabla}^{\hat{k}_u}(x_i)$ and $\nabla_\ell^{\varepsilon(x_i)}(x_i) = \hat{\nabla}^{\hat{k}_\ell}(x_i)$ for all $i \in K'$, and $\Gamma(\cdot)$ is continuous, it follows that if $K'' \subset K$ is any infinite subset such that $\nabla^{\varepsilon(x_i)}(x_i) = \hat{\nabla}^{\hat{k}_u}(x_i)$, or $\nabla^{\varepsilon(x_i)}(x_i) = \hat{\nabla}^{\hat{k}_\ell}(x_i)$, or $\nabla^{\varepsilon(x_i)}(x_i) = \mathrm{co}\{\hat{\nabla}^{\hat{k}_u}(x_i) \cup \hat{\nabla}^{\hat{k}}\}(x_i)$ for all $i \in K''$, then $h(x_i) \xrightarrow{K''} \hat{h}$, with

$$\hat{h} = \Gamma(\hat{x})\hat{h}_f + (1-\Gamma(\hat{x}))\hat{h}_\psi \tag{3.29}$$

where, for $\hat{\nabla}_\infty = \lim_{i \in K''} \nabla^{\varepsilon(x_i)}(x_i)$,

$$\hat{h}_\psi = \mathrm{Nr}(\hat{\nabla}_\infty) \tag{3.30a}$$

and

$$\hat{h}_f = \mathrm{Nr}(\mathrm{co}\{\nabla f(\hat{x}), \hat{\nabla}_\infty\}) \tag{3.30b}$$

Now, if $\nabla^{\varepsilon(x_i)}(x_i) = \mathrm{co}\{\hat{\nabla}^{\hat{k}_u}(x_i) \cup \hat{\nabla}^{\hat{k}_\ell}(x_i)\}$ for all $i \in K''$, then, clearly, by continuity,

$$\partial\psi(\hat{x}) \subset \hat{\nabla}_\infty \tag{3.31}$$

If $\nabla^{\varepsilon(x_i)}(x_i) = \hat{\nabla}^{k_u}(x_i)$ for all $i \in K''$, then we must have that $\psi_u(x_i) - \psi_\ell(x_i) > \varepsilon(x_i)$ for all $i \in K''$, Since by Corollary 3.1, there exists an $\hat{\varepsilon} > 0$ such that $\varepsilon(x_i) \geq \hat{\varepsilon}$ for all $i \geq K''$, and ψ_u, ψ_ℓ are continuous, $\psi_u(\hat{x}) - \psi_\ell(\hat{x}) \geq \hat{\varepsilon}$ and again (3.31) holds. A similar argument holds when $\nabla^{\varepsilon(k_j)}(x_i) = -\hat{\nabla}^{\hat{k}_\ell}(x_i)$. Finally, since $\theta(\hat{x}) < 0$, it follows from Corollary (3.1) that $\|\hat{h}\| \neq 0$.

It now follows from (3.30a,b) and (3.31) that

$$\langle \hat{h},\xi\rangle \geq \|\hat{h}\|^2 \quad \forall \xi \in \partial\psi(x) \tag{3.32}$$

Now, by construction,

$$\psi(x_{i+1}) - \psi(x_i) \leq -\alpha s_i\|h(x_i)\|^2 \quad \forall i \tag{3.33}$$

Since $h(x_i) \xrightarrow{K'} \hat{h}$, there exists an i_0 such that

$$\psi(x_{i+1}) - \psi(x_i) \leq -\alpha s_i\|h(x_i)\|^2 \leq -\alpha s_i\|\hat{h}\|^2/2 \quad \forall i \in K' \quad i \geq i_0 \tag{3.34}$$

and hence, since $\psi(x_i) \searrow \psi(\hat{x})$ we must have that $s_i \xrightarrow{K'} 0$. By construction of s_i, we must have

$$\frac{\psi(x_i-\beta^{-1}s_ih(x_i))-\psi(x_i)}{\beta^{-1}s_i} \geq -\alpha\|h(x_i)\|^2 \tag{3.35}$$

And hence by the mean value theorem of Lebourg [L.3]

$$-\langle \xi_i,h(x_i)\rangle \geq -\alpha\|h(x_i)\|^2 \quad \forall i \in K' \quad i \geq i_0 \tag{3.36}$$

where $\xi_i \in \partial\psi(x_i-\lambda\beta s_ih(x_i))$ with $\lambda \in [0,1]$. It now follows from the u.s.c. of $\partial\psi(\cdot)$ that $\{\xi_i\}_{i\in K'}$ must have convergent subsequences such that if $\xi_i \xrightarrow{K''} \hat{\xi}$, with $K'' \subset K'$, then $\xi \in \partial\psi(x)$. Hence (3.36) yields, in the limit that

$$\langle \hat{\xi},\hat{h}\rangle \leq \alpha\|h(\hat{x})\|^2 \tag{3.37}$$

which contradicts (3.32). Hence $\psi(x_i) > 0$ for all i is not possible.

b) Suppose that $-\varepsilon(x_i) \leq \psi(x_i) \leq 0$ for all $i \geq i_0$. Then $\Gamma(x_i) = 0$ for all $i \geq 0$. By construction in Step 2 of the conceptual Algorithm, we must have that $f(x_{i+1}) < f(x_i)$ $\forall i \geq i_0$ and hence, by continuity of $f(\cdot)$ we must have that $f(x_i) \searrow f(\hat{x})$. For the same reasons as in a) we must have that $h(x_i) \xrightarrow{K''} \hat{h}$ for some $K'' \subset K$, and

$$\hat{h} = \hat{h}_f = Nr(co\{\nabla f(\hat{x}),\nabla_\infty\}) \tag{3.38}$$

where $\partial\psi(\hat{x}) \subset \hat{\nabla}_\infty$. Consequently, we must have that

$$\langle \hat{h}, \nabla f(\hat{x})\rangle \geq \|\hat{h}\|^2 \tag{3.39a}$$

$$\langle \hat{h},\xi\rangle \geq \|\hat{h}\|^2 \quad \forall \xi \in \partial\psi(\hat{x}) \tag{3.39b}$$

Since $\|\hat{h}\| > 0$ and since $\|h(x_i)\| \geq \|\hat{h}\|/2$ for all $i \in K''$, $i \geq i_0$, for some $i_0 \in K''$ and $f(x_i) \downarrow f(\hat{x})$, it follows from the fact that

$$\frac{f(x_{i+1})-f(x_i)}{s_i} \leq -\alpha s_i\|h(x_i)\|^2$$

$$\leq \quad s_i\|\hat{h}\|^2/2 \quad \forall i \in K' \quad i \geq i_0 \tag{3.40}$$

that $s_i \xrightarrow{K''}$ as $i \to \infty$. By construction of s_i, we must have for all $i \in K''$ either

$$\frac{f(x_i-\beta^{-1}s_ih(x_i))-f(x_i)}{\beta^{-1}s_i} \geq -\alpha\|h(x_i)\|^2 \tag{3.41a}$$

or $\psi(x_i-\beta^{-} s_ih(x_i)) > 0$ so that, since $\psi(x_i) \leq 0$,

$$\frac{\psi(x_i-\beta^{-1}s_ih(x_i))-\psi(x_i)}{\beta^{-1}s_i} \geq -\alpha\|h(x_i)\|^2 \tag{3.41b}$$

or both. Taking limits in (3.41a,b) as $i \xrightarrow{K''} \infty$, we must have either

$$\langle \nabla f(\hat{x}),\hat{h}\rangle \leq \alpha\|\hat{h}\|^2 \tag{3.42a}$$

or

$$\langle \hat{\xi};\hat{h}\rangle \leq \alpha\|\hat{h}\|^2 \tag{3.42b}$$

for some $\hat{\xi}$ in $\partial\psi(\hat{x})$ (c.f. case a)). Either way we get a contradiction of (3.41) and hence b) cannot take place.

c) Suppose that $\psi(x_i) < -\varepsilon(x_i) \leq -\hat{\varepsilon}$ $\forall i \in K$. Then clearly, $\psi(x_i) \leq 0$ for all $i \geq i_0$, where i_0 is smallest integer in K and hence, by continuity of $f(\cdot)$ and construction in step 2, $f(x_i) \searrow f(\hat{x})$ (with monotonicity for $i \geq i_0$). Now, for all $i \in K$, $h(x_i) = \nabla f(x_i)$ and $\nabla f(\hat{x}) \neq 0$. Clearly, there exists an $\hat{s} > 0$ such that $\psi(x_i - \hat{s}h(x_i)) \leq 0$ for all $i \in K$. Hence, (as in the case of the ordinary Armijo gradient method) there exists a $\hat{\delta} > 0$ such that

$$f(x_{i+1}) - f(x_i) \leq -\hat{\delta} < 0 \qquad \forall i \in K \tag{3.43}$$

But this contradicts the convergence of $\{f(x_i)\}$ and thus c) cannot take place. This exhausts all possibilities. The fact that $\psi(\hat{x}) \leq 0$ follows from Assumption 3.1 and thus we are done. ¤

We are now ready to state our implementable algorithm.

Implementable Algorithm 3.2.

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

Parameters: $\alpha, \beta, \eta \in (0,1)$, $b \gg 1$.

Step 0: set $i = 0$.

Step 1: Use a proximity algorithm, such as the one in Appendix C, to compute $\tilde{h}_{fi}$, $\tilde{h}_{\psi i}$ and $\tilde{\varepsilon}_i \in E$ (approximations to $h_{f\varepsilon(x_i)}(x_i)$, $h_{f,\varepsilon(x_i)}(x_i)$ and $\varepsilon(x_i)$) such that

$$\langle \tilde{h}_{\psi_i}, \xi \rangle \geq \|\tilde{h}_{\psi i}\|^2 \quad \forall \xi \in \nabla^{\tilde{\varepsilon}_i}(x_i) \tag{3.44a}$$

$$\langle \tilde{h}_{fi}, \xi \rangle \geq \|\tilde{h}_{fi}\|^2 \quad \forall \xi \in \text{co}\{\nabla f(x_i), \nabla^{\tilde{\varepsilon}_i}(x_i)\} \tag{3.44b}$$

$$\|h_{\psi,\tilde{\varepsilon}_i}(x_i)\|^2 - \tilde{\varepsilon}_i/2 \leq \|\tilde{h}_{\psi_i}\|^2 \leq \|h_{\psi,\tilde{\varepsilon}_i}(x_i)\|^2 \tag{3.44c}$$

$$\|h_{f,\tilde{\varepsilon}_i}(x_i)\|^2 - \tilde{\varepsilon}_i/2 \leq \|\tilde{h}_{fi}\|^2 \leq \|h_{f,\tilde{\varepsilon}_i}(x_i)\|^2 \tag{3.44d}$$

$$\tilde{\varepsilon}_i \geq \varepsilon(x_i) \tag{3.44e}$$

and set

$$\tilde{h}_i = \Gamma(x_i)\tilde{h}_{f,i} + (1-\Gamma(x_i))\tilde{h}_{\psi_i} \tag{3.45}$$

Step 2: Compute the largest step size $s_i = \eta^k \in [0,M]$, with k an integer, such that

If $\psi(x_i) > 0$

$$\psi(x_i - s_i\tilde{h}_i) - \psi(x_i) \leq -s_i\alpha\|h_i\|^2 \tag{3.46a}$$

If $\psi(x_i) \leq 0$,

$$f(x_i - s_i\tilde{h}_i) - f(x_i) \leq -s_i\alpha\|h_i\|^2 \tag{3.46b}$$

and

$$\psi(x_i - s_i\tilde{h}_i) \leq 0. \tag{3.46c}$$

Step 3: Set $x_{i+1} = x_i - s_i h_i$, set $i = i+1$ and go to step 1. ¤

Theorem 3.2: a) Suppose that $x_i \in X$ is such that $h(x_i) \neq 0$ (i.e. $\theta(x_i) < 0$) then s_i as constructed in step 2 of Algorithm 3.2 satisfies $s_i > 0$, i.e., the algorithm is well defined. b) If $\hat{x}$ is an accumulation point of a sequence $\{x_i\}$ constructed by Algorithm 3.2, then $\theta(\hat{x}) = 0$ and $\psi(\hat{x}) \leq 0$. ¤

We omit a proof of this theorem since its proof is entirely analogous to the proofs of Lemma 3.1 and Theorem 3.1.

To complete this section, we shall state the conceptual and implementable algorithms for the simplest general case of problem P_k (2.1)-(2.3), characterized by a single matrix $G(x,\omega)$, i.e., one $Q(x,\omega)$, and a single function $\phi(x,\nu)$, viz:

$$P_{\Omega_k}: \min\{f(x) | g^j(x) \leq 0,\ j \in J;\ \phi(x,\nu) \leq 0 \quad \forall \nu \in N,$$
$$\ell(\omega)^2 \leq y^i(x,\omega) \leq u(\omega)^2 \quad \forall \omega \in \Omega_k\}$$

where $\Omega_k \subset \Omega$ is finite. As in [G.3], we must make the following

<u>Assumption 3</u>: For every $x \in X$, the set $N(x) \subset N$ of local maximizers of $\phi(x,\cdot)$ is finite. ¤

Next, we define

$$\psi(x,\omega) \triangleq \max\{y^i(x,\omega)-u(\omega)^2,\ \ell(\omega)^2-y^i(x,\omega)\} \tag{3.47}$$

$$\zeta(x) \triangleq \max_{\nu \in N} \phi(x,\nu) \tag{3.48}$$

and

$$\Phi(x) \triangleq \max\{g^j, j \in J;\ \psi(x,\omega),\ \omega \in \Omega_k;\ \zeta(x)\} \tag{3.49}$$

Next, for any $x \in X$, $\varepsilon \geq 0$, we define

$$J_\varepsilon(x) \triangleq \{j \in J | g^j(x) \geq \Psi(x) - \varepsilon\} \tag{3.50}$$

$$\Omega_{k,\varepsilon}(x) \triangleq \{\omega \in \Omega_k | \psi(x,\omega) \geq \Psi(x) - \varepsilon\} \tag{3.51}$$

$$N_\varepsilon(x) \triangleq \{\nu \in N(x) | \phi(x,\nu) \geq \Psi(x) - \varepsilon\} \tag{3.52}$$

We now add the argument ω to $\nabla^\varepsilon(x)$ defined in (3.10) so that it becomes $\nabla^\varepsilon(x,\omega)$ and, finally, we define

$$\nabla^{\varepsilon}_{P_k}(x,\omega) = \mathrm{co}\{\nabla g_j, j \in J_\varepsilon(x);\ \nabla\phi(x,\nu),\ \nu \in N_\varepsilon(x);$$

$$\nabla^{\varepsilon}(x,\omega),\ \omega \in \Omega_{k,\varepsilon}(x)\} \tag{3.53}$$

Next, we define (c.f. (3.12))

$$h_{\psi,\varepsilon}(x) \triangleq \mathrm{Nr}(\nabla^{\varepsilon}_{P_k}(x,\omega)) \tag{3.54}$$

$$h_{f,\varepsilon}(x) \triangleq \mathrm{Nr}(\nabla f(x), \nabla^{\varepsilon}{}_{P_k}(x,\omega)) \tag{3.54b}$$

and, with $\Gamma(x)$ defined as in (3.13), but with $\Psi(x)$ replacing $\psi(x)$, we define $h_\varepsilon(x)$, $\theta_\varepsilon(x)$, $\varepsilon(x)$, $h(x)$ and $\bar{\theta}(x)$ as in (3.14)-(3.18). The conceptual algorithm for P_{Ω_k} is the same as algorithm 3.1 except that $\Psi(x)$ replaces $\psi(x)$. To obtain an implementable algorithm, if we assume a high level of precision in the computation of the set $N_\varepsilon(x)$, we simply substitute $\nabla_{P_k}^{\tilde{\varepsilon}_i}(x_i)$ for $\nabla^{\tilde{\varepsilon}_i}(x_i)$ and $\Psi(x_i)$ for $\psi(x_i)$. If we wish to use adaptive precision calculations in defining $N_\varepsilon(x)$, then the implementable algorithm statement becomes more cumbersome. The specific manner in which this can be done is stated in [G.3].

We now turn to the task of decomposing the problem P into a sequence P_k.

4. A Master Outer Approximations Algorithm

Again, for the sake of clarity, we shall consider the design problem in its simplest form P (2.4), since this form contains all the relevant difficulty and information as far as decomposition by means of outer approximations is concerned. Since we are no longer working at a single frequency ω and single fixed subset $\Omega_k \subset \Omega$, we shall introduce the quantities ω, Ω_k in all of our relevant notation. Thus, P becomes

$$P_\Omega : \min\{f(x) | \ell^2(\omega) \leq y^i(x,\omega) \leq u(\omega)^2 \ \forall i \in \underline{m},\ \omega \in \Omega\} \tag{4.1}$$

P_k becomes

$$P_{\Omega_k} : \min\{f(x) | \ell(\omega)^2 \leq y^i(x,\omega) \leq u(\omega)^2 \ \forall i \in \underline{m},\ \omega \in \Omega_k\} \tag{4.2}$$

for any $\omega \subset \Omega$, we define

$$\psi(x,\omega) \triangleq \max\{y^i(x,\omega) - u(\omega)^2,\ \ell(\omega)^2 - y^i(x,\omega),\ i \in \underline{m}\} \tag{4.3}$$

and, for any $\omega \in \Omega$ and $\varepsilon \geq 0$, we define $\nabla^\varepsilon(x,\omega)$ as in (3.10), for the given value ω. Next, for any $\Omega' \subseteq \Omega$, we define

$$\overline{\psi}_{\Omega'}(x) \triangleq \max_{\omega\in\Omega'} \psi(x,\omega) \tag{4.4}$$

and for any $\Omega' \subseteq \Omega$ and $\varepsilon \geq 0$, we define

$$\Omega'_\varepsilon(x) \triangleq \{\omega \in \Omega' | \psi(x,\omega) \geq \psi_{\Omega'}(x)_+ - \varepsilon\} \tag{4.5}$$

where $\overline{\psi}_{\Omega'}(x)_+ \triangleq \max\{\overline{\psi}_{\Omega'}(x),0\}$. Next, we define

$$\nabla^\varepsilon_{\Omega'}(x) = \mathrm{co}\{\bigcup_{\omega\in\Omega'_\varepsilon(x)} \nabla^\varepsilon(x,\omega)\} \tag{4.6}$$

We are now ready to define the optimality functions for problems $P_{\Omega'}$. For any $x \in X$ and $\varepsilon \geq 0$, we define

$$h_{\psi,\Omega'}(x,\varepsilon) \triangleq \mathrm{Nr}\{\nabla^\varepsilon_{\Omega'}(x)\} \tag{4.7}$$

and

$$h_{f,\Omega'}(x,\varepsilon) \triangleq \mathrm{Nr}\ \mathrm{co}\{\nabla f(x), \nabla^{\varepsilon}_{\Omega'}(x)\}) \tag{4.8}$$

we then define

$$\theta_{\Omega',\varepsilon}(x) \triangleq -\max\{\|\Gamma_{\Omega'}(x)h_{f,\Omega'}(x,\varepsilon)\|^2, \|(1-\Gamma_{\Omega'}(x)h_{\psi,\Omega'}(x,\varepsilon)\|^2\} \tag{4.9}$$

where $\Gamma_{\Omega'}(x) \triangleq e^{-\gamma\psi_{\Omega'}(x)_+}$ and $\overline{\psi}_{\Omega'}(x)_+ \triangleq \max\{0,\psi_{\Omega'}(x)\}$. It should be clear from the analysis in Section 3 that if $\hat{x}$ is optimal for $P_{\Omega'}$, then $\theta_{\Omega',0}(\hat{x}) = 0$. As before, we define

$$\varepsilon_{\Omega'}(x) \triangleq \max\{\varepsilon \in \varepsilon \mid \theta_{\Omega',\varepsilon}(x) \leq -\varepsilon\} \tag{4.10}$$

and

$$\overline{\theta}_{\Omega}(x) \triangleq \theta_{\Omega',\varepsilon_{\Omega'}(x)}(x) \tag{4.11}$$

We are now ready to state an outer approximations algorithm for decomposing P_Ω (see [G.1])

<u>Outer Approximations Algorithm (4.1)</u>

<u>Data</u>:

(i) $\Omega_0 \subset \Omega$ a finite set.

(ii) A sequence $\{\varepsilon_{kj}\}$[†] such that

a) $\varepsilon_{jj} = 0$ for all j and $\varepsilon_{kj} > 0$ for all $j > k$.

b) $\varepsilon_{kj} \nearrow \hat{\varepsilon}_j$ as $k \to \infty$.

c) $\hat{\varepsilon}_j \searrow 0$ as $j \to \infty$.

(iii) $\beta \in (0,1)$.

<u>Step 0</u>: Set $k = 0$.

<u>Step 1</u>: Solve P_{Ω_k} to the extent of finding an $\overline{x}_k$ such that

$$\overline{\psi}_{\Omega_k}(\overline{x}_k) \leq \beta^k \tag{4.12}$$

[†]A typical sequence with the required properties is defined by

$$\varepsilon_{kj} = 100\{(\tfrac{1}{1+j})^{1/10} - (\tfrac{1}{1+k})^{1/10}\}.$$

$$\overline{\theta}_{\Omega_k}(\overline{x}_k) \geq -\beta^k \tag{4.13}$$

Step 2: Compute an $\omega_k \in \Omega_k$ such that

$$\underline{\overline{\Psi}}_{\Omega}(x_k) = \phi(\overline{x}_k,\omega_k) \tag{4.14}$$

Step 3: Set

$$\Omega_{k+1} = \{\omega_k\} \cup \{\omega_j \in \Omega_k | \phi(\overline{x}_j,\omega_j) > \varepsilon_{kj}\} \tag{4.15}$$

Set k = k+1 and go to Step 1. ¤

Since the ε_{kj} increase as k increases, a particular ω_j will be retained for a certain number of iterations and then, quite likely, dropped, never to be used again. Thus, the cardinality of Ω_k need not grow indefinitely, indeed, it can usually be kept quite low, particularly if the computation is carried out interactively. The reader should refer to [G.2] for a detailed discussion of outer approximations methods.

Theorem 4.1. If x is an accumulation point of a sequence $\{\overline{x}_k\}$ constructed by Algorithm 4.1, then $\underline{\overline{\Psi}}_{\Omega}(\hat{x}) \leq 0$ and $\overline{\theta}_{\Omega}(\hat{x}) = 0$, i.e. $\hat{x}$ is feasible and stationary for P_{Ω}. ¤

To prove this theorem we need a number of preliminary results. We define

$$\tilde{\Omega}_{\varepsilon}(x) \triangleq \{\omega \in \Omega | \psi(x,\omega) \geq -\varepsilon\} \tag{4.16}$$

$$\tilde{\nabla}_{\varepsilon}(x) \triangleq \text{co}\{\bigcup_{\omega \in \tilde{\Omega}_{\varepsilon}(x)} \nabla^{\varepsilon}(x,\omega)\} \tag{4.17}$$

and we define

$$\tilde{h}_{\psi,\varepsilon}(x) \triangleq \text{Nr}(\tilde{\nabla}_{\varepsilon}(x)) \tag{4.17a}$$

$$\tilde{h}_{f,\varepsilon}(x) \triangleq \mathrm{Nr}(\mathrm{co}\{\nabla f(x),\tilde{\nabla}_\varepsilon(x)\}) \tag{4.17b}$$

and, finally, we define

$$\tilde{\theta}_{\Omega',\varepsilon}(x) \triangleq -\max\{\|\Gamma_{\Omega'}(x)\tilde{h}_{f,\varepsilon}(x)\|^2,\ \|(1-\Gamma_{\Omega'}(x))\tilde{h}_{\psi,\varepsilon}(x)\|^2\} \tag{4.18}$$

We note that if $\underline{\psi}_\Omega(\hat{x}) \leq 0$, then $\tilde{\theta}_{\Omega,\varepsilon}(x) = \overline{\theta}_\Omega(x)$ for $\varepsilon = \varepsilon_\Omega(x)$. Also we note that because $\Omega'_\varepsilon(x) \subset \tilde{\Omega}_\varepsilon(x)$, we have $\overline{\theta}_{\Omega'}(x) \leq \tilde{\theta}_{\Omega',\varepsilon}(x) \leq 0$ for all $x \in X$, $\varepsilon = \varepsilon_{\Omega'}(x)$ and for all $\Omega' \subseteq \Omega$.

It follows directly from Theorem 3 in [G.2], that if $\hat{x}$ is an accumulation point of a sequence $\{\overline{x}_k\}$ constructed by Algorithm 4.1, then $\underline{\psi}_\Omega(\hat{x}) \leq 0$. Furthermore, by construction,

$$-\beta^k \leq \overline{\theta}_{\Omega_k}(\overline{x}_k) \leq \tilde{\theta}_{\Omega_k,\varepsilon_k}(\overline{x}_k) \leq 0 \tag{4.19}$$

where $\varepsilon_k = \varepsilon_{\Omega_k}(\overline{x}_k) \to 0$ as $k \to \infty$. Hence $\tilde{\theta}_{\Omega_k,\varepsilon_k}(\overline{x}_k) \to 0$ as $k \to \infty$. Now, because $\underline{\psi}_\Omega(\hat{x}) \leq 0$, it follows that $|\Gamma_{\Omega_k}(\overline{x}_k) - \Gamma_\Omega(\overline{x}_k)| \to 0$ as $k \to \infty$, and hence (4.17) implies also that

$$\tilde{\theta}_{\Omega,\varepsilon_k}(\overline{x}_k) \to 0 \text{ as } k \to \infty. \tag{4.20}$$

It remains to show that $\tilde{\theta}_{\Omega,0}(\hat{x}) = 0$. Our first result is obvious:

<u>Proposition 4.1</u>:

a) For any $x \in X$,

(i) $\tilde{\Omega}_\varepsilon(x)$ is monotone increasing in ε,

(ii) $\tilde{\Omega}_\varepsilon(x)$ is closed for all $\varepsilon > 0$ and

(iii) $\bigcup_{\varepsilon>0} \tilde{\Omega}_\varepsilon(x) = \tilde{\Omega}_0(x)$.

(b) For any $\varepsilon \geq 0$ $\tilde{\Omega}_\varepsilon(\cdot)$ is u.s.c. ¤

This leads to the following result:

Proposition 4.2:

a) For any $x \in X$,

(i) $\tilde{\nabla}^{\varepsilon}(x)$ is monotone increasing in ε,

(ii) $\bigcap_{\varepsilon>0} \tilde{\nabla}^{\varepsilon}(x) = \tilde{\nabla}^{0}(x)$.

b) For any $\varepsilon \geq 0$, $\tilde{\nabla}^{\varepsilon}(\cdot)$ is u.s.c. ¤

The proof of this proposition is given in Appendix 2

Lemma 4.1: If $\hat{x} \in X$ is such that $\|\tilde{h}_f(\hat{x},0)\|^2 > 0$, then there exist a $\hat{\rho} > 0$ and an $\hat{\varepsilon} > 0$ such that $\|\tilde{h}_f(\hat{x},\varepsilon)\|^2 \geq \frac{1}{2}\|\tilde{h}_f(\hat{x},0)\|^2$ for all $(x,\varepsilon) \in B(\hat{x},\hat{\rho}) \times [0,\varepsilon]$.

Proof: Suppose that $\|\tilde{h}_f(\hat{x},0)\|^2 > 0$. Then, by Proposition 4.2a(i)

There exists an $\hat{\varepsilon} > 0$ such that $\|\tilde{h}_f(\hat{x},\hat{\varepsilon})\|^2 \geq \hat{\varepsilon}$. It now follows from Proposition 4.2b that there exists a $\hat{\rho} > 0$ such that for all $x \in B(\hat{x},\hat{\rho})$ $\|\tilde{h}_f(\hat{x},\hat{\varepsilon})\| \geq \hat{\varepsilon}/2$. It now follows from Proposition 4.2a(i) that $\|\tilde{h}_f(x,\varepsilon)\|^2 \geq \|\tilde{h}_f(x,\hat{\varepsilon})\|^2 \geq \hat{\varepsilon}/2$ for all $(x,\varepsilon) \in B(\hat{x},\hat{\rho}) \times [0,\hat{\varepsilon}]$.

Proof of Theorem 4.1: It follows directly from Theorem 3 in [G.2] that $\hat{x}$ is feasible, i.e. that $\psi_{\Omega}(\hat{x}) \leq 0$. Also, it has been shown earlier that $\tilde{\theta}_{\Omega,\varepsilon_k}(\bar{x}_k) \to 0$ and $\varepsilon_k \to 0$ as $k \to \infty$.

Now, suppose that $\underline{\Psi}_{\Omega}(\hat{x}) < 0$ and that $\bar{x}_k \xrightarrow{K} \hat{x}$, for $K \subset \{a,1,2,\ldots\}$. Then there exists a k_0 such that $\underline{\Psi}_{\Omega_k}(\bar{x}_k) \leq \underline{\Psi}_{\Omega}(\bar{x}_k) < 0$ for all $k \geq k_0$, $k \in K$, and hence, since $\varepsilon_k \to 0$ as $k \to \infty$, there exists a $k_1 \geq k_0$ such that $\underline{\psi}_{\Omega_k}(\bar{x}_k) < -\varepsilon_k$ for all $k \geq k_1$, $k \in K$. Therefore, for all $k \geq k_1$, $k \in K$, $\tilde{\theta}_{\Omega,\varepsilon_k}(\bar{x}_k) = \|\nabla f(\bar{x}_k)\|^2$ which leads to $\nabla f(\hat{x}) = 0$, by the continuity $\nabla f(\cdot)$. Thus $\theta_{\Omega,0}(\hat{x}) = 0$.

Now suppose that $\underline{\Psi}_{\Omega}(\hat{x}) = 0$. Then $\Gamma_{\Omega}(\hat{x}) = 1$ and $\theta_{\Omega,0}(\hat{x}) = -\|\tilde{h}_{f,0}(\hat{x})\|^2$. For the sake of contradiction, suppose that $\theta_{\Omega,0}(\hat{x}) < 0$. Then, from

Lemma 4.1, since $\varepsilon_k \to 0$ as $k \to \infty$, we conclude that ther exist a k_2 such that for all $k \in K$, $k \geq k_2$,

$$\|\tilde{h}_{f,\varepsilon_k}(\bar{x}_k)\|^2 \geq \|\tilde{h}_{f,0}(\hat{x})\|^2/2. \tag{4.21}$$

Now $\Gamma_\Omega(x_k) \xrightarrow{K} 1$ as $k \to \infty$ and hence, since $\tilde{h}_{f,\varepsilon_k}(\bar{x}_k)$ and $\tilde{h}_{\psi,\varepsilon_k}(\bar{x}_k)$ must be bounded, it follows that

$$\overline{\lim_{k\to\infty}}\ \tilde{\theta}_{\Omega,\varepsilon_k}(\bar{x}_k) \leq -\overline{\lim_{k\in K}}\ \|\tilde{h}_{f,\varepsilon_k}(\bar{x}_k)\|^2 \leq -\frac{1}{2}\|\tilde{h}_{f,0}(\hat{x})\|^2 < 0 \tag{4.22}$$

which contradicts the fact that $\tilde{\theta}_{\Omega,\varepsilon_k}(\bar{x}_k) \to 0$ as $k \to \infty$. Hence we are done. ¤

5. CONCLUSION

To conclude, it may be worthwhile to summarize in what respect the algorithm presented in this paper is different from a general purpose nondifferentiable optimization algorithm. First, in general, one has no idea whether one is approaching a point of nondifferentiability or not. In the case of singular values, the distance between them serves as a "distance to probable collision," i.e., to a nondifferentiable point. Second, in general, one only has at one's disposal either generalized gradients or approximations to smeared generalized gradients, both of which are only U. S. C. In the case of singular values, we were able to use sets ∇ which are continuous. Thus, our algorithm exploits the structure of the singular value problem to a considerable extent. In addition, it avoids the difficulties caused by the fact that singular vectors cannot be computed with any kind of precision near a multiple singular value (only the subspace can be computed accurately). As a result, our new algorithm should be considerably superior to earlier ones based on general nondifferentiable optimization algorithms.

Appendix A: Proof of Proposition 3.1

We prove Proposition 3.1 by establishing a sequence of facts.

Fact 1: Consider the special case of problem P_k (and consequently of Proposition 3.1) where $x \in \mathbb{R}^1$. Let $\hat{x} \in \mathbb{R}^1$. Let $\frac{d\psi_u}{dX^+}(\hat{x})$ denote the right derivative of ψ_u at $\hat{x}$, and let $\frac{d\psi_\ell}{dX^+}(\hat{x})$ denote the right derivative of ψ_ℓ at $\hat{x}$, let $Q'(\hat{x}) = \frac{dQ}{dx}(\hat{x})$.

Let U_u and U_ℓ be, respectively, $m \times k_n(\hat{x},0)$ and $m \times m+1-k_\ell(\hat{x},0)$ complex matrices such that

$$U_u{}^*U_u = I_{ku(x,0)},\quad U_\ell{}^*U_\ell = I_{m+1-k_\ell(x,0)}, \tag{A.1}$$

and

$$U_u{}^*Q(x)U_u = y^1(x)I_{ku}(x,0),\quad U\,{}^*Q(x)U\ = y^m(x)I_{m+1-k_\ell}(x,0) \tag{A.2}$$

Then

$$\frac{d\psi_u}{dx^+}(\hat{x}) = \text{the largest eigenvalue of } U_u{}^*Q'(\hat{x}) \tag{A.3a}$$

and

$$\frac{d\psi_\ell}{dx^+}(\hat{x}) = \text{the smallest eigenvalue of } U_\ell{}^*Q'(\hat{x})U_\ell \tag{A.3b}$$

Proof: We provide only the proof for $\frac{d\psi_u}{dx^+}(x)$, the Proof of $\frac{\partial\psi_\ell}{\partial X^+}(\hat{x})$ is entirely analogous. From analytic function theory [A.1] and from [R.1] we know that there exists $\delta > 0$ such that:

(i) The maps $x \to y^i(x)$ are analytic $\forall i \in \underline{k_u(\hat{x},0)}$ on $[\hat{x},\hat{x}+\delta]$

(ii) There exists an analytic function $V : [\hat{x},\hat{x}+\delta] \to C^{mxk_u(\hat{x},0)}$ (where $C^{mxk_u(\hat{x},0)}$ is the space of $m \times k_u(\hat{x},0)$ complex matrices) such that for all $x \in [\hat{x},\hat{x}+\delta]$ $V^*(x) \cdot V(x) = I_{k_u}(\hat{x},0)$, and

$$V^*(x)Q(x)V(x) = \mathrm{diag}(y^1(x),y^2(x),\ldots y^{k_u(\hat{x},0)}(x)). \tag{A.4}$$

Finally define

$$\Lambda(x) \triangleq \mathrm{diag}(y'(x),\ldots y^{k_u(\hat{x},0)}(x)) \tag{A.5}$$

Then for all $x \in (\hat{x},\hat{x}+\delta)$ we have from (A.4) that

$$\frac{dV^*}{dx}(x)\ V(x) + V^*(x)\ \frac{dV}{dx}(x) = 0 \tag{A.6}$$

and

$$\frac{d\Lambda}{dx}(x) = \frac{dV^*}{dx}(x)Q(x)V(x) + V^*(x)Q(x)\ \frac{dU}{dx}(x) + V^*(x)Q'(x)V(x). \tag{A.7}$$

Letting $x \to \hat{x}$ we get that

$$\frac{d\Lambda}{dx}(\hat{x}) = y'(x)[\frac{dV^*}{dx}(\hat{x})\ V(\hat{x}) + V^*(x)\ \frac{dV}{dx}(\hat{x})] +$$

$$+ V^*(\hat{x})Q(\hat{x})V(\hat{x}) = V^*(\hat{x})Q'(\hat{x})V(\hat{x}), \tag{A.8}$$

and the last equality follows from (A.7). From (i) and (ii) we conclude that $\dfrac{dy^1(x)}{dx^+}$ = the largrst element on the diagonal of $\dfrac{d\Lambda}{dx}(\hat{x})$ $= V^*(\hat{x})Q'(\hat{x})V(\hat{x})$. Since $\dfrac{d\psi_u}{dx^+}(\hat{x}) = \dfrac{dy'(\hat{x})}{dx^+}$, and $V^*(x)Q(x) \cdot V(x)$ is diagonal, it follows that $\dfrac{d\psi_u(\hat{x})}{dx^+}$ = the largest eigenvalue of $V^*(\hat{x})Q'(\hat{x})V(\hat{x})$. But $V(\hat{x})$ and U_u are related by $V(\hat{x}) = U_uW$, where W is a $k_u(\hat{x},0) \times k_u(\hat{x},0)$ complex matrix, and $W^*W = WW^* = I_{k_u(\hat{x},0)}$. Thus $V^*(\hat{x})Q'(\hat{x})V(\hat{x})$ $= W^*U_uQ'(\hat{x})U_uW$, and therefore $V^*(\hat{x})Q'(\hat{x})V(\hat{x})$ and $U_u{}^*Q'(\hat{x})U_u$ have the same eigenvalues. The desired result now follows directly. ¤

The following fact is a direct corollary of Fact 1. Hence a proof is omitted.

<u>Fact 2</u>: Let X be the Banach space of Hermitian $m \times m$ complex matrices. Let $\mu_u : X \to \mathbb{R}^1$ be a functional such that $\forall\ A \in X\ \mu_u(A) \triangleq$ the largest eigenvalue of A. Given A and B in X we order the eigenvalues of A in decreasing order, i.e. $y_A^1 \geq y_A^2, \ldots \geq y_A^m$ where $\{y_A^i\}_{i=1}^m$ are the eigenvalues of A. Assume that $y_A^1 = y_A^{k_u} \neq y_A^{k_u+1}$ for some $k_u \in \underline{m}$. Let U_u be

an $m \times k_u$ complex matrix such that $U_u{}^*U_u = I_{k_u}$ and $U_u{}^*AU_u = y_A^1 \cdot I_{k_u}$. Finally, let $\mu_u^0(A,B)$ denote the directional derivative of μ_u at A in direction B. Then $\mu_u^0(A,B)$ = the largest eigenvalue of $U_u{}^*BU_u$.

Similarly let $\mu_\ell(A) \triangleq$ the smallest eigenvalue of A. Assume that $y_A^{k_\ell} = y_A^m$ and $y_A^{k_\ell-1} \neq y_A^{k_\ell}$ for some $k_\ell \in \underline{m}$. Let U_ℓ be an $m \times m+1-k_\ell$ complex matrix such that $u_\ell{}^*U_\ell = I_{m+1-k_\ell}$ and $U_\ell{}^*AU_\ell = y_A{}^m I_{m+1-k_\ell}$. Finally, let $\mu_\ell{}^0(A,B)$ denote the directional derivative of μ_ℓ at A in direction B. Then $\mu_\ell{}^0(A,B)$ = the smallest eigenvalue of $U_\ell{}^*BU_\ell$.

¤

The following fact is recalled from [P.8],

Fact 3: Let X and y be two Banach spaces. Let $S:X \to y$ be a frêchét differentiable map and let $\mu:y \to \mathbb{R}$ be a (nonlinear) locally Lipschitz continuous, directionally differentiable functional. Let $\mu^0(y,r)$ denote the directional derivative of μ at y in direction e, and let S_x' denote the frêchét derivative of S at x. Let $\phi(x) \triangleq \mu(S(x))$ and let $\phi^0(x,e)$ denote the directional derivative of ϕ at x in direction e. Then $\forall x$, $e \in X$ $\phi^0(x,e)$ exists and $\phi^0(x,e) = \mu^0(X(x),S_x(e))$.

¤

We now return to the general setting of Problem P_k (and consequently of Proposition 3.1), i.e. we consider the case where $X \in \mathbb{R}^n$.

The following fact is a direct corollary of facts 2 and 3.

Fact 4: Given $\hat{x}$ and $e \in \mathbb{R}^n$, $k_u(\hat{x},0)$ and $k_\ell(\hat{x},0)$ as in (3.7). Let U_u and U_ℓ, be, respectively, $m \times k_u(\hat{x},0)$ and $m \times m+1-k_\rho(\hat{x},0)$ complex matrices, such that $U_u{}^*U_u = I_{k_u(\hat{x},0)}$, $U_\ell{}^*U_\ell = I_{m+1-k_\ell(\hat{x},0)}$, $U_u{}^*Q(\hat{x})U_u = y^1(\hat{x})I_{k_u(\hat{x},0)}$ and $U_\ell{}^*Q(\hat{x})U_\ell = y^m(\hat{x})I_{m+1-k_u(\hat{x},0)}$. Finally let $\psi_u^0(\hat{x},e)$ denote the directional derivative of ψ_u at $\hat{x}$ in direction e, and let $\psi_\ell^0(x,e)$ denote the directional derivative of ψ_ℓ at $\hat{x}$ in direction e.

Then $\psi_U^0(\hat{x},e) =$ the largest eigenvalue of $U_u^*(\sum_{1=1}^{n} e^i \frac{\partial A}{\partial X^i}(\hat{x}))U_u$ and $\psi_\ell^0(\hat{x},e)$ = the smallest eigenvalue of $U_\ell^*(\sum_{1=1}^{n} e^i \frac{\partial A}{\partial X^i}(\hat{x}))U_\ell$, where $e = (e^1,\ldots e^n)^T$ and $x = (x^1,\ldots x^n).^T$ ¤

At present, we shall discuss ψ_u only, a treatment of ψ_ℓ will come later.

<u>Fact 5</u>: $\forall\, x \in \mathbb{R}^n, \forall\, e \in \mathbb{R}^n$

$$\psi^0(x,e) = \mathrm{Max}\{\langle \xi,e\rangle \,|\, \xi\in\nabla_u^0(x)\}. \tag{A.9}$$

<u>Proof</u>: For all $z \in C^{k_u(x,0)}$, $\|z\| = 1$ we must have by Fact 4, that

$$\langle (\langle Z,U_u^* \frac{\partial Q}{\partial x^1}(x)U_u z\rangle ,\ldots\langle Z,U_u^* \frac{\partial Q}{\partial x^n}(x)Uz\rangle),e\rangle = \langle Z,U_u^*(\sum_{1=1}^{n} e^i \frac{\partial Q}{\partial x^i}(x))U_u z\rangle$$

$$\leq \text{the largest eigenvalue of } U_u^*(\sum_{1=1}^{n} e^i \frac{\partial Q}{\partial x^i}(x))U_u = \psi^0(x,e) \tag{A.10}$$

Therefore

$$\mathrm{Max}\{\langle \xi,e\rangle \,|\, \xi\in\nabla_u^0(x)\} \leq \psi^0(x,e) \tag{A.11}$$

On the other hand $\mathrm{Max}\{Z,U_u^*(\sum_{1=1}^{n} e^i \frac{\partial Q}{\partial x^i}(x))U_u Z \,|\, Z\in \mathbb{C}^{k_u(x,0)}, \|z\| = 1\}$

= the largest eigenvalue of $U_u^*(\sum_{1=1}^{n} e^i \frac{\partial A}{\partial x^i}(x))U_u$ and hence we are done.

¤

We state the following fact without a proof, since the proof is straightforward:

<u>Fact 6</u>: The map $x \to \nabla_u^0(x)$ is upper semicontinuous in the sense of Berge ([B.2]). ¤

The "generalized directional derivative $\psi'(x,e)$ of ψ at x in direction e" is defined in [C.1] as .

$$\psi'(x,e) \triangleq \overline{\lim_{\substack{\alpha\downarrow 0\\ h\to 0}}} \frac{\psi(x+h+re)-\psi(x+h)}{\lambda} \tag{A.12}$$

It is shown in [C.1] that

$$\psi'(x,e) = \mathrm{Max}\{\langle \xi,e\rangle \,|\, \xi\in\partial\psi(x)\} \tag{A.13}$$

The following fact holds when $n = 1$, i.e. $x \in \mathbb{R}^1$ and follows directly from the smoothness properties of eigenvalues.

Fact 7: If $n = 1$ then $\psi_u^0(x,1) = \psi_u'(x,1)\ \forall\, x \in \mathbb{R}^1$. ¤

We also need the following auxiliary result.

Fact 8: Given E and F, two compact convex sets in $\mathbb{R}^n$, if for every

$$e \in \mathbb{R}^n \quad \mathrm{Max}\{\langle \xi,e\rangle \,|\, \xi\in E = \mathrm{Max}\{\langle \xi,e\rangle \,|\, \xi\in F\} \tag{A.14}$$

Then $E = F$.

Proof: If $E \neq F$, then, either there exists an $f \in F - E$, or an $f \in E - F$. Suppose that $f \in F - E$. Let $e_f = \arg\min\{\|f-g\| \,|\, g\in E\}$ and let $e = f - E_f$. Then for all $g \in E$, we have that

$$\langle g - f,\ e_f - f\rangle \geq \|e_f - f\|^2 \gneq 0 \tag{A.15}$$

Which implies that

$$\langle g,\ E_f - f\rangle \gneq \langle f,\ E_f - f\rangle \tag{A.16}$$

and hence $\langle g,e\rangle \gneq \langle f,e\rangle$, contradicting (A.14). ¤

From now on we assume that $n \geq 1$ and $x \in \mathbb{R}^n$. First, we prove that the point to set map $x \to \nabla_u^0(x)$ has the mean value property with respect to the function ψ.

Fact 9: For all $x,e \in \mathbb{R}^n$ there exist $\tau \in [0,1]$ and $\xi \in \nabla^0_u(x+\tau e)$ such that

$$\psi_u(x+e) - \psi_u(x) = \langle e,\xi \rangle. \tag{A.17}$$

Proof: For all $\tau \in [0,1]$ define $C(\tau) \triangleq A(x+\tau e)$, and define $\tilde{\psi}_u(\tau) \triangleq$ highest eigenvalue of $C(\tau)$. From Fact 5, (A.13) and Fact 8 we have that

$$\nabla^0_u \tilde{\psi}_u(\tau) = \partial \tilde{\psi}_u(\tau) \ \forall \ \tau \in [0,1]. \tag{A.18}$$

From the mean value theorem of Lebourg [L.3] we have that

$$\psi_u(x+e) - \psi_u(x) = \tilde{\psi}_u(1) - \tilde{\psi}_u(0) = \langle \xi,e \rangle \tag{A.19}$$

for some $\xi \in \partial\tilde{\psi}_u(\tau)$, and some $\tau \in [0,1]$. From (A.18) $\xi \in \nabla^0_u \tilde{\psi}_u(\tau)$ $= \mathrm{co}\{\langle Z, U_u^T \cdot C'(\tau) U_u z\rangle | z \in C^{k_u(x+\tau e,0)}, \|z\| = 1\}$, where $U_u^* U_u = I_{k_u}(x+\tau e)$ and $U_u^* C(\tau) U_u = y^1(x+\tau e) I_{k_u}(x+\tau e)$. Thus $\xi \in \mathrm{co}\{\langle z, U_u^*(\Sigma e^i \frac{\partial Q}{\partial x^i}(x+\tau e)) U_u z\rangle | \|z\| = 1\} = \mathrm{co}\{\langle (\langle z, U_u^* \frac{\partial A}{\partial x^1}(x+\tau e) U_u z\rangle, \ldots,$ $\langle z, U_u^* \frac{A}{x^n}(x+ e) U_u z\rangle)^T, e\rangle | \|z\| = 1\} = \{\langle \xi,e\rangle | \xi \in \nabla^0_u(x+\tau e)\}$, which concludes our proof. ⌑

Corollary: $\forall \ x,e \in \mathbb{R}^n$ we have that

$$\psi_u'(x,e) = \mathrm{Max}\{\langle \xi,e\rangle | \xi \in \nabla^0_u(x)\} \tag{A.20}$$

Proof: $\psi_u'(x,e) = \overline{\lim}_{\substack{\lambda \downarrow 0 \\ h \to 0}} \frac{\psi(x+h+\lambda e) - \psi(x+h)}{\lambda} \geq \overline{\lim}_{\substack{\lambda \downarrow 0 \\ h \to 0}} \langle \xi_{h,\lambda}, e\rangle$ (where $\xi_{h,\lambda} \in \nabla^0_u(x+h+s\lambda e)$ for all $s \in [0,1]$) $= \mathrm{Max}\{\langle \xi,e\rangle | \xi \in \nabla^0_u(x)\}$, where the inequality follows from Fact 9 and the equality follows from Fact 6.

⌑

Fact 10: For all $x \in X$ $\nabla_u^0(x) = \partial\psi_u(x)$.

Proof: This fact follows directly from Fact 5, (A.13) and Fact 8.

Fact 11: For all $x \in X$ $-V_\ell^0(x) = \partial\psi_\ell(x)$.

Proof: Define $\overline{\psi}_\ell(x) \triangleq$ the highest eigenvalue of $-Q(x)$. Then $\ell^2 + \overline{\psi}_\ell(x) = \psi_\ell(x)$ and hence $\partial\overline{\psi}_\ell(x) = \partial\psi_\ell(x)$. But by Fact 10, $\partial\overline{\psi}_\ell(x) = -\nabla_\ell^0(x)$. Consequently $\partial\psi_\ell(x) = -\nabla_\ell^0(x)$. ¤

We can now conclude the proof of Proposition 3.1 if $\psi_u(x) > \psi_\ell(u)$ then $\partial\psi(x) = \partial\psi_u(x) = \nabla_u^0(x)$. If $\psi_u(x) < \psi_\ell(x)$ then $\partial\psi(x) = \partial\psi_\ell(x) = -\nabla_\ell^0(x)$. Hence, for these two cases we are done. It remains to consider the case where $\psi_u(x) = \psi_\ell(x)$.

In this case, it follows from [P.8], Theorem 3.2 that

$$\psi^0(x,e) = \text{Max}\{\psi_h^0(x,e),\ \psi_\ell^0(x,e)\}\ \forall\ e \in \mathbb{R}^n \tag{A.21}$$

and from [P.8], Theorem 3.4 it follows that

$$\text{co}\{\nabla_u^0(x)\ -\nabla_\ell^0(x)\} = \{\xi \in R^u \mid \langle \xi,e\rangle \le \psi^0(x,e)\ \forall\ e \in \mathbb{R}^n\}. \tag{A.22}$$

Now for x and e $\mathbb{R}^n$, we have that

$$\psi'(x,e) = \lim_{\substack{\lambda\downarrow 0\\ h\to 0}} \frac{\psi(x+h+\lambda e)-\psi(x+h)}{\lambda} \le \text{Max}\{\psi_u'(x,e);\ \psi_\ell'(x,e)\}$$

$$= \text{Max}\{\psi_u^0(x,e);\psi_\ell^0(x,e)\} = \text{Max}\{\langle \xi,e\rangle \mid \xi \in \text{Co}\{\nabla_u^0(x),-\nabla_\ell^0(x)\}\}. \tag{A.23}$$

with the last equality holding because of Fact 5,(A.13) and Facts 10 and 11. From Fact 8 it now follows that

$$\partial\psi(x) = \text{Co}\{\nabla_u^0(x),-\nabla_\ell^0(x)\}.$$

¤

Appendix B: Proof of Proposition 3.2, Corollary 3.1 and Assumption 3.2.

Proof of Proposition 3.2:

a) If $0 \leq \varepsilon \leq \varepsilon'$ then $\nabla^{\varepsilon}(x) \subset \nabla^{\varepsilon'}(x)$. Hence, $\|h_{\psi,\varepsilon}(x)\| \geq \|h_{\psi,\varepsilon'}(x)\|$ and $\|h_{f,\varepsilon}(x)\| \geq \|h_{f,\varepsilon'}(x)\|$. Consequently, $\theta_{\varepsilon}(x) \leq \theta_{\varepsilon'}(x)$.

b) First we show that the point to set map $x \to \nabla_u^{\varepsilon}(x)$ is upper semi-continuous. Let $\{x_i\}_{i=1}^{\infty} \subset X$ be such that $x_i \to \hat{x}$, as $i \to \infty$. We know that $k_u(x_i,\varepsilon) \in \underline{m}$. Let $K' \subset \mathbb{N}_+$ be any infinite set of integers such that for every $i \in K'$, $k_u(x_i,\varepsilon) = \hat{k}$, a fixed integer in $\underline{m}$. From the definition of $k_u(\cdot,\varepsilon)$ it follows that

$$y^k(x_i) - y^{k+1}(x_i) < \varepsilon \quad \forall (i,k) \in K' \times \underline{\hat{k}-1} \tag{B.1}$$

and that

$$y^{\hat{k}-1}(x_n) - y^{\hat{k}}(x_n) > \varepsilon \geq 0 \tag{B.2}$$

From (B.1) we conclude that $y^k(\hat{x}) - y^{k+1}(\hat{x}) \leq \varepsilon$ for all $k \in \underline{\hat{k}-1}$, and hence that $k_u(\hat{x},\varepsilon) \geq \hat{k}$. From analytic function theory ([A.1]) and from [R.1], we know that there exists $\hat{U}_u$ and $\tilde{U}_{u,i}$, complex $m \times k_u(\hat{x},\varepsilon)$ matrices, and $I \in \mathbb{N}_+$ such that for every $i \geq I$

$$\hat{U}_u^* \hat{U}_u = I_{K_u(\hat{x},\varepsilon)} \tag{B.3a}$$

and

$$\tilde{U}_{u,i}^* \tilde{U}_{u,i} = I_{K_u(\hat{x},\varepsilon)} \tag{B.3b}$$

$$\hat{U}_u^* Q(\hat{x}) \hat{U}_u = \mathrm{diag}(y^1(\hat{x}),\ldots,y^{k_u(\hat{x})}(\hat{x}))$$

and (B.4)

$$\tilde{U}_{u,i}^* \, Q(X_u) \tilde{U}_{u,i} = \mathrm{diag}(y^1(x_u),\ldots,y^{K_u(\hat{x})}(x_u))$$

and $\tilde{U}_{u,i} \underset{n\to\infty}{\to} \hat{U}_u$.

Let $U_{u,i}$ be the $m\mathrm{x}\hat{k}$ matrix consisting of the first $\hat{k}$ columns of $\tilde{U}_{u,i}$. Now, for every sequence of unit vectors $Z_i \in C^k$ such that $Z_i \xrightarrow{K'} \hat{Z}$ we have

$$(\langle Z_i, U_{u,i}^* \frac{\partial A}{\partial x^T}(x_i)\, U_{u,i}Z_i\rangle, \ldots, \langle Z_i, U_{u,i}^* \frac{\partial A}{\partial x^n}(x_i)\, U_{u,i}Z_i\rangle)$$

$$\underset{i\to\infty}{\xrightarrow{K'}} (\langle \hat{Z},(\hat{U}_uT)^* \frac{\partial A}{\partial x^T}(\hat{x})(\hat{U}_uT)\hat{Z}\rangle \ldots \langle \hat{Z},(U_uT)^* \frac{\partial A}{\partial x^u}(\hat{x})(\hat{U}T)\hat{Z}\rangle)\ , \tag{B.5}$$

where T is the $k_u(\hat{x},\varepsilon) \times \hat{k}$ matrix, the upper $\hat{k}\mathrm{x}\hat{k}$ block of which is $I_{\hat{k}}$, and the lower $(k_u(\hat{x},\varepsilon)-\hat{k}) \times \hat{k}$ block of which consists of zeros only. Let $\hat{Z}_0 = T\hat{Z}$. Then $Z_0' \in C^{k_u(\hat{x},\varepsilon)}$ and $\|\hat{Z}_0\| = 1$. It follows that

$$(\langle \hat{Z},(\hat{U}_uT)^* \frac{\partial A}{\partial x^T}(\hat{x})(\hat{U}_uT)\hat{Z}\rangle \ldots \langle \hat{Z},(\hat{U}_uT)^* \frac{\partial A}{\partial x^u}(\hat{U}_uT)\hat{Z}\rangle) \in \nabla^{\varepsilon}(\hat{x}).$$

since the above analysis holds for every sequence $\{x_i\}_{i=0}^{\infty}$ such that $x_i \to \hat{x}$ and every $Z_i \in C^{K_u(x_{u,\varepsilon})}$, we conclude that the point to set map: $x \to \nabla_u^{\varepsilon}(x)$ is upper semi-continuous.

In a similar way, one can show that the point to set map: $x \to \nabla_{\ell}^{\varepsilon}(x)$ is upper semi-continuous. It follows therefore that the point to set map: $x \to \nabla^{\varepsilon}(x)$ is also upper semi-continuous.

Proposition 3.2.b now follows directly from the fact that the map: $x \to \nabla^{\varepsilon}(x)$ is upper semi-continuous and that $\Gamma(\cdot)$ is continuous.

c) If x solves P_k

$$\theta_{\varepsilon}(\hat{x}) = -\|h_{f,\varepsilon}(\hat{x})\|^2. \tag{B.6}$$

If $\theta_0(\hat{x}) < 0$ then $\|h_{f,0}(x)\|^2 > 0$. Consider first the case where $\psi(\hat{x}) = 0$. From upper semi-continuity of $\nabla^0(\cdot)$, it follows that there exist $\hat{\rho} > 0$ and $\hat{\mu} > 0$ such that for every $x \in B(\hat{x},\hat{\rho})$ $\|h_{f,0}(x)\|^2 > \hat{\mu}$. Now for every

$\lambda > 0$ we have that

$$\psi(\hat{x}-\lambda h_{f,0}(\hat{x})) - \psi(\hat{x}) = -\lambda\langle h_{f,0}(\hat{x}),\xi\rangle \tag{B.7}$$

for some $\xi \in \nabla^0(x)$ and some $x \in B(\hat{x},h_{f,0}(\hat{x})\lambda)$.

We have therefore that

$$\overline{\lim_{\lambda\downarrow 0}} \frac{\psi(\hat{x}-\lambda h_{f,0}(x))-\psi(\hat{x})}{\lambda} \leq \|h_{f,0}(\hat{x})\|^2 \tag{B.8}$$

Therefore, there exists $\bar{\lambda} > 0$ such that for every $\lambda \in [0,\bar{\lambda}]$ we have that $\psi(\hat{x}-\lambda h_{f,0}(\hat{x})) < \psi(\hat{x})$. We also have the fact that

$$\lim_{\lambda\downarrow 0} \frac{f(\hat{x}-\lambda h_{f,0}(\hat{x}))-f(\hat{x})}{\lambda} = -\langle h_{f,0}(\hat{x}),\nabla f^0(\hat{x})\rangle \leq -\|h_{f,0}(\hat{x})\|^2 < 0. \tag{B.9}$$

Hence there exists $\bar{\bar{\lambda}} > 0$ such that for every $\lambda \in [0,\bar{\bar{\lambda}}]$, $f(\hat{x}-\lambda h_{f,0}(\hat{x})) < f(\hat{x})$ and $\psi(\hat{x}-\lambda h_{f,0}(\hat{x})) < \psi(\hat{x})$. Therefore $\hat{x}$ is not optimal.

If $\psi(\hat{x}) < 0$ then $h_{f,0}(\hat{x}) = \nabla f^0(\hat{x})$. If $\nabla f^0(\hat{x}) \neq 0$ then we can clearly find $\lambda > 0$ such that $f(\hat{x}-\lambda\nabla f^0(\hat{x})) < f(\hat{x})$ and (by continuity of ψ) $\psi(x-\lambda\nabla f^0(\hat{x})) \leq 0$, and $\hat{x}$ can not be optimal. ¤

d. Suppose that $\theta(\hat{x}) < 0$. Then there exists an $\bar{\varepsilon} > 0$ such that $\theta_{\bar{\varepsilon}}(\hat{x}) < -\varepsilon$. From (b) we know that there exists a $\hat{\rho} > 0$ such that for every $x \in B(\hat{x},\hat{\rho})$ $\theta_{\bar{\varepsilon}}(x) < -\bar{\varepsilon}$. But for every $x \in B(\hat{x},\hat{\rho})$ and $\varepsilon \in [0,\bar{\varepsilon}]$ we have that $\theta_{\varepsilon}(x) \leq \theta_{\bar{\varepsilon}}(x) \leq -\bar{\varepsilon}$. ¤

<u>A Proof of Corollary 3.1</u>.

If the statement of the corollary is not true, then there exists a sequence $\{x_i\}_{i=1}^{\infty}$, such that $x_i \xrightarrow[i\to\infty]{} \hat{x}$ and $\varepsilon(x_i) \leq \beta^2\,\varepsilon(\hat{x})$. Consequently, $\theta_{\beta\varepsilon(\hat{x})}(x_i) > -\beta\varepsilon(x)$. Thus $\theta_{\beta\varepsilon(\hat{x})}(\hat{x}) \geq -\beta\varepsilon(\hat{x})$. However, $\theta_{\beta\varepsilon(\hat{x})} \leq \theta_{\varepsilon(\hat{x})}(\hat{x}) \leq -\varepsilon(\hat{x}) < -\beta\varepsilon(\hat{x})$, which leads to a contradiction, unless $\varepsilon(\hat{x}) = 0$. But this implies that $\theta(\hat{x}) = 0$. Hence we are done. ¤

A Proof of Lemma 3.2:

We prove only the case of $\hat{\nabla}^{\hat{k}_u}$. We know that $y^{\hat{k}_u}(\hat{x}) \neq y^{\hat{k}_u+1}(\hat{x})$. Hence there exists an $\hat{\rho} > 0$ such that for every $x \in \hat{B}(\hat{x},\hat{\rho})$ $y^{\hat{k}_u}(x) \neq y^{k_u+1}(x)$. For every $x \in B(\hat{x},\hat{\rho})$, $\hat{\nabla}^{k_u}(x)$ is independent of U, provided that $U^*U = I_{\hat{k}_u}$ and $U^*Q(x)U = \mathrm{diag}(y^1(x),\ldots,y^{\hat{k}_u}(x))$. Moreover, from [R.1] we can find for every $x \in B(\hat{x},\hat{\rho})$ a $U(x)$ such that $U(x)$ is analytic in x on $B(\hat{x},\hat{\rho})$. From continuity of Q_v it follows that $\nabla^k(\cdot)$ is continuous in x on $B(\hat{x},\hat{\rho})$. ¤

Appendix C: A Proximity Algorithm.

In Step 1 of Algorithm 3.2 we need to compute $\tilde{h}_{fi}$, $\tilde{h}_{\psi i}$ and $\tilde{\varepsilon}_i$ which satisfy (3.4a). First we state a suitable proximity algorithm.

Algorithm C.1.

Data: $x_i \in \mathbb{R}^n$, $\varepsilon > 0$, $\tilde{h}^0_{fi}(x_i) = \nabla f(x_i)$ and $\tilde{h}^0 \psi_i(x_i) \in \nabla^{\varepsilon}(x_i)$.

Step 0: Set $j = 0$.

Step 1[†]: Set $g^j_{fi} = \arg\min\{\langle g, \tilde{h}^j_{fi}\rangle \mid g \in \nabla^{\varepsilon}(x_i)\}$ (C.1)

and set $g^j_{\psi_i} = \arg\min\{\langle g, \tilde{h}^j_{\psi_i}\rangle \mid g \in \nabla^{\varepsilon}(x_u)\}$ (C.2)

Step 2: a) If $\langle g^j_{fi}, h^j_{fi}\rangle \geq \|h^j_{fi}\|^2 - \varepsilon/4$ (C.3)

If then $\langle g^J_{fi}, h^J_{fi}\rangle \leq \langle \nabla f(x_i), \tilde{h}^j_{fi}\rangle$ (C.4)

then set $\tilde{h}_{fi,\varepsilon} = \dfrac{\langle g^j_{fi}, h^j_{fi}\rangle}{\|\tilde{h}^J_{fi}\|^2} \; h^j_{fi}$. (C.5)

Else set

$$\tilde{h}_{fi,\varepsilon} = \frac{\langle \nabla f(x_i), \tilde{h}^j_{fi}\rangle}{\|\tilde{h}^J_{fi}\|^2} \; \tilde{h}^j_{fi}. \tag{C.6}$$

If $\langle g^J_{\psi_i}, \tilde{h}^J_{\psi_i}\rangle \geq \|\tilde{h}^J_{fi}\|^2 - \varepsilon/4$ (C.7)

then set $\tilde{h}_{\psi_{i,\varepsilon}} = \dfrac{\langle g^J_{\psi_i}, h^J_{\psi_i}\rangle}{\|h^J \psi_i\|^2}$, $h^J \psi_i$ (C.8)

Step 3: Set $h^{j+1}_{fi} = Nr(Co\{\tilde{h}^j_{fi}, g^j_{fi}\})$

set $g^{j+1}_{\psi_i} = Nr(Co\{\tilde{h}^J_{\psi_i}, g^J_{\psi_i}\})$,

set $j = j+1$ and go to Step 1. ¤

[†]A discussion of the computation in (C.1) and (C.2) will follow.

Lemma C.1: If $\varepsilon > 0$ then the proximity algorithm C.1 yields $\tilde{h}_{fi}$ and $\tilde{h}_{\psi_i}$ in finitely many iterations.

Proof: If h_{fi} is not construed in finitely many iterations then for all $j \in \{0,1,2,\ldots\} \triangleq \mathbb{N}_+$ we have in Step 2 that

$$\langle g^{\dot{j}}_{fi}, \tilde{h}^{\dot{j}}_{fi} \rangle < \| h^{\dot{j}}_{fi} \|^2 - \varepsilon/4. \tag{C.3}$$

Since for all $\dot{j} \in \mathbb{N}_+$, $h^{\dot{j}}_{fi} \in \nabla^{\varepsilon}(x_i)$, and $\nabla^{\varepsilon}(x_i)$ is a compact set, it follows that there exists $r \in (0,1)$ such that for all $\dot{j} \in \mathbb{N}_+$ $\|\tilde{h}^{\dot{j}}_{fi}\|^2 - \varepsilon/4 \le r\|\tilde{h}^{\dot{j}}_{fi}\|^2$. From Proposition 5.8 of [P.6] it follows now (by setting $\tilde{M} = \{-\tilde{h}^{\dot{j}}_{fi}\}$, $\Gamma = \{-g^{\dot{j}}_{fJ}\}$ and $\hat{M} = \mathrm{Co}\{\nabla f(x_i), \nabla^{\varepsilon}(x_i\})$ that there exists a constant $C \in (0,1)$ such that $\|\tilde{h}^{\dot{j}+1}_{fi}\|^2 \le \mathrm{Max}\{r,c\} \cdot \|\tilde{h}^{\dot{j}}_{fi}\|^2$. This implies that $\tilde{h}^{\dot{j}}_{fi} \xrightarrow[\dot{j}\to\infty]{} 0$. But then $\min\{\langle g, \tilde{h}^{\dot{j}}_{fi}\rangle \mid g \in \nabla^{\varepsilon}(x_i)\} \xrightarrow[\dot{j}\to\infty]{} 0$. Thus there exists a $\dot{j}_0 \in \mathbb{N}_+$ such that for every $j \ge \dot{j}_0$ we have that $\langle g^{\dot{j}}_{fi}, \tilde{h}^{\dot{j}}_{fi} \rangle \ge \|h^{\dot{j}}_{fi}\|^2 - \varepsilon/4$, contradicting (C.3), and hence $\tilde{h}_{fi}$ is construeded in a finite number of iterations.

The proof for $\tilde{h}_{\psi}$ is similar and hence will be omitted. ¤

We now indicate an efficient way for solving the problems in (C.1) and in (C.2),

Assume for simplicity that $\nabla^{\varepsilon}(x_i) = \nabla^{\varepsilon}_u(x_i)$. For a given vector $a \in \mathbb{R}^n$, we need to compute $\arg\min\{\langle g,a\rangle \mid g \in \nabla^{\varepsilon}_u(x_i)\}$.

We note that

$$\min\{\langle g,a\rangle \mid g \in \nabla^{\varepsilon}(x_i)\} = \min\{\langle Z, U_u^*(\sum_{j=1}^{n} a^j \frac{\partial A}{\partial x^j}(x_i))U_u Z\rangle \mid \|Z\| = 1\}$$

(where $a = (a^1, a^2, \ldots a^u)^T$) = the smallest eigenvalue of

$$U_u^*(\sum_{j=1}^{n} a^{\dot{j}} \frac{\partial A}{\partial x^{\dot{j}}}(x_i))U_u.$$

Let z_a be any unit eigenvector for the above eigenvalue. Then

$$\arg\min\{\langle g,a\rangle \mid g \in \nabla^{\varepsilon}(x_i)\} = (\langle Z_a, U_u^* \frac{\partial A}{\partial x^1}(x_2)U_u Z_a\rangle, \langle Z_a, U_u^* \frac{\partial A}{\partial x^2}(x_i)U_u Z_a\rangle, \ldots \langle Z_a, U_u^* \frac{\partial A}{\partial x^u}(x_i)U_u Z_a\rangle)^T.$$

To compute $\tilde{\varepsilon}_i$ and hence $\tilde{h}_{\psi_i} = \tilde{h}_{\psi_i,\tilde{\varepsilon}_i}$ and $h_{fi} = h_{fi,\tilde{\varepsilon}_i}$, we need to find the largest $\varepsilon \in \Sigma$ such that $\max\{\|\Gamma(x_i)h_{fi,\varepsilon}\|^2, \|(1-\Gamma(x_i)h_{\psi_i,\varepsilon}\|^2) \geq \varepsilon$. Since in general, $\{\tilde{\varepsilon}_i\}$ is an (almost) monotonically decreasing sequence, we initialize the search for $\tilde{\varepsilon}_i$, by applying the proximity algorithm with $\varepsilon = \beta^2\tilde{\varepsilon}_{i-1}$. If this value of ε is too small, then the computation in Algorithm C.1 can be continued with $\beta\varepsilon$ replacing ε and $\tilde{h}_{\psi_i,\varepsilon}$, $\tilde{h}_{fi,\varepsilon}$ as a starting point. This will prove to be much more efficient then starting with $\varepsilon = 1$ and decreasing it by a factor of β over and over again.

REFERENCES

A.1. Ahlfors, Complex Analysis, McGraw-Hill, N.Y., 1953.

B.1. R. Bellman, Introduction to Matrix Analysis, McGraw-Hill, N.Y., 1960.

B.2. C. Berge, Topological Spaces, Macmilan Co., N.Y., 1963.

C.1. F. H. Clarke, "Generalized gradients and applications," Trans. Amer. Math. Soc., Vol. 205, pp. 247-262, 1975.

C.2. J. Collumn, W. E. Donath and P. Wolfe, "The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices," Mathematical Programming Study, No. 3, November 1975, pp. 35-55.

D.1. C. A. Desoer and M. J. Chen, "Design of multivariable control systems with stable plant," 1980 JACC.

D.2. J. C. Doyle, "Robustness of multiloop linear feedback systems," in Proc. 17th IEEE Conf. on Decision and Control, 1970.

G.1. G. Golub and C. Reinsh, "Singular value decomposition and least square solutions," Num. Math., Vol. 14, pp. 403-420, 1970.

G.2. C. Gonzaga and E. Polak, "On constraint dropping schemes and optimality functions for a class of outer approximations algorithms," Siam J. Contr. and Optimization, Vol. 17, pp. 477-493, 1979.

G.3. C. Gonzaga, E. Polak and R. Trahan, "An improved algorithm for optimization problems with functional inequality constraints," IEEE Trans. on Auto. Control, Vol. AC-25, No. 1, Feb. 1980.

L.1. A. J. Laub, "Linear multivariable control: numerical considerations," Report No. ESL-P-833, Electronics Systems Lab, MIT 197.

L.2. A. J. Laub, "An inequality and some computations related to the robust stability of linear dynamic systems," IEEE Trans. Auto. Contr., Vol. AC-24, No. 2, pp. 318-320, 1979.

L.3. G. Lebourg, " Valeur moyenne pour gradient généralisé," C. R. Acad. Sci., Paris, Vol. 281, pp. 795-797, 1975.

L.4. C. Lemarechal, "An extension of Davidon methods to nondifferentiable problems," in Mathematical Programming Studies 3, M. L. Balinski and R. Wolfe Eds., North Holland, Amsterdam, pp. 95-109, 1975.

M.1. D. Q. Mayne and E. Polak, "Algorithms for the design of control systems subject to singular value inequalities," Proc. NSF Workshop on Numerical Methods in Engineering, Lexington, KY, June 16-20, 1980.

M.2. D. Q. Mayne, E. Polak and R. Trahan, " On outer approximations algorithm for computer aided design problems," JOTA, 28, July 1979, pp. 331-352.

M.3. R. Mifflin, "An algorithm for constrained optimization with semi-smooth functions," Math. of Oper. Research, Vol. 2, No. 2, pp. 191-207, 1977.

M.4. A. G. J. MacFarlane (editor), Frequency-Response Methods in Control Systems, IEEE Press.

P.1. B. N. Parlett, The Symmetric Eigenvalue Problem, Prentice Hall, N.Y., 1980.

P.2. E. Polak and D. Q. Mayne, "On the solution of singular value inequalities over a continuum of frequencies," Memorandum No. UCB/ERL M80/8, January 1980; 20th IEEE Conf. on Dec. and Control, Albuquerque, N.M., Dec. 12-14, 1980.

P.3. E. Polak, D. Q. Mayne and Y. Wardi, "On the extension of constrained optimization algorithm to semi-smooth problems," to appear.

P.4. E. Polak, R. Trahan and D. Q. Mayne, "Combined phase I-phase II methods of feasible directions," Mathematical Programming, Vol. 17, pp. 61-73, July 1979.

P.4.A. E. Polak and D. Q. Mayne, " An algorithm for optimization problems with functional inequality constraints," IEEE Trans. on Auto. Control, Vol. AC-21, No. 2, April 1976.

P.4.B. E. Polak, "On a class of computer-aided design problems," A Link Between Science and Applications of Automatic Control, edited by A. Niemi, Pergamon Press, Oxford and New York, 1979.

P.5. E. Polak and R. Trahan, "An algorithm for computer aided design problems of control systems," Proc. IEEE Conf. on Decision and Control, 1976.

P.6. E. Polak and A. Sangiovanni-Vincentelli, "Theoretical and computational aspects of optimal design centering, tolerances and tuning problems," IEEE Trans. on Auto. Control, Vol. CAS-26, No. 9, pp. 295-318, 1979.

P.7. E. Polak, Computational Methods in Optimization: A Unified Approach, Academic Press, New York, 1971.

P.8. B. N. Pshenichnyi, Necessary Conditions for an Extremum, translation edited by L. W. Neustadt, translated by K. Makovski, Marcel Dekker, Inc., N.Y., 1971.

R.1. F. Rellich, Perturbation Theory of Eigenvalue Problems, Gordon and Breach, N.Y., London, Paris.

R.2. W. Rudin, Principles of Mathematical Analysis, McGraw-Hill, N.Y., 1964.

S.1. N. R. Sandel, "Robust stability of systems with applications to singular perturbations," Automatica, Vol. 15, p. 467, 1979.

W.1. J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1961.

Z.1. V. Zakian and V. Al-Naib, "Design of dynamical and control systems by the method of inequalities," Proc. IEE, Vol. 120, No. 11, pp. 1421-1427, 1973.