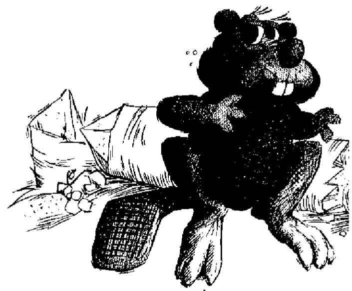
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



Nondifferentiable Optimization and Worse

Andrew R. Conn
Philip F. O'Neill

CS-83-05

March, 1983

Nondifferentiable Optimization and Worse

Andrew R. Conn

Department of Computer Science
University of Waterloo

Philip F. O'Neill

Department of Combinatorics and Optimization
University of Waterloo

ABSTRACT

A general approach to a class of nonsmooth constrained optimization problems is presented. The functions in the problem may be nondifferentiable and even noncontinuous; however, it must be possible to partition the domain of each function into a collection of sets, called cells, which are the feasible regions of systems of smooth constraints, such that the function is smooth over each cell. In these cases, the nonsmooth problem can be decomposed into a collection of smooth subproblems. In order to solve the nonsmooth problem, one might propose to find a descent direction at each iteration of a subgradient-type algorithm; we propose to examine a sequence of descent subproblems. The main advantage in this, is that established algorithms for smooth optimization can be applied to the subproblems.

March 30, 1983

Nondifferentiable Optimization and Worse *

Andrew R. Conn

Department of Computer Science
University of Waterloo

Philip F. O'Neill

Department of Combinatorics and Optimization
University of Waterloo

1. Introduction

We want to solve problems of the type:

$$NCNLP \quad \begin{cases} \inf & f^0(x) \\ \text{subject to} & f^i(x) = 0,\ i \in I_{eq} \\ & f^i(x) \geq 0,\ i \in I_{in} \end{cases}$$

where $f^i \in \{0\} \bigcup I_{eq} \bigcup I_{in}$ are a finite collection of functions $\mathrm{R}^n \rightarrow \mathrm{R}$ that may be nondifferentiable or even noncontinuous. However, we want to solve these problems by decomposing them into a collection of smooth subproblems; hence, we will consider functions of a piecewise smooth nature.

The constrained problem can be transformed into an unconstrained problem by using an exact penalty function. The basic principles of the method can be developed with reference to the unconstrained case. Because it is our purpose to explore the fundamental ideas, we will address only the problem of finding an unconstrained local infimum of a piecewise smooth function.

Throughout, we will use the Euclidean norm, along with the Euclidean distance function and Euclidean neighbourhoods:

For $x \in \mathrm{R}^n$, the Euclidean norm of x is

$$\| x \|_2 \equiv (x^{\mathrm{T}} x)^{1/2}$$

For $A \subseteq \mathrm{R}^n$, the distance function d_A is defined for every $x \in \mathrm{R}^n$ by

$$d_A(x) \equiv \inf \{ \| y - x \|_2 \mid y \in A \}$$

$d_A(x)$ is the Euclidean distance from x to the closest point in A.

For $A \subseteq \mathrm{R}^n$ and $\epsilon > 0$, the set of all points y whose distance from A is less than ϵ is

* This research is supported in part by Natural Science and Engineering Research Council Grant No. A8639

$$N_\epsilon(A) \equiv \{ y \in \mathrm{R}^n \mid d_A(y) < \epsilon \}$$

In the special case when A consists of a single point x, this reduces to

$$N_\epsilon(x) \equiv \{ y \in \mathrm{R}^n \mid \|y - x\|_2 < \epsilon \}$$

For $f : \mathrm{R}^n \to \mathrm{R}$ we define $\underline{f}(x) \equiv \liminf_{x^j \to x} f(x)$, where the "lim inf" is taken with respect to all sequences $\{x^j\}$ which converge to x. By definition, $f(\hat{x})$ is a local infimum of f if and only if $\underline{f}(\hat{x})$ is a local minimum of $\underline{f}$.

The notation $C^i(G)$, $i = 1, \ldots, \infty$ refers to the class of functions which have continuous partial derivatives of order i over the set G. $C^0(G)$ is the class of functions which are continuous over G.

This paper is divided into six sections. In §2 we will discuss the conditions which enable us to *partition* a function, in the neigbourhood of a point, into smooth pieces. We will discuss the consequences of this *partition property* with regard to finding the infimum of a function. In §3 we present a class of functions which has the *partition property* at every point in an open set and give some results for this class. In §4 we show that the class of functions defined in §3 contains all functions that have the *partition property* everywhere in a open set. A *conceptual* algorithm for finding a local infimum of a function in this class is given in §4. We prove that this algorithm converges under relatively weak assumptions and we indicate how this *conceptual* algorithm can be used to derive a *practicable* numerical algorithm. The final section contains a few brief conclusions.

2. The Partition Property

Our fundamental approach is to decompose the problem of finding the infimum of a nonsmooth function into a collection of smooth subproblems. We want to set conditions that are as weak as possible such that the problem of deciding if f has an infimum in the neighbourhood of a point x is equivalent to deciding whether or not x is a local minimizer of a finite collection of smooth constrained problems. For the moment, the only degree of smoothness we will require in the subproblems is continuity.

In addition we will require that:

(i) there exists $\xi > 0$ such that $N_\xi(x)$ can be partitioned into a finite collection of subsets such that the closure of each subset is the feasible region of a finite system of smooth constraints.

(ii) for each subset, there exists a function which is smooth over its closure and which has the same values as f for points in the subset.

This motivates the following definition.

2.1. Definition

Given, a function f defined over an open set G and a point $x \in G$, f is said to have the partition property at x with respect to G if there exist disjoint sets F_i, $i \in I(x)$ and functions f_i, $i \in I(x)$ (where $I(x)$ is a finite index set and the functions f_i, $i \in I(x)$ are not necessarily distinct) such that for some $\xi > 0$

(i) $N_\xi(x) \subset \bigcup_{i \in I(x)} F_i$

(ii) $x \in \overline{F}_i$ for every $i \in I(x)$

(iii) $\overline{F}_i \bigcap N_\xi(x) =$
$\{ y \in N_\xi(x) \mid c^l(y) \leq 0,\ l \in L_i;\ c^e(y) = 0,\ e \in E_i;\ c^g(y) \geq 0,\ g \in G_i \}$
where L_i, E_i and G_i are finite index sets (which may be null and are assumed without loss of generality to be disjoint) and $c^j \in C^0(\overline{F}_i \bigcap N_\xi(x))$, for every $j \in L_i \bigcup E_i \bigcup G_i$ and all $i \in I(x)$.

(iv) For all $i \in I(x)$, and every $y \in F_i \bigcap N_\xi(x)$, $f_i \in C^0(\overline{F}_i \bigcap N_\xi(x))$ and $f_i(y) = f(y)$. □

Note that:

(a) there may be many choices of sets F_i and functions f_i that satisfy the definition. The index set $I(x)$ depends on the particular choices of F_i and f_i. In some cases, it may be nontrivial to find sets and functions that satisfy the definition. However, for many practical problems, the conditions can be satisfied in a straightforward way.

(b) (iv) implies that discontinuities of f in $N_\xi(x)$ may occur only on the boundary relative to $N_\xi(x)$ of some F_i, $i \in I(x)$. In other words, the discontinuities of f occur at points $\{ x \mid c^j(x) = 0,\ j \in J \}$ for $J \subset E_i \bigcup G_i \bigcup L_i$ where the functions c^j, $j \in J$ are continuous over $\overline{F}_i \bigcap N_\xi(x)$.

(c) the function f_i in definition 2.1 is not the same as $f|(F_i \bigcap N_\xi(x))$, The domain of definition of $f|(F_i \bigcap N_\xi(x))$ is $F_i \bigcap N_\xi(x)$ whereas the domain of definition of f_i is $\overline{F}_i \bigcap N_\xi(x)$.

(d) the partition property is defined with respect to an open set G, at *a point* in G.

2.2. Example

$Z - \{0\} = \{\pm 1, \pm 2, \cdots\}$. Consider the function f represented in figure 1.

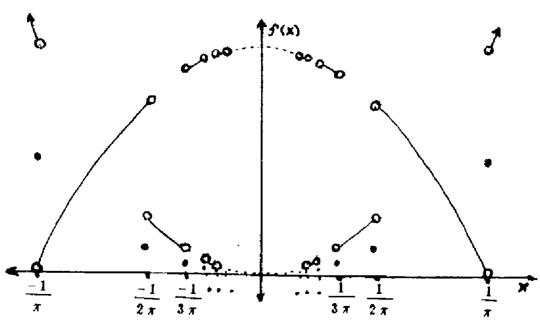


Fig. 1 f has the partition property w.r.t. R at each $x \in$ R

$$f(x) = \begin{cases} -x^2 + \frac{4}{\pi^2} & , \frac{1}{2k\pi} < x < \frac{1}{(2k-1)\pi} \quad ; k \in Z - \{0\} \\ 0 & \begin{cases} , \frac{1}{(2k+1)\pi} \le x \le \frac{1}{2k\pi} & ; k = \pm 100, \ldots \\ , x = 0 \end{cases} \\ \frac{x^2}{2} & , x = \frac{1}{k\pi} \quad ; k = \pm 1, \ldots, \pm 199 \\ x^2 & \begin{cases} , \frac{1}{(2k+1)\pi} < x < \frac{1}{2k\pi} & ; k = \pm 1, \ldots, \pm 99 \\ , -\infty < x < -\frac{1}{\pi} \, , \frac{1}{\pi} < x < \infty & ; k \in Z - \{0\} \end{cases} \end{cases}$$

One might propose to take the sets:

$$[0], \ (-\infty, -\tfrac{1}{\pi}), \ (\tfrac{1}{\pi}, \infty), \ [\tfrac{1}{k\pi}], \ (\tfrac{1}{(2k+1)\pi}, \tfrac{1}{2k\pi}), \ (\tfrac{1}{2k\pi}, \tfrac{1}{(2k-1)\pi}); \ k \in Z - \{0\}$$

as the F_i. Then, definition 2.1 is satisfied at any point except $x = 0$. Using this choice of F_i, there are infinitely many sets in every neighbourhood of $x = 0$.

Consider, however, the function g defined by:

$$g(x) = \begin{cases} x \sin(\frac{1}{x}) & , x \neq 0 \\ 0 & , x = 0 \end{cases}$$

g is continuous over R, [1, p.77]. Furthermore,

$$\{x \mid g(x) = 0\} = \bigcup_{k=1}^{\infty} \frac{1}{k\pi} \bigcup \{0\}$$

$$\{x \mid g(x) > 0\} = \bigcup_{k=1}^{\infty} (\frac{1}{(2k+1)\pi}, \frac{1}{2k\pi}) \bigcup (-\infty, -\frac{1}{\pi}) \bigcup (\frac{1}{\pi}, \infty)$$

$$\{x \mid g(x) < 0\} = \bigcup_{k=1}^{\infty} (\frac{1}{2k\pi}, \frac{1}{(2k-1)\pi})$$

Let

$$F_{-1} = \{\, x \mid g(x) < 0 \,\},\ F_0 = \{\, x \le \tfrac{1}{200\pi} \mid g(x) \ge 0 \,\}$$

$$F_1 = \{\, x > \tfrac{1}{200\pi} \mid g(x) > 0 \,\},\ F_2 = \{\, x > \tfrac{1}{200\pi} \mid g(x) = 0 \,\}$$

$$f_{-1}(x) = -x^2 + \frac{4}{\pi^2},\ f_0(x) = 0,\ f_1(x) = x^2,\ f_2(x) = \frac{x^2}{2}$$

$$c^0(x) = g(x),\ c^1(x) = \frac{1}{200\pi}$$

$$E_{-1} = E_1 = \emptyset,\ E_0 = E_2 = \{\, 0 \,\}$$

$$L_{-1} = \{\, 0 \,\},\ L_0 = \{\, 1 \,\},\ L_1 = L_2 = \emptyset$$

$$G_{-1} = G_0 = \emptyset,\ G_1 = \{\, 0, 1 \,\},\ G_2 = \{\, 1 \,\}$$

Then, for any $x \in F_{-1}$, $I(x) = \{-1\}$; for any $x \in F_1$, $I(x) = \{1\}$; for any $x \in F_0$, $I(x) = \{0, 1\}$, if $x = \frac{1}{k\pi}$, $k = \pm 200, \ldots$, $I(x) = \{0\}$, otherwise; for any $x \in F_2$, $I(x) = \{-1, 1, 2\}$. Then, for some $\xi > 0$ sufficiently small, definition 2.1 is satisfied at every $x \in \mathrm{R}$.

Even though f has infinitely many discontinuities in any neigbourhood of $x = 0$, f has the partition property at $x = 0$. This is because the sets F_i of definition 2.1 need not be connected.

Note that f has a global infimum of 0 at any $x \in F_0$, and a local infimum of $\frac{x^2}{2}$ at any $x \in F_2$.

2.3. Lemma

If f has the partition property at $\hat{x} \in G$, then f has the partition property at each $x \in N_\xi(\hat{x})$ for some $\xi > 0$ sufficiently small. (f cannot have the partition property at an isolated point.)

Proof: Follows from definition 2.1. □

2.4. Theorem

If f has the partition property at $x \in G$, (where F_i and f_i, $i \in I(x)$ are sets and functions, respectively, that satisfy definition 2.1.) then $\underline{f}\,(x) = \min\limits_{i \in I(x)} \{\, f_i(x) \,\}$.

Proof:

$$\underline{f}\,(x) = \liminf_{x^j \to x} f(x^j)$$

$$= \inf_{i \in I(x)} \left\{ \lim_{\substack{x^j \to x \\ \{x^j\} \in (F_i \cap G)}} f(x^j) \right\}$$

$$= \inf_{i \in I(x)} \left\{ \lim_{\substack{x^j \to x \\ \{x^j\} \in (F_i \cap G)}} f_i(x^j) \right\}$$

$$= \inf_{i \in I(x)} \{ f_i(x) \}$$

$$= \min_{i \in I(x)} \{ f_i(x) \} \quad \square$$

2.5. Corollary

Suppose f has the partition property at $\hat{x} \in G$. Then for some $\xi > 0$, $\underline{f}\,(x) \geq \min \{ f_i(x),\ i \in I(\hat{x}) \}$, for every $x \in N_\xi(\hat{x})$.

Proof: From lemma 2.3, if f has the partition property at $\hat{x}$, then for some $\xi > 0$, f has the partition property for every $x \in N_\xi(\hat{x})$ and for an appropriate partition at $\hat{x}$, $I(x) \subseteq I(\hat{x})$.

From theorem 2.4, $\underline{f}\,(x) = \min \{ f_i(x),\ i \in I(x) \}$, for every $x \in N_\xi(\hat{x})$. But since $I(x) \subseteq I(\hat{x})$, $\min \{ f_i(x),\ i \in I(x) \} \geq \min \{ f_i(x),\ i \in I(\hat{x}) \}$, for every $x \in N_\xi(\hat{x})$. $\square$

2.6. Definition

For f with the partition property at $x \in G$, define

$$\underline{I}(x) \equiv \{ i \mid f_i(x) = \underline{f}\,(x),\ i \in I(x) \}$$

where $I(x)$ satisfies definition 2.1. Note, as a result of theorem 2.4, $\underline{I}(x)$ is never empty.

2.7. Lemma

Let f have the partition property at x and $y \in G$. Let F_i and f_i, $i \in I(x)$ satisfy definition 2.1 at x for some $\xi > 0$ sufficiently small. If $y \in \overline{F_i}$ and $f_i(y) < f_i(x)$ for some $i \in \underline{I}(x)$ then $\underline{f}\,(y) < \underline{f}\,(x)$.

Proof: $\underline{f}\,(y) \leq f_i(y) < f_i(x) = \underline{f}\,(x)$. $\square$

2.8. Theorem

If f has the partition property at $\hat{x} \in G$, then $\hat{x}$ is a local infimum of f if and only if $\hat{x}$ is a constrained local minimum for each of the problems

$$\begin{cases} \min & f_i(x) \\ \text{subject to} & x \in \overline{F_i} \end{cases} \quad , i \in \underline{I}(\hat{x}) \qquad 2.8(a)$$

Proof: ($\Leftarrow$) Let $\hat{x}$ be a constrained local minimum for each of the continuous nonlinear programming problems 2.8(a).

(i) Consider $i \in I(\hat{x}) - \underline{I}(\hat{x})$. Since $f_i \in C^0(N_\xi(\hat{x}) \bigcap \overline{F_i})$ we have that $f_i(x) > \underline{f}\,(\hat{x})$ for every $x \in N_{\xi_i}(\hat{x}) \bigcap \overline{F_i}$ for some $\xi_i > 0$.

(ii) Consider $i \in \underline{I}(\hat{x})$. Since $\hat{x}$ is a constrained local minimum of

{minimize $f_i(x)$ subject to $x \in \overline{F}_i$}, $i \in \underline{I}(\hat{x})$ we have $f_i(x) \geq \underline{f}\,(\hat{x})$ for every $x \in N_{\xi_i}(\hat{x}) \cap \overline{F}_i$ for $\xi_i > 0$.
(iii) We have that $f_i(x) \geq \underline{f}\,(\hat{x})$, $i \in I(\hat{x})$, for every $x \in N_{\xi_i}(\hat{x}) \cap \overline{F}_i$ and all $i \in I(\hat{x})$ $\Rightarrow$ $f_i(x) \geq \underline{f}\,(\hat{x})$ for every $x \in N_{\xi}(\hat{x})$ where $\xi = \min\{\xi_i,\ i \in I(\hat{x})\}$. By corollary 2.5, $\underline{f}\,(x) \geq \underline{f}\,(\hat{x})$ for every $x \in N_{\gamma}(\hat{x})$ for some $\gamma > 0$ sufficiently small.
($\Rightarrow$) Assume $\underline{f}\,(\hat{x})$ is local infimum of f. Then, $\underline{f}\,(\hat{x}) \leq \underline{f}\,(x)$ for every x in $N_{\xi}(\hat{x})$ for some $\xi > 0$. Hence, $\underline{f}\,(\hat{x}) \leq \underline{f}\,(x)$ for every x in $N_{\xi}(\hat{x}) \cap F_i$ and for every $i \in I(\hat{x})$ when $\xi > 0$ is sufficiently small. $f_i(\hat{x})$, $i \in \underline{I}(\hat{x})$ gives the infimum of f_i over $N_{\xi}(\hat{x}) \cap F_i$ and thus $\hat{x}$ is a constrained local minimum for the problems 2.8(a). □

In special cases, theorem 2.8 can be used to derive more specific optimality conditions. The conditions in [2] and [4] are examples.

Theorem 2.8 implies that the partition property at a point x means that, conceptually at least, we can use principles of continuous nonlinear programming to decide if x is a local minimum of $\underline{f}$. Lemma 2.7 implies that if the problems 2.8(a) are sufficiently well-behaved, we can use a nonlinear programming algorithm to decrease $\underline{f}$ when x is not a local minimum. This suggests that if f has the partition property at every point in its domain of definition, then sequential algorithms for finding a local infimum of f will be possible.

3. A Class of Noncontinuous Functions

In order to investigate the possibility of sequential algorithms for locating a local infimum of a noncontinuous function, we wish to identify a class of functions that have the partition property at each point in the domain of definition. The following definition gives such a class. # and * represent properties of continuous functions. If H is an arbitrary subset of R^n, the notation $C^0 \cap *(H)$ refers to the class of continuous functions which posses the property * over H. For example, it could represent the class of functions which are linear over H, or the functions which are continuous and convex over H.

3.1. Definition

Let f be a function defined over an open set G. f is said to be *#-PIECEWISE** over G if there exists a collection of sets F_α, $\alpha \in A$ and functions f_α, $\alpha \in A$ (not necessarily distinct) such that for some $\xi > 0$:

(i) each $x \in G$ belongs to exactly one F_α, $\alpha \in A$

(ii) for each $x \in G$, $N_{\xi}(x) \cap G$ contains points of only a finite number of the F_α, $\alpha \in A$

(iii) for every $\alpha \in A$, $\overline{F}_\alpha \cap G =$
$\{x \in G \mid c^l(x) \leq 0,\ l \in L_\alpha;\ c^e(x) = 0,\ e \in E_\alpha;\ c^g(x) \geq 0,\ g \in G_\alpha\}$
where L_α, E_α and G_α are disjoint finite index sets (which may be null) and $c^j \in C^0 \cap \#\,(\overline{F}_\alpha \cap G)$, for every $l \in L_\alpha \cup E_\alpha \cup G_\alpha$

(iv) for every $\alpha \in A$, $f_\alpha \in C^0 \cap *(\overline{F}_\alpha \bigcap G)$ and $f_\alpha(x) = f(x)$, for every $x \in F_\alpha \bigcap G$. □

Note that:

(a) for some $\alpha \in A$ we may have $E_\alpha = L_\alpha = G_\alpha = \emptyset$ in which case $\overline{F}_\alpha \bigcap G = G$.

(b) the property # applies to the functions c^l; $l \in L_\alpha \bigcup G_\alpha \bigcup E_\alpha$, $\alpha \in A$ whereas the property * applies to the functions f_α, $\alpha \in A$. This means, that at any $x \in G$ the nonlinear programming problems defined by equation 2.8(a) will have a * objective function and # constraints. There is a two-fold purpose in this :

   (1) the numerical algorithm proposed in [3] applies an unspecified algorithm for smooth optimization to subproblems of the type defined by equation 2.8(a). In order to use a particular smooth algorithm, we will require additional assumptions on the constraints and/or the objective function of the problems given in equation 2.8(a).

   (2) in certain cases, the problems 2.8(a) will have some special structure or some simplifying features that can be exploited by a sequential algorithm.

In each case, the unspecified properties # and * allow us to anticipate such circumstances.

The sets F_α, $\alpha \in A$ will be called *cells* of f. Because the boundary relative to G of each F_α, $\alpha \in A$ is determined by the functions c^l, $l \in L_\alpha \bigcup G_\alpha \bigcup E_\alpha$, we will refer to these functions as *cell boundary functions*. The functions f_α, $\alpha \in A$ will be called *component functions* of f.

#*-piecewise* * will be abbreviated #*-PW**. If $f : G \to \mathrm{R}$ is #*-PW** over G, we will write $f \in \#\text{-}PW*(G)$.

3.2. Example

The function f of example 2.2 is C^0-PWC^∞ over R. The sets F_{-1}, F_0, F_1, and F_2 given in example 2.2 suffice as cells, having $c^0(x) = g(x)$ and $c^1(x) = \frac{1}{200\pi}$ as continuous cell boundary functions. The functions f_{-1}, f_0, f_1 and f_2 of example 2.2 suffice as C^∞ component functions. □

3.3. Lemma

It seems both natural and desirable that functions in the class $C^0 \cap *(G)$ should be considered piecewise * over G. It is easy to show that, indeed, functions which are continuous and * over an open set G are #*-PW** over G, for any #. *i.e.*,

$$f \in C^0 \cap *(G) \implies f \in \#\text{-}PW*(G), \text{ for any } \#$$

Proof: Let $A = \{1\}$, $F_1 = \mathrm{R}^n$, $f_1 = f$. Then definition 3.1 is satisfied for any #. □

3.4. Lemma

The class C^0–PWC^0 is the most general class of $\#$–$PW*$ functions in that all other classes are a subclass of C^0–PWC^0. *i.e.*,

$$f \in \#\text{-}PW*(G) \implies f \in C^0\text{-}PWC^0(G)$$

Proof: Trivial. □

Because the class C^0–PWC^0 subsumes all other classes of $\#$–$PW*$ functions, we will use the expression "piecewise continuous", to mean C^0–PWC^0 .

3.5. Theorem

If f is C^0–PWC^0 over an open set G, then f has the partition property at each $x \in G$.

Proof: Let f be $\#$–$PW*$ over an open set G. Since each $x \in G$ belongs to exactly one $F_\alpha, \alpha \in A$, then the $F_\alpha, \alpha \in A$ are disjoint. For $\xi > 0$ sufficiently small, $N_\xi(x)$ contains points of only a finite number of the F_α, $\alpha \in A$ and thus, $I(x)$ is a finite set. We now prove (i) - (iv) of definition 2.1. Consider any $x \in G$ and let $I(x) = \{\alpha \in A \mid x \in \overline{F}_\alpha\}$.

(i) (a) If $A - I(x) = \emptyset$, and if ξ is sufficiently small that $N_\xi(x) \subset G$, it follows that $N_\xi(x) \subset \bigcup_{\alpha \in I(x)} F_\alpha$.

(b) If $A - I(x) \neq \emptyset$, let $\sigma = \inf_{\alpha \in A - I(x)} \{ d_{F_\alpha}(x) \}$.

Clearly, $\sigma > 0$. Let $\gamma > 0$ be sufficiently small that $N_\gamma(x) \subset G$ and let $\xi = \min(\gamma, \frac{\sigma}{2})$, then $N_\xi(x) \subset \bigcup_{\alpha \in I(x)} F_\alpha$.

(ii) $x \in \overline{F}_\alpha$ for every $\alpha \in I(x)$, hence $x \in \overline{F}_\alpha \bigcap G$.

(iii) and (iv)

If $\xi > 0$ is sufficiently small that $N_\xi(x) \subset G$, then because $I(x) \subseteq A$, (iii) and (iv) of definition 3.1 follow from (iii) and (iv) of definition 2.1. □

3.6. Definition

Let f in $\#$–$PW*$ over an open set G, with cells $F_\alpha, \alpha \in A$ and component functions $f_\alpha, \alpha \in A$. For each $x \in G$,

$$I(x) = \{\alpha \in A \mid x \in \overline{F}_\alpha\}$$

define

$$\underline{I}(x) \equiv \{\alpha \in I(x) \mid f_\alpha(x) = \underline{f}(x)\}$$

For $H \subseteq G$ define

$$I(H) \equiv \{\alpha \in A \mid H \bigcap \overline{F}_\alpha \neq \emptyset\}$$

Thus, $A = I(G)$.

3.7. Lemma

(i) $\alpha \in I(x)$ if and only if $d_{F_\alpha}(x) = 0$.

(ii) $I(H) = \bigcup_{x \in H} I(x)$

Proof: Trivial. □

3.8. Theorem

For $f \in \#\text{-}PW*(G)$, the collection of sets F_α, $\alpha \in A$, the collection of functions F_α, $\alpha \in A$ and the collection of functions c^l, $l \in E_\alpha \bigcup G_\alpha \bigcup L_\alpha$, $\alpha \in A$ are each countable.

Proof: For $f \in \#\text{-}PW*(G)$ let $\xi > 0$ be sufficiently small to satisfy definition 3.1.

R^n can be covered be covered by closed hypercubes of side $\frac{\xi}{4}$. It is clear that the hypercubes in this grid are countable. Therefore, because $G \subseteq \mathrm{R}^n$, G is covered by a countable collection of hypercubes.

Let x^1, x^2,... be the centers of the hypercubes which cover G. By definition 3.1, $N_\xi(x^i)$, $i=1,2,...$ contains points of only a finite number of the F_α, hence each hypercube contains points of only a finite number of the F_α. Furthermore, each F_α intersects at least one hypercube. So, the cardinality of the entire collection of F_α is at most the cardinality of a countable number of finite subcollections of the F_α. Hence, the F_α are countable, [1, p.36]. □

Because there is only one f_α for each F_α, $\alpha \in A$ and a finite number of c^l, $l \in E_\alpha \bigcup G_\alpha \bigcup L_\alpha$ for each F_α, it follows that the F_α, $\alpha \in A$ and the c^l, $l \in E_\alpha \bigcup G_\alpha \bigcup L_\alpha$ are also countable.

3.9. Corollary

IF G is bounded, the collection of sets F_α, $\alpha \in A$, the collection of functions F_α, $\alpha \in A$ and the collection of functions c^l, $l \in E_\alpha \bigcup G_\alpha \bigcup L_\alpha$ are each finite.

Proof: Trivial. □

It is clear that if H and G are open sets and $H \subseteq G$ then $f \in \#\text{-}PW*(G) \implies f \in \#\text{-}PW*(H)$. If G_α, $\alpha \in A$ is a collection of open sets such that $f \in \#\text{-}PW*(G_\alpha)$ for every $\alpha \in A$ is it true that $f \in \#\text{-}PW*(\bigcup_{\alpha \in A} G_\alpha)$? This question is examined in theorem 3.10 and its corollaries 3.11 to 3.13. The main result of this study, is a complete characterization of all functions that have the partition property at each point in an open set. The characterization will be given in section 4.1.

3.10. Theorem

Given, two open neighbourhoods $N_{\sigma_1}(x^1)$ and $N_{\sigma_2}(x^2)$ where $\sigma_1, \sigma_2 > 0$, and an open set G, if $f \in C^0\text{-}PW*(G \bigcap N_{\sigma_1}(x^1))$ and $f \in C^0\text{-}PW*(G \bigcap N_{\sigma_2}(x^2))$ then $f \in C^0\text{-}PW*(G \bigcap (N_{\sigma_1}(x^1) \bigcup N_{\sigma_2}(x^2)))$.

Proof: Let $G_a = G \bigcap N_{\sigma_1}(x^1)$ and $G_b = G \bigcap N_{\sigma_2}(x^2)$. Let $\xi > 0$ be sufficiently small to satisfy definition 3.1 for f over both G_a and G_b.

To prove the theorem, we will use cells and component functions for f over G_a and G_b to construct cells for f over $G_a \bigcup G_b$. A set of cells for f over G_a may have a nonempty intersection with a set of cells for f over G_b, even if $G_a \bigcap G_b \neq \emptyset$. By using the function $d_{x^1}(x)$, we can modify the boundaries of any set of cells for f over G_a and any set of cells for f over G_b to obtain cells for f that are disjoint in $G_a \bigcup G_b$.

When $G_a \bigcap G_b \neq \emptyset$, it can be covered by modified cells of G_a or modified cells of G_b, (but not both). By adding the constraint $d_{x^1}(x) < \sigma_1$ to the cell boundary constraints of each cell of f over G_a, we obtain modified cells for G_a that cover $G_a \bigcap G_b$; by adding the constraint $d_{x^1}(x) \geq \sigma_1$ to the cell boundary constraints of each cell of f over G_b we obtain modified cells for $G_b - (G_a \bigcap G_b)$. The additional inequality constraint in each case ensures that the modified cells of G_a do not intersect the modified cells of G_b in $G_a \bigcup G_b$.

The component functions of f over $G_a \bigcup G_b$ can be taken as the component functions for the unmodified cells. Using the modified cells, ξ sufficiently small, and the original component functions, it is straightforward to verify that (i) - (iv) of definition 3.1 are satisfied. □

3.11. Corollary

Given, a finite number of open neighbourhoods $N_{\sigma_i}(x^i)$, $i = 1,2,...,k$, where $\sigma_i > 0$, $i = 1,2,...,k$ and an open set G, if $f \in C^0\text{-}PW*(G \bigcap N_{\sigma_i}(x^i))$ for every $i \in \{1,2, \ldots, k\}$ then $f \in C^0\text{-}PW*(G \bigcap (\bigcup_{i=1}^{k} N_{\sigma_i}(x^i)))$.

Proof: By induction.

The case $j = 2$ is theorem 3.10.

For $j \in 3, \ldots, k$ let ξ be sufficiently small to satisfy definition 3.1 for each neighbourhood, $N_{\sigma_1}(x^1), \ldots, N_{\sigma_j}(x^j)$. Modify the cells of $G \bigcap N_{\sigma_j}(x^j)$ by adding the constraints $d_{x^1}(x) \geq \sigma_1, \ldots, d_{x^{j-1}}(x) \geq \sigma_{j-1}$ to the cell boundary constraints of $G \bigcap N_{\sigma_j}(x^j)$. Use these cells together with the unmodified cells of $\bigcup_{i=1}^{j-1} (G \bigcap N_{\sigma_i}(x^i))$. □

3.12. Corollary

Given, an open set G, a set H of isolated points, and $\sigma > 0$, if f is $C^0\text{-}PW*$ over $N_\sigma(x) \bigcap G$, for every $x \in H$, then $f \in C^0\text{-}PW*(G \bigcap (\bigcup_{x \in H} N_\sigma(x)))$.

Proof: Because H is a set of isolated points, for any $x \in H$, $N_\sigma(x) \bigcap N_\sigma(y) \neq \emptyset$ for only a finite number of $y \in H$. Thus, we can give a set of cells and component functions for f over the entire set $G \bigcap (\bigcup_{x \in H} N_\sigma(x))$ by using a construction similar to that of corollary 3.11. The cells and component functions satisfy definition 3.1 with $\# = C^0$ and $\xi = \sigma$ for the entire set $G \bigcap (\bigcup_{x \in H} N_\sigma(x))$. Hence, f is C^0–$PW*$ over this set. □

Corollary 3.12 includes the special case $G \subseteq \bigcup_{x \in H} N_\sigma(x)$. If the collection $N_\sigma(x)$, $x \in H$ satisfies the hypothesis of corollary 3.12 , we can assert that $f \in \#$–$PW*(G)$ since $G \bigcap (\bigcup_{x \in H} N_\sigma(x)) = G$, in this case.

3.13. Corollary

If $f \notin C^0$–$PW*(G)$ for some open set G, then $f \notin C^0$–$PW*(N_\sigma(\hat{x}) \bigcap G)$ for any $\sigma > 0$ arbitrarily small and some $\hat{x} \in G$. (If $f \notin C^0$–$PW*(G)$, definition 3.1 cannot be satisfied for f over G because of the behaviour of f at *a point* in G).

Proof: We will prove that $f \in C^0$–$PW*(N_\sigma(x) \bigcap G)$ for every $x \in G$ and fixed $\sigma > 0 \implies f \in C^0$–$PW*(G)$.

Let H be a set of isolated points in G such that $G \subseteq \bigcup_{x \in H} N_\sigma(x)$. By corollary 3.12, $f \in C^0$–$PW*(N_\sigma(x) \bigcap G)$ for every $x \in H \implies f \in C^0$–$PW*(G)$. □

4. A Complete Characterization of Piecewise Continuous Functions

By theorem 3.7, functions which are C^0–PWC^0 over an open set G have the partition property at each $x \in G$. We now show that the converse statement is true. That is, we will show that the only functions with the partition property at each point in an open set G is the class C^0–$PWC^0(G)$.

4.1. Theorem

If f, defined over an open set G, has the partition property at each $x \in G$ then $f \in C^0$–$PWC^0(G)$.

Proof: We will prove that $f \notin C^0$–$PWC^0(G) \implies$ the partition property does not hold at each $x \in G$.

We have shown, in corollary 3.13, that $f \notin C^0$–$PWC^0(G) \implies$ there exists $x \in G$ such that $f \notin C^0$–$PWC^0(N_\sigma(x) \bigcap G)$ for any $\sigma > 0$ arbitrarily small. We now prove that such an x is a point at which the partition property does not hold.

Let $f \notin C^0$–$PWC^0(N_\sigma(x) \bigcap G)$ for any $\sigma > 0$. Assume that the partition property holds at x, with $\xi > 0$ satisfying definition 2.1. We now show that $f \in C^0$–$PWC^0(N_\xi(x) \bigcap G)$ by using (i) - (iv) of definition 2.1 to show that (i) - (iv) of definition 3.1 hold over $N_\xi(x) \bigcap G$ with $\#$ and $* = C^0$:

(i) Because the F_i, $i \in I(x)$ are disjoint and $N_\xi(x) \subset \bigcup_{i \in I(x)} F_i$ each $y \in N_\xi(x)$ belongs to exactly one F_i, $i \in I(x)$.

(ii) $N_\xi(x) \subset \bigcup_{i \in I(x)} F_i$, where $I(x)$ is a finite index set $\Longrightarrow$ $N_\xi(x) \bigcap G$ contains points of only a finite number of the F_i, $i \in I(x)$.

(iii) and (iv)
are immediate from (iii) and (iv) of definition 2.1 when $G = N_\xi(x)$ and $\# = C^0$.

Thus, the assumption that the partition property holds at x such that $f \notin C^0\text{-}PWC^0(N_\sigma(x) \bigcap G)$ for arbitrary $\sigma > 0$ $\Longrightarrow$ $f \in C^0\text{-}PWC^0(N_\xi(x) \bigcap G)$ for some $\xi > 0$. □

4.2. Corollary

Let G_α, $\alpha \in A$ be an arbitrary collection of open sets such that $f \in C^0\text{-}PW*(G_\alpha)$ for every $\alpha \in A$. Then $f \in C^0\text{-}PW*(\bigcup_{\alpha \in A} G_\alpha)$.

Proof: Let $G = \bigcup_{\alpha \in A} G_\alpha$. Because f has the partition property with respect to G_α at every $x \in G_\alpha$ for all $\alpha \in A$ then f has the partition property with respect to G for all $x \in G$. Thus, by theorem 4.1 $f \in C^0\text{-}PWC^0(G)$.

Let F_i, $i \in I$ and f_i, $i \in I$ be cells and component functions respectively, that satisfy definition 3.1 for f over G. Since each $x \in G$ belongs to some G_α, $\alpha \in A$ such that $f \in C^0\text{-}PW*(G_\alpha)$, there exists $\xi > 0$ such that $f_i \in C^0 \cap *(N_\xi(x) \bigcap \overline{F}_i \bigcap G)$, $i \in I(x)$ for every x in G. Thus, $f_i \in C^0 \cap *(\overline{F}_i \bigcap G)$ for all $i \in I(x)$. □

5. Finding the Infimum of a Piecewise Continuous Function

This section concerns a conceptual algorithm for finding an unconstrained infimum of a $\#\text{-}PW*$ function. Lemma 2.7 and theorem 2.8 are the main underlying ideas.

Let $f \in \#\text{-}PW*(\mathrm{R}^n)$, with cells F_i, $i \in I$ and component functions f_i, $i \in I$. We define a collection of continous nonlinear programming problems:

5.1. Definition

For each $i \in I$ define

$$NLP_i \equiv \begin{cases} \min & f_i(x) \\ \text{subject to} & x \in \overline{F}_i \end{cases}$$

5.2. A Conceptual Algorithm

From theorem 2.8 we have that $\hat{x}$ is a local infimum of f if and only if $\hat{x}$ is a constrained local minimum for each NLP_i, $i \in \underline{I}(\hat{x})$.

Lemma 2.7 states that if given $\hat{x}$, another feasible point y can be found for NLP_j, $j \in I(\hat{x})$ such that $f_j(y) < f_j(\hat{x})$, then $\underline{f}(y) < \underline{f}(\hat{x})$.

This suggests a conceptual way of minimizing $\underline{f}$:

(0) Choose a starting point x^0 and let k be such that $x^0 \in \overline{F}_k$. Set $j \leftarrow 1$.

(1) Find a local minimizer x^j for NLP_k and set $b \leftarrow f_k(x^j)$.

(2) Scan $I(x^j)$ for a different subproblem NLP_l with $f_l(x^j) \leq b$ and such that x^j is not a local minimum for NLP_l .

(3) If no such subproblem exists $STOP$; otherwise set $k \leftarrow l$, $j \leftarrow j + 1$ and go to step (1). □

5.3. Example

Suppose we were to apply the conceptual algorithm to f of example 2.2, starting at $x^0 = 10$ where $I(x^0) = \{1\}$.

Step(1) will bring us either to $x^1 \in \{\pm \frac{1}{\pi}, \pm \frac{1}{3\pi}, \pm \frac{1}{5\pi}, \pm \frac{1}{7\pi}, \ldots, \pm \frac{1}{99\pi}\}$ where $b \leftarrow (x^1)^2$ and $I(x^1) = \{-1, 1, 2\}$, or to $x^1 \in F_0$ where $b \leftarrow 0$ and $I(x^1) = \{0\}$ or $\{0, -1\}$.

No suitable descent subproblem will be found at *step*(2). Hence, the algorithm will terminate with x^1. The infimum of f is $\underline{f}(x^1)$ which is 0, if $x^1 \in F_0$ or $\frac{(x^1)^2}{2}$, otherwise. □

5.4. Assumptions

In what follows, we assume that:

(i) $\{x^j\}$ $j = 0,1,...$ is generated by the conceptual algorithm, and $\{x^j\} \in S$ where $S \subset G$ is a compact set.

(ii) for each subproblem, NLP_i, $i \in I$, such that $F_i \bigcap S \neq \emptyset$, and for some finite integer l, there is a finite set of objective function values in the range $(-\infty, f(x^l))$ for which constrained local minima exist. □

5.5. Theorem: Convergence of the Conceptual Algorithm

For an open set $G \subseteq \mathrm{R}^n$ and a function $f \in C^0\text{–}PWC^0(G)$, the conceptual algorithm converges to a local infimum of f given an arbitrary starting point in G provided the assumptions 5.4 are satisfied.

Proof: $I(x^j)$ is finite for each x^j. The result follows from the observation that if x^j is not optimal for NLP_k then at *step*(1) we will find x^{j+1} with $f(x^{j+1}) < f(x^j)$. Thus, b is always reduced at *step*(1). Since there are a finite number of subproblems NLP_k such that $F_k \bigcap S \neq \emptyset$ (by Corollary 3.9) and since each of these subproblems has only a finite set of optimal function values, *step*(1) can be executed only a finite number of times. Therefore, for some j sufficiently large, no suitable subproblem will be found at *step*(2) and thus the algorithm will terminate at *step*(3). When termination occurs, theorem 2.8 ensures that x^j is a local minimizer of $\underline{f}$. □

5.6. Deriving a Practicable Algorithm

Step(1) of the conceptual algorithm will, in general, require an algorithm that generates an infinite sequence of points; hence the title *conceptual*. Intuitively, it seems that in some cases it should be possible to find a local infimum for f by solving subproblems at *step*(1) to within, possibly very loosely, relaxed tolerances. Furthermore, at *step*(2), rather than choosing the next subproblem from among only those problems NLP_l, $l \in I(x^j)$, $l \neq k$ such that $f_l(x^j) \leq b$, we could consider problems NLP_l such that $l \in I(x^j)$, $l \neq k$, and $f_l(x^j) \leq b + \delta$. The relaxed tolerances could be tightened whenever an approximate stationary point is found, until, after a finite number of steps, a subproblem giving the lowest bound on f could be solved to convergence. If the final subproblem were chosen carefully enough among all subproblems with feasible sets "near" x^j, and if the other subproblems in that neighbourhood were sufficiently well-behaved, it is not unreasonable to expect that by solving the chosen subproblem to convergence, we will locate a local infimum of f.

We remark that the numerical stability of a *practicable* algorithm based on these considerations would depend only on the numerical stability of the algorithm applied to the subproblems.

For a practicable algorithm, therefore, we will need a set of approximate optimality conditions and an algorithm for the subproblems that can find a point which satisfies these conditions after a finite number of steps. In the conceptual case, x^j is a feasible solution for all subproblems indexed by $I(x^j)$; in the practicable case, it will be necessary to consider all subproblems whose feasible sets intersect a neighbourhood of x^j. The numerical algorithm proposed in [3] follows this approach.

5.7. Example

We will illustrate how in principle, the practicable approach is applied to the function f in example 2.2. An actual realization of this approach is considerably more sophisticated.

A point $\hat{x}$ satisfies the *relaxed feasibility criterion* for subproblem NLP_i, $i \in I$ if $\hat{x} \in \bar{F}_i^\epsilon = \{\, x \mid c^l(x) \leq \epsilon,\, l \in L_i;\ |c^e(x)| \leq \epsilon,\, e \in E_i;\ c^g(x) \geq -\epsilon,\, g \in G_i \,\}$ where $\epsilon > 0$ is a given tolerance.

Cell boundary constraint j is *active* at $\hat{x}$ if $|c^j(\hat{x})| \leq \epsilon$, $j \in E_i \bigcup L_i \bigcup G_i$, $i \in I$.

If the j-th cell boundary constraint is active at $\hat{x}$, then $\hat{x}$ satisfies the *relaxed optimality* criterion for NLP_i, $i \in I$ if $\hat{x}$ satisfies the relaxed feasibility criterion for NLP_i and $\nabla f_i(\hat{x}) = \lambda \nabla c^j(\hat{x})$ where $\lambda \geq -\gamma$ if $c^j \in G_i$, $\lambda \leq \gamma$ if $c^j \in L_i$, and $\gamma > 0$ is a given tolerance. If $\hat{x}$ is feasible for NLP_i, $i \in I$, and no cell boundary constraints are active at $\hat{x}$, then $\hat{x}$ satisfies the *relaxed optimality* criterion for NLP_i if $|\nabla f_i(\hat{x})| \leq \epsilon$.

For the purpose of illustration, we will use the following *bisection* method to find an approximate local minimum for NLP_i, $i \in I$.

(0) Input starting point x^j and initial stepsize α_0. Set $\alpha \leftarrow \alpha_0$.

(1) Set $d \leftarrow sign\,(\nabla f_i(x^j))$.
If $f_i(x^j + \alpha d)$ satisfies the relaxed feasibility criterion for NLP_i and $f_i(x^j + \alpha d) < f_i(x^j)$ go to step (3).

(2) Set $\alpha \leftarrow \alpha/2$ and go to step (1).

(3) Set $x^{j+1} \leftarrow x^j + \alpha, \ j \leftarrow j + 1$.
If x^j satifies the relaxed optimality criterion for NLP_i, stop;
otherwise go to step (1). □

The following is a statement of a first-order-type *practicable algorithm* that one could apply to f in this particular case. $LIST$ is a list of subproblems that remain to be considered. m is the index of the subproblem which gives the best bound on the infimum of f. N is the maximum number of times the tolerances are allowed to be be reduced before subproblem m is solved to convergence.

(0) Input $N, \epsilon, \delta, \gamma, \alpha_0, x^0$
Set $j \leftarrow 0, b \leftarrow \infty$

(1) $LIST \leftarrow \{ i \mid x^j \in \overline{F}_i^\epsilon, i \in I \}$

(2) If $LIST = \emptyset$ go to step (7)
Select $k \in LIST$.
If $f_k(x^j) > b + \delta$ go to step (5)
If $f_k(x^j) < b$ set $b \leftarrow f_k(x^j)$ and $m \leftarrow k$.

(3) Use the bisection algorithm to find x^{j+1} that satisfies the relaxed optimality criterion for NLP_k.

(4) If $f_k(x^{j+1}) < b$ go to step (6).

(5) Delete k from $LIST$ and go to step (2)

(6) Set $b \leftarrow f_k(x^{j+1}), m \leftarrow k, j \leftarrow j + 1$ and go to step (2).

(7) Set $i \leftarrow i + 1$. If $i > N$ go to step (8).
Set $\epsilon \leftarrow \epsilon/4, \delta \leftarrow \delta/4, \gamma \leftarrow \gamma/4$.
If $x^j \notin \overline{F}_m^\epsilon$, use bisection algorithm to find $x^{j+1} \in \overline{F}_m^\epsilon$,
then set $x^{j+1} \leftarrow x^j$.
Go to step (1)

(8) Solve NLP_m to convergence. □

Using $x^0 = 10$, $\epsilon = \gamma = .4$, $\delta = .04$ and $N = 4$, the above naive method was carried out with $\alpha_0 = \pi, 1, \frac{1}{3}, \frac{1}{4}$.

In the case $\alpha_0 = 1$, $x^1 = 0.$; no reduction is made in the tolerances. For $\alpha_0 = \frac{1}{3}$, $x^1 = 0$; again, no reduction was made in the tolerances. When $\alpha_0 = \frac{1}{4}$, the tolerances are reduced at $x^1 = .375$; $x^2 = 0$. In the case $\alpha_0 = \pi$, the tolerances are reduced at $x^1 = .331$, $x^2 = -.0139$, $x^3 = .0023$ and $x^5 = -.0056$. $x^5 \in \overline{F}_0^\epsilon$. Thus, since $f_0(x^5) = 0$, $m \leftarrow 0$ and NLP_0 will be solved to convergence. The method will converge to some $x \in F_0$. In all four cases, the global infimum, $\underline{f}(x) = 0$, is found.

6. Conclusions

We have given an approach to nondifferentiable and noncontinuous optimization that is based on the principle of decomposition into smooth subproblems. Furthermore, we have characterized the functions to which this approach applies. A conceptual algorithm that converges to a local infimum under weak assumptions motivates the basic strategy underlying the numerical method in [3].

7. Acknowledgements

The authors would like to thank N.I.M. Gould for his careful reading of this manuscript and for his constructive comments.

References

[1] T.M. Apostol, Mathematcal analysis a modern approach to advanced calculus (Addison-Wesley, Reading, MA, 1964).

[2] T.F. Coleman and A.R. Conn, "Second order conditions for an exact penalty function", *Mathematical Programming,* 19(1980)178-185.

[3] A.R. Conn and P.F. O'Neill, "A numerical algorithm for piecewise differentiable and piecewise continuous optimization", in preparation.

[4] G. Papavassilopoulos, "Algorithms for a class of nondifferentiable problems", *Journal of Optimization Theorey and Applications* 34(1981)41-82.