GENERALIZED PENALTY-FUNCTION CONCEPTS IN MATHEMATICAL OPTIMIZATION

M. Bellmore

*The Johns Hopkins University, Baltimore, Maryland*

H. J. Greenberg

*Southern Methodist University, Dallas, Texas*

and

J. J. Jarvis

*Georgia Institute of Technology, Atlanta, Georgia*

(Received June 17, 1968)

Given a mathematical program, this paper constructs an alternate problem with its feasibility region a superset of the original mathematical program. The objective function of this new problem is constructed so that a penalty is imposed for solutions outside the original feasibility region. One attempts to choose an objective function that makes the optimal solutions to the new problem the same as the optimal solutions to the original mathematical program.

**Consider** the mathematical program:

\[
\begin{align*}
P1: \text{Maximize } & f(x) \text{ subject to } \\
g_i(x) & \leq b_i \text{ for } i = 1, \ldots, k, \\
g_i(x) & = b_i \text{ for } i = k + 1, \ldots, m, \text{ and } x \in X \text{ (a subset of } \mathbb{R}^n). 
\end{align*}
\]

**Definition 1.** The feasibility region \((FR)\) for \(P1\) is the set

\[
FR = \{x : g_i(x) \leq b_i \text{ for } i = 1, \ldots, k \text{ and } g_i(x) = b_i \text{ for } i = k + 1, \ldots, m\} \cap X.
\]

**Definition 2.** The optimality region \((OR)\) for \(P1\) is the set

\[
OR = \{x^* : f(x^*) \geq f(x) \text{ for all } x \in FR\} \cap FR.
\]

**Definition 3.** A solution to \(P1\) is: (1) Ascertain that \(OR\) is empty because (a) \(f(x)\) is unbounded over \(FR\), (b) the supremum cannot be obtained, or (c) \(FR\) is empty (i.e., there is no feasible solution). Or, (2) find at least one point in \(OR\) (and determine if it is unique, this uniqueness determination being optional).
We generally consider closed feasibility regions with an interior. Further, if the mathematical program models an economic problem, then there will be at least one point in the optimality region. However, this assumption is not necessary to all of our discussion.

At times we shall be concerned with special classes of P1 problems. Although our discussion applies to any mathematical program, we will point out some of the advantages that accrue when certain properties are known.

**PROPERTY 1.** P1 is a convex program if (1) \( f(x) \) is concave over \( FR \), (2) \( g_i(x) \) is convex for \( i = 1, \ldots, k \), (3) \( g_i(x) \) is linear for \( i = k+1, \ldots, m \), and (4) \( X \) is a closed convex set.

**PROPERTY 2.** P1 is separable if \( f(x) = \sum f_i(x_i) \) and \( FR = FR_1 \times FR_2 \), where \( x \in FR \) if and only if \( x_1 \in FR_1 \) and \( x_2 \in FR_2 \). The symbol \( \otimes \) denotes a 'Mitten operator' (see Definition 4).

**DEFINITION 4.** A Mitten operator satisfies the monotonicity relations (1) \( x \geq y \Rightarrow x \otimes z \geq y \otimes z \) and (2) \( x \geq y \Rightarrow z \otimes x \geq z \otimes y \) for all \( x, y, z \) in a specified set.

The advantage that accrues when Property 1 is satisfied is deferred until we discuss a related penalty function problem. Mitten operators will also be discussed later. However, we can immediately see an advantage of Property 2: it can easily be shown that \( OR = OR_1 \times OR_2 \), where

\[
OR_1 = \{ x_1^* : f(x_1^*) \geq f(x_1) \} \text{ for all } x_1 \text{ in } FR_1 \Lambda FR_1
\]

and

\[
OR_2 = \{ x_2^* : f(x_2^*) \geq f(x_2) \} \text{ for all } x_2 \text{ in } FR_2 \Lambda FR_2.
\]

In other words, we solve two reduced P1 problems as (1) maximize \( f_1(x_1) \) subject to \( x_1 \) in \( FR_1 \) and (2) maximize \( f_2(x_2) \) subject to \( x_2 \) in \( FR_2 \).

Now consider the mathematical program

\[
P_2: \text{Maximize } F[f(x), g(x)] \text{ subject to } x \text{ in } X.
\]

The nature of the function \( F \) is to penalize large values of \( g_i(x) \) for \( i \leq k \) and large values of \( |g_i(x)| \) for \( i > k \). The choice of \( F \) (the penalty function) is very general, but from a practical standpoint should be continuous in \( f(x) \) and \( g(x) \) in order to solve the P2 problem without augmenting more constraints. There are four basic questions of interest:

1. What is the relation of P2 to P1?
2. Does there exist a function \( F \) such that the solution to P2 solves P1?
3. Can P2 be solved more easily than P1?
4. Given that a function \( F \) exists, is there an efficient way to find it?

Clearly one can 'force' the solution to P2 to be the same as P1 if the solution is known (we allow a discontinuous penalty function at this point). Hence, question (2), existence, is trivial in this context. However, question (4) raises the important aspect of search. We shall therefore use
parameters in order to achieve convergent search schemes. That is, a class of penalty functions may be specified and parameters used to vary within that class. Consider the problem

$$P3: \text{Maximize } F[f(x), g(x); y] \text{ subject to } x \text{ in } X,$$

where $y$ is chosen from the parameter space $Y$ and $F$ is specified.

The analogous four questions of interest are:

(1) What is the relation of $P3$ to $P1$?
(2) Does there exist a parameter $y$ in $Y$ such that the solution to $P3$ solves $P1$?
(3) Can $P3$ be solved more easily than $P1$?
(4) Given that a parameter $y$ in $Y$ exists, is there an efficient way to find it?

Question (2) in this new list is no longer trivial.

Before discussing each question individually, let us consider a well known class of parametric penalty functions known as Lagrangians. The usual Lagrangian is

$$F = f(x) - \sum_{i=1}^{m} y_i g_i(x)$$

and $Y = \{y: y_i \geq 0 \text{ for } i = 1, \ldots, k\}$.

**Kuhn-Tucker** theory may be placed in this framework. Their saddle-point theorem answers question (1) for the class of $P1$ problems they considered (i.e., which satisfies their qualifications: all convex programs with $X = E^n$ and $FR$ having an interior are among these—see also W. P. Pierskalla, where monotonic constraint functions are considered). The existence question in (2) is also answered. Arrow, Hurwicz, and Uzawa give a search procedure for finding $x^*$ and $y^*$ that provides an answer for question (4) with respect to convex programs.

The use of the Lagrangian was extended to arbitrary $P1$ problems by H. Everett. He partially answered question (1) by proving that when $P3$ is solved for $y = y^*$, $P1$ is solved for $b = g(x^*)$.

Everett addressed question (2) by demonstrating that not every $b$ vector can be obtained no matter how $y$ is chosen from $Y$. These values of $b$ Everett termed 'gaps.' We generalize this in our discussion of question (2).

Brooks and Geoffrion showed that, if $y$ exists, then $(x^*, y^*)$ is a saddle point of the Lagrangian. They also weakened Everett’s main theorem so that if $y_i = 0$ for $i = 1, \ldots, u$ (where $u \leq k$), then any $P1$ defined with $b_i \geq g_i(x^*)$ for $i = 1, \ldots, u$ and $b_i = g_i(x^*)$ for $i = u + 1, \ldots, m$ has the solution $x^*$. This agrees with our intuition, since $y_i = 0$ often means that the constraint $g_i(x) \leq b_i$ is not binding (i.e., if removed from $P1$, $OR$ remains the same). Brooks and Geoffrion also showed a method of searching $y$ for Lagrangians. This is reviewed in our discussion of question (4).
The nonparametric Lagrangian is defined as
\[ F = f(x) - \sum_{i=1}^{\infty} G_i(x), \]
where \( G_i \) is monotone nondecreasing in \( g_i(x) \) for \( i = 1, \ldots, k \).

Gould proved the existence of \( G_i \) such that the solution to \( P2 \) solves \( P1 \) by allowing discontinuous functions. He also proved that the selection \( G_i \) from the class of monotone functions stated has the minimizing property of a saddle point. That is, \( (x^*, G^*) \) is a saddle point of \( F \). Our work is principally directed toward parametric penalty functions, but the reader is referred to Gould for further reading in nonparametric theory.

We define a class of parametric penalty functions as Lagrangians if
\[ F = G_0[f(x)] - \sum_{i=1}^{m} y_i G_i[g_i(x)], \]
where
- \( G_0 \) is continuous and monotone increasing in \( f(x) \),
- \( G_i \) is continuous and monotone nondecreasing in \( g_i(x) \) for \( i = 1, \ldots, k \),
- \( G_i \) is continuous for \( i = k+1, \ldots, m \),
- \( Y = \{ y : y_i \geq 0 \text{ for } i = 1, \ldots, k \} \cap E^m \).

Nunn has reported that one may transform the \( P1 \) problem from \( g_i(x) \leq b_i \) to \( -G_i[g_i(x)] \leq -G_i(b_i) = b_i' \), where \( G_i \) is monotone decreasing, without changing the feasibility region. In addition, \( \max G_0[f(x)] = G_0[\max f(x)] \), so that the Lagrangians are really the usual Lagrangians of a transformed \( P1 \) problem. The motivation for doing this is to eliminate gap regions.

**Example.** Consider the problem \( P1: \) maximize \( f(x) = \sum_{i=1}^{m} x_i^2 \) subject to \( x \leq b \). Using the Lagrangian, we have \( P3: \) maximize \( F = f(x) - y_g(x) = \sum_{i=1}^{m} [x_i^2 - y_i x_i] \) subject to \( x \geq 0 \). The \( P3 \) problem is unbounded so that all values of \( b \) lie in an Everett gap. Now consider another Lagrangian as \( P3': \) maximize \( F' = f(x) - y e^{\lambda(x)} = \sum_{i=1}^{m} x_i^2 - y \exp(\sum_{i=1}^{m} x_i) \), subject to \( x \geq 0 \). For \( y = 2b \exp(-b) \), we obtain the solution as \( x_j^* = b \) for any \( j \) and \( x_i^* = 0 \) for \( i \neq j \). We thus have the solution for all values of \( b \) of interest (i.e., positive).

The point here is that the weighted difference of \( f(x) \) and \( g(x) \) is insufficient penalty, since \( f(x) \) will always ‘outweigh’ \( g(x) \) with linear weighting.

Moreover, a restricted Lagrangian is a Lagrangian with its parameter space \( Y \) restricted. For example, Carroll used the penalty function
\[ F = f(x) + r \sum_{i=1}^{m} 1/[|g_i(x) - b_i|], \]
where \( r \geq 0 \) and \( m = k \) (i.e., no equality constraints). This is a restricted Lagrangian with \( y_i = r \) for \( i = 1, \ldots, m \). Fiacco and McCormick generalized and formalized this into what is known as SUMT.

A more general class of penalty functions of interest is a monotone penalty
function. It is defined as one in which \( F \) is monotone increasing in \( f(x) \) and nonincreasing in \( g_i(x) \) for \( i = 1, \ldots, k \).

A **separable-monotone** penalty function is of the form

\[
F = G_0[f(x)] \bigcirc G_1[g_1(x)] \bigcirc \cdots \bigcirc G_m[g_m(x)],
\]

where \( F \) is monotone and \( \bigcirc \) is a Mitten operator. Clearly, any Lagrangian is separable-monotone.

We shall now discuss each question individually. In each case, we are unable to answer the question completely, but some theorems are presented that address classes of penalty functions in an effort to add insight.

**Question 1: What is the Relation of \( P_3 \) to \( P_1 \)?**

**Theorem 1.1.** Let \( F \) be a monotone penalty function for a given \( y \). Let \( x^* \) be an optimal solution to \( P_3 \). Then, \( x^* \) solves \( P_1 \) for \( b = g(x^*) \).

The proof of Theorem 1.1 is in the appendix. It basically follows the proof of Everett for the Lagrangian. We can weaken Theorem 1.1 with the following:

**Corollary 1.1.** If \( F[f(x), g(x); y] = C \) (a constant) implies \( F[f(x), g(x) + \epsilon e_i; y] = C \) for all \( \epsilon \), where \( i \leq k \), then the solution to \( P_3 \) (viz., \( x^* \)) solves \( P_1 \) for \( b_j = g_j(x^*) \) for \( j \neq i \) and \( b_i \geq g_i(x^*) \).

The vector \( e_i \) has components each 0 except the \( i \)th component, which is 1.

**Corollary 1.1.** (see the appendix for a proof) extends Brooks and Geoffrion's weakening of Everett's theorem. Basically, if \( F \) is invariant with \( g_i(x) \) (as the hypothesis states), then no penalty is placed upon it so that it cannot bind our solution. Therefore, it is as though the constraint were not present, and any \( b_i \) that allows \( x^* \) to be feasible (for \( P_1 \)) is acceptable.

We thus have a relation of \( P_3 \) to \( P_1 \) for all monotone penalty functions. Now we wish to prove some bounding theorems for a more general class of penalty functions.

**Theorem 1.2.** Let \( F \) be monotone increasing in \( f(x) \) and such that a contour, say \( F = K \) (a constant), implies a continuous function as \( f(x) = Q_x[g(x); y] \). Further, let \( Q^* \) be that function corresponding to \( F = F^* \) (the maximum in \( P_3 \)). Then, for all \( x \) we have \( f(x) \leq Q^*[g(x); y] \).

For \( f(x) = Q_x[g(x); y] \), see Courant,[6] page 114, for sufficient conditions (viz., the implicit function theorem). Note that any Lagrangian satisfies these conditions.

We prove Theorem 1.2 in the appendix. Geometrically, it says that all points in the space \( f(x) \) vs. \( g(x) \) lie below this contour resulting from the maximization. If we let \( f^*(b) \) be \( f(x^*) \), where \( x^* \) solves \( P_1 \), then we have a point of \( f^*(b) \) for some \( b \) and an upper bound for all other \( b \).
As we solve more P3 problems, these bounds may be strengthened. Corollary 1.2.1 addresses itself to this strengthening.

**Corollary 1.2.1.** Let \( F_1, \ldots, F_s \) be \( s \) penalty functions satisfying the conditions of Theorem 1.2 with maxima \( F_1^*, \ldots, F_s^* \). Corresponding to each \( F_i^* \) there is a \( Q_i^* \), and
\[
f(x) \leq \min_{i=1}^s Q_i^*[g(x); y].
\]

The proof of Corollary 1.2.1 follows directly from Theorem 1.2 and is omitted.

We thus have related P3 to P1 by showing that we are approximating \( f^*(b) \). (The usual Lagrangian is a linear approximation.) Each solution to P3 gives an exact point on \( f^*(b) \) and upper bounds for all other \( b \).

**Question 2: Does There Exist a \( y \) in \( Y \) for P3 to Solve P1?**

Suppose all functions are differentiable. A necessary condition for \( x^* \) to solve P3 is then
\[
F_i \nabla f + \sum_{i=1}^m F_i \nabla g_i = 0,
\]
if \( x^* \) is in the interior of \( X \).

Therefore, a first necessary condition for \( x^* \) to solve P1 is that there exist a \( y \) in \( Y \) such that the above equations hold at the solution to P1. For the Lagrangian, this is the usual Lagrange regularity, which usually requires some constraint qualification on P1.

In this section we seek to be more specific and generalize the gap concept described by Everett.\(^8\)

**Theorem 2.1.** Let \( F \) be a Lagrangian as
\[
F = G_0[f(x)] - \sum_{i=1}^m y_i G_i[g_i(x)].
\]

Let problem P3 for \( y = y^* \) have a finite (or denumerable) number of solutions, say \( x_1^*, \ldots, x_t^* \), where \( t \geq 1 \). Let the set \( B \) be defined as
\[
B = \{ b : G_i(b) = \sum_{i=1}^m w_i G_i[g_i(x_i^*)], \quad i = 1, \ldots, m; \quad \sum_{i=1}^m w_i = 1 \text{ and } w \geq 0 \}.
\]

Further, let \( B^- \) be defined as
\[
B^- = B - \{ b : b = g(x_j^*) \text{ for } j = 1, \ldots, t \}.
\]

Then, there exists no value of \( y \) in \( Y \) with a solution \( x^{**} \) such that \( g(x^{**}) \) is in \( B^- \).

This theorem is proved in the appendix. Basically, it says that, if several alternate maxima exist for P3, then the values of \( b \) that can be expressed as a convex combination of those generated cannot be obtained (except, of course, those that have just been generated). Figure 1 depicts an Everett gap in the one-constraint case.
There are several corollaries of interest whose proofs follow directly from Theorem 2.1.

**Corollary 2.1.1.** A gap arises if some choice of \( y \) in \( Y \) produces at least two solutions (and at most a denumerable number) to \( P3 \) with distinct \( g(x^*) \).

**Corollary 2.1.2.** If \( f^*(b) \) is discontinuous at \( b^* \), then there exists \( d > 0 \) such that all \( b \) in the range \( b^* - d < b < b^* \) lie in a gap—i.e., no values of \( y \) in \( Y \) have yet obtained all values of \( b \) in some left-sided neighborhood of \( b^* \).

---

**Fig. 1. Illustration of an Everett gap and its relation to Theorem 2.1.**

Corollary 2.1.1 states a consequence that would help in anticipating gap regions. Corollary 2.1.2 is illustrated by Fig. 2. An important class of \( P1 \) problems that have such discontinuities is discrete optimization—i.e., \( X \) is a subset of integers (we thank G. Burch and W. Nunn for pointing the discontinuous case out to us). Although there will always be gaps theoretically, from a practical view some of these are not of interest. For example, consider a linear constraint of the form \( \sum_{i=1}^{n} a_i x_i \leq b \), where \( a_1, \ldots, a_n \) are integers. The only values of \( b \) of interest are those that are multiples of the greatest common divisor of \( a_1, \ldots, a_n \).

---

**Fig. 2. A discontinuity at \( b^* \).**

The more general question of the existence of \( y \) when \( X \) is a subset of integers requires more investigation. If \( P1 \) is a convex program and the integer constraint is not binding for some \( b \) of interest—i.e., the solution to \( P1 \) when \( x \) is allowed to be noninteger turns out to be integer—then the \( y \) exists for the Lagrangian penalty function (NEMHAUSER AND ULL-
Fig. 3. Unbounded Lagrangian.

\[ f(x) \]

\[ g(x) \]

\[ \nabla F \]

\[ \text{GAP} \]

MAN\textsuperscript{[21]} show the if-and-only-if case for the 0–1 linear integer program. When the integer constraint is binding, we cannot anticipate, in general, whether or not a gap will arise.

Gaps may arise even though no \( y \) in \( Y \) yields alternate optima to \( P3 \). One example of this arises when the penalty function is unbounded. Consider a Lagrangian where the \( P1 \) problem has a single constraint \( g(x) \leq b \). Suppose there are points for which \( g(x) \to -\infty \) and \( f(x) \) is finite. In this

\[ f(x) \]

\[ \text{slope} = y > 0 \]

\[ g(x) \]

\[ h(b) \]

\[ h(b) \text{ discontinuous} \]
case, for any \( y > 0 \), the penalty function is unbounded, and there is not a maximum (see Fig. 3).

A second example arises when \( h(b) \) has a discontinuity, where \( h(b) \) is the envelope bounding the points of \( f(x) \) versus \( g(x) \) from above. Again consider one constraint and the usual Lagrangian. Figure 4 shows a case in which a gap arises at \( b_0 \), yet there are no alternative maxima to \( P_3 \).

In addition, one might have a continuum of maxima to \( P_3 \) and either have a gap (see Fig. 5) or not have a gap (see Fig. 6).

However, the isolated point in Fig. 4 is pathological and does not arise in practice. If our \( P_3 \) is bounded, gaps tend to arise when alternative maxima exist for a given parameter. The gap then physically resembles a dip in the response surface where no support contour of \( F \) exists. This is the more common type of gap occurrence, and resolution of this type of gap can be termed 'gap filling.' We shall discuss methods of gap resolution heuristically. It is important to refer to Question 3 to realize that the approaches suggested here may be computationally infeasible at present.

We have seen that, from a practical view, there are two basic sources of gaps: the unboundedness of \( P_3 \) and alternative maxima to \( P_3 \). The former is more easily disposed of by the following rule of procedure. Let us suppose our objective in \( P_3 \) is of the form \( G_0[f(x)] + G[g(x)] \), which is identical to \( H_0(x) - H(x) \). Thus, selecting \( G_0 \) and \( H \) specifies \( H_0 \) and \( H \). Let \( H_0 \) and \( H \) be monotone increasing in at least one component of \( x \). In order that \( P_3 \) be bounded, it is therefore necessary to select \( G_0 \) and \( H \) (recall that \( G_0 \) and \( G \) are continuous) such that the order (see Courant, page 189) of \( H_0 \) is less than the order of \( H \). Intuitively, we are saying that the penalty must be 'sufficiently strong.'

Thus, the example in the introductory section has a gap of this type for the Lagrangian. This is completely resolved by following the above rule of procedure. It is recognized that this resolution may generate computational problems (see Question 3).

The second source of gaps is more common and requires more complicated methods to resolve. In theory, there is no problem, since penalty functions can be chosen to 'dip' into the gap region. For example, Fig. 1 shows a gap between \( g(x_1^*) (= b_1) \) and \( g(x_2^*) (= b_2) \). If we seek \( b: b_1 < b < b_2 \), then we have a gap using the usual Lagrangian.

Various alternatives are available (see Gould). One of these is to use a parabolic of the form

\[ F = f(x) - y_1[g(x) - b]^2 + y_2, \quad \text{where} \quad y_1 \geq 0, \quad y_2 \geq 1. \]

The effect of this procedure is to probe the gap region near \( b \) and obtain a nearer solution, or at least strengthen the upper bound on \( f^*(b) \). As \( y_1 \to \infty \) we force \( g(x^*) = b \). This, of course, may not be the solution if the envelope
is not monotone nondecreasing. Even so, the ability to solve this problem is lessened if certain properties (see Question 3) are not present.

If we choose $y_1 = 1$ and $y_1$ such that $[f^*(b_1), b_1]$ and $[f^*(b_1), b]$ are on the same contour (as pictured in Fig. 7, where, in regions A and B, the La-
grangian yields a gap, region A being resolved by the parabola, and region B having stronger bounds to the point denoted by C), then the gap is resolved in one step if \( f^*(b) = f^*(b_i) \). Otherwise, a nearer feasible solution and stronger bounds are obtained. This may be repeated until the gap is resolved.

In closing our discussion of Question 2, let us point to the future. A sophisticated penalty function program would use the techniques of artificial intelligence to select the penalty function \( F \) and the parameter space \( Y \). As the search progresses, learning would dictate an 'optimal' selection to generate as much information as possible about \( f^*(b) \). Some of the known results in search under uncertainty may be applied in this 'optimal selection.'

However, before concluding our comments on Question 2, we wish to consider a non-Lagrangian penalty function. Our purpose is to offer a gap theorem that may prove beneficial in determining methods of resolution.

We consider an exponential of the form given in the following:

**Theorem 2.2.** Let \( F(x; y) \) be given by

\[
F(x; y) = G_0[f(x)] - \prod_{i=1}^m G_i[g_i(x)]^{y_i},
\]

where \( y \) is in \( Y = \{y : y \geq 0\} \). We assume \( G_0 \) and \( G_i \) are monotone increasing, and \( G_i[g_i(x) \geq 1] \) for \( i = 1, \cdots, m \). Then, if \( F(x; y) \) has alternate maxima \( x_1^*, \cdots, x_i^* \), there exists no \( y \) in \( Y \) such that

![Fig. 7. A parabolic gap filler.](image-url)
\[ G_i[g_i(x^*)] = \prod_{i \in J} G_i[g_i(x^*_i)]^{v_i} \]
for \( i = 1, \ldots, m \) where \( v \geq 0 \) and \( \sum_{i \in J} v_i = 1 \),
where \( J \) is the set of indices \( j \) satisfying
\[
\prod_{i \in J} G_i[g_i(x^*_i)]^{v_i} \geq \prod_{i \in J} G_i[g_i(x^*_i)]^{v_i}.\]

The proof of Theorem 2.2 is in the appendix. Note the similarity to Theorem 2.1 where \( F \) is linear over \( y \) (the Lagrangian). For the Lagrangian any arithmetic mean of the \( g(x,*) \) is in a gap; when \( F \) is geometric over \( y \), any geometric mean of some of the \( g(x,*) \) is in a gap. Of course, the term ‘gap’ is used here in the Everett sense. For \( y = 0 \), Corollary 1.1.1 may be invoked to reduce the gap region.

**Question 3:** Can \( P3 \) Be Solved More Easily than \( P1 \)?

The very nature of the question suggests that a qualitative discussion is presented. The authors wish to stress two desirable properties and cite some classes of \( P1 \) problems that would benefit from them. The two properties of interest are: (1) \( F \) is concave in \( x \) over \( X \), and (2) \( F \) is separable in \( X \) over \( X \), where \( X \) can be expressed as a cartesian product of the partitioned vector, \( (x_1, x_2) \).

Besides eliminating the gap problem pointed out in Question 2, the ability to solve \( P3 \) is enhanced when \( F \) is concave. First, the local maximum is also the global maximum. Second, schemes presently exist to maximize a concave objective when \( X \) is convex.

Moreover, even if \( F \) is not concave in \( x \), it is advantageous to have \( F \) unimodal in \( x \) over \( X \). For example, consider the \( P1 \) problem as: Maximize \( f(x) = \sum_{i=1}^n x_i^2 \), subject to \( g(x) = \sum_{i=1}^n x_i \leq b \) and \( x \geq 0 \). The objective \( f(x) \) is not concave. Suppose we use a monotone-separable penalty function as:

\[ P3: \text{Maximize } F = \min \left[ f(x), \frac{y}{g(x)} \right] \]
subject to \( x \geq 0 \), where \( Y = \{y : y \geq 0\} \).

Since \( f \) and \( g \) are both monotone increasing in \( x \), the solution to \( P3 \) satisfies \( f(x^*) = y/g(x^*) \). Therefore, if \( y = b \), then \( x_j^* = b \) for any \( j \) and \( x_i^* = 0 \) for \( i \neq j \). We thus have the solution to \( P1 \) for all nonnegative values of \( b \) (note that for \( b < 0 \) there is no feasible solution).

The Lagrangian preserves concavity properties in that, if \( P1 \) is a convex program, then \( P3 \) will have a concave objective. An example to illustrate the advantage of this is the quadratic program. Let \( P1 \) be: maximize \( -\frac{1}{2}x' D x + cx \) subject to \( Ax \leq b \) and \( X = E^n \), where \( D \) is positive definite and symmetric (there is no loss in generality from assuming \( D \) symmetric—\( d_{ij} = d_{ji} \), then replace \( D \) with \( D' \) given by \( d'_{ij} = (d_{ij} + d_{ji})/2 \)). Then the \( P3 \) problem is: maximize \( \left( -\frac{1}{2} \right) x' D x + (c - yA)x \), where \( Y = \{y : y \geq 0\} \).
Necessary and sufficient conditions for \( x^* \) to be the solution to \( P_3 \) are that all first derivatives vanish. Hence, we can find \( x^* \) by solving the equations given by \( x^* = (c - yA)D^{-1} \). Thus, solving \( P_3 \) when \( y \) is specified reduces to performing these matrix operations.

The second property of interest is separability. To illustrate its advantage, consider the above quadratic program when \( D \) is the identity matrix. The solution to \( P_3 \) is then \( x^* = c - yA \), or \( x_j^* = c_j - ya_j \) for \( j = 1 \ldots, n \), where \( a_j \) is the \( j \)th column of \( A \).

If \( X \) restricts \( x \) to be nonnegative, we can trivially extend this to be \( x_j^* = \max (0, c_j - ya_j) \). One can verify that \( x^* \) given above is the solution by checking the Kuhn-Tucker conditions.\(^{[18]}\)

Suppose further that \( X \) restricts \( x \) to be integer. The cell property is such that we solve \( n \) univariate integer objectives. For a unimodal univariate function, the optimal integer solution is one of the neighboring integers of the optimal continuous solution. Hence, only \( 2n \) comparisons need be made (rather than \( 2^n \)).

In general, if the penalty function is Mitten-separable—i.e., if \( F = H_1(x_1) \bigcirc H_2(x_2) \), where \( \bigcirc \) is a Mitten group operator, and \( X = X_1 \times X_2 \) where \( x \) is in \( X \) if and only if \( x_1 \) is in \( X_1 \) and \( x_2 \) is in \( X_2 \); then \( P_3 \) may be solved as two separate problems given by:

\[
\begin{align*}
P_3(1): & \quad \text{maximize } H_1(x_1), \quad \text{subject to } x_1 \in X_1; \\
P_3(2): & \quad \text{maximize } H_2(x_2), \quad \text{subject to } x_2 \in X_2.
\end{align*}
\]

Consider the Lagrangian applied to a \( P_1 \) problem that is sum-separable so that we have:

\[
\begin{align*}
P_1: & \quad \text{maximize } \sum_{i=1}^{n} f_i(x_i), \\
& \quad \text{subject to } \sum_{i=1}^{n} g_{i_j}(x_i) \leq b_j \quad \text{for } j = 1, \ldots, k, \\
& \quad \sum_{i=1}^{n} g_{i_j}(x_i) = b_j \quad \text{for } j = k+1, \ldots, m, \\
& \quad x_i \text{ in } X_i \quad \text{for } i = 1, \ldots, n.
\end{align*}
\]

The \( P_3 \) problem is

\[
\begin{align*}
P_3: & \quad \text{maximize } \sum_{i=1}^{n} h_i(x_i; y), \\
& \quad \text{subject to } x_i \text{ in } X_i \text{ for } i = 1, \ldots, n,
\end{align*}
\]

where \( h_i(x_i; y) = f_i(x_i) - \sum_{j=1}^{m} g_{i_j}(x_i) y_j \).

Everett terms this a 'cell problem,' where each \( h(x_i; y) \) is a cell.

Instead of solving \( P_3 \) for a given \( y \) and repeating this until \( P_1 \) is solved (or a gap is reached), we can use recursion as follows: Let \( Q_i(y^i) \) be defined as \( Q_i(y^i) = \sum_{i=1}^{n} \max h_i(x_i; y^i) \). We thus have the recursion \( Q_i(y^i) = Q_{i-1}(y^i) + \max h_i(x_i; y^i) \). For each value of \( y^i \) there is an associated \( s^{(i)} \) such that \( (x_1^*, \ldots, x_i^*) \) is the solution to the subproblem:
Maximize \[ \sum_{i=1}^{k} f_i(x_i), \]
subject to \[ \sum_{i=1}^{k} g_{ij}(x_i) \leq s_j(\tau) \quad \text{for} \quad j=1, \ldots, k, \]
\[ \sum_{i=1}^{k} g_{ij}(x_i) = s_j(\tau) \quad \text{for} \quad j=k+1, \ldots, m, \]
\[ x_i \in X_i \quad \text{for} \quad i=1, \ldots, t. \]

We select values of \( y \) to generate a ‘useful’ range of \( s(\tau) \) values and continue until \( \tau = n \). Everett reports\(^9\) that this approach requires only a few iterations of solving \( P3 \) problems when the functions are monotone increasing.

Observe the similarity to dynamic programming, where stages are equivalent to cells. We are searching a parameter space \( Y \) rather than a state space containing the \( s(\tau) \) vectors.

Thus, separability provides special structure which may be taken to advantage. It is therefore desirable to select penalty functions where \( P3 \) is separable.

**Question 4:** Given that a Parameter \( y \) in \( Y \) Exists, Is There an Efficient Way to Find It?

Let us first consider a restricted Lagrangian with a single parameter as
\[ F = G_0[f(x)] - y \sum_{i=1}^{m} G_i[g_i(x)], \]
where \( m=k \) (i.e., no equality constraints).

Fiacco and McCormick\(^10\) use the following search scheme for \( G_0[f(x)] = f(x) \) and \( G_i[g_i(x)] = -1/[g_i(x) - b_i] \) for \( i=1, \cdots, m \): (1) Select \( y > 0 \) and let \( x_0 \) be any feasible solution in the interior of \( FR \) (i.e., \( g(x_0) < b \)). (2) Starting at \( x_{j-1} \), we generate \( x_j \) as follows (for \( j=1, 2, \cdots \)): reduce \( y \), where \( 0 < y' < y^{j-1} \), and use a gradient search procedure to maximize \( P3 \); this new maximum is \( x_j \); repeat step 2 until the desired accuracy is obtained.

They prove convergence when \( P1 \) is a convex program. Geometrically, the nature of the penalty function chosen prevents the gradient search from crossing the boundary by imposing an infinite penalty. That is, \( F \) diverges to negative infinity when \( g_i(x) \) approaches \( b_i \) from below for any \( i \).

Other penalty functions having the same divergence property were investigated by Zangwill.\(^25\) His results are somewhat more general than those of Fiacco and McCormick, but still only apply to restricted Lagrangians. He uses a single-parameter search, where the initial point is usually infeasible, and hence uses an ‘outside-in’ search.

Another approach (developed independently for \( m \) constraints by Everett and Furman\(^19\)) is given in Bellmore, Greenberg, and Jarvis\(^4\) for \( m=1 \) and \( F=f(x) - yg(x) \); they assume that \( F \) is bounded for all \( y \geq 0 \):

**Step 1.** Solve \( P3 \) for \( y=0 \) and \( y=M \), where \( M \) is arbitrarily large; let \( x_i^* \)
and \( x_1^* \) be the two solutions, respectively; further, let \( b_{i}^* = g(x_i^*) \) for \( i = 1, 2 \). **Step 2.** Let \( b_L = \max_i [g(x_i^*) \geq b] \) and \( b_u = \min_i [g(x_i^*) \leq b] \) if \( b_L = b_u \), then stop; otherwise set \( y = [f'(b_L) - f'(b_u)]/[b_L - b_u] \) and solve P3; if the solution \( x^* \) is such that \( g(x^*) = b_L \) or \( b_u \), terminate, since \( b \) is in a gap; otherwise, repeat step 2.

Computational experience indicates that these search procedures can be very efficient for certain P1 problems (see Bellmore, Greenberg, and Jarvis). Now consider \( m \)-parameter Lagrangians. Everett uses a step-wise procedure whereby \( y_i^{(j)} = y_i^{(j-1)}(1 + d_i^{(j)}) \), where \( d_i^{(j)} \) may contain positive, negative, and zero components. The sign of \( d_i^{(j)} \) is determined by the following rule: if \( g_i(x_{i-1}^*) > b_i \), then \( d_i^{(j)} > 0 \); if \( g_i(x_{i-1}^*) = b_i \), then \( d_i^{(j)} = 0 \); or if \( g_i(x_{i-1}^*) < b_i \), then \( d_i^{(j)} < 0 \).

Arrow, Hurwicz, and Uzawa also used a step-wise procedure in finding saddle points. Brooks and Geoffrion showed that we indeed are searching for \( y = y^* \), where \( (x^*, y^*) \) is a saddle-point of \( F \).

Whenever \( g_i(x) \) 'crosses' \( b_i \)—i.e., when \([g_i(x_{i-1}^*) - b_i][g_i(x_{i-1}^*) - b_i]<0\), then the magnitude of \( d_i^{(j)} \) should be decreased. Everett reports that a reduction of \( y_i \) has empirically proven successful.

Everett also reports that convergence may be accelerated by adding an extra parameter as:

\[
F = G_0[f(x)] - y_0 \sum_{i=1}^{k} y_i G_i[g_i(x)]
\]

where

\[
Y = \{(y_0, y) : y_i \geq 0 \quad \text{for} \quad i = 0, \cdots, k\} \cap E^{m+1}.
\]

The value of \( y_0 \) is adjusted by applying the rules to the redundant constraint

\[
\sum_{i=1}^{k} y_i G_i[g_i(x)] \leq \sum_{i=1}^{m} y_i G_i(b_i).
\]

Brooks and Geoffrion used linear programming as another method of searching for \( y \) in \( Y \), where \( F \) is the Lagrangian. Recall that one could solve a convex program by linearizing the functions and solving the linear program (see, for example, Hadley, page 106). The Brooks-Geoffrion scheme may be viewed as such a device, except that the location of the break points need not be specified in advance. One approximates the P1 problem by solving the case:

**Maximize**

\[
\sum_{i=1}^{k} f(x_i^*) u_i,
\]

**subject to**

\[
\sum_{i=1}^{k} g_i(x_i^*) u_i \leq b_i \quad \text{for} \quad i = 1, \cdots, k,
\]

\[
\sum_{i=1}^{k} g_i(x_i^*) u_i = b_i \quad \text{for} \quad i = k+1, \cdots, m,
\]

and (for normalization) \( \sum_{i=1}^{k} u_i \leq 1, \quad u_i \geq 0 \),

where the vector \( x_i^* \) solves the P3 problem for some previously chosen \( y = y_i \). The new value of \( y \) has components equal to the \( m \) 'true' dual variables to this linear program. That is, the dual is:
Minimize $\sum_{i=1}^{\infty} b_i w_i + w_0$, subject to $w_0 + \sum_{i=1}^{\infty} w_i g_i(x^*) \geq f(x^*)$ for $t=1, \cdots, T$, $w_i \geq 0$ for $i=0, \cdots, k$.

The value of $y$ is $w^*$. This choice of $y$ either gives the solution or generates a new dual constraint. Hence, one uses the dual simplex, for example, to proceed with the search.

To illustrate, suppose we consider the problem: Maximize $-\frac{1}{2} (x_1^2 + x_2^2) + x_1$ subject to $x_1 - x_2 \geq 0$ and $x_1 + x_2 \leq 1$. Using the Lagrangian, we have: Maximize $F = -\frac{1}{2} x_1^2 + (1 - y_1 - y_2) x_1 - \frac{1}{2} x_2^2 + (y_1 - y_2) x_2$ subject to $x_1, x_2 \geq 0$. The solution is: $x_1^* = \max(0, 1 - y_1 - y_2)$, $x_2^* = \max(0, y_1 - y_2)$.

Suppose we start with $y = (0, 0)$. The solution is $x^* = (1, 0)$. This is not feasible for $F_1$, yielding $b^* = (1, 1)$. We solve the linear program: Minimize $w_0 + w_2$, subject to $w_0 + w_1 + w_2 \geq \frac{1}{2}$ and $w_0, w_2 \geq 0$. An optimal solution is $w_0^* = w_2^* = 0$ and $w_1^* = \frac{1}{2}$. Using $y = (w_1^*, w_2^*)$ yields $x^* = (\frac{1}{2}, \frac{1}{2})$, which is the optimal solution to our original program.

We thus see that some results are available to search over $Y$, but only when $F$ is a Lagrangian. Some theorems are needed to gain insight into structure. To this end, Theorem 4.1 is an extension of Everett's monotonicity theorem.

**Theorem 4.1.** Let $F = F_i[f(x), g(x)] + r_i[1 - G_i(g_i(x))]$, where $r \neq 0$, $r G_i(g_i(x))$ is monotone nonincreasing in $g_i(x)$ and $Y = [y: y^1 \geq 0]$. Then:

1. If $y^*_1 = y^*_2 > 0$ and if $y_1^* > y_2^*$, then $g_i(x^*) \leq g_i(x^{**})$.
2. If $y_1^* = y_2^* > 0$, $y_2^* > y_1^*$, $G_i(g_i(x^*)) > 1$, and $G_i(g_i(x^{**})) > 1$, then $g_i(x^*) \leq g_i(x^{**})$.

The proof of Theorem 4.1 is in the appendix. The result provides information on how to change $y$ in order that $g_i(x^*)$ converge to $b_i$.

An example of a monotone penalty function with $G_i > 1$ is $F = F_1 + y_2[1 + \exp[-g_i(x)]]$, where $F_1$ is monotone.

It is hoped that this theorem will provide a method for searching $y_1$ and $y_2$ analogous to the Lagrangian search schemes.

Now consider the following:

**Theorem 4.2.** Let $F_1, \cdots, F_s$ be $s$ penalty functions with maxima $F_1^*, \cdots, F_s^*$. Further, let $F$ be defined as $F = H(F_1, \cdots, F_s)$, where $H$ is monotone increasing in $F_1, \cdots, F_s$. Then max $F \leq H(F_1^*, \cdots, F_s^*)$. Further, if $x_1^* = \cdots = x_s^* = x^*$, then max $F = H(F_1^*, \cdots, F_s^*)$, with solution $x^*$.

**Corollary 4.2.1.** Let $F = a F_1 + (1 - a) F_2$, where $0 \leq a \leq 1$. Further, let $F_i = f(x) - \sum_j y_{j}^{(i)} g_i(x)$ for $i = 1, 2$.

Then, if $x_1^* = x_2^*$, values of $y$ satisfying $y = a y^{(1)} + (1 - a) y^{(2)}$ cannot yield any new solution.

The proof of Theorem 4.2 is in the appendix. Corollary 4.2.1 follows.
directly from Theorem 4.2. This result should help in searching parameters for non-Lagrangian penalty functions.

In conclusion, it is useful to note when the $y$ sought is a saddle point of $F(f, g - b; y)$. One could then, for example, employ the Arrow, Hurwicz, and Uzawa search scheme.

Brooks and Geoffrion have shown that $y^*$ is a saddle point of $F$ when $F$ is the usual Lagrangian. Gould also proved a saddle-point theorem dealing directly with functionals. Our question is: “Under what circumstances will we have $F[f(x^*), g(x^*) - b; y^*] \leq F[f(x^*), g(x^*) - b; y]$ for all $y$ in $Y$?”

We shall consider $F$ monotone, and $(x^*, y^*)$ satisfying the conditions in Corollary 1.1.1. Further, let $F$ be differentiable in $y$. From Taylor's theorem, we have

$$F[f(x^*), g(x^*) - b; y] = F[f(x^*), g(x^*) - b; y^*] + (y - y^*) \nabla_y F + R.$$  

In order for $y^*$ to minimize $F^*$, we require $(y - y^*) \nabla_y F + R \geq 0$ for all $y$ in $Y$. A sufficient condition is:

**Lemma 4.1.** If $\nabla_y F^* = 0$ and $F$ is convex in $y$, then $y^*$ is a saddle point of $F$.

**Lemma 4.2.** If (1) $y_i^* F_{y_i} = 0$ for $i = 1, \ldots, m$, and (2) $y \nabla_y F$ is monotone nondecreasing, then $(x^*, y^*)$ is a saddle point of $F$.

In both lemmas, the saddle-point property holds if $y^*$ exists (i.e., no gaps). Note that the Lagrangian is a special case of both lemmas.

The presence of a saddle point then provides a dual program to $P_1$. If $F$ is concave in $x$, then a dual to $P_1$ is: Minimize $F(f, g; y)$: (1) $y \in Y$, (2) $\nabla_x F = 0$, where we assume that $Y$ is a convex set. This relates to Dantzig, Eisenberg and Cottle's more general duality (i.e., not relying on the Lagrangian) in that we can deal with programs which are not convex. [For example, Pierskalla's monotonic case has a dual if $F$ is chosen by the rule of procedure discussed in Question 2 and $F$ is chosen such that it is concave in $x$ over $X$ and convex in $y$ over $Y$ (a closed, convex set).]

**Further Generalizations**

The preceding discussion is a special case of the general concept of ‘problem transformation.’ The general case allows transformations of variables as well.

**Definition 5.** A problem transformation of $P_1$, say $P_2$, is an optimization problem where $FR2 = R(x1, f1, g1, X1)$, $f2 = B(x1, f1, g1, X1)$, and $x2 = V(x1, f1, g1, X1)$, where $FRi$ is the feasibility region for $Pi$. Similarly, $fi, gi, Xi$ and $xi$ denote the objective, subsidiary conditions, the set $X$ and the decision variables, respectively for $Pi$.

**Definition 6.** A valid problem transformation of $P_1$ (denoted $VPT$), say $P2$, is a problem transformation where $OR1 = T(OR2)$. 

**Generalized Penalty-Function Concepts**
In other words, we solve a substitute problem in which the original optimality region is a transform of the new optimality region. Penalty functions use the transformation given by: \( FR2 = X1 \), \( f2 = F(f1, g1) \), and \( x2 = x1 \).

Whether or not \( P2 \) is valid is not known in advance. Instead, we discover that \( OR2 \) is the optimality space for some \( P1 \) problem, but not necessarily the one we wish to solve. However, information is obtained to determine our next choice of \( P2 \).

Classically, variable transformation has used elementary results. For example, if \( x2 = t(x1) \) and \( x1 = t^{-1}(x2) \), then \( P2 \) becomes

\[
\begin{align*}
\text{maximize} & \quad f2(x2) = f1[t^{-1}(x2)], \\
\text{subject to} & \quad g2i(x2) = g1i[t^{-1}(x2)] \leq b_i \quad \text{for} \quad i = 1, \ldots, k, \\
& \quad g2i(x2) = g1i[t^{-1}(x2)] = b_i \quad \text{for} \quad i = k + 1, \ldots, m, \\
& \quad x2 \in X2 = Q(X1),
\end{align*}
\]

where \( Q \) is given by: \( x2 \) is in \( X2 \), if and only if, \( t^{-1}(x2) \) is in \( X1 \).

To illustrate, consider the problem given by Kuhn and Tucker\(^{[10]} \) as: maximize \( x1 \), subject to \(-(1-x1)^3 + x2 \leq 0, \ x \in X = [x : x \geq 0] \).

The solution, \( x^* = (1, 0) \), is at a cusp so that the Kuhn-Tucker constraint qualification is not met. However, let us consider the variable transformation \( x = t(x) \) given by: \( z1 = -(1-x1)^3 + x2, \ z2 = x2 \), with the inverse \( x1 = (z1-z2)^{1/3} + 1, \ x2 = z2 \). The \( P2 \) problem is: maximize \( f2(x) = (z1-z2)^{1/3} + 1 \) subject to \( z1 \leq 0 \) and \( x \) in \( Z = [z : (z1-z2)^{1/3} + 1 \geq 0, \ z2 \geq 0] \).

Further transformations discussed earlier give the equivalent problem: maximize \( z1 - z2 \), subject to \( z1 \leq 0, \ z1 - z2 \geq -1 \) and \( z2 \geq 0 \). This now is a linear program with \( x^* = (0,0) \). Using the inverse transformation we obtain \( x^* = (1,0) \).

The point of this exercise is to illustrate how problem transformations other than penalty functions may be applied to solve a given problem efficiently.

**CONCLUDING REMARKS**

This paper has placed the concept of penalty functions in a general theory that unifies concepts and presents some new theorems to gain insight into structure.

Moreover, there is still the more general concept of problem transformation in optimization theory, of which the penalty-function approach is a special case of interest. However, there are other transformations that reduce the complexity of the problem.
APPENDIX

PROOFS OF THE THEOREMS

THEOREM 1.1. Let $F$ be a monotone penalty function for a given $y$. Let $x^*$ be an optimal solution to $P_3$. Then, $x^*$ solves $P_1$ for $b = g(x^*)$.

Proof. Since $x^*$ maximizes $P_3$, we have

$$F[f(x^*), g(x^*); y] \geq F[f(x), g(x); y]$$

for all $x$ in $X$. Now, suppose we restrict $x$ so that $g_i(x) \leq b_i = g_i(x^*)$ for $i = 1, \ldots, k$ and $g_i(x) = b_i = g_i(x^*)$ for $i = k+1, \ldots, m$. Suppose an $x$ satisfying the above, say $x^1$, exists such that $f(x^1) > f(x^*)$. Then, by the monotonicity property:

$$F[f(x^1), g(x^1); y] > F[f(x^*), g(x^*); y],$$

which contradicts (1).

COROLLARY 1.1.1. If $F[f(x), g(x); y] = C$ (a constant) implies $F[f(x), g(x) + a\varepsilon_i; y] = C$ for all $a$, where $i \leq k$, then the solution to $P_3$ (viz., $x^*$) solves $P_1$ for $b_i = g_i(x^*)$ for $j \neq i$ and $b_i \geq g_i(x^*)$.

Proof. From Theorem 1.1 we have that $x^*$ solves $P_1$ for $b = g(x^*)$. Further,

$$F[f(x^*), g(x^*); y] = F[f(x), g(x) + a\varepsilon_i; y].$$

Therefore, $x^*$ maximizes $F[f(x), g(x) + a\varepsilon_i; y]$ over $X$. From Theorem 1.1, $x^*$ solves $P_1$ for $b = g(x^*) + a\varepsilon_i$ for all $a$. Therefore, the only restriction on $b$, is that $x^*$ be feasible—i.e., $b_i \geq g_i(x^*)$.

THEOREM 1.2. Let $F$ be monotone increasing in $f(x)$ and such that a contour, say $F = K$ (a constant), implies a continuous function as $f(x) = Q^*[g(x); y]$. Further, let $Q^*$ be the function corresponding to $F = F^*$ (the maximum in $P_3$). Then, for all $x$ we have $f(x) \leq Q^*[g(x); y]$.

Proof. Suppose there exists $x^1$ such that $f(x^1) > Q^*[g(x^1); y]$. Let $F^1 = F[f(x^1), g(x^1); y]$, which then implies the existence of $Q^1$ such that $f(x^1) = Q^1[g(x^1); y]$. We thus have

$$F[Q^1[g(x^1); y], g(x^1); y] > F[Q^*[g(x^1); y], g(x^1); y],$$

and since $Q^*$ is the contour for $F = F^*$, we have

$$F[Q^*[g(x^1); y], g(x^1); y] = F[Q^*[g(x^*); y], g(x^*)].$$

Hence, $F^1 > F^*$. This contradicts the fact that $F^*$ is the maximum.

THEOREM 2.1. Let $F$ be a Lagrangian as:

$$F = G_0[f(x)] - \sum_{i=1}^m y_i G_i[g_i(x)].$$

Let problem $P_3$ for $y = y^*$ have a finite (or denumerable) number of solutions, say $x_1^*, \ldots, x^*_t$, where $t \geq 1$. Let the set $B$ be defined as

$$B = [b: G_i(b_i) = \sum_{j=1}^m w_j G_i(g_i(x_j^*)) \text{ for } j = 1, \ldots, m, \sum_{j=1}^m w_j = 1 \text{ and } w \geq 0].$$

Further, let $B^-$ be defined as $B^- = B - [b: g(x_j^*) = b_j \text{ for } j = 1, \ldots, t]$. Then, there exists no value of $y$ in $Y$ with a solution $x^{**}$ such that $g(x^{**})$ is in $B^-$. 

Proof. Assume there exists $y^{**} = y^* + \varepsilon$ in $Y$ such that $x^{**}$ is the solution to
problem $P_3$, and $b^{**} = g(x^{**})$. Clearly, if $x^{**} = x_j^*$ for any $j = 1, \ldots, t$, then $b^{**} = b^*$ so that $b^{**}$ is not in $B^-$. Hence, we consider $x^{**} \neq x_j^*$ for any $j = 1, \ldots, t$.

Since $x^{**}$ solves $P_3$ for $y^* = y^{**}$, we have

$$G_0[f(x^{**})] - \sum_{i=1}^{m} v_i [G_i(x^{**})] \geq G_0[f(x_j^*]) - \sum_{i=1}^{m} v_i [G_i(x_j^*])$$

for $j = 1, \ldots, t$. Substituting $y^* + d$ for $y^{**}$, we have

$$G_0[f(x^{**})] - \sum_{i=1}^{m} v_i [G_i(x^{**})] - \sum_{i=1}^{m} d_i G_i(x^{**})$$

$$\geq G_0[f(x_j^*]) - \sum_{i=1}^{m} v_i [G_i(x_j^*]) - \sum_{i=1}^{m} d_i G_i(x_j^*)$$

for all $j = 1, \ldots, t$. Further, since $x_j^*$ is a solution to $P_3$ for $y = y^*$ and since $x^{**}$ is not a solution to $P_3$ for $y = y^*$, we have

$$G_0[f(x^{**})] - \sum_{i=1}^{m} v_i [G_i(x^{**})] \geq G_0[f(x_j^*]) - \sum_{i=1}^{m} v_i [G_i(x_j^*])$$

for all $j = 1, \ldots, t$. Thus, in order for $x^{**}$ to be a solution to $P_3$ for $y = y^* + d$ and not be a solution to $P_3$ for $y = y^*$, it is necessary that $d$ satisfy the following inequality:

$$\sum_{i=1}^{m} d_i [G_i(x_j^*) - G_i(x^{**})] > 0$$

for all $j = 1, \ldots, t$.

Mottokin's transposition theorem\cite{Mottokin1955} states that $d$ exists to satisfy the above inequality if and only if there is no solution to the system:

$$\sum_{i=1}^{m} v_i [G_i(g_i(x^*)) - G_i(g_i(x^{**}))] = 0, \quad v_i \geq 0, \quad v_i \neq 0 \quad \text{for all} \quad i = 1, \ldots, m.$$

Setting $w_j = v_j / \sum_{i=1}^{m} v_i$, we have that $g(x^{**})$ cannot be in $B^-$.\n
**Theorem 2.2.** Let $F(x, y)$ be given by $F(x, y) = G_0[f(x)] - \prod_{i=1}^{m} G_i(x)^{v_i}$, where $y$ is in $Y = [y, y_0]$. We assume $G_0$ and $G_i$ are monotone increasing, and $G_i(x) \geq 1$ for $i = 1, \ldots, m$. Then, if $F(x, y)$ has alternate maxima $x_1^*, \ldots, x_t^*$, there exists no $y$ in $Y$ such that

$$G_i(x_i^*) = \prod_{j=1}^{t} G_i(x_j^*)^{v_j}$$

for $v_0 \geq 0$ and $\sum_{j=1}^{t} v_j = 1$, where $J$ is the set of indices $j$ satisfying

$$\prod_{i=1}^{m} G_i(x_i^*)^{v_i} \geq \sum_{i=1}^{m} G_i(x_i^*)^{v_i}.$$

**Proof.** Consider $y^* = y + d$, and let $x^*$ maximize $F(x, y^*)$. For $x^* = x_j^*$ for some $j = 1, \ldots, t$ we have no new value of $g(x^*)$. Hence, we shall suppose that $x^*$ does not maximize $F(x, y^*)$.

Since $x^*$ maximizes $F(x, y^*)$, we have

$$G_0[f^*(b)] - \prod_{i=1}^{m} G_i(b_i)^{v_i} + d_i \geq G_0[f^*(b)] - \prod_{i=1}^{m} G_i(b_i)^{v_i}$$

for $j = 1, \ldots, t$. Let us subtract $\prod_{i=1}^{m} G_i(b_i)^{v_i} + \prod_{i=1}^{m} G_i(b_i)^{v_i}$ from each side, and we obtain

$$F(x^*; y) - \prod_{i=1}^{m} G_i(b_i)^{v_i} - \prod_{i=1}^{m} G_i(b_i)^{v_i} \leq F(x_j^*; y) - \prod_{i=1}^{m} G_i(b_i)^{v_i} - \prod_{i=1}^{m} G_i(b_i)^{v_i}$$

for $j = 1, \ldots, t$. Since $F(x_j^*; y) > F(x^*, y)$, we have

$$\prod_{i=1}^{m} G_i(b_i)^{v_i} \geq \prod_{i=1}^{m} G_i(b_i)^{v_i} + \prod_{i=1}^{m} G_i(b_i)^{v_i} < \prod_{i=1}^{m} G_i(b_i)^{v_i} - \prod_{i=1}^{m} G_i(b_i)^{v_i},$$

or

$$\prod_{i=1}^{m} G_i(b_i)^{v_i} (\prod_{i=1}^{m} G_i(b_i)^{d_i} - 1) < \prod_{i=1}^{m} G_i(b_i)^{v_i} (\prod_{i=1}^{m} G_i(b_i)^{d_i} - 1).$$
For each \( j \) satisfying \( \prod_{i=1}^{n} G_i(b_i) v_i^j \geq \prod_{i=1}^{n} G_i(b_i') v_i^j \), we have [since \( G_i(\cdot) \geq 1 \)]
\[ \prod_{i=1}^{n} G_i(b_i) v_i^j < \prod_{i=1}^{n} G_i(b_i') v_i^j, \text{ or } \sum d_i[\log G_i(b_i) - \log G_i(b_i')] < 0. \]
Such a \( d \) exists if and only if there is no solution to the system given by \( \sum_{j} w_j[\log G_i(b_i) - \log G_i(b_i')] = 0 \), \( w_j \geq 0 \), \( \sum_j w_j = 1 \), where the sum is over those \( j \) for which \( \prod_{i=1}^{n} G_i(b_i) v_i^j \geq \prod_{i=1}^{n} G_i(b_i') v_i^j \). This becomes \( G_i(b_i) = \prod_{i=1}^{n} G_i(b_i') v_i^j \) for \( v_i \geq 0 \) and \( \sum_{i=1}^{n} v_i = 1 \), where \( v_j = w_j / \sum_{i=1}^{n} w_i \). Thus, if \( b \) satisfies \( \prod_{i=1}^{n} G_i(b_i) v_i^j \geq \prod_{i=1}^{n} G_i(b_i') v_i^j \) for \( j \) in \( J \), then \( b \) cannot satisfy \( G_i(b_i) = \prod_{i=1}^{n} G_i(b_i') v_i^j \); for any \( v : v_i \geq 0 \) and \( \sum_{i=1}^{n} v_i = 1 \).

**Theorem 4.1.** Let \( F = F[f(x), g(x)] + r y_1[G[g_i(x)]]^y \) where \( r \neq 0, r G[g_i(x)] \) is monotone nonincreasing in \( g_i(x) \) and \( Y = [y : y \geq 0] \). Then:

1. If \( y_2 > y_1 > 0 \) and \( y_1 > y_1^* \), then \( g_i(x_1^*) \leq g_i(x_2^*) \).
2. If \( y_1^* > 0, y_2^* > y_2^*, G[g_i(x_2^*)] > 1 \), and \( G[g_i(x_1^*)] > 1 \), then \( g_i(x_2^*) \leq g_i(x_1^*) \).

where \( x^* \) solves \( PZ \) for \( y = y^* \), and \( x^*_2 \) solves \( PZ \) for \( y = y^*_2 \).

**Proof.** Conclusion (1) follows directly from Everett’s monotonicity theorem, but we present a proof for completeness.

We have \( F[f(x_2), g(x_1)] + r y_1[G[g_i(x_2)]]^y \geq F[f(x_1^*), g(x_1^*)] + r y_1[G[g_i(x_1^*)]]^y \), since \( x_1^* \) is the solution to \( PZ \) for \( y = y_1^* \). This implies

\[ F[f(x_2), g(x_2)] + r y_1[G[g_i(x_2)]]^y \geq F[f(x_1^*), g(x_1^*)] + r y_1[G[g_i(x_1^*)]]^y + r d[G[g_i(x_1^*)]]^y. \]

Since \( x_1^* \) solves \( PZ \) for \( y = y^* \), we have that

\[ r d[G[g_i(x_1^*)]]^y \geq r[G[g_i(x_1^*)]]^y. \]

Now suppose \( r > 0 \). Since \( y_2^* > 0 \), we have \( G[g_i(x_2^*)] \geq G[g_i(x_1^*)] \). Since \( r G \) is monotone nonincreasing in \( g_i(x) \) and \( r > 0 \), we have that \( G \) is monotone nonincreasing in \( g_i(x) \). Hence, \( g_i(x_1^*) \leq g_i(x_2^*) \).

If \( r < 0 \), then \( G[g_i(x_2^*)] < G[g_i(x_1^*)] \) and \( G \) is monotone nondecreasing in \( g_i(x) \), which gives the desired result.

Now consider conclusion (2). Again, since \( x^* \) is a solution to \( P3 \) when \( y = y^* \), we have

\[ F[f(x_2), g(x_2)] + r y_1[G[g_i(x_2)]]^y \geq F[f(x_1^*), g(x_1^*)] + r y_1[G[g_i(x_1^*)]]^y. \]

We set \( y_2^* = y_1^* + d \), where \( d > 0 \), and note that \( y_1^* = y_1^* > 0 \). Thus, we have

\[ F[f(x_2), g(x_2)] + r y_1[G[g_i(x_2)]]^y \geq F[f(x_1^*), g(x_1^*)] + r y_1[G[g_i(x_1^*)]]^y + r d[G[g_i(x_1^*)]]^y. \]

Let us add

\[ r y_1[G[g_i(x_2)]]^y + r y_1[G[g_i(x_1^*)]]^y \]

to both sides of the inequality. We thus have
Since $x^{**}$ solves P3 for $y = y^{**}$, we have
\[ ry^{**} + ry^{**} (G[g_i(x^{**})]) \geq F_i(F(x^{**}), y^{**}) + ry^{**} (G[g_i(x^{**})]) \]
which can be written as
\[ ry^{**} (G[g_i(x^{**})]) \geq 1 - [G[g_i(x^{**})]^4] \geq 1 - [G[g_i(x^{**})]^4]. \]
Further, since $y^{**} > 0$, we have
\[ ry^{**} (G[g_i(x^{**})]) \geq 1 - [G[g_i(x^{**})]^4] \geq ry^{**} [G[g_i(x^{**})]] \geq 1 - [G[g_i(x^{**})]^4]. \]
Suppose $r > 0$. We then wish to show that $G[g_i(x^{**})] \geq G[g_i(x^{**})]$. Upon so doing, the monotonicity of $G$ then yields the desired result.

Consider the contrary—i.e., $G[g_i(x^{**})] < G[g_i(x^{**})]$. Then, since $G[g_i(x^{**})] > 1$, and $G[g_i(x^{**})] > 1$, we have
\[ 0 > 1 - [G[g_i(x^{**})]^4] \geq 1 - [G[g_i(x^{**})]^4] \]
or
\[ 0 < -1 - [G[g_i(x^{**})]^4] < -1 - [G[g_i(x^{**})]^4]. \]
Thus,
\[ -1 - [G[g_i(x^{**})]^4] \geq -1 - [G[g_i(x^{**})]^4] \]
which is a contradiction.

For $r < 0$, $G$ is monotone nondecreasing, and
\[ G[g_i(x^{**})] \geq 1 - [G[g_i(x^{**})]^4] \geq 1 - [G[g_i(x^{**})]^4]. \]
Again, we need only show that $G[g_i(x^{**})] \leq G[g_i(x^{**})]$. We assume the contrary—i.e., suppose $G[g_i(x^{**})] > G[g_i(x^{**})]$. Then,
\[ -1 - [G[g_i(x^{**})]^4] > -1 - [G[g_i(x^{**})]^4] \]
so that we have
\[ -1 - [G[g_i(x^{**})]^4] > -1 - [G[g_i(x^{**})]^4] \]
which is a contradiction.

**Theorem 4.2.** Let $F_1, \ldots, F_s$ be $s$ penalty functions with maxima $F_1^*, \ldots, F_s^*$. Further, let $F$ be defined as $F = H(F_1, \ldots, F_s)$, where $H$ is monotone increasing in $F_1, \ldots, F_s$. Then $\max F \leq H(F_1^*, \ldots, F_s^*)$. Further, if $x_1^* = \cdots = x_s^* = x^*$, then $\max F = H(F_1^*, \ldots, F_s^*)$ with solution $x^*$.

**Proof.** Since $H$ is monotonic, we have $H(F_1, \ldots, F_s) \leq H(F_1^*, \ldots, F_s^*)$ if $F \leq F^*$. Hence, $H(F_1^*, \ldots, F_s^*) \geq H(F_1, \ldots, F_s) = F$, since $F_i \leq F_i^*$ for $i = 1, \ldots, s$.

Further, if $x_i^* = x^*$ for $i = 1, \ldots, s$, then, $\max F = H(F_1^*, \ldots, F_s^*)$.

**ACKNOWLEDGMENT**

The authors wish to express special thanks to H. Everett III for his many helpful comments.
REFERENCES


9. ———, private conversation.


